# CHAPTER 4
# HYBRID IMMUNE CLONAL NETWORK ALGORITHM

## 4.1 BACKGROUND

This chapter discusses our proposed method for detection and classification phase, and how AIS algorithms and dataset are used for this research. Besides, this chapter reviews existing method using WEKA, a machine learning tool for data mining. A new algorithm named "Hybrid Immune Clonal Network Algorithm" is also presented in this chapter.

## 4.2 EXPERIMENTS IN DETECTION PHASE

Detection is the first phase in our IMSM model. Detection is the process of identifying the SMS messages received either they are classified as spam or ham. The purpose for this phase is to reduce the number of spam messages entering to our mobile phone and also help us in recognizing and differentiating between the characteristics of ham and spam messages. The first section details out method and procedures of the conduct of experiment, with the following sub-section presented our initial work for enhancing Danger Theory by introducing three (3) parameters for effective detection.

### 4.2.1 PROCEDURES AND METHODS

In the spam managing model, Danger Theory is used to detect the incoming SMS messages that already show that they are spam messages by understanding the content and meaning of message and for the Negative Selection, incoming SMS message will be filtered first to identify it is spam or ham.

Three different datasets were used for detecting text spam messages using AIS algorithms and the dataset are discussed below:-

a. GrumbleText (GT)

   A website to submit complains of SMS Spam. People, who receive SMS Spam, voluntarily submit the SMS on this site.

b. Dublin Institution of technology (DIT)

   The DIT SMS spam dataset is a corpus of 1353 unique spam SMS text messages collected by scraping messages from two UK public consumer complaints website.

c. UCI Machine Learning Repository

   This corpus site stored a collection of public set of SMS messages labelled as spam and ham (non-spam) for the use by many researchers.

Table 4.1 shows the number of spam and ham messages for each dataset and they will be tested using the Danger Theory and Negative Selection. A comparison is made in terms of correctly classified, incorrectly classified, ROC (Receiver Operating Characteristics)

AREA, Accuracy and time taken to build the model. WEKA is used to test the performance of these algorithms in detecting spam messages. The measurement of detected performance using WEKA is discussed below:-

- Correctly Classified (CC): SMS is **correctly classified** as spam or ham message.
- Accuracy: The rate for accuracy of a classifier or algorithm is determined by the formula as stated below:

$$Accuracy = (TP + TN) / (TP + FN + TN + FP)$$

- Incorrectly Classified (IC): SMS is **incorrectly classified** as spam or ham messages.
- ROC Area: For analysing and illustrating the performance of various systems.

**Table 4. 1:** Number of messages in each dataset

| | Dataset A (DIT) | Dataset B (GT) | Dataset C (UCI) |
|---|---|---|---|
| **Ham** | 579 | 232 | 211 |
| **Spam** | 774 | 47 | 1443 |
| **Total** | 1353 | 279 | 1654 |

Experiments were conducted to compare the performance between Negative Selection and Danger Theory and also to find which algorithm is better using WEKA (i.e. open source tool for machine learning). Only three datasets were used due to limited dataset available since we wanted to find the performance of algorithms using different number of messages. From the experiment, results showed that Negative Selection perform well than Danger Theory in all aspects such as accuracy and time. Detail discussion can be found in Chapter 5, Section 5.4.1.

### 4.2.2 ENHANCING AIS DETECTION

Having understood that Danger Theory needs to be improved, three features were proposed; namely based on the length of messages that is more than 100 characters, messages containing special characters such as symbols and numbers and keywords of spam messages and our justification of using these are highlighted as following:

1. *Length of messages* (i.e. greater than 100 characters).

From our finding in analysing SMS spam datasets, we found SMS spam tends to be longer. Besides, SMS spam normally uses standard and formal language in order to attract users so that they can understand and put their interest to those particular messages. Therefore, in this thesis, it is assumed that messages that are more than 100 in length could potentially be classified as spam due to the maximum length of any SMS is up to 160 characters in length (Derek Johnson, 2013). Table 4.2 shows the difference between spam and ham messages based on length.

**Table 4. 2:** The difference between spam and ham messages

| | |
|---|---|
| SPAM | PRIVATE! Your 2003 Account Statement for shows 800 un-redeemed S. I. M. points. Call 08715203694 Identifier Code: 40533 Expires 31/10/04. |
| HAM | Everybody had fun this evening. Miss you. |

2. *Special Characters* (i.e. numbers and symbols).

From our study, it is revealed that the usage of special characters do exist and are common in SMS spams. For instance, spammers prefer using numbers or digits such as phone number, code number to claim, service code and sum of money (i.e. 300 pounds). In addition to these, special characters or symbols such as ****, $, XXX are commonly found to be used in SMS spams. Table 4.3 shows example of spam messages containing special characters.

**Table 4. 3:** Spam messages containing special characters using numbers

| |
|---|
| PRIVATE! Your 2003 Account Statement for shows 800 un-redeemed S. I. M. points. *Call 08715203694 Identifier Code: 40533* Expires 31/10/04. |

3. *Keywords of spam and ham.*

From our analysis, words that being used in spam messages are common and similar across platforms. Thereby, our approach also considers common keywords for detecting spam. Examples of keywords are like – 'Free', 'Call' and 'Claim'. Besides, ham keywords are also investigated and used in our detection algorithms. Example of keywords for ham and spam messages can be found in Appendix B.

By detecting SMS spam using their lengths, used characters and keywords, we postulate it will make the detection rate higher; and at the same time minimizing device resources. To achieve what we have claimed for, three aforementioned detection features were stipulated into five different algorithms as shown in Table 4.4 and Figure 4.1 presents the

pseudo-code of the proposed method. These three features were embedded into five algorithms because we want to see how they work on detection process using different environment (i.e. using one or more than one features).

**Table 4. 4:** Different features in each algorithm

| DETECTION OF HAM/SPAM | | | | |
|---|---|---|---|---|
| **Algorithm** | **Phase 1** | **Phase 2** | **Phase 3** | **Phase 4** |
| 1 | Spam keywords | Ham keywords | (NA) | (NA) |
| 2 | Length+ Spam keywords | Ham keywords | (NA) | (NA) |
| 3 | Number/digit+ Spam keywords | Ham keywords | (NA) | (NA) |
| 4 | Length | Number/digit+ Spam keywords | Ham keywords | (NA) |
| 5 | Length | Number/digit + Spam keywords | Ham keywords | Spam keywords |

The first three algorithms use only two aforementioned features with the remaining algorithms combining all features. Specifically, Algorithm 1 uses only keywords, Algorithm 2 uses the combination of message length and keywords, and Algorithm 3 uses the features of special characters and keywords respectively. Both Algorithms 4 and 5 use message length, special characters and keywords. Figures 4.2 to 4.6 show the flow diagram for each algorithm with the explanation on how they work.

```
Input; M:SMS messages
Output; H:Ham SMS or S: Spam SMS

1. Begin (while TRUE)
2. Read all SMS messages, M
3. Detect S (i.e. using five algorithms as in Table 4.4)
4. Print the results
5. Copy to H or S folder
6. End
```

**Figure 4. 1:** Generalized Pseudo-code for five algorithms

1. The process starts.
2. SMS message content is read and viewed.
3. SMS messages are scanned using five algorithms to detect spam and ham messages.
4. The results of ham and spam messages are shown.
5. Ham messages are copied to folder ham while spam messages will go to folder spam.
6. The process ends.

Three different datasets used in this experiment are discussed below and Table 4.5 details out different numbers of spam and ham messages in each dataset. DIT dataset is used although it only has spam messages because we want to see the performance of algorithms using variation of dataset.

a. UCI Machine Learning (UCI)

   This dataset has been discussed in Section 4.2.1.

b. British English SMS Corpora (BEC)

   The messages are collected from GrumbleText website for SMS spam and Caroline Tag's PhD Thesis for SMS ham.

c. Dublin Institute of Technology (DIT).

   This dataset has been discussed in Section 4.2.1.

Pre-processing or cleaning process for each dataset is not necessary because the characters (i.e. numbers and symbols) in these messages with the length of messages may help with the detection process.

**Table 4. 5:** Characteristics of used datasets

| | Dublin Institute Of Technology (DIT) | British English SMS Corpora (BEC) | UCI Machine Learning (UCI) |
|---|---|---|---|
| HAM | - | 450 | 4825 |
| SPAM | 1353 | 425 | 747 |
| TOTAL | 1353 | 875 | 5572 |

The measurement of detection performance is based on Correctly Classified Messages, True Positive and True Negative.

- Correctly Classified (CC): SMS is correctly classified as spam or ham message.
- Accuracy: The rate for accuracy of a classifier or algorithm is determined by the formula as below.

$$Accuracy = (TP + TN) / (TP+FN + TN + FP)$$

- True Positive (TP): SMS messages are correctly classified as ham messages.
- True Negative (TN): SMS messages are correctly classified as spam messages.
- False Positive (FP): SMS messages are incorrectly classified as ham messages.
- False Negative (FN): SMS messages are incorrectly classified as spam messages.

ALGORITHM 1

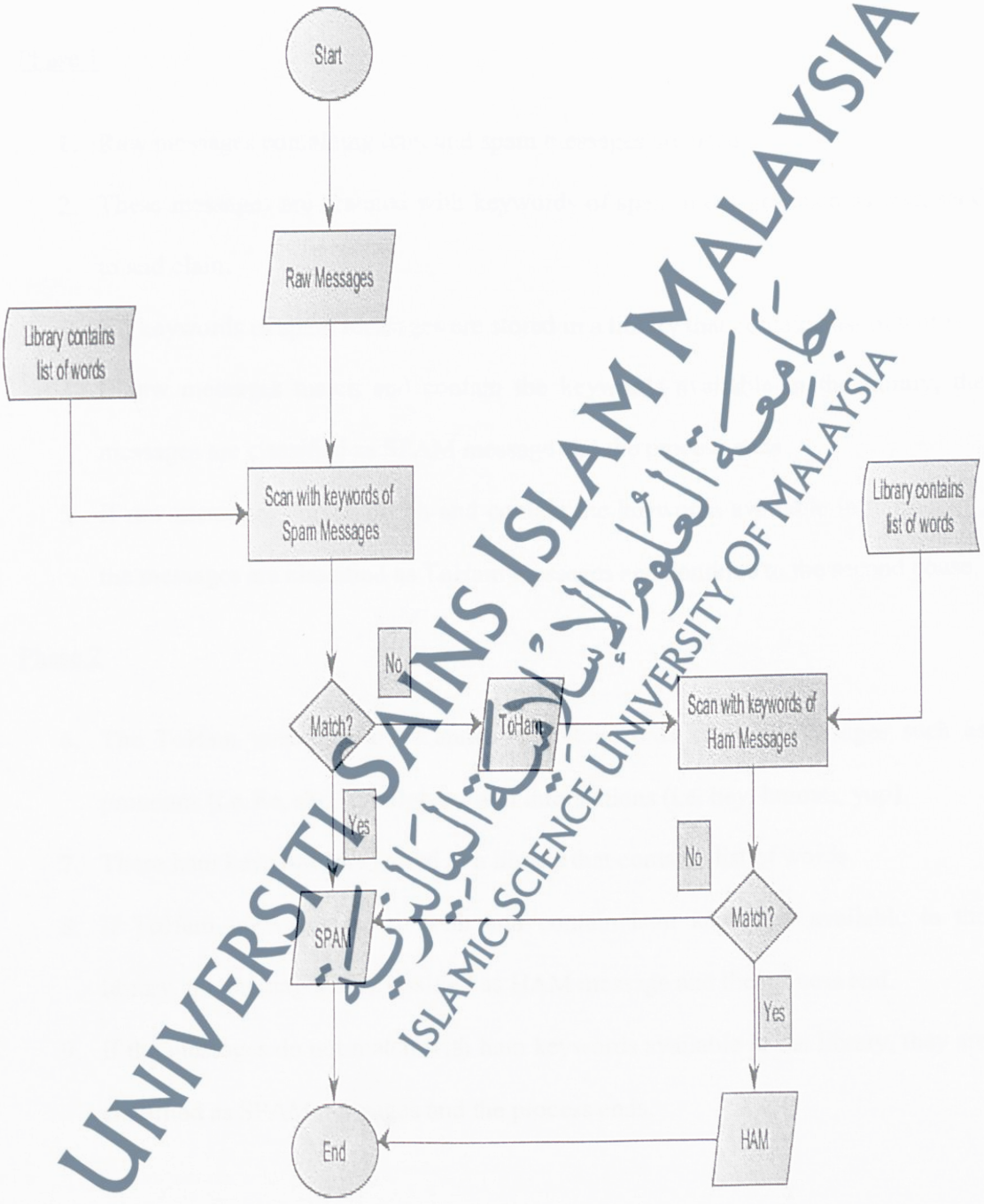




**Figure 4. 2:** Algorithm 1

Figure 4.2 shows the flowchart of Algorithm 1. This algorithm requires two phases and each phase is discussed as follows:-

Phase 1

1. Raw messages containing ham and spam messages are used.
2. These messages are scanned with keywords of spam messages such as text, send to and claim.
3. All keywords of spam messages are stored in a library that contains list of words.
4. If raw messages match and contain the keywords available in the library, the messages are classified as SPAM message and the process ends.
5. If raw messages do not match and contain the keywords available in the library, the messages are classified as ToHam messages and continue to the second phase.

Phase 2

6. The ToHam messages are scanned with keywords of ham messages such as pronouns (i.e. he, she, themselves) and interjections (i.e. hey, hmmm, yup).
7. These ham keywords are stored in a library that contains list of words.
8. If ToHam messages match with and contain ham keywords available in the library, the messages are classified as HAM message and the process end.
9. If the messages do not match with ham keywords available in the library, they are classified as SPAM messages and the process ends.

ALGORITHM 2

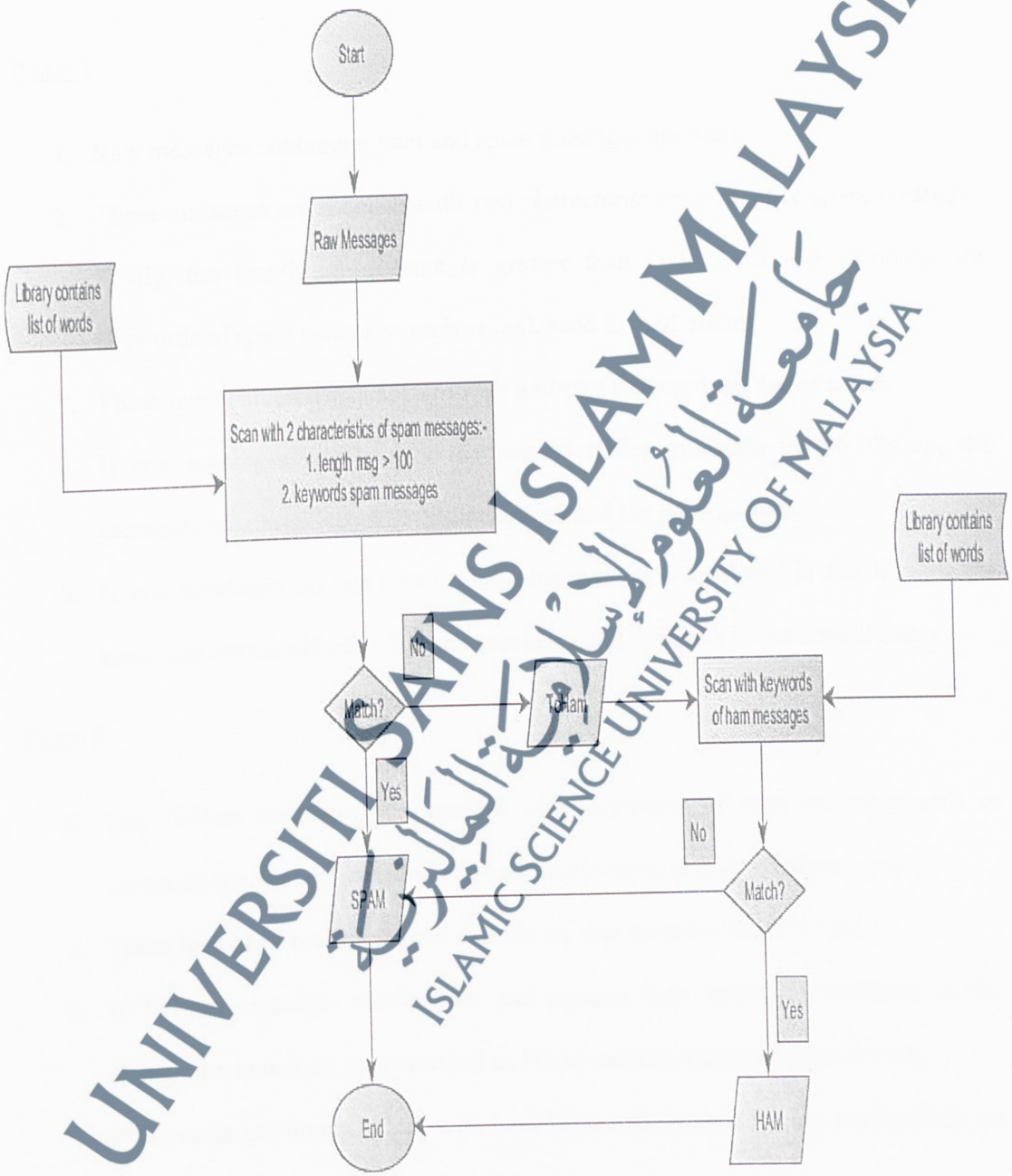


**Figure 4. 3:** Algorithm 2

Figure 4.3 shows the flowchart of Algorithm 2. This algorithm requires two phases and each phase is discussed as follows:-

Phase 1

1. Raw messages containing ham and spam messages are used.
2. These messages are scanned with two characteristics to identify spam messages. Firstly, the length of message is greater than 100 words and secondly, the keywords of spam messages such as text, send to, and claim.
3. These two characteristics are stored in a library that contains list of words.
4. If raw messages match with the characteristics available in the library, the messages are classified as SPAM messages and the process ends.
5. If raw messages do not match with characteristics available in the library, the messages are classified as ToHam messages and continue to the second phase.

Phase 2

6. The ToHam messages are scanned with keywords of ham messages such as pronouns (i.e. he, she, themselves) and interjections (i.e. hey, hmmm, yup).
7. These ham keywords are stored in a library that contains list of words.
8. If ToHam messages match with and contain ham keywords available in the library, the messages are classified as HAM messages and the process ends.
9. If the messages do not match with ham keywords available in the library, they are classified as SPAM messages and the process ends.

ALGORITHM 3

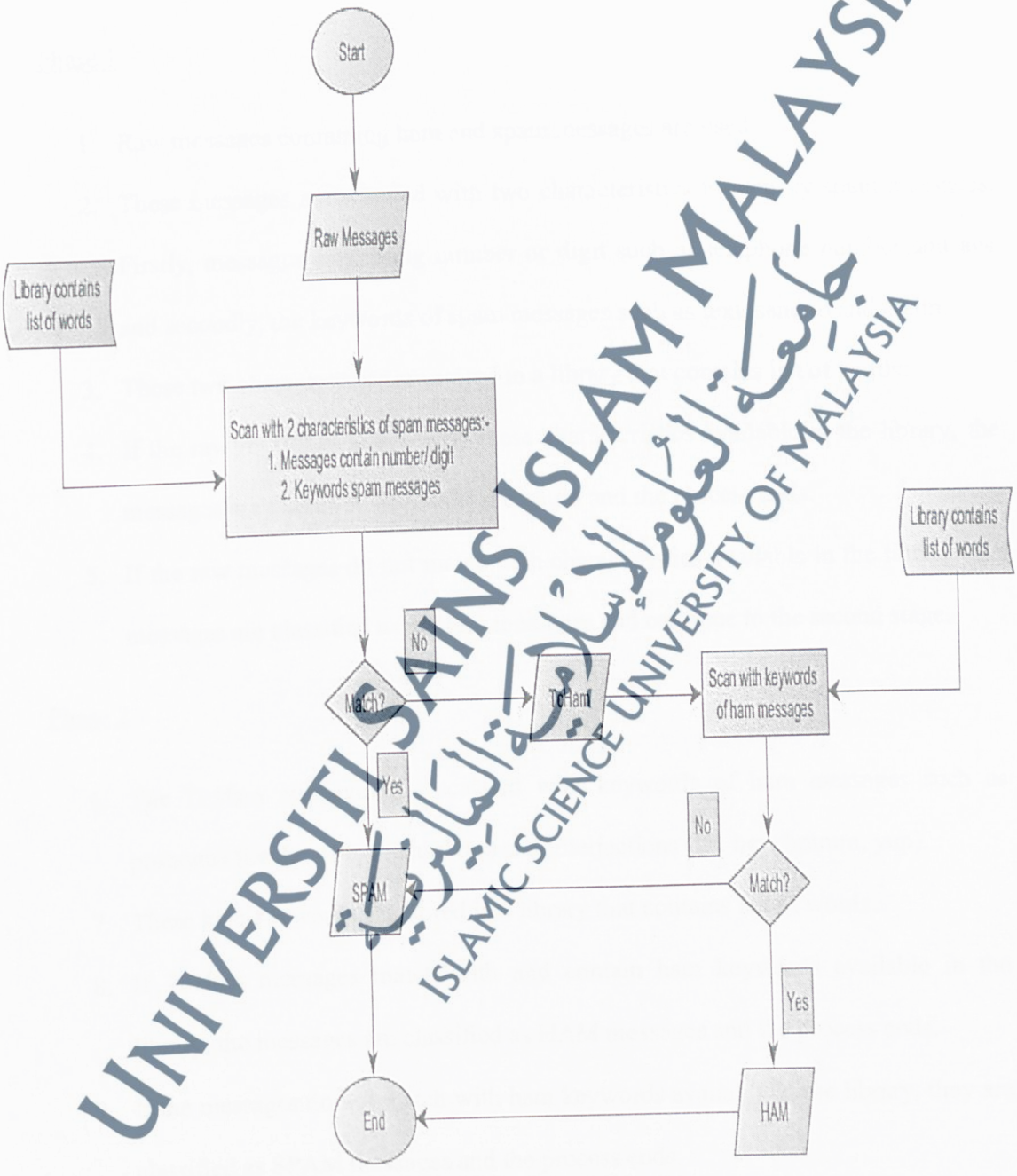


**Figure 4. 4:** Algorithm 3

Figure 4.4 shows the flowchart for Algorithm 3. This algorithm requires two phases and each phase is discussed as follows:-

Phase 1

1. Raw messages containing ham and spam messages are used.
2. These messages are scanned with two characteristics to identify spam messages. Firstly, messages containing number or digit such as telephone number and age and secondly, the keywords of spam messages such as text, send to and claim.
3. These two characteristics are stored in a library that contains list of words.
4. If the raw messages match with these characteristics available in the library, the messages are classified as SPAM messages and the process ends.
5. If the raw messages do not match with characteristics available in the library, the messages are classified as ToHam messages and continue to the second stage.

Phase 2

6. The ToHam messages are scanned with keywords of ham messages such as pronouns (i.e. he, she, themselves) and interjections (i.e. hey, hmmm, yup).
7. These ham keywords are stored in a library that contains list of words.
8. If ToHam messages match with and contain ham keywords available in the library, the messages are classified as HAM messages and the process ends.
9. If the messages do not match with ham keywords available in the library, they are classified as SPAM messages and the process ends.

ALGORITHM 4

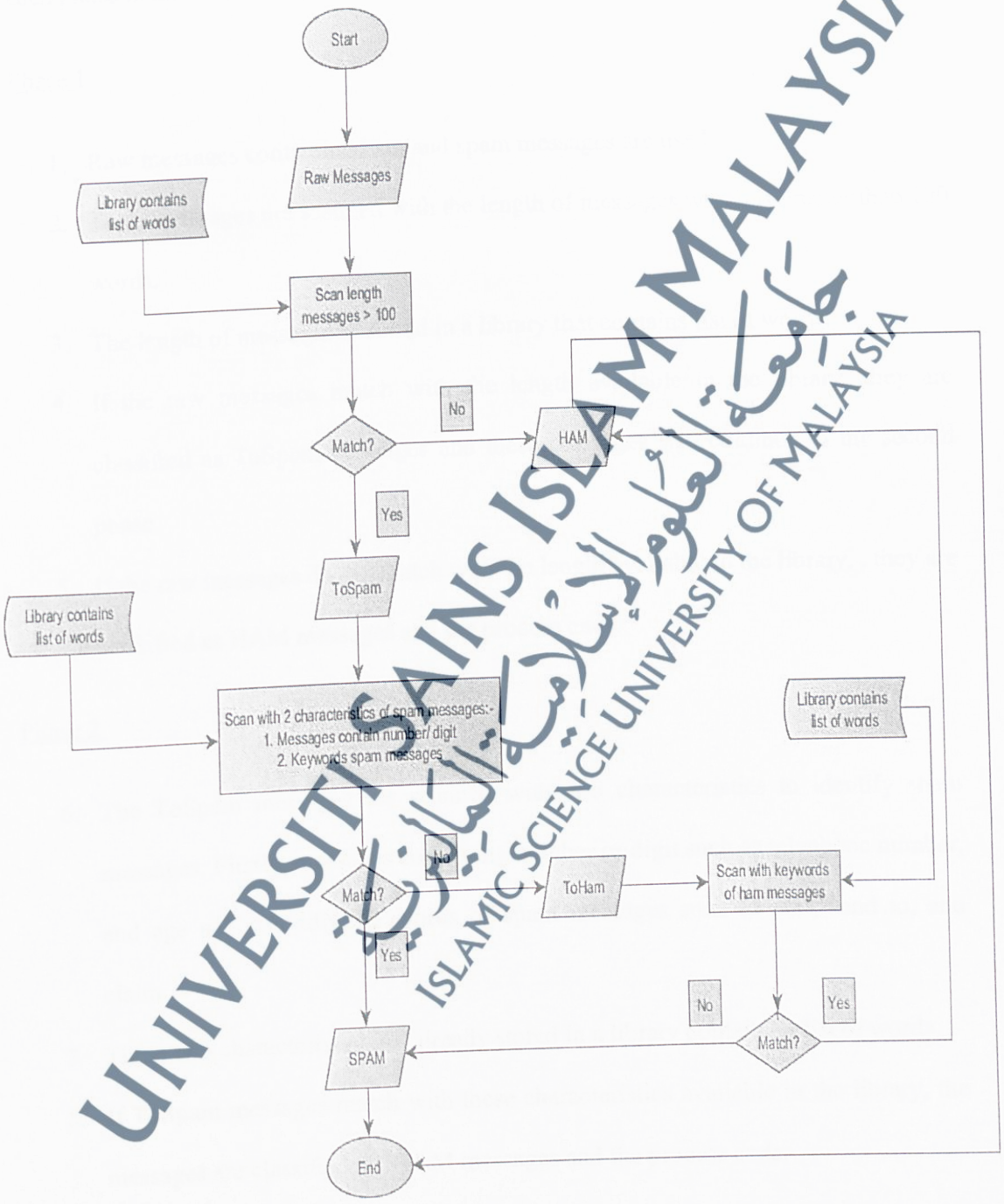




**Figure 4. 5:** Algorithm 4

Figure 4.5 shows the flowchart of Algorithm 4. This algorithm requires three phases and each phase is discussed as follows:-

Phase 1

1. Raw messages containing ham and spam messages are used.
2. These messages are scanned with the length of messages which is greater than 100 words.
3. The length of messages is stored in a library that contains list of words.
4. If the raw messages match with the length available in the library, they are classified as ToSpam messages and these messages will continue to the second phase.
5. If the raw messages do not match with the length available in the library, , they are classified as HAM messages and the process ends.

Phase 2

6. The ToSpam messages are scanned with two characteristics to identify spam messages. Firstly, messages containing number or digit such as telephone number, and age and secondly, keywords of spam messages such as text, send to, and claim.
7. These two characteristics are already stored in a library containing list of words.
8. If ToSpam messages match with these characteristics available in the library, the messages are classified as SPAM messages and the process ends.

9. If ToSpam messages do not match with these characteristics available in the library, the messages are classified as ToHam messages and continue to the third phase.

Phase 3

10. The ToHam messages are scanned with keywords of ham messages.
11. These keywords are stored in a library that contains list of words.
12. If ToHam messages match with the keywords of ham messages available in the library, the messages are classified as HAM messages and the process ends.
13. If ToHam messages do not match with the keywords of ham messages available in the library, they are classified as SPAM messages and the process ends.

ALGORITHM 5

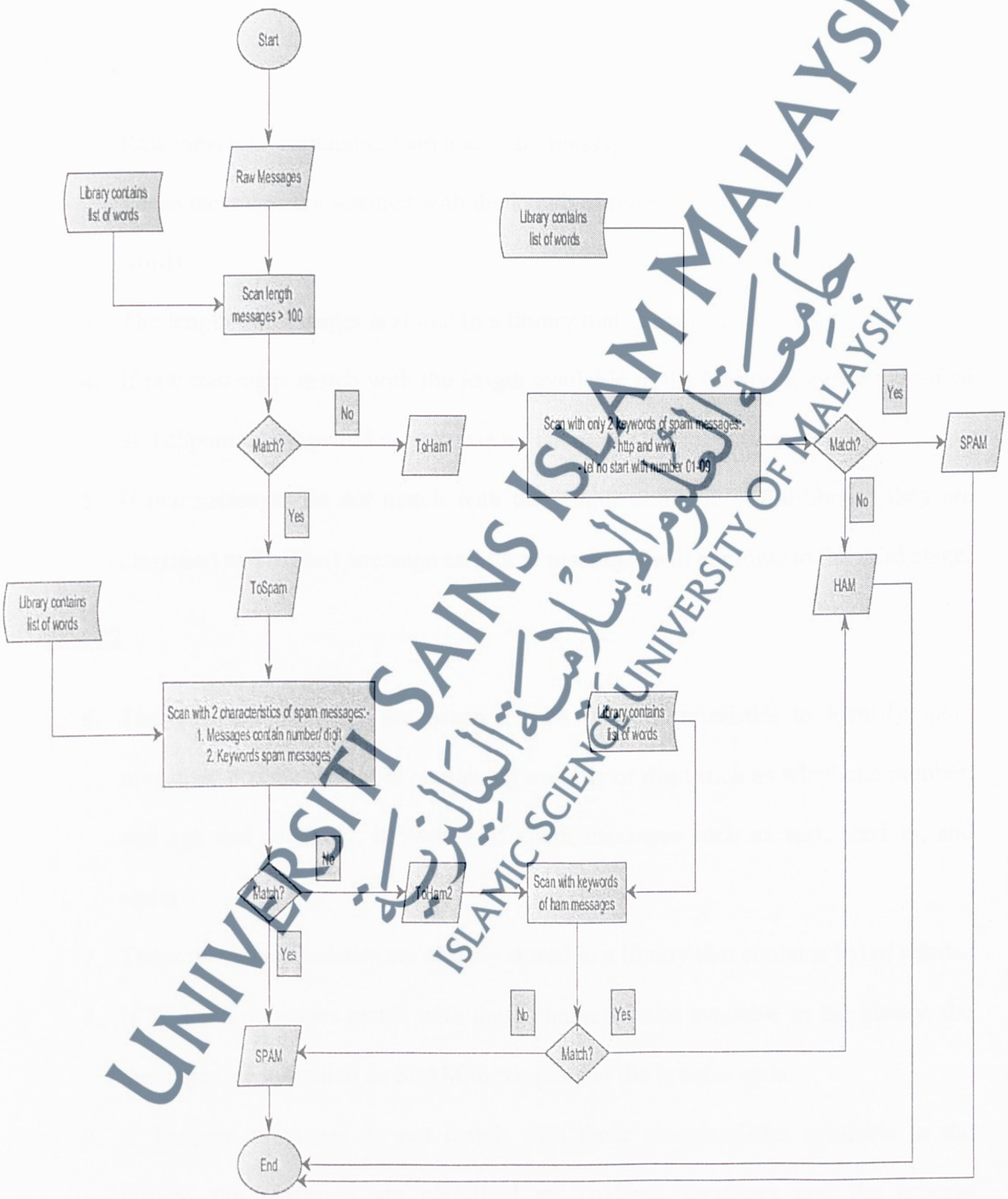


**Figure 4. 6:** Algorithm 5

Figure 4.6 shows the flowchart of Algorithm 5. This algorithm requires four phases and each phase is discussed as follows:-

Phase 1

1. Raw messages containing ham and spam messages are used.
2. These messages are scanned with the length of messages which is greater than 100 words.
3. The length of messages is stored in a library that contains list of words.
4. If raw messages match with the length available in the library, they are classified as ToSpam message and the process will continue to the second stage.
5. If raw messages do not match with the length available in the library, they are classified as ToHam1 message and these messages will continue to the third stage.

Phase 2

6. The ToSpam messages are scanned with two characteristics to identify spam messages Firstly, messages containing number or digit such as telephone number, and age and secondly, keywords of spam messages such as text, send to, and claim.
7. These two characteristics are already stored in a library that contains list of words.
8. If ToSpam messages match with these characteristics available in the library, the messages are classified as SPAM messages and the process ends.
9. If ToSpam messages do not match with these characteristics available in the library, the messages are classified as ToHam2 messages and the process continues to the fourth stage.

Phase3

10. The ToHam1 messages are scanned with only two keywords of spam messages. Firstly, the link of website and email such as http, www, @ and .com. Secondly is number of telephone that starts with 01-09. This stage chooses only two keywords of spam messages because most spammers ask users to dial the number provided or visit the link in order for them to steal the important information and get money. Besides, some spam messages used short size of messages for visiting website and dialling numbers.
11. These keywords are stored in a library that contains list of words.
12. If ToHam1 messages match with the spam keywords available in the library, the messages are classified as SPAM messages and the process ends.
13. If ToHam1 messages do not match with the spam keywords available in the library, they are classified as HAM messages and the process ends.

Phase 4

14. The ToHam2 messages in stage three are scanned with keywords of ham messages.
15. These keywords are stored in a library that contains list of words.
16. If ToHam2 messages match with the keywords of ham messages available in the library, the messages are classified as HAM messages and the process ends.
17. If ToHam2 messages do not match with the keywords of ham messages available in the library, they are classified as SPAM messages and the process ends.

## 4.3 EXPERIMENTS IN CLASSIFICATION (OR CLUSTERING) PHASE

Clustering is the process of grouping data or documents into similar groups. In this phase, clustering process only involves spam messages because we want to identify which category each spam message belongs to. There are various types of spam messages sent by spammers to attract users and most of us do not realize what the group of the messages is. Part of spam messages will make users interested and believe with their contents and part of them will be ignored by users. The aim of this phase is to cluster text spam messages into categories and identify which types of spam has higher number of messages so that we know the motive of spammer in sending the spam messages.

### 4.3.1 HYBRID IMMUNE CLONAL NETWORK ALGORITHM (HICNA)

A new algorithm named "Hybrid Immune Clonal Network Algorithm (HICNA)" is introduced. It combines the theory of Clonal Selection and Immune Network Theory for clustering spam messages. There is no previous research and paper introduced on method and algorithm for clustering spam messages using the combination of these two algorithms (Nazri et al., 2009; 2010). Table 4.6 shows the mapping component between AIS with proposed algorithm while Figure 4.7 shows the algorithm for HICNA. The explanation of the algorithm shows in Table 4.7. The detail process of HICNA is shown in Figure 4.8 with its explanation of the process is shown in Table 4.8.

**Table 4. 6:** Mapping component between AIS and HICNA

| AIS component | HICNA |
|---|---|
| Pattern,S | SMS spam messages |
| Set Antibodies | Library |
| Affinity | keywords |
| Clones | Cluster of spam messages |

Input: S = set of patterns to be recognised, $n$ = the number of worst elements to select for removal, nt = network affinity threshold, p1=common keywords, p2=uncommon keywords, p3=expert judgement

Output: M = Final set of memory detectors capable of classifying unseen patterns (i.e. identified cluster), N= set of unknown memory detectors capable of classifying unseen patterns (i.e. Cluster Miscellaneous), O= set of undefined memory detectors capable of classifying unseen patterns (i.e. Cluster Others),

1. **Begin**
2. Create an initial random set of antibodies, A
3. P1 : Determine the affinity of S with each antibody in *A*
4. Generate clones of a subset of the antibodies in *A* with the highest affinity.
5. The number of clones for an antibody is proportional to its affinity
6. If affinity for each element in A is higher than nt, clones are generated and move to M
7. (i.e. identified cluster).
8. Mutate attributes of these clones to the set *A* , and place a
9. copy of the highest affinity antibodies in *A* into the memory
10. set, *M (i.e. identified cluster)*
11. else eliminate the *n* lowest affinity antibodies in *A into N (i.e. cluster Miscellaneous)*
12. *New randomly antibodies are generated in A.*
13. ***then***
14. P2: Determine the affinity of N with each antibody in A
15. If affinity for each element in A is higher than nt, clones are generated and move to M
16. *(i.e. identified cluster).*
17. Mutate attributes of these clones to the set *A* , and place a
18. copy of the highest affinity antibodies in *M* into the memory
19. set, *M*(i.e. identified cluster).
20. else the *n* lowest affinity antibodies in *A are eliminated into O (i.e. cluster Other)*
21. **then**
22. P3: Determine the affinity of O with FP using expert judgement
23. If the affinity is higher than nt, clones are generated and move to M
24. else the *n* lowest affinity *are eliminated into O*
25. **end**

**Figure 4. 7:** Algorithm for HICNA

**Table 4. 7:** Explanation of the algorithm

| No | Activity |
|---|---|
| 1 | The process starts. |
| 2 | Creates an initial random set of network antibodies, A. |
| 3 | In phase 1, patterns, S binds with each antibody in A to determine the affinity (i.e. spam messages are scanned using common keywords). |
| 4 | Clones are generated for subset of the antibodies in A if the affinity is highest (i.e. if SMS messages match with common keywords available, they will be clustered). |
| 5 | The number of clones for an antibody is proportional to its affinity. |
| 6,7 | If the affinity for each element in A is higher than nt, clones are generated and move to M (identified cluster). |
| 8,9,10 | Mutate attributes of these clones to the set A, and place a copy of the highest affinity clones into a set M. |
| 11 | Eliminates the n lowest affinity antibodies into N (i.e. cluster Miscellaneous) if the affinity is less than nt. |
| 12 | New randomly antibodies are generated in A. |
| 13 | The process continue. |
| 14 | In phase 2, patterns, S in M binds with each antibody in A to determine the affinity (i.e. spam messages are scanned using uncommon keywords). |
| 15, 16 | If the affinity for each element in A is higher than nt, clones are generated and move to M (identified cluster). |
| 17,18,19 | Mutate attributes of these clones to the set A, and place a copy of the highest affinity clones into a set M. |
| 20 | Eliminates the n lowest affinity antibodies into O (i.e. cluster Others) if the affinity is less than nt. |
| 21 | The process continue. |
| 22 | In phase 3, patterns, S in O with FP value are determined their affinity using expert judgement. |
| 23 | If the affinity for each element in A is higher than nt, clones are generated and move to M (identified cluster). |
| 24 | Eliminates the n lowest affinity antibodies into O (i.e. cluster Others) if the affinity is less than nt. |
| 25 | The process end. |

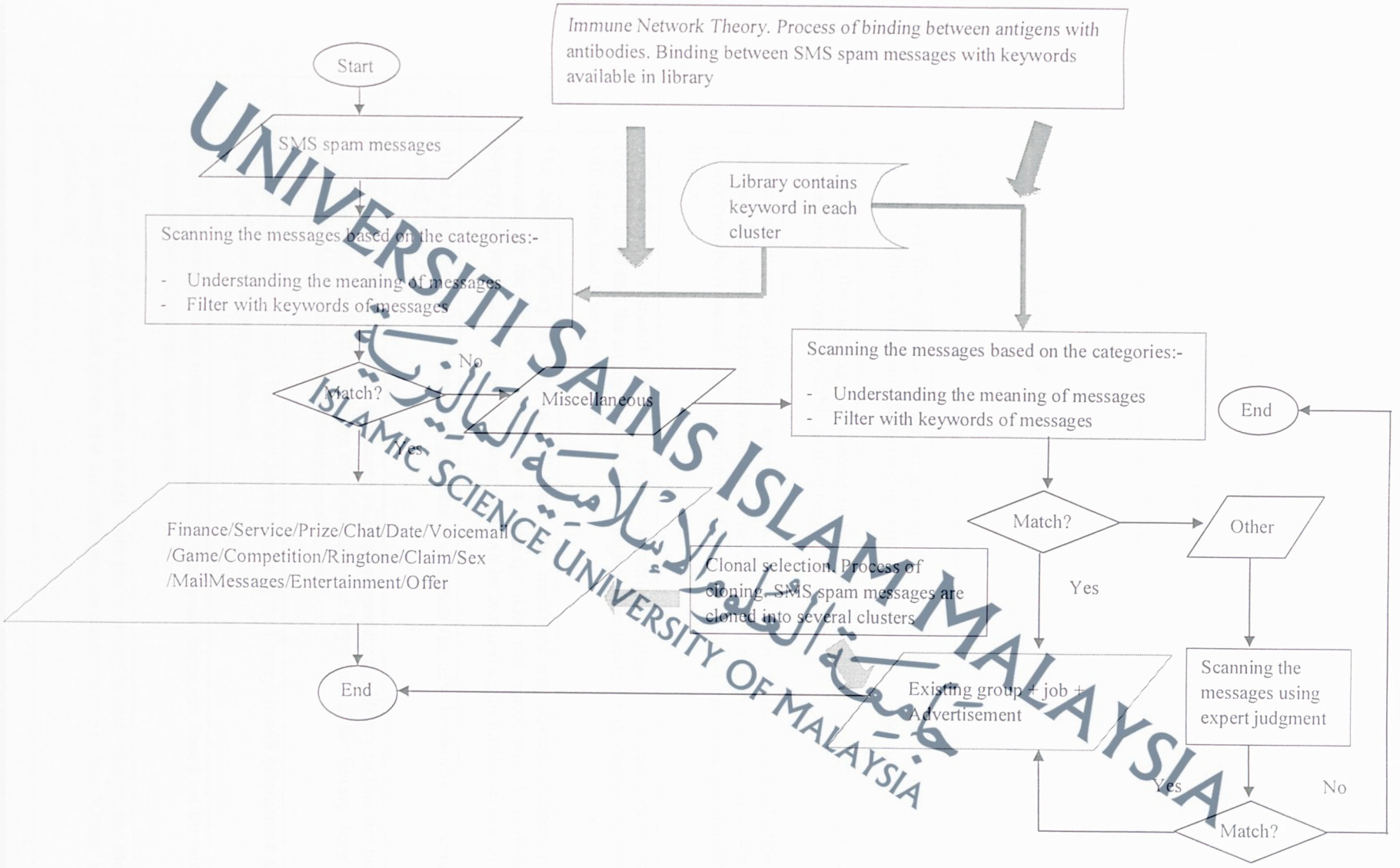


**Figure 4. 8:** Hybrid Immune Clonal Network Algorithm

Figure 4.8 shows the proposed algorithm for the process of clustering SMS spam messages using Clonal Selection and Immune Network Theory. The process is discussed below. Then the summary of the process in HICNA with AIS theory is shown in Table 4.9.

**Table 4. 8:** The explanation of the process in HICNA

| No | Activity |
|---|---|
| 1 | SMS spam messages act as an antigen or foreign microorganisms such as virus and bacterial will be scanned and destroyed by antibodies which is the library. |
| 2 | In the library, the meaning and content of the messages will be scanned and filtered with the common keywords available in the library. |
| 3 | The keywords are the affinity of binding between SMS spam messages with the library. If a message has similar keywords available in the library or a message has closer meaning with the keywords, it means that the affinity of binding between the messages and library is higher. |
| 4 | In the Immune Network Theory, if epitope in antigen matches with paratope in antibody, they will bind and the strength of affinity is higher. The binding will destroy the antigen or will divide into other types of cell. |
| 5 | The higher affinity of binding will clone the messages into several clusters of spam messages. If the affinity is lower (i.e. a message does not match with keywords in the library), the messages will not be cloned and will be scanned again for the second phase. |
| 6 | At the same time, the library will generate and update the new keywords for scanning in second phase. |
| 7 | In the second phase, the spam messages will be scanned by the library to find the matching uncommon keywords. If a message contains the same keywords in the library, they will be bound and cloned into existing cluster or new cluster. |
| 8 | If a message does not have the same keywords as in the library, it will be scanned again in third phase using expert judgement. |
| 9 | Expert judgement is the process of scanning the spam messages based on the view and understanding the content of messages. |
| 10 | If the process of expert judgment is match with the meaning in each cluster, the messages are clustered into defined group, else the messages are clustered into cluster 'Other' and the process end. |

| | |
|---|---|
| 11 | In Clonal Selection, the cell will divide into two types of cell which are plasma cell and memory cell. Memory cell will remember the structure and characteristics of antigen when the second attack occurs and the plasma cell will fight the antigen again. |
| 12 | In our proposed clustering process, the function of memory cell is located in the first process of scanning because when messages are scanned with the keywords in library and then they match and go into the correct cluster, it means that the library remembers the characteristics of the messages. |
| 13 | Plasma cell occurs when the SMS spam messages cannot match with the keywords available in the library and they will go to the second phase for the second process. |
| 14 | New style and updated of SMS spam messages cause the library to generate new keywords, that is why our proposed model is needed for the second phase of clustering process. |
| 15 | Expert judgement is needed to ensure all messages are able to be clustered. |

**Table 4. 9**:Hybrid Immune Clonal Network Algorithm with AIS understanding theory

| Process | Description | AIS Understanding Theory |
|---|---|---|
| **Message Reading** | Messages are read from the server or users' mobile phone. | Antigens or foreign microorganisms such as virus and bacteria enter the body. |
| **Messages Scanning into three phases:-**<br>**P1 = Scanning using common keywords**<br>**P2 = Scanning using uncommon keywords**<br>**P3 = Using expert judgement** | Messages are scanned with keywords available for each group. The list of keywords located in the library contains word-list. There are three phases in scanning the messages. The keywords are the affinity of binding between SMS spam with the library and the library is an antibody. | Antigen will bind with antibody with the help of epitope and paratope. If epitope in antigen matches with paratope in antibody, they will bind and the strength of affinity is higher. The binding will destroy the antigen or will clone into two types of cell which are plasma cell and memory cell. If the affinity is lower, the antigen will be bound again with another antibody. The binding of antigen with antibody is the process of Immune Network Theory. If the binding occurs and produces higher affinity, the cell will be cloned into two types of cell which are plasma cell and memory cell. The process of cloning is using the theory of Clonal Selection. The function of memory cell is located in the first process of scanning because when a message is scanned with the keywords in library and then it matches and goes into the correct cluster, it means that the library remembers the characteristics of the message. Plasma cell occurs when the SMS spam messages cannot match with the keywords available in the library and they will go to the second phase for the second process. Memory cell also occurs in the phase 3 when expert judgement is needed to remember the meaning of message in order to group into the correct cluster. |
| | **Phase 1**: Scanning using common keywords available in the library. If any messages match with the keyword, then the message will go to the correct group. If it does not match for any group, then it will scan again into the second phase. At the same time, library will update and generate new keywords because of new style and format of SMS spam. This library will be used for scanning in second phase. | |
| | **Phase 2**: In the second phase, the SMS spam will be scanned by the library to find the matching uncommon keywords. If a message contains the same keywords in the library, they will be bound and cloned into existing cluster or new cluster. If the matching does not occur, the messages will be eliminated and the process ends. | |
| | **Phase 3**: Expert judgement is needed to scan the group that contains False positive (FP) results (i.e. messages that are incorrectly classified) and the messages that cannot be grouped from phase two (i.e. cluster Other). The purpose of this phase is to make sure that all messages are correctly grouped. | |
| **Print** | Number of messages in their defined category. | Number of cells have been cloned. |

### 4.3.2 CLUSTERS OF SPAM

Delany et al., (2012) introduced ten (10) clusters of spam text messages and each cluster contains its own keywords and meaning. Their finding helps this research to identify what cluster and category of spam messages are available. From our analysis, we have found that there are limitations in terms of keywords, where redundancy of clusters identified in which it was found that the definition of each cluster is unclear. For example, keywords in the cluster ringtones are "send', "text", "free", "SMS" and "reply". Therefore, spam messages which supposed to be grouped into cluster prizes can also be grouped into cluster ringtone due to the redundancy of keywords used. Having identified this limitation, we enhanced new clusters using different keywords for each cluster and add another cluster category. Table 4.10 shows the comparison of clusters between Delany et al., (2012) with the proposed clusters and their meaning.

**Table 4. 10:**Comparison of clusters available between Delaney et al., (2012) with proposed clusters.

| No | Delany et al., (2012) | Proposed clusters | Meaning |
|---|---|---|---|
| 1 | Finance | Finance | Msg that involve activity of money like payment of bills, loans and card credit . |
| 2 | Service | Service | Msg that provide any service for phone and service to upgrade, and service for order. |
| 3 | Prize | Prize | Msg telling users that they win and selected to get prize and money. |
| 4 | Competition | Game/ Competition | Msg that provide quiz and game and user need to answer or take part in order to get the prize. |
| 5 | Ringtone | Ringtone | Msg offering ringtone and asking user to download the ringtone. |
| 6 | Chat | Chat | Msg that provide chatting service, msg asking to be a friend. |
| 7 | Date | Date | Msg that provide a dating service or msg asking for the date. |
| 8 | Voicemail | Voicemail | There are messages or urgent messages in voicemail and user need to dial or call. |
| 9 | Claim | Claim | Msg that tells the recipients to claim compensation for accidents. |
| 10 | Miscellaneous | Other | Msg that cannot be categorized. |
| 11 | | Mail messages | There are messages in mailbox. |
| 12 | | Entertainment | Msg that provide services related to entertainment like video, music, picture and others. |
| 13 | | Advertisement | Msg that advertise sale, new or any activities. |
| 14 | | Offer | Msg that offer applications, gadget or voucher and free call with half price or by activating the offer and registering. |
| 15 | | Sex | Msg that offer sex service and related to the sex. |
| 16 | | Job | Msg that provide employment service, companies who is looking employee and msg asking for resume. |

The limitation of cluster in Delany et al.,(2012) inspired us to add new types of cluster and retaining the existing cluster. From the Table 4.10, another 6 clusters have been added that are Mail messages, Entertainment, Advertisement, Offer, Sex and Job. The adding of 6 clusters with the proposed cluster for this research was based on basic frequency-based term selection to identify and select frequent terms occurring in each message as to identify the group or cluster based on the term. Besides, by doing our own

review towards the meaning of each message, it helps in developing the cluster of spam. We made a test to compare results between Delany et al., (2012) with our proposed cluster using the same dataset named BEC. There are different numbers of spam messages in each cluster between both of them as shown in Table 4.11.

**Table 4. 11:**Results comparing Delany et al., (2012) with the proposed method

| Category | Delany et al., (2012) | | Proposed cluster | |
|---|---|---|---|---|
| | TP | FP | TP | FP |
| Finance | 39 | - | 22 | - |
| Prize | 70 | - | 59 | 5 |
| Service | 136 | - | 18 | - |
| Game/Competition | 76 | - | 34 | - |
| Ringtone | 25 | - | 44 | 3 |
| Chat | 27 | - | 48 | 3 |
| Date | 25 | - | 10 | 1 |
| Voicemail | 18 | - | 8 | - |
| Claim | 1 | - | 2 | - |
| Advertisement | - | - | 91 | - |
| Entertainment | - | - | 6 | - |
| Sex | - | - | 30 | 3 |
| Offer | - | - | 36 | - |
| Mail message | - | - | 2 | - |
| Job | - | - | - | - |
| Other | 8 | - | 91 | - |

Table 4.11 shows the results between Delany et al., (2012) with the proposed clusters. TP or True Positive means the SMS spam messages are correctly classified in the right cluster while FP or False Positive means the messages are wrongly classified to the actual cluster. From Table 4.11, Delany et al., (2012) resulted with no FP as compared to the proposed method. There are reasons why there are significant differences between both results, as explained in following,

a. The meaning of each cluster for Delaney et al., (2012) is different with the proposed method, thus we cannot identify whether all messages in each cluster is correctly classified or not.

b. Keywords that are used in Delaney et al., (2012) for each cluster is different with keywords in the proposed method. Thus the result is different.

c. Delaney et al., (2012) contained only 10 clusters while the proposed method contains 16 clusters.

### 4.3.3 CLASSIFICATION USING HICNA

The process of classification using HICNA involves three phases as shown in Figure 4.9


**Figure 4. 9:** Process involved in HICNA

Figure 4.9 shows three process involved in clustering spam messages. In the phase one, we cluster the spam messages using common keywords. Common keywords refer to the

messages that have familiar and frequent keywords across different datasets. Phase two uses uncommon keywords to cluster updated spam messages. Uncommon keywords are the keywords that are new and not familiar. Every year, spammers will update their style of spam messages by using new contents and term so that they can avoid any software to filter spam messages. By having this phase, it can cluster the messages although new versions of spam messages are produced. The last phase is for the expert judgement process and this is the final phase for the clustering process. Expert judgment is needed when the spam messages are incorrectly classified into the correct cluster. In this phase, we ask opinion about the meaning of messages from the experts from various fields. Examples of common and uncommon keywords are in Appendix C.

Five different datasets are used in this phase (i.e. DIT, UCI, BEC, SMSv.0.1 and FadhilahSpam) and each of them only contains spam messages. The explanation of each dataset is discussed below.

a. UCI Machine Learning (UCI)

   This dataset has been discussed in Section 4.2.1.

b. British English SMS Corpora (BEC)

   This dataset has been discussed in Section 4.2.2

c. Dublin Institute of Technology (DIT).

   This dataset has been discussed in Section 4.2.1.

d. SMSv.0.1

A set of SMS tagged messages that have been collected for SMS spam research. It contains to collections of SMS messages in English, tagged according being legitimate (ham) or spam.

e. FadhilahSpam dataset

This dataset contains the collection of spam messages from various sources (i.e. DIT, UCI, BEC and SMSv.0.1). The purpose of this dataset is to gather all spam messages due to limitation of the messages (i.e. making number of spam messages larger) besides for verification process in clustering for this research.

Table 4.12 shows the number of spam messages in each dataset. Each of the spam messages is required for cleaning process to remove the unimportant information like telephone number of the sender and types of telephone used because we only need the text messages. However, punctuation marks or any symbols will still remain in the text messages as they do not affect clustering process.

Table 4.12 :Number of spam messages in each dataset

| | Number of spam messages |
|---|---|
| Dataset A (DIT) | 1353 |
| Dataset B (UCI) | 747 |
| Dataset C (BEC) | 425 |
| Dataset D (SMSv.0.1) | 322 |
| FadhilahSpam | 2847 |

The measurement of clustering performance is based on True Positive (TP) and False Positive (TN).

- True Positive (TP): Spam messages are correctly clustered into identified group.
- False Positive (FP): Spam messages are incorrectly clustered into identified group.

## 4.4 DETECTION AND CLUSTERING USING WEKA

WEKA is a collection of machine learning algorithms for data mining tasks (Witten & Frank, 2005). The algorithms can either be applied directly to a dataset or called from your own Java code. WEKA stands for Waikato Environment for Knowledge Learning and it was developed by the University of Waikato, New Zealand. WEKA contains tools for data pre-processing, classification, regression, clustering, association rules and visualization. The supported data formats in WEKA are ARFF, CSV, C4.5 and binary. Figure 4.10 shows the front view of WEKA tool.


**Figure 4. 10:** Front view of WEKA tool

WEKA can be used in this research for the first phase (i.e. detection) and second phase (i.e. classification) for IMSM. The purpose of using WEKA is to validate the dataset used in this experiment besides it can help to determine which classifying algorithms or clustering algorithms are good in performance.

In WEKA, detection is known as classification. It is a supervised learning and requires training data and testing data. The training set contains data that have been previously categorized. Hence classification is used to predict the target that must be categorical. Some classifying algorithms available in WEKA are Bayes, Functions, Lazy, Meta, Misc, Rules and Trees (Kumar and Chatterjee, 2016).

Clustering is unsupervised learning and it is a process of making group data into similar classes. Cluster is a group of objects that belong to the same class. Canopy, CLOPE, Cobweb, EM, FarthestFirst, FilteredClusterer, HierarchicalClusterer, LVQ, MakeDensityBasedCluster and SimpleKMeans are example of clusterer algorithms in WEKA.

## 4.5 SUMMARY

This chapter discussed methods that are used in detection and classification phases. In the first phase of spam management model which is detection, the Danger Theory and Negative Selection are used to filter messages into spam or ham using WEKA. Another proposed method for detection is by using three features; length of messages, messages containing special characters and keywords. These features are proposed as the improvement for performance of Danger Theory. As the research focuses in the second phase which is classification, a new algorithm is proposed by combining the Clonal Selection and Immune Network Theory named Hybrid Immune Clonal Network Algorithm (HICNA). A number of dataset were used for testing and validating the proposed method.