An Immunological-Based Simulation: A Case Study of Risk Concentration for Mobile Spam Context Assessment

Kamahazira Zainal[#], Mohd Zalisham Jali[#]

[#] *Faculty of Science and Technology, Universiti Sains Islam Malaysia, 71800 Nilai, Negeri Sembilan, Malaysia.*
E-mail: arizah78@yahoo.com; zalisham@usim.edu.my

***Abstract* - Over the past two decades, there has been a substantial increase in spam messages that caused critical impact loss. Besides the factor of integration of Internet and mobile technology, this issue is also due to the human's reaction towards spam. This paper presents RiCCA or Risk Concentration for Context Assessment model that performs a risk classification of text spam messages in Short Message Service (SMS) format. The identified risk levels will assist users in anticipating the potential impact of the spam message that they have been receiving. Danger Theory, a prominent theory from Artificial Immune Systems (AIS), inspires the developed model. During the simulation phase, an immunological-based testing lifecycle is applied, with the deployment of the dataset that is shared at UCI Machine Learning Repository and self-collected messages. The performance of the testing revealed a distinctive result, which more than 80% of true positive rate is achieved, employed with two variants algorithm from the Danger Theory; Dendritic Cell Algorithm (DCA) and Deterministic Dendritic Cell Algorithm (dDCA). This simulation demonstrated that the Danger Theory as a feasible model to be applied in measuring the risk of spam. The further articulation on how this immunological-based testing lifecycle is applied in computer simulation and adopting mobile spam as the case study is clarified thoroughly.**

***Keywords*— danger theory; risk classification; SMS spam messages; mobile spam; knowledge discovery; information retrieval; immunological simulation.**

I. INTRODUCTION

Biological Immune System or BIS can be defined as the mechanism on how human body identifies and protect itself against foreign cells or substances that could bring harm to the body. In BIS, white blood cells are the well-known and important components in defending the body against foreign cells that include bacteria and viruses. Some of the defense mechanism may directly attack foreign substances in the body, and some others collaborate with other body cells to initiate the immune response [1]. The BIS is divided into three (3) main layers, which is including anatomic barrier, innate immunity, and adaptive immunity. These layers provide different types of defense mechanisms for detection, recognition, and responses to foreign or dangerous substances [2].

Theory inspired by this human body immunology and developed artificially, or Artificial Immune Systems (AIS) is the designation of signified signal processing and self-adapting system. Since its emergence on 1990's, AIS has become known as a new branch of computational intelligence. It gets significant success in the field of computational intelligence. Numerous evidence of success in various research applying AIS has been recorded and this theory keeps expanding in its understanding, immunologically and computationally. The employment of AIS in computational intelligence includes spam classification, virus detection [3], anomaly detection [4], intrusion detection [5], [6] and optimization. In this paper, application of one of the eminent theory from AIS, Danger Theory as an approach for classifying the risk of spam is discussed.

The rise of spam has evolved into element that is unprecedented such as instant messenger services, fax to email services, Voice over Internet Protocol (VoIP), mobile and smartphones, social networks and mobile instant messenger applications [7]. Spammers are easily adapted to use the available technology to reap the potential illicit revenue. Even though there are many mechanisms has been applied to curb this threat, its adverse effect is still kept rising persistently, with no sign to be any lesser.

The vast proliferation and advancement of mobile devices have attracted the attention of cyber-criminals, who exploited the functionality of the device for malevolent purposes. In 2013, an annual threat report by Cloudmark showed an alarming situation in the United States of America (USA) and United Kingdom (UK) which already expand its effects throughout the world [8]. The advancement of mobile devices such as a smartphone that is integrated with Internet technology has made the usefulness of Short Message Service or SMS has become more extensive. The link provided in SMS is easily clickable and accessible online

using smartphone. Through this facility, users are prone to cybercrime activities.

The paper focuses on SMS spam issue since it has been intruding in our daily life that also has caused severe losses. This threat has become more mature and even sophisticated regarding its form in spam dissemination [7] and its unpleasant impact [9], which can be affected by many aspects such as money loss, discredit of reputation, and even time wasting. These adverse effects are borne by individuals, organizations and also governments [7]. It also corresponds to the innovation in mobile technology, booming together with the Internet advancement [10]. Many reports have been documented due to this threat's impact loss, and its advent seems unending [8], even getting almost stealth and resistance from any mechanism of anti-spam solutions. In addition to the factors above, human behavior also has been identified in increasing the loss due to their trust and interested with the spam's contents [11], [12] and unawareness of the impact [13]. A well-known hacker, Kevin Mitnick suggested that human is the common factors that caused security breach because a human has been identified as the weakest link in security chain [14]. Due to these reasons, they are secure is exploited by spammers.

On top of that, the study for SMS spam classification is still small. In 2013, it was pointed out that the SMS and Twitter usage globally makes 37.83% and 0.16% respectively [15]. However, the research to study SMS only accounts for 14.29%, while 75% for Twitter. This statistic somehow showed that there are still lacking figure in studying SMS even though the utilization of the services is high. Hence, there is a critical need to help users in curbing against the SMS spam by giving out the awareness by exposing the potential harm that could cause by SMS spam messages.

In addition to that, most studies in the field of managing spam have only focused on spam filtering or differentiating between spam and valid messages [16]–[18]. However, less attention has been paid to measuring the risk that could cause by this spam. This is unlikely occurred in intrusion detection area in which there are numerous studies in measuring the risks and ranked the identified intrusion for further action [19], [20]. Due to that, this paper intended to focus on assessing the potential risk that may be unforeseen in the spam message. The focus of this paper is to develop a solution in prescriptive mode; predict a potential risk in a spam message that could be an implicit decision maker for users. With the assistance from such tool, users should be able to react and respond wisely once they realized the unforeseen impact that possibly could cause harm.

This paper is outlined in four (4) main sections. Section II elaborates insight of the previous literature that has been conducted in this specific study. This section covers both on the AIS application in computational intelligence and the issue of spam that has been threatening users around the globe. The immunological-based simulation that inspired this study specifically in assessing the risk concentration for a spam message also articulated with its proposed algorithm. Subsequently, the experimental setup to implement the study is described. Section III discusses all the results and findings of the simulation. Finally, the conclusion and potential future works related to this study are described in Section IV.

II. Material and Method

A. Danger Theory as the Approach

1) Abstraction View of the Danger Theory: Danger Theory is the most recent development in AIS. Polly Matzinger found this theory in 1994, which focuses on what is dangerous instead of self and non-self discrimination idea. It implies that the immune system can discriminate between danger and non-danger [21]. Danger Theory suggests that foreign invaders that are dangerous stimulate the production of cellular molecules (danger signals) by commencing cellular stress or cell death [22]. The prominent Danger Theory that has been established is the Dendritic Cell Algorithm (DCA), which is signal processing algorithm that is inspired by the behavior of dendritic cells [23].

Dendritic cells or DCs are the fundamental possession of Danger Theory, which also is an innate immune system, and indeed an intrusion or anomaly detection agent in the human body. These DCs are Antigen Presenting Cell (APC) that responsible to digest antigen material and forward it on the cell surface to the T-cells of the immune system. It is acting as messengers between the innate and the adaptive immune systems.

These DCs are possessing in the body tissues and gather antigen and other (danger) signals that give a picture of the current condition of the tissues. It is representing the current state and figures out if the antigen has been gathered in a safe or dangerous context, and causes DCs to change into a semi-mature or mature state. The mission of the DCs is to distinguish antigens as being either benign (harmless) or malignant (harmful) in nature [24]. Dendritic Cell Algorithm (DCA) is a prominent algorithm emerged from Danger Theory, and the entire process is depicted as in Fig. 1.

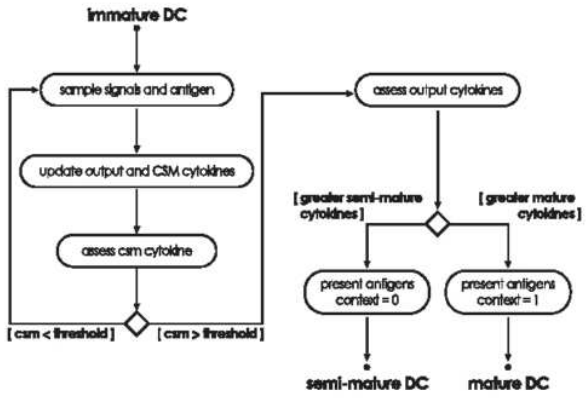

Fig. 1 Activity diagram is representing the key features of DC biologically. The processes on the left occur in the tissue, on the right on the lymph node [24]

It is emphasized that within the biological systems, Pathogen Associated Molecular Patterns (PAMPs) are molecules released exclusively by pathogens [25]. Then, danger signals are released from tissue cells following unplanned necrotic cell death, while safe signals are released from frequently dying cells as an indicator of healthy tissue. Inflammation is classed as the molecules of an inflammatory response to tissue injury.

The signals that migrated to the lymph node are divided into two (2) types of signals as regards to the degree of concentration of danger detected. Apoptotic alerts or semi-

mature brings the safe signal, while necrotic alerts bring the mature signal. Semi-mature indicate a 'safe' context and mature indicate a 'dangerous' context. These signals are a reflection of the state of the surrounding.

It is elaborated that the generated output signals are at a certain concentration and proportional to the received input signals [24]. The calculation can be measured as in (1).

$$Oj = \sum_{i=0}^{2}(Wij \times Si) \quad (1)$$

where O*j* is the output signals, S*i* is the input signals and W*ij* is the transforming weight from S*i* to O*j*.

The collected data of antigen by DCs is measured using the Mature Context Antigen Value (MCAV) which this is the mean value of context per antigen type, or in another word, determine the intensity or degree of the detected danger [26]. This can be measured by (2):

$$MCAV\ (antigen\ type) = \frac{mature\_count}{antigen\_count} \quad (2)$$

Throughout time, DCA is keep evolved with numerous versions that is applied in the various field. In this case study, the original version of DCA and its initial evolved version, deterministic DCA (dDCA) are employed as the classifier. Both of these algorithms have previously verified as suitable and feasible for classification task [24]. Besides that, the developed version other than DCA and dDCA has loads of stochastic decisions (randomly determined), complex algorithms [27] and only feasible for another task such as for robotics problem area [28].

The calculation scheme for the context assessment to determine its anomalous level is a bit different between DCA and dDCA. In the dDCA, the anomaly metric, K_α is implemented, and the magnitude of k value is used. This generates real-valued anomaly scores and may assist in the polarization of normal and anomalous processes [29]. The process of calculating this anomaly score is shown in (3), where k_m is the k value for DC_m, α_m is the number of antigens presented of type α by DC_m.

$$K_\alpha = \frac{\sum_m k_m}{\sum_m \alpha_m} \quad (3)$$

Other than that, there is a minor differentiation in translating the value of anomalous level in dDCA. The outcome of K_α in dDCA is tagged as anomalous when it is returned as a positive value, $K_\alpha > 0$ and tagged as normal when the returned value is negative, $K_\alpha < 0$ [30].

2) Applications of Danger Theory in Computational Intelligence: Since its recognition in the world of research, many works have been done, and AIS is proved to be appropriate and suitable to solve many real-world problems. The characteristics of an immune system adapted in AIS for instance; pattern recognition, learning, memory, and self-organization has built numerous solution by establishing mathematical and computational modeling [31]. A useful resource for the latest developments in AIS also can be found via a web of International Conference on Artificial Immune Systems, ICARIS [32].

Intrusion detection [20], [33] is the primary field of AIS employment. However, it is not limited to this area only, and AIS has been considerably applied in various other fields such as malware detection [34], classification [35] and optimization [36].

3) Dendritic Cell Algorithm in Spam Classification: Some works of other researchers for applying DCA in spam classification can be referred as to guide this study. It is proposed that a predictive model classify email spam based on DCA [37]. They used three (3) different machine learning algorithms to produce the input signals, which are k-Nearest Neighbour (kNN) to produce PAMP signal, Naïve Bayesian (NB) to produce danger signal and Support Vector Machine (SVM) to produce a safe signal. They analyzed the email header and message body to extract the related features. Term Frequency-Inverse Document Frequency (TF-IDF) is utilized in assigning a weight to each term.

While for SMS spam messages, applied NB and SVM are used to produce input signals, whereby PAMP signal are generated when both classifiers agree that the SMS message is spam. Feature extracted include URL link, spam words, emotion symbols, special characters, message metadata and function words or grammatical words. The combination of spam words and metadata produced a high accuracy rate for all three (3) classifiers.

Even though both of these studies are focussing on spam detection, it may facilitate the understanding of DCA employment for this particular task, assessing the risk level of text spam messages. Hence, the initial version of the algorithm as depicted in the following Fig. 2 is expected to facilitate the understanding of Danger Theory employment for this study.

```
1   Input  :        S = set of spam messages to be labelled as high, medium or low risk
2   Output :        L = set of spam messages labeled as high, medium or low risk
3   Begin
4                   Create a database of spam with risk indicator, D (database library)
5                   Create a folder to contain risk-labeled spam, M (folder for spam with risk level indicator)
6   Identified      for all spam messages, S in P do
    risk
    →
7                       Create a set of spam messages from the sample in D, P (spam folder)
8                       for all spam messages in P do
9                           Add data item in D
10                          Update information on the high, medium and low-risk level
11                          Update risk level of spam cluster with related spam term
12                          Migrate S from P to M and create a new data item in D, if current information not available
13                          S then becomes L
14                      end
15                  end
16                  for all L in M do
17                      Label L as to be a high, medium or low risk
18                  end
19  The assessed    for all spam messages, S do
    risk
20  level       →       Calculate the risk value accordingly with the spam content
21  Prioritize          Label spam messages with the high, medium or low risk
    risk
    →
22                      Add spam messages with risk level indicator into M
23                  end
24  end
```

Fig. 2 Initial version of spam risk-labeled algorithm based on generic DCA

Based on this initial version, the required process flow for the algorithm is identified and is further elaborated in Section II. *D. Experimental Setup*. The model proposed in this study, **Ri**sk **C**oncentration for **C**ontext **A**ssessment (RiCCA) is designed and developed from the requirement of measuring the 'risk concentration' for the content of messages, which 'assessment of the context'. The RiCCA characteristics are adapted as the following:

- Danger Theory is not only able to detect danger but in addition to that is its capability of measuring the maliciousness [24]. Hence, risk concentration can be translated as the severity density (dangerousness) of a spam text message.
- Context assessment is referring to the phase where antigen (SMS messages) is being assessed to identify the malicious level or dangerousness [23], [25].

4) Additional Requirements: In addition to the Danger Theory as a basis, there are additional necessities to ensure the algorithm is well functioning as designed. The general algorithm depicted in Fig. 2 is enhanced by the integration of other techniques. These techniques are suggested to fulfill the requirement of the original algorithm, which includes:

- Application of text mining that covers two (2) main tasks:
  i. the pre-processing phase to break the antigen (spam message) into peptide (tokenized words)
  ii. the use of term weighting schemes to calculate the importance of relevant words (in spam category) that is represented as an input signal
- The application of risk scale range to distinguish every level of the input signal (PAMP, danger and safe) and the output signal (calculated risk as high, medium or low)

The details of these requirements are elaborated in Section II. *D. Experimental Setup*.

B. Dataset

The mandatory item for this study is the availability and accessibility of the SMS messages corpus. Without dataset, the developed algorithm is impossible to be tested and implemented, and as a result, could not be verified for its significance. It is insisted that accessibility to a requisite dataset constitute one of the challenges researchers often face in successfully researching filtering or classifying SMS spam messages [38]. Authors also explored and refined a list of credible research dataset used by researchers for the study in the field that require SMS or short text messages. A good corpus of the dataset for spam experiment required to verify its accuracy detection empirically [39].

The most substantial English corpus for SMS messages is shared publicly at UCI Machine Learning Repository [40] and deployed in this study. This collection consists of total 5,574 SMS messages which 13.4% is spam messages. These messages are collected from four (4) different sources; Grumbletext, National University of Singapore (NUS) Corpus, Tagg's Ph.D. and SMS Spam's Corpus v0.1 Big. In addition to UCI dataset, there are 1,012 self-collected spam messages combined with the UCI dataset deployed in this study.

C. Methodology

In 2005, Stepney suggested a Conceptual Framework before designing and developing AIS algorithm. It has been proposed that bio-inspired algorithms are best developed and analyzed in a multidisciplinary conceptual framework that provides for sophisticated biological models and well-founded analytical principles [41]. The developed theory usually is tested with a real-world case study to verify the computational algorithms. Some of the notable prototype developed in AIS are including libtissue [42] and LISYS [43] or known as Lightweight Intrusion detection SYStem. Both of these prototypes are applied and focused on intrusion detection study and was developed with the guidance of the Conceptual Framework.

In addition to the Conceptual Framework, a terminology of broad study categories for biological experiments also commonly applied in the bio-inspired algorithm. *In silico*, a term in Latin refers to characterize biological experiments carried out entirely in a computer. This means that an algorithm is performed on the computer or via computer simulation to verify its functionality [44].

1) Immunological-based Simulation Lifecycle: Figueredo and associates [45] have introduced a quick guide containing the main steps for modeling and simulation in immunology. This life cycle simulation guide has been implemented in this study since the applied theory in this study; the Danger Theory is the approach that is emerged from the immunological field. This guide is applied for simulation in operational research problems and adapted for simulations of text spam risk assessment. To adapt the guidelines developed by these authors, an observation of similarities and differences, if any is identified with the operational research (spam risk assessment). The general steps for building an immune simulation is depicted as in Fig. 3. These steps represent a life-cycle of simulation in an iterative mode.

2) Simulation of the Case Study: It has bee articulated how a simulation in immunology could be translated as one of the approaches applied in problem investigation [45]. This approach has been referred to as a general framework to design and develop **Ri**sk **C**oncentration for **C**ontext **A**ssessment (RiCCA) prototype in this study. It mainly covers the initial testing to verify the functionality of the RiCCA design is implemented as projected. The process is iteratively implemented until the simple design with the useful result is achieved convincingly. This study is particularly guided and coordinated via these processes as depicted in Fig. 2 and further description is tabulated in Table 1. The elaboration on how every step in the life-cycle process conducted is described further in various sections in this paper and also can be found in published articles as outlined in Table 1.

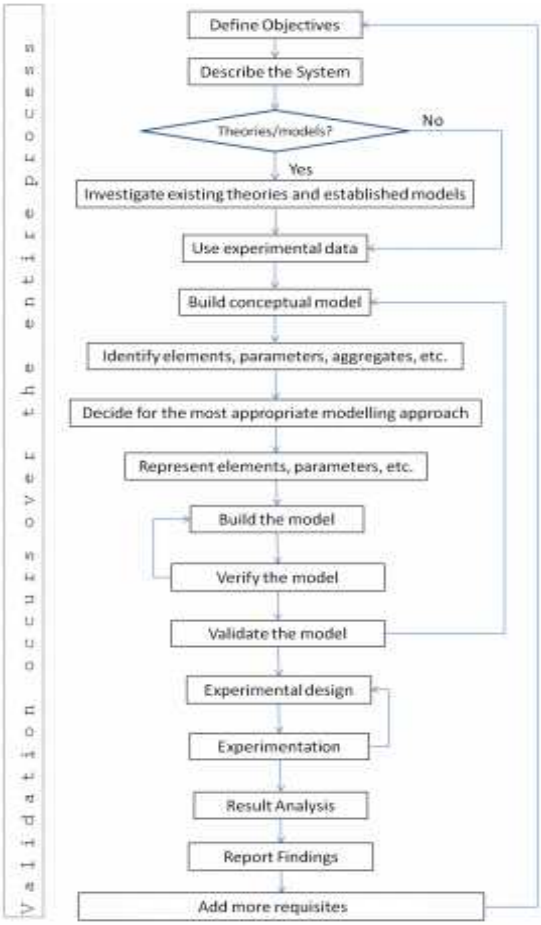

Fig. 3 The life cycle of simulation study processes [48]

TABLE I
THE APPLICATION OF SIMULATION STUDY PROCESSES IN RICCA DESIGNATION AND DEVELOPMENT

Process in Immunological Simulation (as depicted in Fig. 3)	Process for RiCCA Implementation	References for further Description
Define the objectives	Objectives of this research are defined and discussed based on the identified scenario as the motivation of the research. According to the formulated problem statement, the research objectives are developed with the intention to have an outcome of risk assessment tool.	Section I: Introduction, [46]
Describe the system	The description of the system is elaborated and defined by the scope of research works. The functionalities and limitations of the system are also identified.	Section II. A.3) Dendritic Cell Algorithm in Spam Classification, [47]
Investigate existing theories and established models	Meticulous reviews of the past literature are conducted. All possible related papers and journals are validated and studied to build a new model or an extended version as an improvement of what has already been established.	Section II. A. Danger Theory as the Approach, [47], [48]
Use experimental data	The dataset from the real-world observation and experimentation is identified. By utilizing the largest collection and publicly shared corpus that is available online, a potential comparison by other researcher is possible in finding the most efficient method to assess the spam risk.	Section II. B. Dataset [47], [48]
Build conceptual model	A conceptual model of how the Danger Theory of AIS is capable of developing solution for spam risk assessment is analyzed. The biological idea that employed in spam risk assessment is articulated, and the RiCCA model is designed.	Section II. A.3) Dendritic Cell Algorithm in Spam Classification
Identify elements, parameters, aggregates, etc. already established in theory and real-world data	All the potential elements in Danger Theory (theoretical framework) are identified and mapped with the designed conceptual model. The biological behavior is articulated and clarified in spam problem environment.	Section II. A.3)Dendritic Cell Algorithm in Spam Classification and 4)Additional Requirements, [49], [50]
Decide on the most appropriate simulation approach	All related elements of the proposed conceptual model to function properly are identified and tested. These additional elements include risk assessment process and text mining.	Section II. C.1)Immunological-based Simulation Lifecycle and 2)Simulation of the Case Study
Represent elements, parameters, etc. using the appropriate simulation approach	Series of experiments are executed to ensure the most appropriate simulation approach is effectively applied for the proposed model. The results from this preliminary simulation determined the design precision.	Section II. D.1) Requirements
Build the simulation model	The findings from series of experiments are used to build the prototype for the purpose of the automation process.	Section II. D. Experimental Setup, [51]
Verify the model	The computational model (*in silico*) is verified to ensure the algorithms are aligned and as delineated in the original DCA algorithm.	Section III. Results and Discussion, [51]
Validate the	The model is re-validated with the	

Process in Immunological Simulation (as depicted in Fig. 3)	Process for RiCCA Implementation	References for further Description
model with existing theories and if available real-world data	biological theory (Danger Theory) and the source of the dataset is confirmed. The authenticated model is established at this point.	
Experimental design	The design of the automation tool is re-validate. The source of the dataset for the initial population and testing phase are confirmed and collected. Series of experiments are determined according to identified objectives.	[51]–[53]
Experimentation	Once the proposed model is completely developed as an automation tool, experiments are run with multiple series to verify multi-objectives.	
Result Analysis	The results are analyzed as an outcome of the model. The efficiency of the model is identified.	
Report Findings	The findings and conclusion are reported.	
Validate and add more requisites	The findings are validated and verified if it is well aligned with the biological inspiration idea and future enhancement is identified if any.	

1) Requirements: The simulation is conducted based on the following Fig. 4. The connection between the process flow in Fig. 4 and initial algorithm in Fig. 2 is depicted as in Table 2.

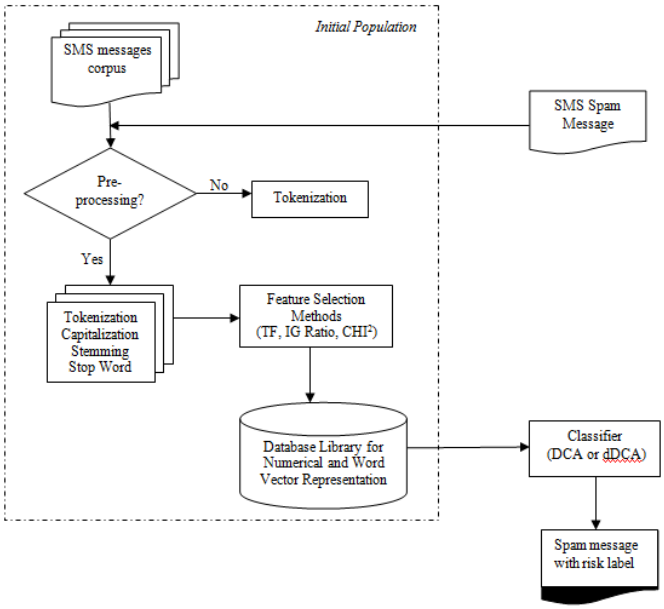

Fig. 4 Overview of the experimental setup

D. Experimental Setup

TABLE III
THE CONNECTION PORTRAYED FOR INTEGRATION OF DCA ALGORITHM IN RICCA MODEL. THE DIFFERENT COLOUR OF CONNECTIONS INDICATE A DIFFERENT PHASE IN RICCA

Process Flow (Fig. 4)		The connection	Initial Algorithm (Fig. 2)		
Phase	Process		Description	Process	Generic DCA
1	Preparation		collect dataset corpus	Create initial population	Line 4-13
			text pre-processing		
2	Pre-Processing		assign input signals value for antigen		
3	Wordlist		store antigen value		
4	Term Weighting Schemes		receive new spam message	Calculate the severity level	Line 20
			text pre-processing		
5	Numerical and Word Vector Representation		map the identified antigen with the stored input signals value (refer library database)		
			calculate the severity level using classifier (DCA or dDCA)		
6	Classification		map the calculated output signals with developed risk scale and identify the level of risk	Prioritize risk	Line 21
			according to identified risk level, response against spam could be delete, escalate to authority body, re-calculate the risk level (for the case of false positive), or do nothing	Action to response	Immune response activated

The flow for conducting the simulation includes:

- SMS messages (ham and spam) are collected and prepared for the initial database.

- The corpus (for initial population) or text message has an option either to run through text pre-treatment or not. Tokenization is the necessary to process if decided not to have the pre-processing in place. Otherwise, the full cycle of text pre-processing is executed, which include tokenization, capitalization, stemming or also known as lemmatization and removal of stop word from the data.
- Term Frequency (TF), Information Gain Ratio (IG Ratio) and CHI-Squared (CHI^2) are the available term weighting schemes and act as feature selection methods. These schemes will calculate the importance of every word in the corpus that indicates its relevance to the spam (risk) category. The value derived statistically from this method is further implemented as an input signal in the classifier. All schemes calculated tokenized word with a range in between 0 to 1 that suggested the closer the value to 1, the closer it is to malicious level (as depicted in Fig. 1).
    i. TF also was known as term strength or word frequency. It measures the number of feature (term) in a category that appears in a corpus;
    ii. IG Ratio used as one of the disparity measures, and the high gain ratio for selected feature implies that the feature is useful for classification. It applies normalization to information gain score by utilizing a split information value; and
    iii. CHI^2 measures the association between the word feature and its associated class or category
- This information is stored as an internal database library.
- Once the initial database has been developed, a new spam message can be processed. The new message is assessed according to the chosen classifier, either DCA or dDCA.
- The calculated value for terms in the message then will be mapped to the risk scale. Every derived value is represented by the term's severity degree, i.e. PAMP, danger or safety. The range for risk scale is user-defined.
- The spam message is risk-marked and labeled accordingly to the systematic risk, i.e. high, medium or low risk.

2) Risk Scale Value Range, S: This testing is applied three (3) different range value for risk scale as tabulated in Table 3.

TABLE IIIII
THREE (3) DIFFERENT RANGE OF RISK SCALE

S1	S2	S3	Risk Level		Description
			Input Signals	Output Signals	
1.00 - 0.70	1.00 - 0.80	1.00 - 0.90	PAMPs	High	Reflects catastrophe effect that is difficult to handle. Spam message with high risk usually consists of more than one spam term. It is most likely also containing text that request users to *access given URL*.
0.69 - 0.40	0.79 - 0.50	0.89 - 0.30	Danger	Medium	The effect is lesser than high but still dangerous. A spam message contains with less weight of spam term in its context.
0.39 - 0.00	0.49 - 0.00	0.29 - 0.00	Safe	Low	The level of hazard could be very minimal or nearly secure with no significant damage. Spam message commonly contains with the low weight of spam term (negligible).

3) Anomaly Threshold, t_m/T_k: To determine a reviewed process tends to be anomalous, a value of threshold needs to be predetermined. A value below this threshold will indicate the process as normal. In this experiment, four (4) different values are applied to verify its effects on assessing the risk level.

- Initial population – based on a total number of deployed messages, anomaly threshold, is derived from the initial population proportion and set as 0.2671;
- Depending on the minimum value of mature class from the risk scale. For instance, anomaly threshold for S1 is 0.4000, S2 is 0.5000 and S3 is 0.3000;
- Based on the equal division of 2 categories (mature and semi-mature) and the value is set as 0.5000; and
- For dDCA, there is one additional value considered as anomaly threshold. Value 0 is suggested to be considered in this experiment as the threshold value, which this value is the minimum value for mature class, retrieved from the risk scale for output signal.

4) Weights for Signal Transformation, WM: Weights to transform input signals into output signals are one of the mandatory prerequisites for signal processing intentionally for the algorithm. Measurement calculates the output signal as in Eq. (1). The transforming weight applied for this testing is tabulated in Table 4.

TABLE IVV
THREE (3) DIFFERENT TRANSFORMING WEIGHTS

Signals	WM1			WM2			WM3		
	P	D	S	P	D	S	P	D	S
CSM	1	0.5	1.5	2	0	2	1	0.5	1
smDC	0	0	1	1	0	1	0	0	1
mDC	1	0.5	1.5	2	3	3	1.5	0.5	1.5

P – PAMPs; D – Danger; S -Safe

5) Antigen Multiplication: The antigen multiplication is applied with the objective to verify its effect in optimizing the accuracy rate. As elaborated in Gu, Greensmith, & Aickelin [59], the antigen multiplication is implemented to overcome the problem of antigen deficiency or insufficient antigens that are supplied to the initial DCs population. This antigen multiplication made several copies of each antigen and calculated for its input signals value.

In this testing, the same spam messages (antigen) are multiplied or copied for 10, 20, and 30 until 100 times before creating the initial population. The process for severity assessment is repeated with multiplied antigens, and the effect of this multiplication is identified.

E. Series of Experiments

The simulation executed with the following series of all three (3) pre-selected terms weighting schemes are applied at all rounds.

TABLE V
NINE (9) SERIES OF EXPERIMENTS EXECUTED FOR THIS SIMULATION

The range of Risk Value Scale (S)	Weight Matrix (WM)
S1	WM1
	WM2
	WM3
S2	WM1
	WM2
	WM3
S3	WM1
	WM2
	WM3

Once the best value for risk value scale, S and weight matrix, WM are identified, the testing as tabulated in Table 5 is repeated with the dDCA classifier. This is purposely to execute some comparable analysis between DCA and dDCA. Finally, the test is conducted with multiplied antigen employing with both classifiers.

F. Performance Measurement

The model developed in this study needs to be evaluated to verify its performance. The performance of this model needs to be identified to ensure the reliability of the proposed model. The evaluation metrics are helpful in analyzing the model performance and also to avoid any misleading prediction.

Based on Table 6, the possible metrics that is practical to be used for this performance measurement [38], [52] include:

- True Positive (TP): spam messages are identified precisely as malicious (high and medium risk);
- True Negative (TN): spam messages are identified precisely as benign (low risk);
- False Positive (FP): benign messages are incorrectly predicted as malicious (high and medium risk); and
- False Negative (FN): malicious spam messages are incorrectly predicted as benign (low risk)
- Accuracy (Acc): the total number of spam messages identified precisely as malicious (high and medium risk) and benign (low risk)

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

TABLE VI
CONFUSION MATRIX

		True Class	
		Mature	**Semi-mature**
Hypothesized Class	**Mature**	TP	FP
	Semi-mature	FN	TN

III. RESULTS AND DISCUSSION

A. Identification of an Optimum Value for Parameters

It is a vital phase in identifying the best value for the required parameters in the classifier. In this particular simulation, it is critical to find the precise term frequency schemes, anomaly threshold, risk scale and weight matrix applied in DCA. Throughout the experiment, TF is identified as the best suited and resulted in the highest rate of TP for DCA as tabulated in the following Table 7. Experiencing that the process of TF without pre-processing has resulted a low accuracy rate (below 80%) and high false positive rate (more than 10%), the experiment is further executed with other two (2) options of term weighting schemes (IG Ratio and CHI^2) with pre-processing phase only.

TABLE VII
PERFORMANCE MEASUREMENT RESULTS (ACCURACY VALUE) USING DCA WITH TF, IG RATIO AND CHI^2

S and WM	Anomaly Threshold, t_m	TF (with pre-processing)	TF (without pre-processing)	IG Ratio (with pre-processing)	CHI^2 (with pre-processing)
		Accuracy			
S1WM1	0.2671	0.8967	0.7484	**0.8283**	0.6183
	0.4000	**0.9067**	**0.7650**	**0.8283**	0.6183
	0.5000	**0.9000**	0.7466	0.7950	0.6183
S1WM2	0.2671	0.8666	0.6200	**0.8683**	0.6183
	0.4000	0.8783	0.6200	**0.8683**	0.6183
	0.5000	0.8714	0.6000	0.8217	0.6183
S1WM3	0.2671	0.8967	0.7467	**0.8283**	0.6183
	0.4000	**0.9067**	**0.7600**	**0.8283**	0.6183
	0.5000	0.8983	0.7433	0.7950	0.6183
S2WM1	0.2671	0.8616	0.7166	0.6983	0.6183
	0.5000	0.8916	0.7566	0.6983	0.6183

S and WM	Anomaly Threshold, t_m	TF (with pre-processing)	TF (without pre-processing)	IG Ratio (with pre-processing)	CHI^2 (with pre-processing)
		Accuracy			
	0.5000	0.8916	0.7566	0.6983	0.6183
S2WM2	0.2671	0.8583	0.6467	0.7533	0.6183
	0.5000	0.8617	0.6467	0.7533	0.6183
	0.5000	0.8617	0.6467	0.7533	0.6183
S2WM3	0.2671	0.8700	0.7166	0.6967	0.6183
	0.5000	0.8950	0.7583	0.6967	0.6183
	0.5000	0.8950	0.7583	0.6967	0.6183
S3WM1	0.2671	0.7717	0.6067	0.8150	0.6183
	0.3000	0.7817	0.6183	0.8150	0.6183
	0.5000	0.7700	0.6150	0.7883	0.6133
S3WM2	0.2671	0.7667	0.5000	0.9050	0.6417
	0.3000	0.7667	0.5000	0.9050	0.6417
	0.5000	0.7550	0.4967	0.8534	0.6217
S3WM3	0.2671	0.7717	0.6067	0.8150	0.6183
	0.3000	0.7817	0.6183	0.8150	0.6183
	0.5000	0.7700	0.6150	0.7816	0.6133

From Table 7, it is shown that for DCA, a combination of S1 and WM1 with a minimum value of mature category from risk scale (0.4000) as the anomaly threshold has produced the highest accuracy rate for TF with pre-processing, 90.67%. The combination of S1 and WM3 also resulted in the same accuracy rate, 90.67% with 0.4000 as the anomaly threshold value.

B. Comparative Analysis between DCA and dDCA

The test is continued using dDCA classifier applying all the term weighting schemes, risk scale, S and anomaly threshold, T_k values with the pre-processing phase. The results are tabulated in the following Table 8.

TABLE VIII
PERFORMANCE MEASUREMENT RESULTS USING DDCA WITH TF, IG RATIO AND CHI^2

Weighting Scheme	Risk Scale	Anomaly threshold, T_k	TP	TN	FP	FN	Accuracy
TF	S1	0.2671	0.3533	0.5733	0.0517	0.0217	**0.9266**
		0.4000	0.2850	0.5733	0.0517	0.0900	0.8583
		0.5000	0.2400	0.5733	0.0517	0.1350	0.8133
		0.0000	0.3683	0.5733	0.0517	0.0067	**0.9416**
	S2	0.2671	0.3117	0.5767	0.0483	0.0633	0.8884
		0.5000	0.2083	0.5767	0.0483	0.1667	0.7850
		0.5000	0.2083	0.5767	0.0483	0.1667	0.7850
		0.0000	0.3533	0.5783	0.0467	0.0217	**0.9316**
	S3	0.2671	0.3583	0.4533	0.1717	0.0167	0.8116
		0.3000	0.3467	0.4533	0.1717	0.0283	0.8000
		0.5000	0.2800	0.4533	0.1717	0.0950	0.7333
		0.0000	0.3733	0.4533	0.1717	0.0017	0.8266
IG Ratio	S1	0.2671	0.1717	0.5967	0.0283	0.2033	0.7684
		0.4000	0.0850	0.5967	0.0283	0.2900	0.6817
		0.5000	0.0417	0.5967	0.0283	0.3333	0.6384
		0.0000	0.3000	0.5967	0.0283	0.0750	0.8967
	S2	0.2671	0.0783	0.6033	0.0217	0.2967	0.6816
		0.5000	0.0150	0.6033	0.0217	0.3600	0.6183
		0.5000	0.0150	0.6033	0.0217	0.3600	0.6183
		0.0000	0.2000	0.6033	0.0217	0.1750	0.8033
	S3	0.2671	0.2233	0.5783	0.0467	0.1517	0.8016
		0.3000	0.1967	0.5783	0.0467	0.1783	0.7750
		0.5000	0.0617	0.5783	0.0467	0.3133	0.6400
		0.0000	0.3467	0.5783	0.0467	0.0283	**0.9250**
CHI^2	S1	0.2671	0.0000	0.6183	0.0067	0.3750	0.6183
		0.4000	0.0000	0.6183	0.0067	0.3750	0.6183
		0.5000	0.0000	0.6183	0.0067	0.3750	0.6183
		0.0000	0.0000	0.6183	0.0067	0.3750	0.6183
	S2	0.2671	0.0000	0.6183	0.0067	0.3750	0.6183
		0.5000	0.0000	0.6183	0.0067	0.3750	0.6183
		0.5000	0.0000	0.6183	0.0067	0.3750	0.6183

Weighting Scheme	Risk Scale	Anomaly threshold, T_k	TP	TN	FP	FN	Accuracy
		0.0000	0.0000	0.6183	0.0067	0.3750	0.6183
	S3	0.2671	0.0033	0.6100	0.0150	0.3717	0.6133
		0.3000	0.0033	0.6100	0.0150	0.3717	0.6133
		0.5000	0.0000	0.6100	0.0150	0.3750	0.6100
		0.0000	0.0400	0.6100	0.0150	0.3350	0.6500

According to the simulation results as tabulated in Table 8, the classifier dDCA resulted in a higher and optimized accuracy rate compared to DCA (Table 7), especially with TF as the weighting scheme. From these two simulations, it is obvious that CHI^2 is not a suitable weighting scheme to be applied in this field of the domain (spam risk assessment).

C. Antigen Multiplication Effects on DCA and dDCA Classifier

Finally, the test is continued with antigens multiplication for creating the initial population with 10, 20, 30, until 100 times of antigen. This is executed for testing both using DCA and dDCA classifiers.

TABLE IX
ACCURACY RATE FOR DCA AND DDCA WITH ANTIGEN MULTIPLICATION

Antigen Multiplication (%)	DCA	dDCA
10	**0.5117**	**0.5050**
20	0.4358	0.4233
30	0.4167	0.4067
40	0.3950	0.3850
50	0.3950	0.3850
60	0.3950	0.3850
70	0.3950	0.3850
80	0.3867	0.3750
90	0.3867	0.3750
100	0.3850	0.3750

Based on the result as tabulated in Table 9, surprisingly, this testing showed that the accuracy rate has been tremendously decreased with the application of antigen multiplication in both DCA and dDCA with TF. This method is inappropriate and does not assist in optimizing the result, which the higher the times of antigen multiplication, the lower the accuracy of both classifiers in assessing the risk level.

IV. CONCLUSION

The creation of AIS involves the translation of basic immunological models into feasible algorithms. This study defined the original DCA into an enhanced version for risk assessment of text spam message and namely as RiCCA. Assessing risk level of a spam message is acknowledged as a rare study in research world since no publication is found for the related works at the time of this writing. Previous studies have been focussing on measuring the hazardous impact of an intrusion attack only. Hence, apart from intrusion attack domain, the study in measuring the potential risk for a text spam message is proposed in this paper.

The simulation is executed by conducting a series of simulation with the deployment of public shared and self-collected dataset for testing with the employment of RiCCA tool. With the right value for the classifier's important parameters, this is potentially seen as a tool that able to decipher a spam message for a real intention of the message that usually malicious. Verified from this testing, the model is expected to be expanded in another field of risk assessment for spam in another format such as email and social media. This tool also potentially to be developed as a mobile application, which may benefit users by assisting them to identify the severity of hidden messages.

ACKNOWLEDGMENT

This research is fully funded by the Ministry of Higher Education of Malaysia and Research Management Centre of USIM via grant research with code USIM/FRGS/FST/32/50315.

REFERENCES

[1] P. Parham, *The Immune System, Fourth Edition*. Taylor & Francis Group, 2014.
[2] D. Dasgupta, "Advances in artificial immune systems," *IEEE Comput. Intell. Mag.*, vol. 1, no. 4, pp. 40–49, Nov. 2006.
[3] T. Lu, K. Zheng, R. Fu, Y. Liu, B. Wu, and S. Guo, "A Danger Theory Based Mobile Virus Detection Model and Its Application in Inhibiting Virus," *J. Networks*, vol. 7, no. 8, pp. 1227–1232, Aug. 2012.
[4] M. Gong, J. Zhang, J. Ma, and L. Jiao, "An efficient negative selection algorithm with further training for anomaly detection," *Knowledge-Based Syst.*, vol. 30, pp. 185–191, Jun. 2012.
[5] L. E. Jim and M. A. Gregory, "A Review of Artificial Immune System Based Security Frameworks for MANET," *Int. J. Commun. Netw. Syst. Sci.*, vol. 9, no. 1, pp. 1–18, 2016.
[6] H. Yang, T. Li, X. Hu, F. Wang, and Y. Zou, "A Survey of Artificial Immune System Based Intrusion Detection," *Sci. World J.*, vol. 2014, pp. 1–11, 2014.
[7] Internet Governance Forumn, "Best Practice Forum on Regulation and Mitigation of Unsolicited Communications," 2014.
[8] Cloudmark, "2013 Global Messaging Threat Report," 2013.
[9] S. Alotaibi, S. Furnell, and N. Clarke, "A Novel Taxonomy for Mobile Applications Data," *Int. J. Cyber-Security Digit. Forensics*, vol. 5, no. 3, pp. 115–121, 2016.
[10] M. Theoharidou, A. Mylonas, and D. Gritzalis, "A Risk Assessment Method for Smartphones," in *International Information Security Conference*, 2012, pp. 443–456.
[11] D. Goel and A. K. Jain, "Mobile phishing attacks and defence mechanisms: State of art and open research challenges," *Comput. Secur.*, vol. 73, pp. 519–544, Mar. 2018.
[12] J. Lin, N. Sadeh, S. Amini, J. Lindqvist, J. I. Hong, and J. Zhang, "Expectation and purpose," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*, 2012, p. 501.
[13] K. H. Haritha, R. S. Kumar, and M. S. Krishnan, "An Analytical Exploration on SMS Spam & User Retort," *Int. J. Pure Appl. Math.*, vol. 114, no. 11, pp. 147–156, 2017.
[14] K. D. Mitnick and W. L. Simon, *The Art of Deception: Controlling the Human Element of Security*. 2003.
[15] T. Chen and M.-Y. Kan, "Creating a live, public short message service corpus: the NUS SMS corpus," *Lang. Resour. Eval.*, vol. 47, no. 2, pp. 299–335, Aug. 2012.

[16] N. Choudhary and A. K. Jain, “Towards Filtering of SMS Spam Messages Using Machine Learning Based Technique,” in *International Conference on Advanced Informatics for Computing Research (ICAICR)*, vol. 712, D. Singh, B. Raman, A. K. Luhach, and P. Lingras, Eds. Singapore: Springer Singapore, 2017, pp. 18–30.

[17] D. Suleiman and G. Al-Naymat, “SMS Spam Detection using H2O Framework,” *Procedia Comput. Sci.*, vol. 113, pp. 154–161, 2017.

[18] S. J. Warade, P. A.Tijare, and S. N.Sawalkar, “Implementation of SMS Spam Detection System,” *Int. J. Res. Advent Technol.*, vol. 4, no. 5, pp. 46–52, 2016.

[19] Y. Ge, H. Liang, L. Chen, and Q. Zhang, “The Designation of Bio-Inspired Intrusion Detection System Model in Cloud Computing Based on Machine Learning,” in *International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE)*, 2015, pp. 1932–1937.

[20] D. Singh and S. S. Bedi, “Novel Intrusion Detection in MANETs Based on Trust,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 4, pp. 3556–3560, 2015.

[21] P. Matzinger, “Tolerance, Danger, and the Extended Family,” *Annu. Rev. Immunol.*, vol. 12, no. 1, pp. 991–1045, Apr. 1994.

[22] P. Matzinger, “The Danger Model: A Renewed Sense of Self,” *Science (80-.).*, vol. 296, no. 5566, pp. 301–305, Apr. 2002.

[23] J. Greensmith, “The Dendritic Cell Algorithm,” University of Nottingham, 2007.

[24] J. Greensmith, U. Aickelin, and S. Cayzer, “Detecting Danger: The Dendritic Cell Algorithm,” in *Robust Intelligent Systems*, London: Springer London, 2008, pp. 89–112.

[25] U. Aickelin and J. Greensmith, “Sensing danger: Innate immunology for intrusion detection,” *Inf. Secur. Tech. Rep.*, vol. 12, no. 4, pp. 218–227, Jan. 2007.

[26] J. Greensmith, U. Aickelin, and J. Twycross, “Articulation and Clarification of the Dendritic Cell Algorithm,” 2006, pp. 404–417.

[27] J. Greensmith and U. Aickelin, “Artificial Dendritic Cells: Multi-faceted Perspectives,” in *Human-Centric Information Processing Through Granular Modelling*, vol. 182, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 375–395.

[28] R. Oates, G. Kendall, and J. M. Garibaldi, “Frequency analysis for dendritic cell population tuning,” *Evol. Intell.*, vol. 1, no. 2, pp. 145–157, Jun. 2008.

[29] J. Greensmith and U. Aickelin, “The Deterministic Dendritic Cell Algorithm,” in *Artificial Immune Systems*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 291–302.

[30] C. J. Musselle, “Insights into the Antigen Sampling Component of the Dendritic Cell Algorithm,” in *International Conference on Artificial Immune Systems (ICARIS)*, 2010, pp. 88–101.

[31] J. Timmis, P. Andrews, N. Owens, and E. Clark, “An interdisciplinary perspective on artificial immune systems,” *Evol. Intell.*, vol. 1, no. 1, pp. 5–26, Mar. 2008.

[32] AISWeb, “AISWeb,” *Systems, The Online Home of Artificial Immune*. .

[33] M. Abdelhaq, R. Alsaqour, M. Ismail, and S. Abdelhaq, “Dendritic Cell Fuzzy Logic Algorithm over Mobile Ad Hoc Networks,” in *International Conference on Intelligent Systems, Modelling and Simulation*, 2015, pp. 64–69.

[34] A. C. M, S. Uma, and M. M. Kumar, “Detaining and Avoiding Mobile Virus Propagation by Considering Human Behavior,” *Int. J. Adv. Comput. Technol.*, vol. 3, no. 3, pp. 619–623, 2014.

[35] A. A. Al-Hasan and E.-S. M. El-Alfy, “Dendritic Cell Algorithm for Mobile Phone Spam Filtering,” *Procedia Comput. Sci.*, vol. 52, pp. 244–251, 2015.

[36] R. A. Mulay and A. S. Chauhan, “Review: Optimized Webpage Classification,” *Int. J. Emerg. Trends Technol. Comput. Sci.*, vol. 3, no. 4, pp. 302–306, 2014.

[37] E.-S. M. El-Alfy and A. A. Al-Hasan, “A novel bio-inspired predictive model for spam filtering based on dendritic cell algorithm,” in *2014 IEEE Symposium on Computational Intelligence in Cyber Security (CICS)*, 2014, pp. 1–7.

[38] S. M. Abdulhamid *et al.*, “A Review on Mobile SMS Spam Filtering Techniques,” *IEEE Access*, vol. 5, pp. 15650–15666, 2017.

[39] T. Oda, “A Spam-Detecting Artificial Immune System,” Carleton University Ottawa, Canada, 2005.

[40] T. A. Almeida and J. M. G. Hidalgo, “UCI Machine Learning Repository,” 2012. .

[41] S. Stepney, R. E. Smith, J. Timmis, A. M. Tyrell, M. J.Neal, and A. N.W.Hone, “Conceptual Frameworks for Artificial Immune Systems,” 2005.

[42] J. Twycross and U. Aickelin, “libtissue - implementing innate immunity,” in *2006 IEEE International Conference on Evolutionary Computation*, 2010, pp. 499–506.

[43] S. A. Hofmeyr and S. Forrest, “Architecture for an Artificial Immune System,” *Evol. Comput.*, vol. 8, no. 4, pp. 443–473, Dec. 2000.

[44] The Marshall Protocol Knowledge Base, “Differences Between In Vitro, In Vivo, and In Silico Studies,” *Autoimmunity Research Foundation*, 2012. [Online]. Available: https://mpkb.org/home/patients/assessing_literature/in_vitro_studies.

[45] G. P. Figueredo, P.-O. Siebers, U. Aickelin, and S. Foan, “A Beginner’s Guide to Systems Simulation in Immunology,” in *International Conference on Artificial Immune Systems, ICARIS 2012*, 2012, pp. 57–71.

[46] K. Zainal, N. F. Sulaiman, and M. Z. Jali, “An Analysis of Various Algorithms For Text Spam Classification and Clustering Using RapidMiner and Weka,” *Int. J. Comput. Sci. Inf. Secur.*, vol. 13, no. 3, pp. 66–74, 2015.

[47] K. Zainal and M. Z. Jali, “A Perception Model of Spam Risk Assessment Inspired by Danger Theory of Artificial Immune Systems,” *Procedia Comput. Sci.*, vol. 59, pp. 152–161, 2015.

[48] K. Zainal and M. Z. Jali, “A Review of Feature Extraction Optimization in SMS Spam Messages Classification,” in *Communications in Computer and Information Science*, vol. 652, 2016, pp. 158–170.

[49] K. Zainal and M. Z. Jali, “The significant effect of feature selection methods in spam risk assessment using dendritic cell algorithm,” in *2017 5th International Conference on Information and Communication Technology (ICoIC7)*, 2017, pp. 1–8.

[50] K. Zainal and M. Z. Jali, “Comparative Analysis of Danger Theory Variants in Rating Risk Concentration via Context Assessment of Text Spam Messages,” 2017.

[51] K. Zainal and M. Z. Jali, “The Design and Development of Spam Risk Assessment Prototype: In Silico of Danger Theory Variants,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 4, pp. 401–410, 2017.

[52] N. Japkowicz and M. Shah, *Evaluating Learning Algorithms*. Cambridge: Cambridge University Press, 2011.

[53] F. Gu, J. Greensmith, and U. Aickelin, “Further Exploration of the Dendritic Cell Algorithm: Antigen Multiplier and Time Windows,” in *Artificial Immune Systems*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 142–153.