Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Applied Soft Computing 24 (2014) 612–622

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Outbreak detection model based on danger theory

Mohamad Farhan Mohamad Mohsin*, Azuraliza Abu Bakar, Abdul Razak Hamdan

Data Mining and Optimization Research Group, Centre for Artificial Intelligence Technology, Faculty of Science & Information Technology, Universiti Kebangsaan Malaysia, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 28 October 2013
Received in revised form 17 June 2014
Accepted 12 August 2014
Available online 22 August 2014

Keywords:
Outbreak detection
Artificial immune system
Danger theory
Dendritic cell algorithm

ABSTRACT

In outbreak detection, one of the key issues is the need to deal with the weakness of early outbreak signals because this causes the detection model to have has less capability in terms of robustness when unseen outbreak patterns vary from those in the trained model. As a result, an imbalance between high detection rate and low false alarm rate occurs. To solve this problem, this study proposes a novel outbreak detection model based on danger theory; a bio-inspired method that replicates how the human body fights pathogens. We propose a signal formalization approach based on cumulative sum and a cumulative mature antigen contact value to suit the outbreak characteristic and danger theory. Two outbreak diseases, dengue and SARS, are subjected to a danger theory algorithm; namely the dendritic cell algorithm. To evaluate the model, four measurement metrics are applied: detection rate, specificity, false alarm rate, and accuracy. From the experiment, the proposed model outperforms the other detection approaches and shows a significant improvement for both diseases outbreak detection. The findings reveal that the robustness of the proposed immune model increases when dealing with inconsistent outbreak signals. The model is able to detect new unknown outbreak patterns and can discriminate between outbreak and non-outbreak cases with a consistent high detection rate, high sensitivity, and lower false alarm rate even without a training phase.

1. Introduction

The aim of an outbreak detection system is to assist epidemiologists monitor the progress of a disease and raise an alert when there is an impending outbreak. An accurate and fast alarm notification system allows healthcare professionals to set up an early prevention plan before an outbreak spreads to a wider geographical area. If an outbreak is uncontrollable, it may lead to high death tolls and worse still; it kills without warning [1]. For example, the spread of severe acute respiratory syndrome (SARS) [2], influenza [3,4], and the avian virus [5] led to high mortality rates shortly after the first case was detected. Beside those viruses, there are seasonal types of diseases such as malaria, dengue, cholera and many more that can be classified as having the potential to escalate to outbreaks according to the World Health Organization and Centers for Disease Control and Prevention portal. One of the indirect effects of an outbreak is that it has a detrimental economic impact on a country. This is exemplified by the SARS and avian virus outbreaks in 2003, which led to a huge loss of around US$20 million for the tourism industry in most affected Asian countries such China, Hong Kong, Singapore and Vietnam because of an ineffective surveillance system [6]. For this reason, we need to be able to rely on an accurate detection system to monitor for signs of an emerging outbreak. The importance of having such a system has been emphasized in much of the literature, not least because it can help authorities to control the spread of an outbreak and reduce the mortality rate [4,7].

An outbreak is defined as a sudden spread of disease with a much greater number cases reported than expected over time, where the point of initial occurrence is a small area but quickly spreads to a wider geographic area [8]. A health surveillance system can monitor the progress of incidences of a disease over time: daily, weekly, or monthly. If there is an extraordinary pattern, this might indicate that an outbreak has started [9]. A sudden increase in the number of suspected cases from previous day is an indicator of a potential outbreak and it changes are viewed as abnormal characteristics; therefore an outbreak detection system is actually looking for anomalies in health data [10,11]. In outbreak detection studies, an outbreak is a collective anomaly where it requires more than one case to be detected before it can be labelled as an actual outbreak [12,13]. For example, if a single patient has been detected as having a certain disease, the alert system will not be activated until the number of similar cases has reached a certain default number.

* Corresponding author. +60 49285125.
E-mail addresses: farhan@uum.edu.my, mfarhan.mmohsin@gmail.com, mfarhan.mmohisn@gmail.com (M.F. Mohamad Mohsin), aab@ftsm.ukm.my (A.A. Bakar), arh@ftsm.ukm.my (A.R. Hamdan).

http://dx.doi.org/10.1016/j.asoc.2014.08.030
1568-4946/

Each disease has a different default number which is determined by epidemiologists. For instance, in the case of a disease such as dengue just more than a suspected cases difference from previous day will raise an alarm while cholera requires up to 20 registered cases before it can be classified as an outbreak [14]. However, when the spread of an outbreak is slow, the number of cases is not the key indicator and further epidemiological investigation is required [15]. Asthe notification of outbreak is based on collective anomalies that need to be tailored with same locality and time, this makes outbreak detection different from other detection tasks such as intrusion, fraud and fault.

In recent years, many researchers have worked on various outbreak detection methods. Those methods can be classified into three main approaches: statistical [1,9,16–18], artificial intelligence [15,19–25], and a combination of these two approaches into hybrid methods [26–28]. The statistical approach was applied in early outbreak detection models and to date it remains the preferred approach in many health surveillance systems [29,30]. One of the key issues that needs to be addressed in existing models is how to obtain a high detection rate while at the same time reducing the false alarm rate. However, due to the weakness of the outbreak signal which always behaves under uncertainties, existing systems produce an inconsistent false alarm rate during detection. Here, uncertainty refers to inconsistent outbreak signals, where the outbreak pattern frequently changes and differs between years. This means that a trained model has less capability in terms of robustness particularly when unseen outbreak patterns vary from the trained model. Robustness is also lost when detection algorithms require a sample from both an outbreak and non-outbreak session for model development. However, in practice, it is a challenging task to define a sample for an outbreak period because most of the data available relate only to non-outbreak periods.

Previously, outbreak detection models were mostly based on the univariate surveillance approach [18,31]. However, due to the weakness of the outbreak signal, the possibility of these models generating an imbalanced result between detection rate and false alarm rate is high because they rely on a single attribute [28]. As a possible solution to this problem of a weak signal, researchers have been investigating the use of multivariate surveillance by injecting the weak signal with a stronger signal, for instance by combining spatial and temporal data [1,16,32]. In addition, efforts have been made to combine multiple syndromic data to boost detection, for instance by combining emergency visit data with weather information, or clinical diagnosis results [33,34] and by investigating social network status and internet tracking search [22,35,36]. Since an outbreak is observed over time, distraction factors such as the seasonal event effect always disturb the detection results [9]. In relation to the data issue, each disease has a different outbreak definition which is determined by its the environment, government policies, outbreak rate and the medium for spread. Information on aspects such as outbreak duration (outbreak period), outbreak magnitude (rate of outbreak cases) and onset date are required and they are different depending on the disease [7]. The above-mentioned factors indirectly influence the effectiveness of an outbreak algorithm. Fig. 1 depicts the general outbreak detection concept.

In this conceptual model, an outbreak is defined based on three parameters: similar time, similar place, and huge number of reported cases. The aim of the outbreak detection model is to generate an accurate alarm which has high sensitivity and fast detection time. Since an outbreak is viewed as an abnormal period, anomaly detection analysis is used to detect the point at which data start to abnormally change. The quality of outbreak data and disease characteristic are the two factors that influence detection performance.

To improve the outbreak detection model's robustness when dealing with uncertain outbreak signals, we sought to identify an appropriate model from various disciplines, but especially from the field of biology, to detect early outbreak signals with good detection results. In this paper, an outbreak detection model based on danger theory is proposed. Danger theory is a bio-inspired method that replicates how the human body fights pathogens and the literature has shown that it can solve problems mainly related to the detection process. While not many works that apply danger theory have addressed the outbreak detection problem, danger theory has been successfully applied in intrusion [37,38], fraud [39,40], and fault detection problems [41] with good detection performance. Since the capability of danger theory as a good detector is proven in other areas, it motivates us to adapt the artificial immune system (AIS) as an outbreak detection model. The robustness of the immune system mostly lies in the ability of the dendrite cell to sense an early death cell (viewed as an outbreak signal), which can be replicated in an outbreak detection model to reduce the high false alarm rate. Moreover, danger theory offers a multivariate detection approach without relying on a training phase, which can improve a model's robustness.

In the proposed model, a signal formalization approach based on cumulative sum (*CUSUM*) is formalized and a cumulative mature antigen contact value (cMCAV) is proposed to suit the outbreak characteristic and danger theory. In experiments, the model is applied to two outbreak datasets; a real-world dengue outbreak and a synthetic SARS outbreak. To evaluate the model, four evaluation criteria are used: detection rate, specificity, false alarm rate, and accuracy. Then, the result is compared with three statistical control chart approaches: the CUSUM, Exponentially-weighted

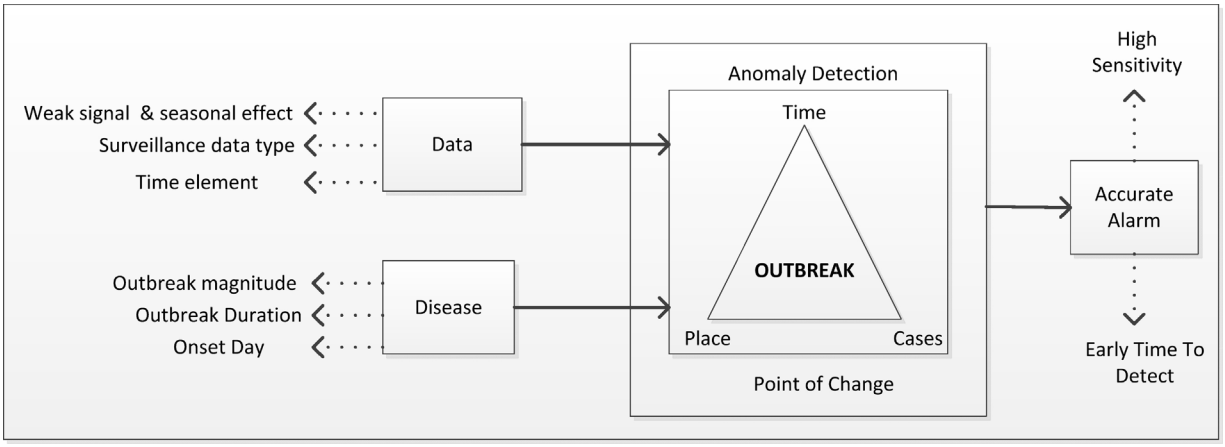


Fig. 1. General concept of outbreak detection.

Moving Average (EWMA), and Moving Average (MA). The proposed model is also compared with two multivariate classifiers, Multilayer Perceptron (MLP) and Naive Bayes (NB).

The remainder of the paper is organized as follows: Section 2 outlines the concept of danger theory, covering the biological background, dendritic cell algorithm, and danger theory as a detector. Section 3 presents the proposed outbreak detection model and Section 4 describes the experiment set-up. Then Section 5 presents the main finding of the study and Section 6 discusses and elaborates on the results. The final section, Section 7, concludes this work.

2. Danger theory

In this section, the concept of danger theory is outlined. It covers the biological background of danger theory, the dendritic cell algorithm, and danger theory application as a detector.

2.1. The biological background

Danger theory relies on the idea that the human immune system is triggered by a danger signal produced by a necrotic cell which unexpectedly dies due to a pathogenic infection [42]. When a cell is infected, the distressed cell establishes a danger zone around itself to mitigate and localize the impact of the attack. The key point of danger theory is that the immune system only reacts to the danger signal which will do harm to health. This goes against conventional immunology theory that the immune system is based on discrimination between self cells and non-self cells that leads the immune system responding to the presence of foreign entities and not to the body's own cells.

In principle, danger theory views all cells in the human body as antigens which have a similar possibility of being infected by harmful antigens. It relies on the function of dendrite cells, a family of cells known as macrophages. As an antigen presenting cell, the dendrite cell controls the activation state of T-cells in the lymph nodes and it has three different states, namely immature state, semi-mature state and mature state. At the beginning of the detection process, the dendrite cells, which are initially born as immature cells in thymus, observe the progress of the body's cells. At this stage, the dendrite cell collects the body cell protein paired with its three signals; pathogen associated molecular patterns (PAMP), danger, and safe signal in cell tissue as shown in Fig. 2. Based on the collected input throughout its life span, the dendrite cell will evolve from being immature into one of two maturation states: either a semi-mature (apoptotic death) or a mature state (necrotic death). At this phase, the dendrite cell is migrated from cell tissue to lymph node. Reaching a 'mature' state indicates that the cell has experienced more danger signals throughout its life span that have been caused by a foreign antigen, wound, etc. If this happen, it indicates that the harmful antigen has been detected and a danger zone will be released. From a mature state, T-cells are activated to release antibody. While a 'semi-mature' state indicates that apoptotic death has occurred and this is seen as part of normal cell function. The semi mature dendrite cells cannot activate T-Cells and they are tolerized to the presented antigen.

By integrating the appropriate mechanisms from danger theory, it is envisioned that danger theory will enhance the efficacy of the currently available outbreak detection systems. Compared with other outbreak detection approaches, danger theory provides inspiration for a robust, highly distributed, adaptive, and autonomous detection mechanism for early outbreak notification with better detection results. Since the identification of the cause of cell distress (either apoptosis or necrosis) is the key for antigen detection in the human body, it provides a solution to overcome the uncertainty problem in early outbreak signals. In this context, all outbreak datasets are considered as antigens and all of them have a similar chance of being affected by foreign antigens. Throughout the monitoring period, the maturity state of the dendrite cell is updated; if the outbreak movement becomes abnormally changed, then a danger/alarm zone will be released indicating that harmful antigens exist.

The characteristics of danger theory are that it is sensitive to any changes occurring inside the body and it can highly discriminate between harmful and normal cells. In line with this, danger theory considers all input data as similar under the assumption that they have a similar chance of being affected by foreign antigens. Therefore when danger theory is applied, the requirement to define input data into either a normal or outbreak class is not required, and moreover, no training phase is required. In a conventional immune system approach, some of the antigens have to be predefined as self cells (the training set) but in practice, the self cells are difficult to define and can change over time. Hence, when changes to these self cells occur during implementation, the error detection rate increases sharply.

2.2. The dendritic cell algorithm

The dendritic cell algorithm (DCA) is derived from the danger theory abstract model Greensmith [43] which is depicted in Fig. 3. In the DCA, the dendrite cell acts as an agent that is responsible for collecting antigen coupled with its three context signals (PAMP, danger, safe) as input to the system. Within computer context, the antigen in DCA presents each record contains in dataset and the signals present the normalized value of selected attributes. Each dendrite cell accumulates the changes that occur in the monitored system and determines which antigen causes the changes. Using the accumulative function in Eq. (1), all input signals are transformed into three cumulative outputs signals; co-stimulatory molecules (CSM), mature, semi-mature.

$$\mathrm{OS}_j(x) = \frac{\sum_{i=0}^{i=3} W_{ij} \times \mathrm{IS}_{ij}(x)}{\sum_{i=0}^{i=3} |W_{ij}|} \tag{1}$$

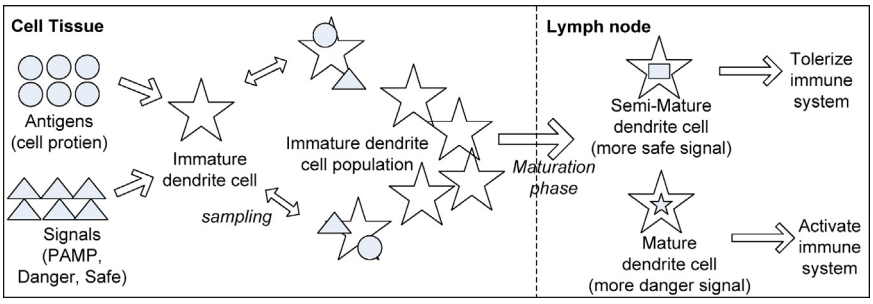


Fig. 2. The biological process of danger theory.

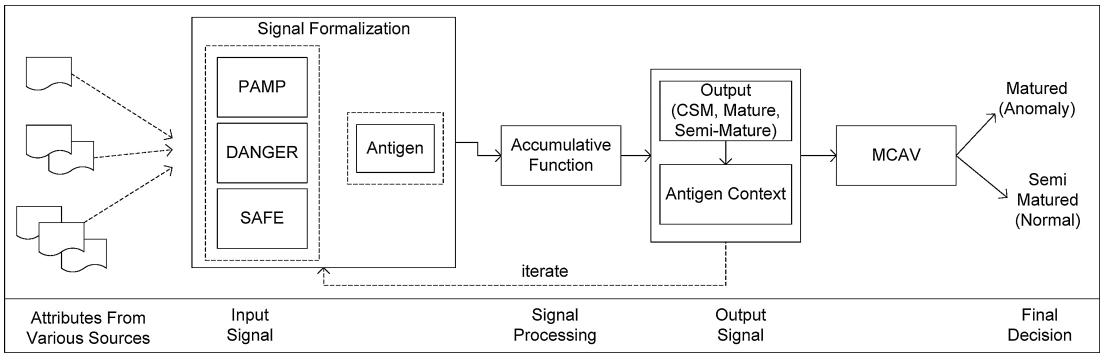


Fig. 3. The concept of danger theory applied in the DCA.

where W is the weight matrix, IS is the input signal, OS is the output signal, i is the input signal categories, and j is the output signal categories.

As a population-based algorithm, the dendrite cell samples input signals and antigens multiple times. This is analogous to sampling a series of suspected antigen in human body such that the dendrite cell will hold several antigens until it matures as shown in Fig. 4. Throughout the sampling process, the experience of each cell is increasing whereby the entire experience is documented in CSM (OS_1), mature (OS_2), and semi-mature (OS_3) as output signals. The sampling process stops when the cell is ready to migrate. This occurs when the OS_1 of that cell reaches the migration threshold and it is then removed from the population for antigen presentation. After migration, the output values of OS_2, and OS_3 are compared in order to derive a context for the presented item. The antigen is termed mature if $OS_2 > OS_3$ or semi mature if $OS_2 < OS_3$. Then the migrated dendrite cell is replaced with a new cell to restart sampling and return to the population. This process is iterated several times.

When learning ends, antigens appear in different contexts. In the last step, the potential anomalous antigen is determined based on the collected context. Termed as the mature context antigen coefficient (MCAV), the anomalous antigen is determined as:

$$\mathrm{MCAV}_i = \frac{\sum \mathrm{Ag_{mi}}}{\sum \mathrm{Ag}_m + \sum \mathrm{Ag_{sm}}} \tag{2}$$

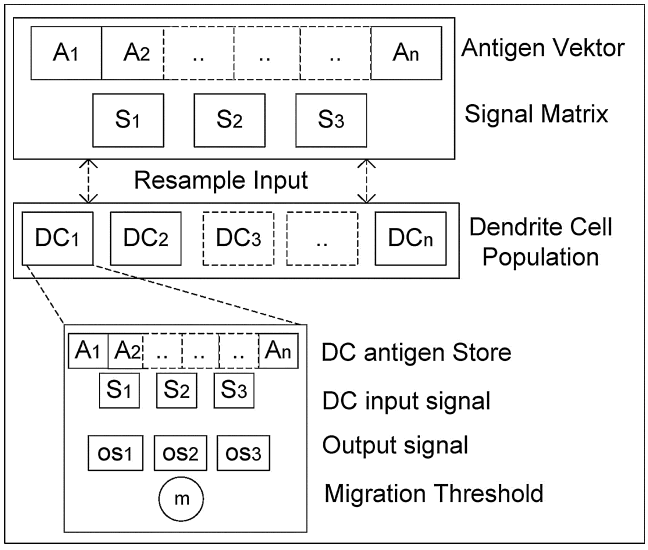


Fig. 4. Dendrite cell resample antigen and signals.

where i refers to the antigen type, $\sum \mathrm{Ag_{mi}}$ refers to the total number of mature antigens of antigen type i, $\sum \mathrm{Ag}_m$ is to the total number of mature antigens, and $\sum \mathrm{Ag_{sm}}$ refers to the total number of semi-mature antigens. If the value of MCAV_i is higher than the anomaly threshold, then the antigen is categorized as an anomalous antigen.

2.3. Danger theory as detector

Danger theory has been used in many studies as one of the anomaly detection techniques. It has been applied mostly in computer networks to detect intruders. The early work in this field was initiated by Greensmith et al. [44], who incorporated the biological danger theory approach into a DCA. This type of algorithm is able to detect network intrusion with better performance compared to negative selection algorithms. Later, a more sensitive version of the DCA was proposed which included new controllable parameters [45]. Following successful trials, danger theory has been applied to address a wide range of computer security network issues such as malicious code detection, misbehaviour in wireless networks, port scanning, spyware, and worm detection [37,38,46–54]. Besides providing accurate detection with a lower false alarm rate, the danger theory framework can be applied as an agent in huge real-time heterogonous networks [55–58]. In line with this, other recent work demonstrates that the danger theory framework can be implemented in real-time flood detection in a big scalable environment with fast response to pattern changes [54]. Moreover, it has been found that detection sensitivity towards intruders increases when danger theory is hybridized with other detection approaches such a negative selection algorithm [38,59].

Beside intrusion detection, a flurry of detection frameworks inspired by danger theory has also been reported in other areas. For example in fraud detection, the danger theory approach is hybridized with a negative selection algorithm to detect fraudulent online video-on-demand transactions [39,60] as well as fraud in runtime malware for Windows processes [40]. Inspired by dendrite cell function, danger theory has also been implemented in the area of fault detection. For example, Ran, et al. [41] adopt the idea of the DCA to diagnose faults in robots and control systems. In line with their work, Prieto, Nino, & Quintana [61] have proved the ability of danger theory as an efficient detector in control systems. They propose the DTAL, which is a goalkeeper strategy in robot soccer. In addition, danger theory has also been successfully applied to detect faults in task scheduling where each job to be scheduled is considered as an antigen and aligned with the cell signal [62]. Other successful applications for fault diagnosis can be seen in Laurentys, Palhares, & Caminhas [63] and Zhou, Fan, & Zhang [64]. As well as acting as an anomaly detector, danger theory has also performed well in other data mining task such as filtering web documents [65] and image classification [66].

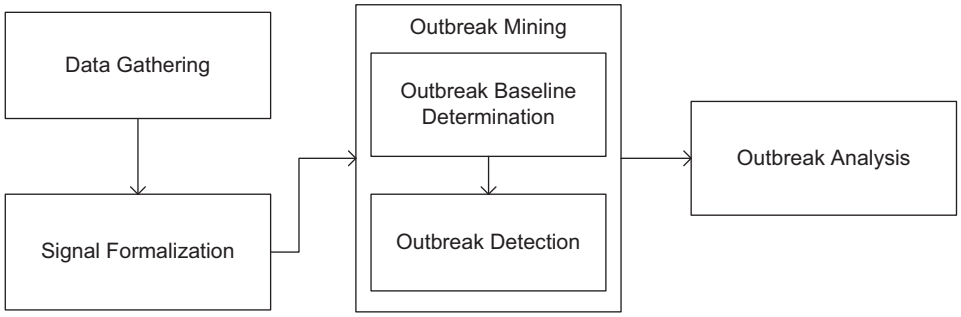


Fig. 5. Outbreak detection model based on the DCA.

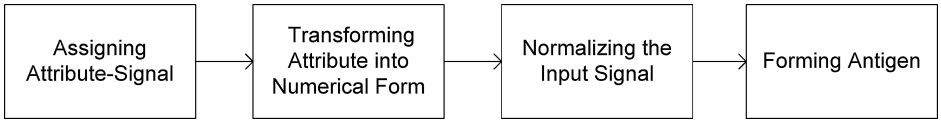


Fig. 6. The signal formalization step.

Based on the successful results in intrusion, fraud, and fault detection, danger theory seems to have the potential to detect outbreak signals with good results. Since, to the best of our knowledge, there does not appear to be a recent outbreak detection model based on danger theory, we propose an outbreak detection model based on danger theory in the next section.

3. Outbreak detection model based on danger theory

The general outbreak detection process consists of a two-step procedure. First, the baseline period that represents the abnormal pattern is obtained and second, the score of the observed data is calculated such that the abnormal cases are detected when they exceed the maximum baseline [11]. Based on this process, the outbreak detection model based on danger theory is presented in Fig. 5. It comprises four processes, data gathering, signal formalization, outbreak mining, and outbreak analysis. The danger theory algorithm (DCA) is used as the detection algorithm.

3.1. Data gathering

This step involves gathering data from original sources. Data can be of different types since the DCA supports real-time multivariate mining. As is usual, the data are pre-processed, where the numerical missing value is replaced with a mean value while Mode for categorical attribute.

3.2. Signal formalization

Signal formalization is the most important step, where data are prepared for the DCA. In this study, data are normalized into an appropriate form to suit outbreak data for the DCA environment. The normalization approach is inspired by the statistical control chart called cumulative sum (CUSUM). Several steps are involved in formalizing the signal: assigning the attribute into an appropriate signal, transforming the data into numerical form, normalizing the input signal, and forming an antigen. Fig. 6 shows the signal formalization steps.

The first step is to select and assign the attribute to an appropriate signal: PAMP, danger, or safe. To match attribute and signal, an expert judgement is required otherwise it can be taken from the literature. The PAMP signal is assigned when the attribute shows an anomalous situation indicator while the safe signal is assigned when no anomalous indicators are present in the attribute. However, where the attribute may or may not indicate an anomalous situation but the probability of an anomaly is higher under normal circumstances and is labelled as a danger signal [43]. Categorical data need to be transformed into numerical data since the DCA only functions in numerical form. For transformation, the categorical value of an attribute is replaced with the total reported cases in a day. For example, the reported symptoms attribute in the WSARE dataset [67] has four categories which are none, respiratory, nausea, and rash. The registered cases in day X are counted according to the attribute value, as shown in Fig. 7.

To make outbreak data suitable for the DCA, a normalization approach based on CUSUM is proposed. CUSUM is a statistical approach primarily used to monitor the planned process in manufacturing operations. This approach monitors the mean of the process and assumes a process remains under control when the cumulative mean is within the control value. The process is considered out of control when a huge shift in movement occurs away from the target value. As recent outbreak activity has a relation with what has happened during the previous day's activities, the cumulative mean shift is taken into consideration for normalization. Since an outbreak is only indicated when there is a sudden increase in cases, the lower side CUSUM is ignored. So, to normalize, the upper side CUSUM is applied, as shown in Eq. (3):

$$C_i^+ = \max[0, x_i - (\mu_0 + K) + C_{i-1}^+] \tag{3}$$

where the C_i^+ is the upper cumulative value at i_{th} observation, x_i is the process at i_{th} observation, μ_0 is the initial mean and K is

Date	Reported Symptom
OCT-09-2003	respiratory
OCT-09-2003	nausea
OCT-09-2003	none
OCT-09-2003	none
OCT-09-2003	nausea
OCT-09-2003	none
OCT-09-2003	respiratory
OCT-09-2003	nausea
OCT-09-2003	respiratory
OCT-09-2003	none

(a) Raw

Date	Reported Symptom	Total Case
OCT-09-2003	None	4
	Respiratory	3
	Nause	3
	Rash	0

(b) Total case based on symptom

Fig. 7. Example of total cases for WSARE dataset.

the allowance value which is chosen between the target μ_0 and out of control value μ_1. K is expressed by $K = (\delta/2)\sigma = |\mu_1 - \mu_0|/2$, where δ is the shift size from standard deviation, σ. The C_i^+ value accumulates deviation from μ_0 that is greater than K which is reset to zero on becoming negative. The starting value is $C_i^+ = 0$.

The normalized value of x_i is C_i^+ if the cumulative sum value is greater or equal to zero, otherwise the value is 0 as applied in Eq. (4):

$$\text{Signal}_j(x_i) = \left\{ \begin{array}{c} C_i^+; C_i^+ >= 0 \\ 0; C_i^+ < 0 \end{array} \right\} \quad \text{where} \quad j = \text{PAMP, danger, @safe} \tag{4}$$

The last step is to form an antigen that functions as a platform to represent the item being analyzed. The antigen is expressed in Eq. (5) as:

$$\text{Antigen(ID,} \quad \text{Antigen}_{\text{ctx}}) \tag{5}$$

where ID refers to the outbreak record to be analyzed and $\text{Antigen}_{\text{ctx}}$ represents the antigen context which is either normal or dangerous. Normally, the monitoring date or registered case ID is chosen as the ID.

3.3. Outbreak mining

In outbreak mining, the normalized outbreak data are mined using the DCA. It has two tasks; first, to determine the outbreak baseline and second, to mine the outbreak dataset with the DCA.

3.3.1.1. Outbreak baseline determination

The outbreak baseline is the default value that must be reached before the DCA can raise an alarm. To determine this value, a previous outbreak dataset covering several years is taken into consideration for pre-mining. The formula for each disease may be different since each disease has a different outbreak definition. If no such definition is available, suggestions from experts can be used as a guide. This value is compared with the MCAV which is generated from the DCA. Eq. (6) shows the general outbreak baseline calculation used in this study:

$$O_{\text{bs}} = \frac{\sum \text{CS}_{\text{ot}B}}{\sum \text{CS}_{\text{ot}B} + \sum \text{CS}_{\text{not}B}} \tag{6}$$

where O_{bs} is an outbreak baseline, $\sum \text{CS}_{\text{ot}B}$ refers to the total cases classified as outbreak, and $\sum \text{CS}_{\text{not}B}$ refers to the total cases classified as non-outbreak

3.3.1.2. Outbreak detection

In this phase, the outbreak data are presented to the DCA for detection. The aim is to generate a MCAV for each antigen that represents the final condition of an outbreak. The general process of this phase consists of setting the initial parameters, updating the input signal and antigen, calculating the MCAV, and categorizing the antigen. To categorize the antigen as outbreak or non-outbreak, a cumulative MCAV (cMCAV is proposed, where the average MCAV for antigens of similar date is calculated, as shown in Eq. (7), and the cMCAV_i is compared with outbreak baseline O_{bs}in Eq. (8):

$$\text{cMCAV}_i = \frac{\sum \text{MCAV of date } j}{\sum \text{case of date } j} \tag{7}$$

$$\text{Status}(x_i) = \left\{ \begin{array}{c} +re; \text{cMCAV}_i > O_{\text{bs}} \\ -re; \text{cMCAV}_i < O_{\text{bs}} \end{array} \right\} \tag{8}$$

where cMCAV_i is the cumulative MCAV, x_i is an antigen at i_{th} observation, $+re$ refers to an anomalous condition or outbreak, and $-re$ refers to normal condition or non-outbreak. The occurrence of an outbreak is notified when $\text{cMCAV}_i > O_{\text{bs}}$.

3.4. Outbreak analysis

Outbreak analysis is under health department control. When DCA raises an alarm, the notification will be verified by a team of health workers. After verification, an immediate prevention plan is arranged. In this stage, the health workers will follow the health surveillance system procedure in managing the outbreak period. A review session will be conducted once the outbreak ends.

4. Experiment set-up

This section explains the experimental data used, how an outbreak is defined and the parameter settings for the DCA. Two outbreak datasets are used to validate the proposed model, dengue outbreak and SARS outbreak data, and are described below:

4.1. Dengue outbreak data

These data are taken from two departments: (1) the emergency visit data from the Vector Control Unit, Seremban Medical Centre, Malaysia and (2) the weather data from Seremban Metrological Centre, Malaysia. The time period for both datasets is from 2003 to 2009. In this study, both datasets are separated into two groups, where the 2003–2005 data is used to determine the outbreak baseline and the 2006–2009 data is used for the monitoring step in the experiment. The emergency visit data consist of 15 mixed attributes while the weather dataset has eight continuous attributes. The attributes selected for mining are shown in Table 1. Both datasets are represented in a weekly-based format. A dengue outbreak in Malaysia is defined as the occurrence of more than one case in the same locality where the onset date between cases is less than 14 days [14]. Therefore, a dengue outbreak is considered to have started if there are additional cases in the observed week at least a case compared to the mean of previous two week cases. The all-clear status is given when no new cases have been reported for 14 days.

4.2. SARS outbreak data

These data are known as WSARE and have been used in [67]. The dataset consists of 100 sets of artificial SARS outbreak datasets with different outbreak release dates. In this study, WSARE7 has been selected for the experiment. The dataset covers the period from 2002 to 2003, and the 2003 records are used for monitoring. WSARE has 12 attributes which are all categorical. The attribute selected for mining is shown in Table 1. In contrast to the dengue dataset, for monitoring purposes the WSARE is represented in a daily-based format. The outbreak definition for WSARE is not clearly defined in

Table 1
Attribute-signal assignment for dengue and WSARE.

Signal	Attribute	
	Dengue	WSARE
PAMP	Emergency visit	Seasons Winter Autumn
Safe	Registered cases	Seasons Summer Spring
Danger	Weather Temperature Humidity Rainfall	Reported Symptom Nausea Respiratory Rash

Table 2
Attribute selection for detection approaches.

Approach	Dengue	WSARE
DCA	Emergency visit & weather	Seasons and reported symptom (nausea, respiratory, rash)
MLP	Emergency visit & weather	Reported symptom (nausea, respiratory, rash)
NB	Emergency visit & weather	Reported symptom (nausea, respiratory, rash)
CUSUM	Emergency visit	Reported symptom (respiratory)
EWMA	Emergency visit	Reported symptom (respiratory)
MA	Emergency visit	Reported symptom (respiratory)

[67]; in his work, a SARS outbreak is defined based on the actual date when the virus is released and it is assumed that the virus will remain in the community for several days, while detection is based on how fast the algorithm can sense the first outbreak day after the release date. Therefore, for the purpose of this study, an outbreak starts a day after the release date and remains present for 14 days, i.e. the algorithm will identify the whole 2 weeks as an outbreak period.

4.3. Selection of attributes and parameter settings

A different approach is used for the datasets in selecting suitable attributes and assigning them to the appropriate signal. For the dengue dataset, attribute selection is based on expert suggestions while in WSARE, the selection is based on [28]. Table 1 summarizes the attribute-signal assignment. The attribute assignment to an appropriate signal is important since it will affect the capability of the DCA to discriminate any anomalies accurately. For example, in the WSARE dataset, the number of cases in winter and autumn (seasonal attribute) is assigned to the PAMP signal because flu incidence rates are high during this period . Conversely, in summer and spring the number of cases falls, therefore both of these seasons are assigned to the safe signal. The general guideline to select and assign attribute to appropriate dendrite cell signal is based on the signal behaviour itself as follows:

- PAMP: Their existence indicates an anomalous situation
- Safe: Their presence indicates no anomalous situation is present
- Danger: Their occurrence may or may not indicate an anomalous situation but the probability of an anomaly is higher than in the normal situation.

For dengue dataset, the emergency visit attribute is selected to represent the PAMP and safe signal. To assign appropriate value for both signals, the dengue outbreak definition is referred such that the total case variant between current week and average previous 2 weeks is used as basis. The following rule shows the process.

```
MaxValue = 100;
Determine Average Cases of the previous two weeks;
If (Cases of Current Week – Average Cases of Previous 2 Weeks) > 1
Case Then
   PAMP signal = 0;
   Safe Signal = MaxValue;
Else
   PAMP Signal = MaxValue × 0.7;
   Safe Signal = 0;
End
```

Table 3
Parameter settings for the DCA.

Parameter	Value
Total DC	100
Iteration	30
Migration threshold	(9) $\frac{\sum_{i=0}^{i=3} \max(IS_{ij}) \times w_{ij}}{2}$
Weight [43]	$W = \begin{bmatrix} w_1 & \frac{w_1}{2} & w_1 \\ 0 & 0 & 1 \\ w_2(1.5) & \frac{w_2}{2} & w_2(-1.5) \end{bmatrix}$; $w_1 = w_2 = 1$
Outbreak threshold	Refer to Eq. (6)

Table 2 summarizes attribute selection for the DCA and for the other five comparative detection approaches examined in this study.

The parameter settings for the DCA are presented in Table 3. The migration threshold is set to 50% of the median value of the input signal, as shown in Eq. (9) in the table.

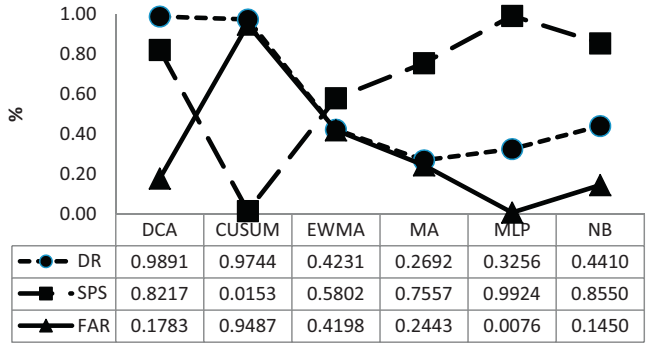
5. Findings

This section presents the performance of the outbreak detection model based on danger theory. The model is evaluated based on four performance metrics: detection rate (DR), specificity (SPS), false alarm rate (FAR), and accuracy (ACC). The DR shows the model accurateness to detect an outbreak while the ability of the model to detect a non-outbreak is evaluated by SPS. The FAR measures how much false detection occurs when a non-outbreak case is detected as an outbreak. Lastly, the ACC checks the accuracy of the model in classifying both classes correctly. For comparison, three statistical detection approaches are selected, namely, CUSUM, EWMA, and MA. These approaches represent the univariate detection approach. In addition, MLP and NB are also compared with the proposed model and represent the multivariate detection approach. MLP and NB rely on a trained model developed from a training phase.

5.1. Dengue outbreak detection

The first evaluation concerns the detection of dengue outbreaks. For the purpose of monitoring, the dataset from 2006–2009 (209 weeks) are presented to the DCA. The dataset is mined 50 times and the average is calculated for further analysis. Fig. 8 summarizes the dengue outbreak detection result in terms of DR, SPS, and FAR using DCA, CUSUM, EWMA, MA, MLP, and NB.

Based on Fig. 6, DCA ranks first and has the highest DR with a score of 0.9891. This indicates that the DCA has high ability to detect a true outbreak week as an outbreak. The DCA is closely followed by CUSUM with a score that is 0.0147 lower than that of



	DCA	CUSUM	EWMA	MA	MLP	NB
DR	0.9891	0.9744	0.4231	0.2692	0.3256	0.4410
SPS	0.8217	0.0153	0.5802	0.7557	0.9924	0.8550
FAR	0.1783	0.9487	0.4198	0.2443	0.0076	0.1450

Fig. 8. DR, SPS, and FAR for dengue outbreak detection using DCA, CUSUM, EWMA, MA, MLP and NB.

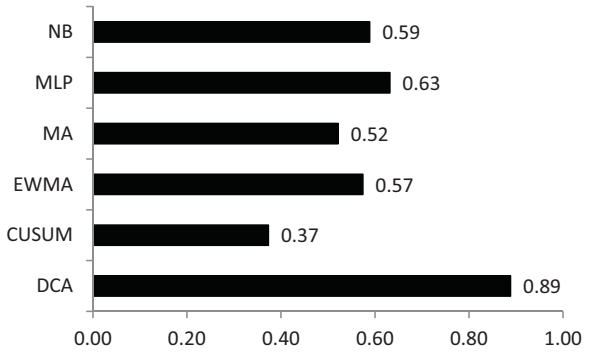


Fig. 9. ACC of the detection approaches applied to the dengue dataset.

Table 4
Standard deviation of DCA and other approaches for dengue dataset.

	DCA	CUSUM	EWMA	MA	MLP	NB
DR	0.013	0.010	0.400	0.509	0.469	0.388
SPS	0.025	0.580	0.171	0.047	0.121	0.024
FAR	0.025	0.545	0.171	0.047	0.121	0.024
ACC	0.016	0.368	0.226	0.262	0.184	0.212

the DCA. Although the DRs of both approaches do not differ much, the DCA performed better than CUSUM as the former consistently detected non-outbreak cases as non-outbreaks. This can be seen from the higher SPS score (0.8217) and lower FAR score (0.1783) for the DCA. In contrast, CUSUM failed to classify many non-outbreak cases as non-outbreaks with a high FAR (0.9847) recorded. When we compare the results of the MLP with the other methods, MLP seems to be the most accurate detector as it is able to detect non-outbreak weeks as non-outbreaks with the highest SPS (0.9924) and lowest FAR (0.0076). However, MLP is less performed in detecting true outbreak weeks (DR is only 0.3256). The performance of NB seems comparable with MLP but with a higher DR and lower SPS. Meanwhile MA also shows similar capability as MLP but its ability to detect true outbreak weeks is the worst among others approaches; a score of only 0.2692 for DR is recorded. In the case of the EWMA's detection performance, an average score between DR, SPS, and FAR are recorded.

In terms of ACC, DCA is the most accurate model with a result of 0.89; it performs better than all the other approaches tested, as shown in Fig. 9.

In every experiment, DCA generates a consistent result for each performance metrics that shows it robust capability towards different dengue outbreak pattern. This can be seen in Table 4 where DCA has the lowest standard deviation in comparison to other approaches. Higher DR and ACC are recorded in other approaches.

In general, the stability between DR (high) and FAR (low) reveals that the DCA is the most accurate detection approach among the other approaches for this dataset. The information in Table 5 depicts the average true positive (TP), true negative (TN), false negative (FN), and false positive (FP) of the DCA after 50 experiments. The total number of outbreak weeks and non-outbreak weeks is 78 and 131, respectively. Fig. 10 presents the outbreak detection specifically from week 150 to week 209. The dengue outbreak threshold is set to 0.34.

Table 5
Average TP, TN, FN, and FP of DCA after 50 experiments for dengue dataset.

TP	TN	FN	FP	Total weeks
77.9	107.4	23.6	0.1	
No. of outbreak days: 78				209
No. of non-outbreak days: 131				

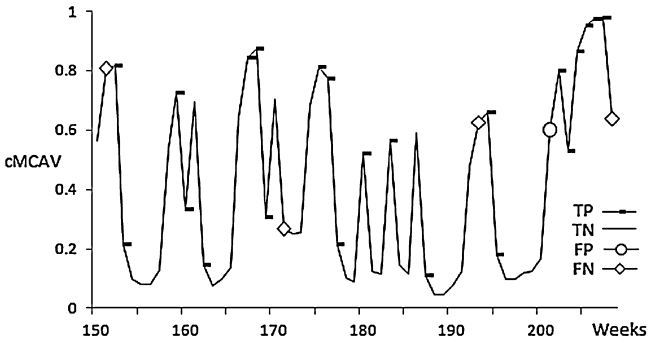


Fig. 10. Outbreak detection for dengue dataset from week 150 to week 209.

5.2. SARS outbreak detection

The second evaluation concerns the detection of SARS outbreaks. For this part of the experiment, the WSARE7 is presented to the DCA and it has to monitor 365 days (2003). Among those 365 days, there are 14 days which are labelled as outbreak days. Fig. 11 summarizes the SARS outbreak detection result in terms of DR, SPS, and FAR using DCA, CUSUM, EWMA, MA, MLP, and NB.

Based on Fig. 9, three approaches score a high DR, which are DCA, CUSUM, and EWMA. CUSUM has the best DR result where it nearly detects all 14 outbreak days as outbreaks (0.9900). It is closely followed by DCA (0.9569) while EWMA ranks in third place (0.9286). The DR for MA and NB are not as good as in the other approaches scoring only 0.5714 and 0.5000, respectively. Nevertheless, they did show better performance in detecting non-outbreak days where they generated the least FAR (MA 0.0171; NB 0.0228) and higher SPS (MA 0.9829; NB 0.9772) compared to the other approaches. In the case of MLP, it has a capability similar to MA and NB in terms of SPS and FAR; however, it failed to detect true outbreak weeks as outbreaks and had the worst DR score (0.2857). A good detection model needs to have balance between a high DR, high SNS, and low FAR, but the results for MA, NB, and MLP fail to meet this condition. In line with this proviso, DCA is found to be the best model for this dataset, generating a lowest FAR (0.0036) and highest SNS (0.9964) while at the same time having a good DR (0.9569). Comparable to DCA, the EWMA also has a similar capability for discriminating outbreak and non-outbreak days when generating a consistently high SNS (0.9459), low FAR (0.0541) and high DR (0.9286).

The model performance in terms of ACC shows that most models generate high classification accuracy with a result of between 0.94 and 0.98. The DCA recorded a highest level of ACC (0.98) which is not significantly different from that of other approaches. However, CUSUM scored the worst ACC with 0.56. The ACC of each approach is shown in Fig. 12.

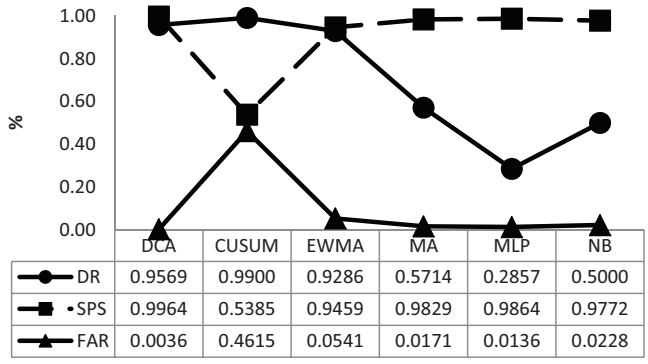


Fig. 11. DR, SPS, and FAR for SARS outbreak detection using DCA, CUSUM, EWMA, MA, MLP, and NB.

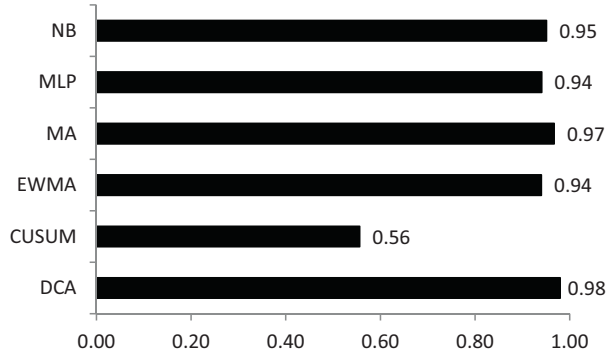


Fig. 12. ACC of the detection approaches applied to the SARS outbreak.

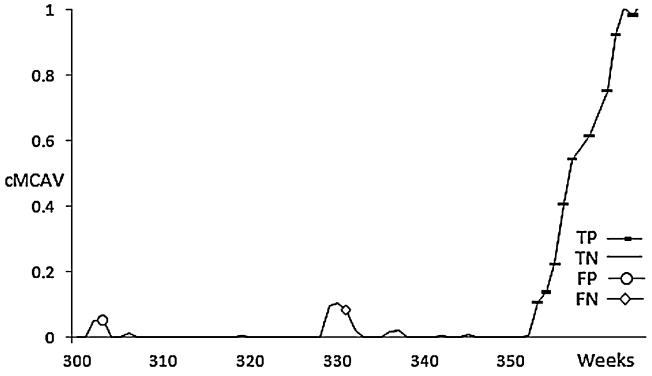


Fig. 13. Outbreak detection for WSARE7 data from day 300 to day 365.

Table 6
Standard deviation of DCA and other approaches for SARS dataset.

	DCA	CUSUM	EWMA	MA	MLP	NB
DR	0.044	0.023	0.020	0.273	0.475	0.323
SPS	0.025	0.324	0.036	0.010	0.007	0.014
FAR	0.025	0.324	0.036	0.010	0.007	0.014
ACC	0.045	0.300	0.028	0.009	0.028	0.021

Table 7
Average TP, TN, FN, and FP of DCA after 50 experiments for SARS outbreak.

TP	TN	FN	FP	Total days
12.44	350.75	1.25	0.56	
No. of outbreak days: 14				365
No. of non-outbreak days: 351				

The information in Table 6 compares the standard deviation between DCA and other approaches. In comparison to dengue outbreak detection, most models have smaller standard deviation mainly for SPS, FAR, and ACC. DCA remains constant as the most consistent approach when produces the lowest standard deviation in every performance metrics.

Table 7 depicts the average TP, TN, FN, and FP of the DCA after 50 experiments. Fourteen days are classified as outbreak days while 351 are non-outbreak days. Fig. 13 presents the outbreak detection from day 300 to day 365. The SARS outbreak threshold is set to 0.055.

6. Discussion

For a model to be a good detection model, it must have the ability to generate a balanced result for DR, FAR, and SPS when detecting anomalies. The above section showed that the proposed outbreak detection model based on danger theory generates a consistent result between high DR, SPS and lower FAR. The preference matrix in Table 8 reveals the ability of danger theory to act as a framework for an outbreak detection model when DCA has the lowest weight score after considering all performance metrics. Through the preference matrix [68], the accumulative score of each of the performance metrics is given based on the priority of each metric; highest (DR, SPS, ACC) or lowest priority (FAR). A score of 1 is given for the best mining and score 6 for the worst result according to the priority metric. The accumulative weight score ($\sum$S1, $\sum$S2) represents the total scores of all priorities. The lowest score indicates the best model. Based on the last column of Table 8 ($\sum$S1 + $\sum$S2), DCA has the lowest score, which indicates the most accurate approach for both outbreaks.

There are several benefits that can be derived from the proposed immune system, as summarized in Table 9. First, danger theory overcomes the inconsistent outbreak signal problem. This can be seen from how danger theory perceives the input data. In danger theory, all data inputs are considered as antigens where they have a similar chance of becoming infected with a harmful pathogen. Based on the generated MCAV that represents the antigen life experience, it determines whether the antigen has died due to infection (outbreak) or not (non-outbreak). Since outbreak and non-outbreak periods are hard to define in real life, danger theory eliminates the assignment of input data to both of these periods during model development. Therefore, there is no need for a training phase and or a pre-developed model in danger theory. This is in contrast to some other detection approaches such as MLP and NB, where the model must first be trained. As a result, the robustness of these models can decrease when the unseen pattern is not the same as that learned during the training phase. Based on the experiments, danger theory can highly discriminate between outbreaks and non-outbreaks with consistent detection results even though no training phase is involved. Moreover, since no training phase is required, the proposed model is able to process real-time input and has high potential for implementation as a real-time system.

Second, danger theory is a multivariate detection approach. In a biological immune system, a dendrite cell processes three input signals (PAMP, danger, safe) into three output signals (CSM, mature, semi-mature). During this process, DCA may accept various inputs in term of cell protein and classify them either into PAMP, danger, or safe. As an outbreak signal is often weak and not consistent, the use of multiple input factors may improve detection performance in comparison to those methods that rely on a single predictive factor

Table 8
The preference matrix.

	Dengue					SARS					$\sum$S1 + $\sum$S2
	DR	SPS	FAR	ACC	$\sum$S1	DR	SPS	FAR	ACC	$\sum$S2	
DCA	1	3	3	1	8	2	1	1	1	5	13
CUSUM	2	6	6	6	20	1	6	6	6	19	39
EWMA	4	5	5	4	18	3	5	5	5	18	36
MA	6	4	4	5	19	4	3	3	2	12	31
MLP	5	1	4	2	12	6	2	2	4	14	26
NB	3	2	1	3	9	5	4	4	3	16	25

Table 9
Relationship between outbreak issues and danger theory as well as the benefits of the outbreak detection model based on danger theory.

Outbreak issues	Effect	Danger theory characteristics	Outbreak detection model based on danger theory	Benefits
Outbreak and outbreak period is hard to determine. Most of the time, only outbreaks period are available for model development	Models lose robustness when unseen outbreak patterns vary from those of the trained model. This reduces detection capability when the unseen patterns vary from those in the trained model	Antigens have a similar chance of becoming infected with a harmful pathogen	No need to define outbreak and non-outbreak class. In other words, no training phase involved	Improves detection performance by balancing the DR and FAR Offers a new method for multivariate detection Has potential to be applied as a real-time system as applied in intrusion detection areas
Weak and inconsistent outbreak signal	Imbalance in the result between DR and FAR	The dendrite cell process uses multiple input signals	Accepts multiple input factors to improve detection performance rather than relying on a single predictive factor	

[28]. As an example, three attributes for a dengue outbreak (average humidity, average temperature and average rainfall) are taken as inputs for the danger signal while the emergency visit statistic is assigned to either a PAMP or safe signal. From this mixture of multiple input factors, the DCA generates a better result detection performance compared to the univariate approach (CUSUM, EWMA, and MA). Moreover, the experimental result for SARS outbreak detection also indicates a similar performance. Hence, the DCA indirectly improves on the CUSUM result. This can be clearly seen from in the results for both outbreak datasets where DCA and CUSUM generated a comparable DR; however, the DCA performs better than CUSUM in terms of FAR and SPS. This indicates that the multivariate approach offered by danger theory can increase outbreak detection performance.

7. Conclusion

Inspired by the success of other detection-based solutions in the literature, this study proposed a novel outbreak detection model based on danger theory. In the proposed model, a signal formalization approach based on CUSUM was formalized and a cumulative mature antigen contact value (cMCAV) was employed to suit the outbreak characteristic and danger theory. Based on experiments performed on two outbreak datasets, dengue and SARS, the proposed model was found to be more robust than other available models when dealing with inconsistent outbreak signals. Although there is no training phase involved, the danger theory-based outbreak detection model is able to handle new unknown outbreak patterns and can highly discriminate outbreak and non-outbreak cases with a consistent high DR, high SNS, and lower FAR. This can be seen in the case of the dengue outbreak dataset, where the proposed model produced significantly better results than CUSUM, EWMA, moving average, multilayer perceptron, and naïve Bayes. Also, in the case of the SARS outbreak dataset, the proposed model produced a performance that was comparable to that of EWMA. Since danger theory is a multivariate approach, it offers a new alternative for outbreak detection that can deal with vague and inconsistent outbreak signals for different types of data. To further evaluate the effectiveness of the proposed model, further analysis will be conducted on different outbreak diseases, where the performance of the proposed model will be investigated when dealing with different outbreak shifts. The proposed outbreak detection model will also be extended as a real-time agent-based model.

Acknowledgements

The authors gratefully acknowledge Seremban Public Health Department, Malaysia and Faculty of Health Sciences, UKM, Malaysia as well as the Exploratory Research Grant Scheme for supporting this research project through grant ERGS/1/2011/STG/UKM/02/49.

References

[1] X. Jiang, G.F. Cooper, A Bayesian spatio-temporal method for disease outbreak detection, Am. Med. Inform. Assoc. 17 (2010) 462–471.
[2] A. Wilder-Smith, The severe acute respiratory syndrome: impact on travel and tourism, Travel Med. Infect. Dis. 4 (2006) 53–60.
[3] K.D. Kochanek, J. Xu, S.L. Murphy, A.M. Minini, H.-C. Kung, Deaths: preliminary data for 2009, in: National Vital Statistics Reports, U.S. Department of Health and Human Services, CDC, 2011.
[4] J.B.S. Ong, M.I.C. Chen, A.R. Cook, H.C. Lee, V.J. Lee, R.T.P. Lin, P.A. Tambyah, L.G. Goh, Real-time epidemic monitoring and forecasting of H1N1-2009 using influenza-like illness from general practice and family doctor clinics in Singapore, PLoS ONE 5 (2010) e10036.
[5] H.-I. Kuo, C.-C. Chen, W.-C. Tseng, L.-F. Ju, B.-W. Huang, Assessing impacts of SARS and Avian Flu on international tourism demand to Asia, Tour. Manag. 29 (2008) 917–928.
[6] M. McAleer, B.-W. Huang, H.-I. Kuo, C.-C. Chen, C.-L. Chang, An econometric analysis of SARS and Avian Flu on international tourist arrivals to Asia, Environ. Model. Softw. 25 (2010) 100–106.
[7] G. Shmueli, H. Burkom, Statistical challenges facing early outbreak detection in biosurveillance, Technometrics 52 (2010) 39–51.
[8] WHO, Disease Outbreak, World Health Organization (WHO), 2013.
[9] L. Hsin-Min, D. Zeng, C. HsinChun, Prospective infectious disease outbreak detection using markov switching models, IEEE Trans. Knowl. Data Eng. 22 (2010) 565–577.
[10] G.F. Cooper, D.H. Dash, J.D. Levander, W.-K. Wong, W.R. Hogan, M.M. Wagner, Chapter 18 – Bayesian methods for diagnosing outbreaks, in: M.M. Wagner, A.W. Moore, R.M. Aryel (Eds.), Handbook of Biosurveillance, Academic Press, Burlington, 2006, pp. 273–288.
[11] W. Wong, A. Moore, Chapter 14 – Classical time series methods for biosurveillance, in: M.M. Wagner, A.W. Moore, R.M. Aryel (Eds.), Handbook of Biosurveillance, Academic Press, Burlington, 2006.
[12] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey, ACM Comput. Surv. 41 (2009) 1–58.
[13] M.F. Mohamad Mohsin, A.R. Hamdan, A. Abu Bakar, A review on anomaly detection in disease outbreak detection, in: The 1st International Conference on Information Science and Management (ICoCSIM), Danau Toba, Indonesia, 2012, pp. 22–28.
[14] S.B. Seng, A.K. Chong, A. Moore, Geostatistical modelling, analysis and mapping of epidemiology of Dengue fever in Johor State, Malaysia, in: 17th Annual Colloquium of the Spatial Information Research Centre (SIRC 2005), Dunedin, New Zealand, 2005.
[15] Z. Awang Long, A. Abu Bakar, A. Razak Hamdan, M. Sahani, in: L. Cao, Y. Feng, J. Zhong (Eds.), Multiple Attribute Frequent Mining-Based for Dengue Outbreak Advanced Data Mining and Applications, Springer, Berlin, Heidelberg, 2010, pp. 489–496.
[16] K. Bork, B. Klein, K.M. Trautner, U. Pedersen, E. Heegaard, Surveillance of ambulance dispatch data as a tool for early warning, Eurosurveillance 11 (2006) 229–233.
[17] W.R. Hogan, F.-C. Tsui, O. Ivanov, P.H. Gesteland, S. Grannis, J.M. Overhage, J.M. Robinson, M.M. Wagner, Detection of pediatric respiratory and diarrheal outbreaks from sales of over-the-counter electrolyte products, J. Am. Med. Inform. Assoc. 10 (2003) 555–562.
[18] R. Watkins, S. Eagleson, B. Veenendaal, G. Wright, A. Plant, Applying cusum-based methods for the detection of outbreaks of Ross River virus disease in Western Australia, BMC Med Inform. Decis. Mak. 8 (2008) 37.
[19] A.A. Bakar, Z. Kefli, S. Abdullah, M. Sahani, Predictive models for dengue outbreak using multiple rulebase classifiers, in: International Conference on Electrical Engineering and Informatics (ICEEI), 2011, pp. 1–6.

[20] N.A. Husin, N. Salim, A.R. Ahmad, Modeling of dengue outbreak prediction in Malaysia: a comparison of neural network and nonlinear regression model, in: International Symposium on Information Technology (ITSim), 2008, pp. 1–4.
[21] Z. Jie, L. Jie, Z. Guangquan, A hybrid knowledge-based prediction method for avian influenza early warning, in: IEEE International Conference on Systems, Man and Cybernetics, 2009, pp. 617–622.
[22] X. Wei, H. Zhen-Wen, M. Jian, A neural netwok based approach to detect influenza epidemics using search engine query data, in: International Conference on Machine Learning and Cybernetics (ICMLC), 2010, pp. 1408–1412.
[23] Y. Wu, G. Lee, X. Fu, H. Soh, T. Hung, Mining weather information in dengue outbreak: predicting future cases based on wavelet, SVM and GA, in: S.-I. Ao, L. Gelman (Eds.), Advances in Electrical Engineering and Computational Science, Springer, Netherlands, 2009, pp. 483–494.
[24] F. Xiuju, C. Liew, H. Soh, G. Lee, T. Hung, N. Lee-Ching, Time-series infectious disease data analysis using SVM and genetic algorithm, in: IEEE Congress on Evolutionary Computation, 2007, pp. 1276–1280.
[25] M.F. Mohamad Mohsin, A.R. Hamdan, A. Abu Bakar, The preliminary design of outbreak detection model based on inspired immune system, in: L.T. Ngo, A. Abraham, L.T. Bui, L.Q. Don, E. Corchado, C. Yun-Hoy, K. Ma (Eds.), Third World Congress on Information and Communication Technologies (WICT 2013), Hanoi, Vietnam, 2013.
[26] K. Das, J. Schneider, D.B. Neill, Anomaly pattern detection in categorical datasets, in: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, Las Vegas, NV, USA, 2008, pp. 169–176.
[27] F.-C. Tsui, M.M. Wanger, J. Espino, Case detection algorithm, in: M.M. Wagner, A.W. Moore, R.M. Aryel (Eds.), Handbook of Biosurveillance, Academic Press, Burlington, 2006, pp. 199–216 (Chapter 13).
[28] W.-K. Wong, A. Moore, G. Cooper, M. Wagner, What's strange about recent events (WSARE): an algorithm for the early detection of disease outbreaks, J. Mach. Learn. Res. 6 (2005) 1961–1998.
[29] B. Lawson, E. Fitzhugh, S. Hall, C. Franklin, L. Hutwagner, G. Seeman, A. Craig, Multifaceted syndromic surveillance in a public health department using the early aberration reporting system, Public Heal. Manag. Pract. 11 (2005) 274–281.
[30] C.A. Bradley, H. Rolka, D.W. Loonsk, BioSense: implementation of a national early event detection and situational awareness system, Morb. Mortal. Wkly. Rep. (2005) 11–19.
[31] A. Earnest, M. Chen, D. Ng, L. Sin, Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore, BMC Heal. Serv. Res. 5 (2005) 1–8.
[32] E.N. Naumova, E. O'Neil, I. MacNeill, INFERNO: a system for early outbreak detection and signature forecasting, Morb. Mortal. Wkly. Rep. (2005) 77–83.
[33] C.-T. Tsai, F.-C. Sung, P. Chen, S.-C. Lin, Exploring the spatial and temporal relationships between mosquito population dynamics and dengue outbreaks based on climatic factors, Stoch. Environ. Res. Risk Assess. (2011) 1–10.
[34] V. Lampos, N. Cristianini, Tracking the flu pandemic by monitoring the social web, in: 2nd International Workshop on Cognitive Information Processing (CIP), 2010, pp. 411–416.
[35] E. Quincey, P. Kostkova, Early warning and outbreak detection using social networking websites: the potential of twitter, in: P. Kostkova (Ed.), Electronic Healthcare, Springer, Berlin, Heidelberg, 2010, pp. 21–24.
[36] Z. Xichuan, Y. Jieping, F. Yujie, Tuberculosis surveillance by analyzing google trends, IEEE Trans. Biomed. Eng. 58 (2011) 2247–2254.
[37] O. Chung-Ming, C.R. Ou, Immunity-inspired host-based intrusion detection systems, in: 5th International Conference on Genetic and Evolutionary Computing (ICGEC), 2011, pp. 283–286.
[38] T.S. Sobh, W.M. Mostafa, A cooperative immunological approach for detecting network anomaly, Appl. Soft Comput. 11 (2011) 1275–1283.
[39] R. Huang, H. Tawfik, A.K. Nagar, On the use of innate and adaptive parts of artificial immune systems for online fraud detection, in: IEEE 5th International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), 2010, pp. 1669–1676.
[40] S. Manzoor, M.Z. Shafiq, S.M. Tabish, M. Farooq, A sense of 'danger' for windows processes, in: P. Andrews, J. Timmis, N.L. Owens, U. Aickelin, E. Hart, A. Hone, A. Tyrrell (Eds.), Artificial Immune Systems, Springer, Berlin, Heidelberg, 2009, pp. 220–233.
[41] B. Ran, J. Timmis, A. Tyrrell, The diagnostic dendritic cell algorithm for robotic systems, in: IEEE Congress on Evolutionary Computation (CEC), 2010, pp. 1–8.
[42] P. Matzinger, Tolerance, danger and the extended family, Annu. Rev. Immunol. (1994) 991–1045.
[43] J. Greensmith, The Dendritic Cell Algorithm, University of Nottingham, 2007.
[44] J. Greensmith, J. Twycross, U. Aickelin, Dendritic cells for anomaly detection, in: IEEE Congress on Evolutionary Computation, 2006. CEC 2006, 2006, pp. 664–671.
[45] J. Greensmith, U. Aickelin, The deterministic dendritic cell algorithm, in: Proceedings of the 7th international conference on Artificial Immune Systems, Springer-Verlag, Phuket, Thailand, 2008, pp. 291–302.
[46] Y. Al-Hammadi, U. Aickelin, J. Greensmith, DCA for bot detection, in: IEEE Congress on Evolutionary Computation (CEC 2008), 2008, pp. 1807–1816.
[47] R. Fu, K. Zheng, T. Lu, D. Zhang, Y. Yang, Biologically inspired anomaly detection for hierarchical wireless sensor networks, J. Netw. 7 (2012) 1214–1219.
[48] A. Iqbal, M.A. Maarof, Danger theory and intelligent data processing, World Acad. Sci. Eng. Technol. 3 (2007) 646–649.
[49] Z. Junmin, L. Yiwen, A novel intrusion detection model based on danger theory, in: Pacific-Asia Workshop on Computational Intelligence and Industrial Application, 2008, pp. 867–871.
[50] J. Kim, P. Bentley, C. Wallenta, M. Ahmed, S. Hailes, Danger is ubiquitous: detecting malicious activities in sensor networks using the dendritic cell algorithm, in: H.a.C.J. Bersini (Ed.), ICARIS 2006, Springer-Verlag, Berlin, 2006, pp. 390–403.
[51] H. Salmon, C. Farias, P. Loureiro, L. Pirmez, S. Rossetto, P.A. Rodrigues, R. Pirmez, F. Delicato, L. Costa Carmo, Intrusion detection system for wireless sensor networks using danger theory immune-inspired techniques, Int. J. Wirel. Inf. Netw. (2012) 1–28.
[52] F.X. Sun, Z.G. Wu, Immune danger theory based model for syn flooding attack situation awareness, J. Adv. Mater. Res. 181–182 (2011) 66–71.
[53] Y. Jie, T. Chengyu, C. Yuxi, X. Yue, G. Yichun, The dendritic cell algorithm optimized by chemokine simulator in spyware detection, in: International Conference on Computer Science and Network Technology (ICCSNT), 2011, pp. 1012–1016.
[54] N.B.I. Al-Dabagh, I.A. Ali, Design and implementation of artificial immune system for detecting flooding attacks, in: International Conference on High Performance Computing and Simulation (HPCS), 2011, pp. 381–390.
[55] O. Chung-Ming, W. Yao-Tien, C.R. Ou, Intrusion detection systems adapted from agent-based artificial immune systems, in: IEEE International Conference on Fuzzy Systems (FUZZ), 2011, pp. 115–122.
[56] C.-M. Ou, Y.-T. Wang, C.R. Ou, Multiagent-based dendritic cell algorithm with applications in computer security, in: N. Nguyen, C.-G. Kim, A. Janiak (Eds.), Intelligent Information and Database Systems, Springer, Berlin, Heidelberg, 2011, pp. 466–475.
[57] F. Hashim, K.S. Munasinghe, A. Jamalipour, On the negative selection and the danger theory inspired security for heterogeneous networks, IEEE Trans. Wireless Commun. 19 (2012) 74–84.
[58] F. Haidong, Y. Xiguo, W. Na, Multi-agents artificial immune system (maais) inspired by danger theory for anomaly detection, in: International Conference on Computational Intelligence and Security Workshops, 2007, pp. 570–573.
[59] Z. Junmin, L. Yiwen, Integrating innate and adaptive immunity for worm detection, in: International Conference on Computational Intelligence for Modelling Control & Automation, 2008, pp. 645–650.
[60] R. Huang, H. Taufik, A.K. Nagar, Artificial dendritic cells algorithm for online break-in fraud detection, in: Second International Conference on Developement in eSystems Engineering, 2009.
[61] C.E. Prieto, F. Nino, G. Quintana, A goalkeeper strategy in robot soccer based on danger theory, in: IEEE Congress on Evolutionary Computation, 2008. CEC 2008 (IEEE World Congress on Computational Intelligence), 2008, pp. 3443–3447.
[62] N. Lay, I. Bate, Improving the reliability of real-time embedded systems using innate immune techniques, Evol. Intell. 1 (2008) 113–132.
[63] C.A. Laurentys, R.M. Palhares, W.M. Caminhas, Design of an artificial immune system based on danger model for fault detection, Expert Syst. Appl. 37 (2010) 5145–5152.
[64] D. Zhou, Z. Fan, W. Zhang, Research on AMU fault detection algorithm based on immune danger theory, in: Control and Decision Conference (CCDC), 2011, pp. 1957–1961 (in Chinese).
[65] U. Aickelin, S. Cayzer, The danger theory and its application to artificial immune systems, in: 1st International Conference on ARtificial Immune Systems (ICARIS-2002), Canterbury, UK, 2002, pp. 141–148.
[66] E. Bendiab, M. Kholladi, The danger theory applied to vegetal image pattern classification, in: P. Liò, G. Nicosia, T. Stibor (Eds.), Artificial Immune Systems, Springer, Berlin, Heidelberg, 2011, pp. 406–418.
[67] W.-K. Wong, Data Mining for Disease Outbreak Detection, Carnigie Mallon University, Pittsburgh, 2004.
[68] L. Al Shalabi, Z. Shaaban, Normalization as a preprocessing engine for data mining and the approach of preference matrix, in: International Conference on Dependability of Computer Systems, DepCos-RELCOMEX '06, 2006, pp. 207–214.