JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY
https://e-journal.uum.edu.my/index.php/jict

How to cite this article:
Mohamad Mohsin, M. F., Abu Bakar, A., Hamdan, A. R., Sahani, M., & Mohd Ali, Z. (2023). Dengue outbreak detection model using artificial immune system: A Malaysian case study. *Journal of Information and Communication Technology, 22*(3), 399-419. https://doi.org/10.32890/jict2023.22.3.4

Dengue Outbreak Detection Model Using Artificial Immune System: A Malaysian Case Study

*[1]Mohamad Farhan Mohamad Mohsin, [2]Azuraliza Abu Bakar,
[3]Abdul Razak Hamdan, [4]Mazrura Sahani &[5]Zainudin Mohd Ali
[1]School of Computing, Universiti Utara Malaysia, Malaysia
[2,3]Faculty of Science & Information Technology,
Universiti Kebangsaan Malaysia, Malaysia
[4]Faculty of Health Sciences,
Universiti Kebangsaan Malaysia, Malaysia
[5]Negeri Sembilan State Health Department,
Negeri Sembilan, Malaysia

*[1]farhan@uum.edu.my
[2] azuraliza@ukm.edu.my
[3]arh@ukm.edu.my
[4]mazrura@ukm.edu.my
[5] drzainudin@gmail.com
*Corresponding author

Received 15/8/2022 Revised 22/2/2023 Accepted 8/5/2023 Published: 24/7/2023

ABSTRACT

Dengue is a virus that is spreading quickly and poses a severe threat in Malaysia. It is essential to have an accurate early detection system that can trigger prompt response, reducing deaths and morbidity.

Nevertheless, uncertainties in the dengue outbreak dataset reduce the robustness of existing detection models, which require a training phase and thus fail to detect previously unseen outbreak patterns. Consequently, the model fails to detect newly discovered outbreak patterns. This outcome leads to inaccurate decision-making and delays in implementing prevention plans. Anomaly detection and other detection-based problems have already been widely implemented with some success using danger theory (DT), a variation of the artificial immune system and a nature-inspired computer technique. Therefore, this study employed DT to develop a novel outbreak detection model. A Malaysian dengue profile dataset was used for the experiment. The results revealed that the proposed DT model performed better than existing methods and significantly improved dengue outbreak detection. The findings demonstrated that the inclusion of a DT detection mechanism enhanced the dengue outbreak detection model's accuracy. Even without a training phase, the proposed model consistently demonstrated high sensitivity, high specificity, high accuracy, and lower false alarm rate for distinguishing between outbreak and non-outbreak instances.

Keywords: Aedes, artificial immune system, danger theory, dengue, outbreak.

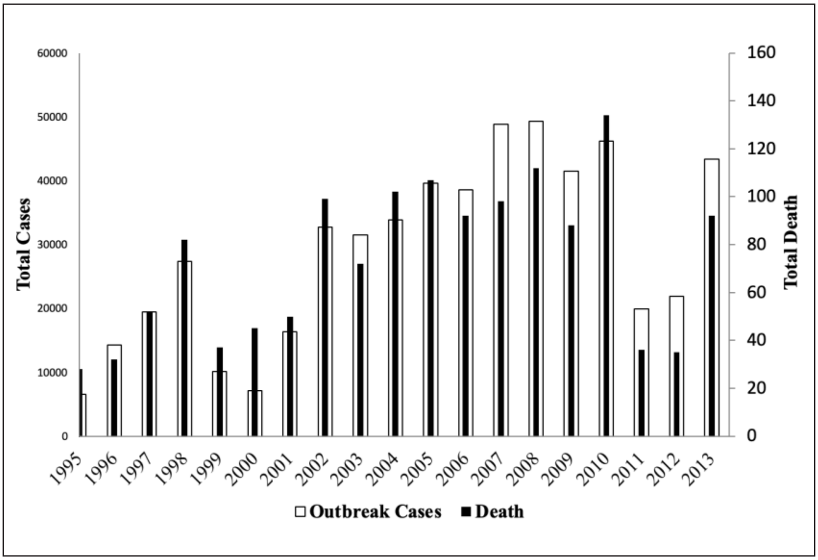
INTRODUCTION

Dengue is a contagious disease that is among the fastest spreading in the world. Female Aedes mosquitoes that feed on human blood spread the disease, which is transmitted through mosquitoes. A person with dengue suffers a sudden high fever accompanied by muscle and joint pains, vomiting, and rashes. These symptoms usually develop from day 4 to day 10 after an infected Aedes mosquito has bitten a person. Normally, dengue requires a 14-day recovery period (WHO, 2021). Dengue haemorrhagic fever (DHF) is a rare, dangerous consequence of dengue fever that includes abnormal fever, plasma leakage, fluid accumulation, respiratory distress syndrome, severe bleeding, or circulatory organ failure.

Dengue fever outbreaks have been more common in recent decades worldwide. It is a disease that affects around 2.5 billion people worldwide, most of whom live in Asia Pacific nations such as Malaysia (Suppiah et al., 2018). Furthermore, 50 million dengue cases, of which 500,000 were DHF, and 12,500 deaths have been reported annually (CDC, 2021). In today's globalised world, disease outbreaks may spread rapidly across borders through trade and tourism (WHO, 2021). In Malaysia, dengue fever was first detected in 1902 and DHF in 1962. As a result of the increase in the population density in urban areas, rapid development, poor environmental cleanliness, a high Aedes breeding index, and changes in the climate, the country has recently experienced an upsurge in the incidence of dengue (Ahmad et al., 2018). There has also been an upward trend in its incidence over the past few decades despite the implementation of various campaigns and prevention control programmes (Mudin, 2015), as shown in Figure 1 as reported by the Malaysian Ministry of Health (Loh, 2016). Dengue outbreaks spiked in 2015, and it remains the most serious type of outbreak in the country due to the effect of climate change (Cheong et al., 2013). In 2020, the reported dengue cases showed a declining pattern after the country's COVID-19 lockdown because most people stayed home and did not go to the hospital (Ong et al., 2021).

Figure 1

Malaysian Dengue Cases and Fatalities from 1995 to 2019 (Loh, 2016)



An outbreak occurs when the number of cases in a certain period exceeds what would be expected. The definition of a dengue outbreak varies among countries. For instance, in China, an outbreak starts after three cases have been identified in a similar location within 15 days (Zhang et al., 2014). In Malaysia, at least one suspected case reported within 14 days is enough to raise the alarm (Seng et al., 2005). At a catastrophic level, an outbreak can lead to a high death toll and a downturn in the affected country's economy. A detection system that can notify authorities about early signs of an outbreak is necessary to avoid a dengue outbreak and its consequences (Masmas & Mohamed, 2021).

Recently, there have been many efforts to produce a more effective outbreak detection model for dengue, particularly focusing on areas such as improving the detection rate, reducing the false detection rate, and speeding up the detection time. The existing models use either a statistical or artificial intelligence (AI) approach. However, because outbreak information is weak, vague, and inconsistent, these models still tend to generate inaccurate results (Baharom et al., 2022). One of the key issues is that the model loses its robustness when dealing with unseen outbreak patterns, i.e., patterns that were unavailable during model development. Although a detection model undergoes an extensive training phase, it fails to detect an incoming outbreak when the signal varies from what it has received as input during training. The existing outbreak data have yet to be analysed intelligently to extract any discernible patterns that would provide insights to enable the accurate prediction of unknown new outbreaks. Therefore, a robust detection system that is capable of reacting to the ever-changing nature of this disease is vital. In achieving this, considerable research over the past decade has focused on the artificial immune system (AIS), a subdivision of artificial intelligence (Timmis et al., 2008). AIS has been proven to be an accurate analytical tool for other detection problems, such as the detection of fraud in business (Huang et al., 2009), faulty machines (Bi et al., 2010), intrusion in a computer network (Aickelin & Greensmith, 2008), and disease in medical (Periasamy et al., 2021). Inspired by those successes, this study has developed an outbreak detection model for dengue by using an AIS approach called danger theory (DT) (Matzinger, 1994). DT can detect hidden anomalies from a dataset without undergoing a training phase and thus does not rely on knowing the number of outbreak and non-outbreak cases that would have been input during training (Greensmith et al., 2006). When DT is compared to other detection techniques, it has been shown to offer a satisfying outcome (Yan et al., 2021) as it can produce a balance score between the actual detection rate and the false detection rate.

RELATED WORKS

Different approaches can be used to model an outbreak: anomaly detection and classification (Bakar et al., 2011). An outbreak is an anomaly detection task if the outbreak case is viewed as a rare item in a dataset, where it behaves very differently from other data. The anomaly detection approach usually involves unsupervised learning, where the information on the outbreak and non-outbreak classes is unknown during training (Toshniwal et al., 2020). Conversely, an outbreak can be a classification problem if the number of outbreaks and non-outbreak cases is almost balanced. The classification process determines the objects that belong in each class according to their specific features. Normally, two types of dengue dataset are used to construct a detection model: retrospective and prospective. The retrospective dataset consists of the results of laboratory diagnoses and is very accurate. Nevertheless, it is very timely as the information only becomes available when incidences of a disease have already occurred. In contrast, the prospective dataset comprises syndromic data that are not based on laboratory diagnosis results but provide insightful information on outbreaks. Accordingly, this study chose to develop an anomaly detection-based model and utilise prospective datasets.

Many studies on dengue outbreak detection models are based on statistical and artificial intelligence approaches. In the early years, the models mostly depended on statistics, where health data were monitored in a timely manner, and any activities that exceeded the upper control line were considered abnormal (outbreak). The China Infectious Disease Automated-alert and Response System (CIDARS) was built a few years prior, in 2008, to identify a dengue outbreak at the province level in China. With a time-series moving-percentile technique, the system compares the reported cases during the current observation period to those during the historical databank that is equivalent for each of the 31 provinces. (Yang et al., 2011). Yan et al. (2008) developed a hybrid predictive model by combining statistical and artificial intelligence approaches, i.e., the support vector machine and regression, to detect dengue in Singapore. Interestingly, they found that climate factors, such as rainfall and humidity, contribute to prediction accuracy. Their findings are consistent with Ahmad et al. (2018), who showed the viability of building a model-based dengue early warning system utilising temperature and rainfall data. Salim et al. (2021) discovered that wind speed effect was a factor in the spike in dengue cases.

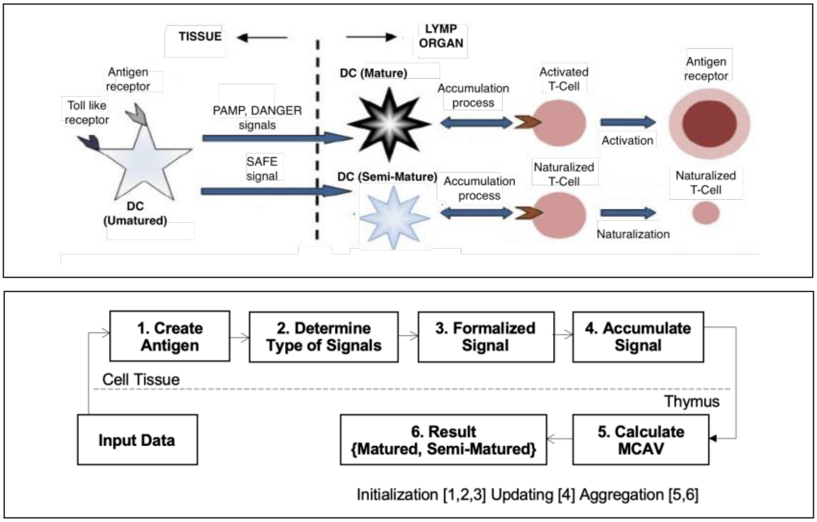
Several studies have attempted to model dengue outbreaks in respect of Malaysia specifically. Pham et al. (2015) developed a statistics-based system that used linear regression to forecast the occurrence of a dengue outbreak in Kuala Lumpur, Malaysia's capital. Based on an evaluation of three factors: daily rainfall, average daily temperature, and improved vegetation index, they claimed that their model could forecast dengue 16 days in advance. An earlier model for the detection of dengue using multiple classifiers was created by Bakar et al. (2011). To integrate the best rules provided by each model into a single classifier, their model included many classifiers, comprising a rough classifier, decision tree, associative classifier, and Naïve Bayes (NB). In a similar vein, Tarmizi et al. (2013) tested three different classifiers – neural network, decision tree, and rough classifier – on a Malaysian dengue dataset and generated a good result. They concluded that the selection of important features determined prediction accuracy. Long et al. (2010) proposed an outbreak detection model based on association rules, where an outbreak was detected using the multiple attribute value (MAV). The idea behind using the MAV was to generate a set of frequent attributes of dengue data, and then an anomaly could be detected from the generated set. An anomaly was considered to be an outbreak when its confidence level exceeded the outbreak threshold. Salim et al. (2021) modelled the dengue prediction model using several machine learning algorithms for the Selangor state. They found that the support vector machine (SVM) model was the best-predicted dengue outbreak model without overfitting. In another study, Husin et al. (2012) employed a neural network as a dengue predictor for five districts in Selangor. The model was hybridised with a genetic algorithm for updating the neural network weight, and the prediction was made based on rainfall information. Mousavi et al. (2013) used the AIS approach to develop an algorithm to detect dengue, which they called the negative selection algorithm. The model demonstrated high prediction accuracy; however, their proposed algorithm had some weaknesses highly related to the selection of attributes. Five machine learning algorithms were tested for their capabilities to forecast dengue outbreaks in Malaysia (Yavari Nejad & Varathan, 2021). They employed climatic risk factors as input and found the Bayes network model, which was capable of accurately identifying new meteorological risk factors. The recent work on dengue prediction was to discover the effect of COVID-19 lockdown towards dengue cases using the seasonal autoregressive integrated moving average (SARIMA) model (Ong et al., 2021). Alternatively, the spread of an outbreak could be monitored using a disease mapping approach, in which the geographical distribution of an outbreak was visualised in specific geographic regions (Mohd Diah & Aziz, 2022).

Danger Theory

Danger theory (DT) is one of the AIS paradigms proposed by Matzinger (1994). It implies that the human immune system is activated in response to a danger signal sent by a necrotic cell that suddenly dies from a pathogenic infection. When a cell in the human body becomes contaminated, it creates a danger zone around itself to reduce infection while an antibody is produced to fight the infection or illness. In contrast to a necrotic cell, an apoptotic cell is one that dies naturally; when it dies, it does not cause harm to the body, and it releases a safe signal rather than a danger signal. The danger theory differs from the conventional immune theory. The latter is founded on the idea that the human immune system distinguishes the self-cells and non-self-cells for it to be able to react appropriately when foreign substances are present while avoiding reacting with or attacking its own body cells (Lei et al., 2013). In DT, all body cells have a similar possibility of being affected by the pathogen. When a pathogen injures a bodily cell, it creates a danger zone around itself to protect the rest of the body. Here, the dendritic cells (DCs) play a crucial role in determining the danger signals by manipulating the type of signals: pathogen-associated molecular pattern (PAMP), danger signal (DS), and safe signal (SS). Based on these signals, the DC can determine if the cell's status is safe or dangerous to the body. Figure 2a illustrates the biological DT.

Figure 2

Biological DT and Artificial DT



Several researchers have developed detection algorithms based on biological DT. The dendritic cell algorithm (DCA) is one such algorithm (Greensmith & Aickelin, 2008). As illustrated in Figure 2b, DCA comprises three major steps: initiation, updating, and aggregation. The algorithm parameters are set during the initialisation process (numbers 1, 2, and 3 in Figure 2b). All DCs are set to immature, each record is marked as an antigen, and the attributes must be assigned and normalised into appropriate DT signals throughout this stage. In the following stage, the data structures derived from the input signals and antigens are updated (number 4 in Figure 2b). The immature DC samples input signals (PAMP, DS, and SS) as well as various antigens, calculate changes and determine which antigen is generating the changes using the accumulative function in Equation 1:

$$O_j(x) = (\sum_{i=0}^{i=3} W_{ij} * IS_{ij}(x))) / (\sum_{1=0}^{i=3} |W_{ij}|) \tag{1}$$

where,

W is the weight matrix,
IS is the input signals,
OS is the output signals,
i represents PAMP, SS, and DS,
j is the output signal categories: costimulation (CSM), mature, and semi-mature (no. 5 in Figure 2b).

All the input signals are then transformed into three cumulative output signals: costimulation (CSM), mature, and semi-mature. Depending on the CSM value, the DC's maturity will evolve over time into one of two states, which is semi-mature/normal or mature/abnormal. The maturity type is identified if the CSM value is larger than the migration threshold: If mature > semi-mature, use 'mature'; otherwise, use 'semi-mature'. The third step, aggregation (no. 6 in Figure 2b), occurs when learning is accomplished. The antigens given by the mature and semi-mature settings are tested for anomalies in this final phase. The mature context antigen value (MCAV) is used to determine an antigen's abnormality, as depicted in Equation 2. The antigen is classified as anomalous if the MCAV is greater than the anomaly threshold; otherwise, it is classified as normal. Artificially, the mature cell with a higher MCAV is an anomaly or outbreak, while the semi-mature cell is seen as normal/non-outbreak.

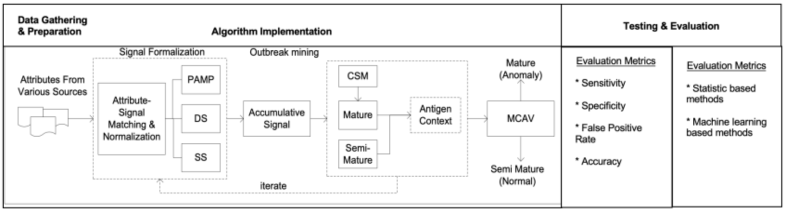
$$(\text{MCAV}) = (\text{Mature}) / (\text{Semi-Mature} + \text{Mature}) \tag{2}$$

METHODOLOGY

This study divided the methodology into three phases: data gathering and preparation, algorithm implementation, and testing and evaluation. Figure 3 provides a visual representation of the three phases.

Figure 3

The Research Methodology



Data Gathering and Preparation

This study combined two prospective datasets: the emergency visit dataset from the Vector Control Unit, Seremban District Hospital, Malaysia and the climate dataset provided by the Meteorological Centre, Malaysia. The original dataset of emergency visits contained 15 mixed attributes representing the demographic and clinical data of dengue patients. In comparison, the original climate dataset had eight continuous attributes representing the information related to temperature, humidity, and rain. Both datasets covered the period from 2003 to 2009, were inclusive, and were presented in a weekly format. In this study, the datasets were combined into a single 'dengue profile dataset', which was then split into two smaller datasets, with the 2003–2005 data used to calculate the baseline for the outbreak and the 2006–2009 data used for the experiment's monitoring phase. The dataset had 23 attributes cumulatively, including the target class, but only four were chosen based on expert recommendations to meet the DT requirement. Table 1 samples the four attributes: accumulated cases, humidity, rainfall, and temperature. Because DCA runs on continuous data types, the data types of each attribute were preserved in their original form.

Table 1

Attributes Chosen for Mining - Accumulated Cases, Temperature, Humidity, and Rainfall

Year	Accumulated Week	Accumulated Cases	Temperature	Humidity	Rainfall	Class
2003	40	32.0	0.38	0.62	0.10	normal
2003	40	32.0	0.38	0.62	0.10	normal
2003	40	32.0	0.38	0.62	0.10	normal
2003	40	32.0	0.38	0.62	0.10	normal
2003	40	32.0	0.38	0.62	0.10	normal
2003	40	32.0	0.38	0.62	0.10	normal

Implementation of DCA

The DCA for dengue outbreak detection modelling consisted of three main steps: signal formalisation, outbreak mining, and outbreak analysis. Signal formalisation involved two main tasks: 1) assigning dengue attributes into appropriate input signals (PAMP, DS, and SS) and 2) normalising them according to the type of input signal. Appropriate attribute-signal matching and signal normalisation were critically important in DCA because each signal represented a different characteristic that affected the result. PAMP and SS were good markers of an abnormal or normal state, whereas DS was more likely to represent an abnormal situation. The consulted experts suggested that the temperature, humidity, and rainfall attributes were set as DS, while PAMP and SS could be represented as the accumulated registered cases, as shown in Table 2. The general principle for matching the attribute to the input signals was that PAMP indicated that an abnormal scenario existed and SS signified that no anomalous situation existed. DS implied that an occurrence might or might not indicate an anomalous condition, although the possibility of an anomaly was higher than in a normal situation.

Table 2

Attribute-Signal Matching for DCA

DCA Signals	Attributes
PAMP	Accumulated Cases
SS	
DS	Weather (Average) - Temperature - Humidity - Rain

PAMP and SS were normalised by using a pre-defined rule based on the abovementioned Malaysian definition of a dengue outbreak. The total case variant between the current week and the average two weeks was used to determine the value of both signals, as depicted in Algorithm 1.

Algorithm 1: The Algorithm to Determine the Signal

```
Start
Set the maximum value to 100;
Determine the average number of dengue cases over the last two weeks1wqsxc
If (Cases of the current weeks – average of dengue cases from the previous 2 weeks > 1 case
    PAMP = 0;
    SS = Maximum value;
Else
    PUMP = Maximum value * 7;
    SS = 0;
End
```

DS was normalised by using the statistical time series formula, i.e., cumulative sum (CUSUM), as defined in Equation 3:

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

where,

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,
K is the allowance value.

The formula normally chose between the centre of the target mean μ_0 and the out-of-control mean μ_1. After that, the average CUSUM of the temperature, humidity, and rainfall attributes was calculated using Equation 4:

$$\mathrm{DS} = \frac{\mathrm{C}^+_{\mathrm{rainfall}} + \mathrm{C}^+_{\mathrm{temperature}} + \mathrm{C}^+_{\mathrm{humidity}}}{\mathrm{n}} \tag{4}$$

After normalising PAMP, SS, and DS, the dengue outbreak dataset was presented to the DCA for mining. In this outbreak mining and outbreak analysis, the outbreak data were mined using DCA to detect an outbreak. The aim was to create an MCAV for each antigen that represented the final state of an outbreak, starting with DC in an immature state and progressing to a mature or semi-mature state. The general process consisted of setting the initial parameters, updating the input signal and antigen, calculating the MCAV, and categorising the antigen. The outbreak was detected using Equation 5. The occurrence of an outbreak was notified when MCAVi > Outbreak baseline.

$$\mathrm{Status}(\mathrm{x_i}) = \begin{Bmatrix} +\mathrm{Outbreak}\,;\ \mathrm{MCAV_i} > \mathrm{Outbreak\ baseline} \\ -\mathrm{Outbreak}\,;\ \mathrm{MCAV_i} < \mathrm{Outbreak\ baseline} \end{Bmatrix} \tag{5}$$

where,

x_i is an antigen at ith observation,
+*Outbreak* refers to an anomalous condition or outbreak,
-*Outbreak* refers to a normal condition or non-outbreak.

Test and Evaluation

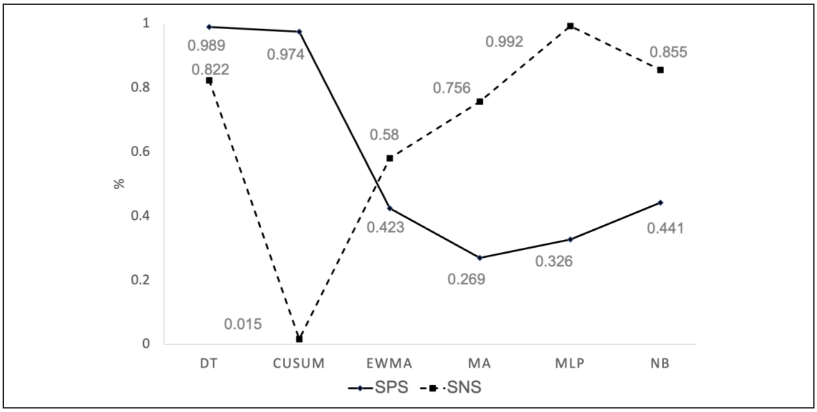
This phase involved evaluating the proposed model based on a binary classification that categorised the outbreak problem into two possible groups: outbreak (anomaly) and non-outbreak (normal). Four evaluation metrics were selected for this study, namely sensitivity (SNS), specificity (SPS), False Positive Rate (FPR), and Accuracy (ACC). The details of the metrics are as follows, each of which has a different aim. SNS demonstrated the model's ability to detect an outbreak, whereas SPS evaluated the model's ability to detect a non-outbreak. Meanwhile, FPR measured the amount of false detection that occurred when a non-outbreak case was mistakenly identified as an outbreak. Finally, ACC validated the modelling's accuracy in correctly classifying both classes. For SNS, SPS, and ACC, the highest value indicated the best result, whereas for FPR, the best result was the lowest value. The performance of the proposed algorithm was compared in this study to three statistics-based methods (CUSUM, moving average (MA), and exponentially weighted moving average (EWMA)), as well as to two machine learning methods (multi-layer perceptron (MLP) and NB). These approaches were chosen based on their frequent use in anomaly detection and classification problems mainly related to time series.

ANALYSIS AND RESULT

The aim of the experiment was to evaluate the performance of the proposed model based on DT in detecting a dengue outbreak. The experiment used a population of 100 cells, with a total cycle cell update of 20. During each cycle, DC was authorised to sample the antigen ten times. The weights of the accumulative function were set to W1 = 1 and W2 = 2. The weight value assigned to each signal represented the influence of a signal on an actual human body cell, whereby the value was relative and derived empirically from immunological data (Greensmith et al., 2008). Each assessment metric (SPS, SNS, FDR, and ACC) was averaged over the course of the experiment's 100 runs and recorded for analysis. The algorithm was used in the Malaysian dengue dataset from 2006 to 2009 for the experiment. Figure 4 shows the SPS and SNS values of the proposed model and that of the compared models. The scale in the figure runs from 0 to 1, with 1 representing 100 percent of the highest scores.

Figure 4

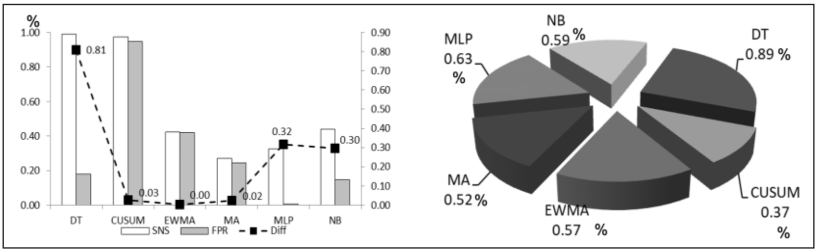
SNS and SPS of DT and Other Approaches



The suggested DT model's effectiveness to correctly predict a dengue outbreak as an outbreak was demonstrated by having the highest SNS among the other techniques (SNS = 0.989). This result was closely followed by CUSUM with SNS = 0.974; the rest of the approaches generated a lower value for SNS. The proposed model was also among the highest-ranking models in terms of accurately detecting a non-outbreak week as a non-outbreak with SNS = 0.822. Although its SPS was ranked behind MLP (SPS = 0.992) and NB (SPS = 0.855), the proposed model seemed to be a better approach because it could detect both an outbreak and a non-outbreak week as it achieved a balanced SNS and SPS result. A model must be able to deliver a balanced score in terms of SNS, SPS, and FPR while detecting anomalies to be deemed an effective detection model. Figure 5a shows the relationship between SNS, FPR, and their difference.

Figure 5

(a) SNS, FPR, and the difference (b) ACC of DT and Other Approaches



The proposed model was the most accurate model compared to the other techniques, as seen by the large difference between SNS and FPR (difference = 0.81), representing the model's ability to discriminate an outbreak week from a non-outbreak week accurately. In other approaches, the difference between SNS and FPR was less than 0.32. The FPR value should be as low as feasible because it signified the number of times an outbreak class was mistaken for a non-outbreak class. Besides that, DT was the most accurate model in terms of ACC, with a score of 0.89. Figure 6b indicates that the proposed model outperformed all other approaches examined.

The dengue dataset in the experiment in this study has been used in several research studies to test the performance of different approaches for dengue outbreak detection. Table 3 provides the result of a performance comparison of the proposed DT model with those reported in previous studies. As shown in the table, the proposed model outperformed the other dengue outbreak detection models. The result revealed that the proposed DT model generated a higher and more balanced result between SNS and SPS.

Table 3

Comparison of the Proposed DT Model and Other Approaches in the Literature

Techniques (Researchers)	SNS	SPS	ACC
a. Danger theory	0.9891$^{\text{Winner}}$	0.8217	0.8900$^{\text{Winner}}$
b. Frequent-outlier mining (Long, 2012)	0.9050$^{\text{W}}$	0.7040$^{\text{W}}$	0.8080$^{\text{W}}$
c. Multi-classifier (Bakar et al., 2011)	-	-	0.7125$^{\text{W}}$
d. Negative selection algorithm (Mousavi et al., 2013)	0.7016$^{\text{W}}$	1.00$^{\text{Winner}}$	0.8306$^{\text{W}}$

Overall, the results of the experiments demonstrated that the proposed DT model could accurately detect an outbreak signal in a dengue dataset. It generated high SNS, SPS, and ACC as well as a lower FPR. The proposed DT model outperformed other detection approaches in all performance matrices and generated a better balance than the other approaches in terms of SNS, SPS, and FPR. This finding demonstrated that the proposed DT model had the ability to distinguish between an outbreak and a non-outbreak with a lower FPR. The other approaches were inconsistent in terms of accuracy as they misclassified many nonoutbreaks as an outbreak. When this happens, the dengue detection system raises a high false alarm, which may result in the responsible parties expending energy and resources needlessly.

There are several reasons why the proposed DT model is a good dengue outbreak detector. First, when monitoring data, DT uses the threat of an antigen (known as the MCAV) rather than a pattern-matching strategy that considers all antigens (in the dengue data record) to have a comparable chance of becoming infected by a virus (in this case an outbreak). Consequently, it does not require a pre-defined dataset when monitoring data to help it classify incidences into an outbreak or a non-outbreak. Second, DT does not require a training phase, so it has increased robustness to deal with uncertainties in new dengue outbreak data, as the limitations of the pre-trained model do not constrain it. Third, DT is a multivariate detection approach that allows a mixture of multiple input features; it uses three types of input signal (PAMP, DS, SS) to represent multiple inputs from different prospective and retrospective sources. Fourth, because outbreak signals are frequently weak and inconsistent, the use of many input factors may explain why the suggested DT model outperformed techniques that rely on a single predictive variable in detecting outbreaks. In this study, the number of registered cases and CD (rainfall, temperature, humidity) were used as input. More interestingly, as seen by its present use in many real-time applications for detecting intruders in a computer network, an AIS-based system such as DT may be implemented in real-time applications with very low CPU processing requirements (González-Patiño et al., 2020). DT is typically a problem-specific technique. In comparison to other DT-based anomaly detection systems, the assigned input to DT's signal and the mechanism used to normalise the input signal would be significantly different.

CONCLUSION

This study developed a DT-based outbreak detection model to detect dengue outbreaks in Malaysia. Based on experimental data, it was discovered that the proposed model was superior to other well-known models for detecting unpredictable dengue outbreak patterns. In addition, when compared to other research that used similar experiment datasets, the proposed model was shown to be a better dengue outbreak detection system with better detection capabilities.

The DT outbreak detection model can handle new unknown outbreak patterns and distinguish between outbreak and non-outbreak cases with a consistently higher specificity, higher sensitivity, and lower false positive rate even without a training phase. In the future, this model will be enhanced to predict which areas would be affected by a dengue outbreak.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Seremban Public Health Department, Malaysia and the 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

Ahmad, R., Suzilah, I., Wan Najdah, W. M. A., Topek, O., Mustafakamal, I., & Lee, H. L. (2018). Factors determining dengue outbreak in Malaysia. *PloS One*, *13*(2), e0193326–e0193326. https://doi.org/10.1371/journal.pone.0193326

Aickelin, U., & Greensmith, J. (2008). Sensing danger: Innate immunology for intrusion detection. *CoRR*, abs/0802.4. http://arxiv.org/abs/0802.4002

Baharom, M., Ahmad, N., Hod, R., & Abdul Manaf, M. R. (2022). Dengue early warning system as outbreak prediction tool: A systematic review. *Risk Management and Healthcare Policy*, *15*, 871–886, https://doi.org/10.2147/RMHP.S361106

Bakar, A. A., Kefli, Z., Abdullah, S., & Sahani, M. (2011, July). Predictive models for dengue outbreak using multiple rulebase classifiers. In *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics* (pp. 1–6). https://doi.org/10.1109/ICEEI.2011.6021830

Bi, R., Timmis, J., & Tyrrell, A. (2010, July). The diagnostic dendritic cell algorithm for robotic systems. In *IEEE Congress on Evolutionary Computation* (pp. 1–8). https://doi.org/10.1109/CEC.2010.5586499

CDC. (2021). *Dengue - Statistics and maps*. Centers for Disease Control and Prevention. https://www.cdc.gov/dengue/statistics-maps/index.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fdengue%2Fepidemiology%2Findex.html

Cheong, Y. L., Burkart, K., Leitão, P. J., & Lakes, T. (2013). Assessing weather effects on dengue disease in Malaysia. *International Journal of Environmental Research and Public Health*, *10*(12), 6319–6334. https://doi.org/10.3390/ijerph10126319

González-Patiño, D., Villuendas-Rey, Y., Argüelles-Cruz, A. J., Camacho-Nieto, O., & Yáñez-Márquez, C. (2020). AISAC: An artificial immune system for associative classification applied to breast cancer detection. *Applied Sciences*, *10*(2), 515. MDPI AG. http://dx.doi.org/10.3390/app10020515

Greensmith, J., Twycross, J., & Aickelin, U. (2006, July). Dendritic cells for anomaly detection. In *2006 IEEE International Conference on Evolutionary Computation* (pp. 664–671). https://doi.org/10.1109/CEC.2006.1688374

Greensmith, J., Aickelin, U., & Cayzer, S. (2008). Detecting danger: The dendritic cell algorithm. In Schuster, A. (Ed.), *Robust intelligent systems* (pp. 89–112). Springer. https://doi.org/10.1007/978-1-84800-261-6_5

Greensmith, J., & Aickelin, U. (2008, August). The deterministic dendritic cell algorithm BT - Artificial immune systems. In *International Conference on Artificial Immune Systems* (pp. 291–302). Springer Berlin Heidelberg.

Huang, R., Tawfik, H., & Nagar, A. K. (2009, December). Artificial dendritic cells algorithm for online break-in fraud detection. In *2009 Second International Conference on Developments in ESystems Engineering* (pp. 181–189). https://doi.org/10.1109/DeSE.2009.59

Husin, N. A., Mustapha, N., Sulaiman, M. N., & Yaakob, R. (2012, September). A hybrid model using genetic algorithm and neural network for predicting dengue outbreak. In *2012 4th Conference on Data Mining and Optimization (DMO)* (pp. 23–27). https://doi.org/10.1109/DMO.2012.6329793

Lei, D., Yu, F., & Zhenghua, Y. (2013). Survey of DCA for abnormal detection. *Journal of Software*, *8*, 2087–2094.

Loh, F. F. (2016, January 12). Numbers of dengue cases to rise further when El Nino hits. *The Star*. https://www.thestar.com.my/news/nation/2016/01/12/a-week-into-2016-1000-extra-cases-numbers-to-rise-further-when-el-nino-hits

Long, Z. A. (2012). Teknik perlombongan corak data terpencil-kerap bagi pengesanan wabak dalam pengawasan kesihatan awam [PhD dissertation, Universiti Kebangsaan Malaysia].

Long, Z. A., Bakar, A. A., Hamdan, A. R., & Sahani, M. (2010, November). Multiple attribute frequent mining-based for dengue outbreak BT. In *Advanced Data Mining and Applications: 6th International Conference, ADMA 2010* (pp. 489–496). Springer Berlin Heidelberg.

Masmas, B. A., & Mohamed, A. (2021). Knowledge-based decision support system for emergency management: The pandemic framework. *Journal of Information and Communication Technology*, *20*(4), 599–628. https://doi.org/10.32890/jict2021.20.4.6

Mousavi, M., Bakar, A. A., Zainudin, A. S., Long, Z. A., & Sahani, M. (2013). Negative selection algorithm for dengue outbreak detection. *Turkish Journal of Electrical Engineering & Computer Sciences*, *1*(1), 2346–2356.

Matzinger, P. (1994). Tolerance, danger, and the extended family. *Annual Review of Immunology*, *12*, 991–1045. https://doi.org/10.1146/annurev.iy.12.040194.005015

Mohd Diah, I., & Aziz, N. (2022). The comparison between standardized mortality ratio, poisson-gamma and stochastic sic model for pneumonia disease mapping in Malaysia. *Journal of Information and Communication Technology*, *21*(4), 549–570. https://doi.org/10.32890/jict2022.21.4.4

Mudin, R. N. (2015). Dengue incidence and the prevention and control program in Malaysia. *IIUM Medical Journal Malaysia*, *14*(1). https://doi.org/10.31436/imjm.v14i1.447

Ong, S.-Q., Ahmad, H., & Mohd Ngesom, A. M. (2021). Implications of the COVID-19 lockdown on dengue transmission in Malaysia. *Infectious Disease Reports*, *13*(1), 148–160. https://doi.org/10.3390/idr13010016

Periasamy, K., Periasamy, S., Velayutham, S., Zhang, Z., Ahmed, S. T., & Jayapalan, A. (2022). A proactive model to predict osteoporosis: An artificial immune system approach. *Expert Systems*, *39*(4), e12708.

Pham, D., Aziz, T., Kohan, A., Nellis, S., Abd. Jamil, J., Khoo, J., Lukose, D., Bakar, S. A., & Sattar, A. (2015, October). An efficient method to predict dengue outbreaks in Kuala Lumpur. In *Proceeding of the 3rd International Conference on Artificial Intelligence and Computer Science (AICS2015), Penang, Malaysia* (pp. 169–178).

Salim, N. A. M., Wah, Y. B., Reeves, C., Smith, M., Yaacob, W. F. W., Mudin, R. N., Dapari, R., Sapri, N. N. F. F., & Haque, U. (2021). Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Scientific Reports*, *11*(1), 939. https://doi.org/10.1038/s41598-020-79193-2

Seng, S. B., Chong, A. K., & Moore, A. (2005, November). 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: A Spatio-temporal Workshop)* (pp. 109–123).

Suppiah, J., Ching, S.-M., Amin-Nordin, S., Mat-Nor, L.-A., Ahmad-Najimudin, N.-A., Low, G. K.-K., Abdul-Wahid, M.-Z., Thayan, R., & Chee, H.-Y. (2018). Clinical manifestations of dengue in relation to dengue serotype and genotype in Malaysia: A retrospective observational study. *PLoS Neglected Tropical Diseases*, *12*(9), e0006817. https://doi.org/10.1371/journal.pntd.0006817

Tarmizi, N. D. A., Jamaluddin, F., Bakar, A. A., Othman, Z. A., Zainudin, S., & Hamdan, A. R. (2013). Malaysia dengue outbreak detection using data mining models. *Journal of Next Generation Information Technology (JNIT)*, *4*, 96-107.

Timmis, J., Hone, A., Stibor, T., & Clark, E. (2008). Theoretical advances in artificial immune systems, *Theoretical Computer Science, 403*(1), 11–32.

Toshniwal, A., Mahesh, K., & Jayashree, R. (2020, October). Overview of anomaly detection techniques in machine learning. In *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (pp. 808–815). https://doi.org/10.1109/I-SMAC49090.2020.9243329

WHO. (2021). Dengue and severe dengue. World Health Organisation. https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue

Yan, W., Lee, G., Fu, X., & T. H. (2008, July). Detect climatic factors contributing to dengue outbreak based on wavelet, support vector machines and genetic algorithm. In *Proceeding of the World Congress on Engineering 2008* (Vol. 1, pp. 2–4).

Yan, S., Li, T., & Zheng, H. (2021, January). Prediction model of sleep membership in drugstores based on immune danger theory. In *2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA), Shenyang, China* (pp. 631-635). https://doi.org/ 10.1109/ICPECA51329.2021.9362558

Yang, W., Li, Z., Lan, Y., Wang, J., Ma, J., Jin, L., Sun, Q., Lv, W., Lai, S., Liao, Y., & Hu, W. (2011). A nationwide web-based automated system for outbreak early detection and rapid response in China. *Western Pacific Surveillance and Response Journal (WPSAR)*, *2*(1), 10–15. https://doi.org/10.5365/WPSAR.2010.1.1.009

Yavari Nejad, F., & Varathan, K. D. (2021). Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction. *BMC Medical Informatics and Decision Making*, *21*(1), 141. https://doi.org/10.1186/s12911-021-01493-y

Zhang, H., Li, Z., Lai, S., Clements, A. C. A., Wang, L., Yin, W., Zhou, H., Yu, H., Hu, W., & Yang, W. (2014). Evaluation of the performance of a dengue outbreak detection tool for China. *PloS One*, *9*(8), e106144–e106144. https://doi.org/10.1371/journal.pone.0106144