

Performance of Hybrid GANN in Comparison with Other Standalone Models on Dengue Outbreak Prediction

Nor Azura Husin, Norwati Mustapha, Md. Nasir Sulaiman,
Razali Yaacob, Hazlina Hamdan and Masnida Hussin

Department of Computer Science, Faculty of Computer Science and Information Technology,
University Putra Malaysia, Selangor, Malaysia

Article history

Received: 24-06-2015

Revised: 23-06-2016

Accepted: 01-07-2016

Corresponding Author:
Nor Azura Husin
Department of Computer
Science, Faculty of Computer
Science and Information
Technology, University Putra
Malaysia, Selangor, Malaysia
Email: support@thescipub.com

Abstract: Early prediction of diseases especially dengue fever in the case of Malaysia, is very crucial to enable health authorities to develop response strategies and context preventive intervention programs such as awareness campaigns for the high risk population before an outbreak occurs. Some of the deficiencies in dengue epidemiology are insufficient awareness on the parameter as well as the combination among them. Most of the studies on dengue prediction use standalone models which face problem of finding the appropriate parameter since they need to apply try and error approach. The aim of this paper is to conduct experiments for determining the best network structure that has effective variable and fitting parameters in predicting the spread of the dengue outbreak. Four model structures were designed in order to attain optimum prediction performance. The best model structure was selected as predicting model to solve the time series prediction of dengue. The result showed that neighboring location of dengue cases was very effective in predicting the dengue outbreak and it is proven that the hybrid Genetic Algorithm and Neural Network (GANN) model significantly outperforms standalone models namely regression and Neural Network (NN).

Keywords: Hybrid GANN, Genetic, Neural Network, Predicting

Introduction

Dengue is a tropical mosquito-borne disease affecting more than 100 countries worldwide. Current estimate by WHO put the number of cases at 50-100 million cases per-year while the most recent estimate by a multinational study just published in the Lancet, tripled the WHO estimate at 360 million cases per year with 40% of the world population at risk (MOH, 2013). This disease may become the most important global health problem in the next decade which can no longer be ignored. This is further aggravated by environmental parameters like global warming, rapid urbanization and international traveling, which are recognized as contributing parameters to the spread of dengue outbreak (MOH, 2013). Dengue disease turns out to be the highest communicable disease compared to other prominent diseases like malaria, HFMD, typhus and yellow fever (MOH, 2012).

In Malaysia, dengue cases were categorized as a notifiable disease in 1971. Since then, it continues to

persist in predominantly urban and semi urban areas throughout the country. Approximately 70-80% of dengue cases are reported in areas where there is a high population density and rapid development activities which contribute to dengue transmission (Mahiran and Ho, 2011). Rapid urbanization has brought about enormous infrastructural build-up indirectly producing breeding areas for mosquito. Consequently, population growth and climate also considered as main parameters that contribute to the spike in dengue cases outbreak (Muhuidin and Jamie, 2015).

Previous research already shows that the accuracy of prediction model can be better over standalone model if we combine several different models. The hybrid models are proved in order to search the suitable parameter and make it model more robust with regard to the possible structure change in the data. Although combining or hybrid model prove to be alternative on solving the previous problem, there is not many existing prediction model using hybrid model especially prediction on dengue outbreak. Therefore, this study propose that the

hybrid model as the best model for predicting outbreak as it can solve previous problems in this research areas and provide better accuracy compared to those generated by other standalone model.

The aim of this paper is to conduct experiments for determining which model structure could deliver the best performance and thus, a more capable model for predicting the spread of dengue outbreak by comparing between a hybrid model of genetic algorithm and neural network and standalone models namely neural network and regression model. The rest of this paper is organized as follows; section 2 describes previous works. Designing model was explained in section 3. Section 4 discusses the implementation of hybrid GANN. Section 5 explained results and discussion of the case study. Section 6 describes our conclusion. Finally, some future works are summarized in Section 7.

Previous Work

Hybrid models are the combination of two or more model in one system that propose to solve the given problem. According to Gray and Kilgour (1997), hybrid model can be categorized by sequential, auxiliary and embedded hybrids which the sequential models make use of model in serial way. Meanwhile in auxiliary hybrid model, one model calls the other as a subroutine to process or manipulate information desired by it. Embedded hybrid models integrate the models in such a manner that they appear intertwined. The combination is so complete that in such manner that it would come out that no model can be used without the others for effective problem solving.

Table 1 summarizes the findings from previous researches on disease prediction models. Studies by Cen and Wang (2008) on predicting *lung cancer* using a hybrid Genetic Algorithm and Neural Network (GA-BP) demonstrated the hybrid model can speed up convergence to the optimal solution and provide an effectual model for

early diagnosis of lung cancer. The result also shows that the hybrid of NN and GA can provide more accurate and efficient prediction compared to standalone model. Aburas *et al.* (2010) conclude that the combination of four parameters namely rainfall, number of dengue cases, mean temperature and relative humidity were very effective for predicting dengue cases. Besides, NN model have been found to be very effective processing systems for predicting dengue outbreak.

Another research by Rachata *et al.* (2008) also proposed NN and entropy model for predicting dengue hemorrhagic fever outbreak and the result demonstrated that it can achieve up to 85.92% of accuracy and concluded that the result is going better when using entropy transformation. Yusof (2011) develop a prediction model that incorporates LS-SVM and NN model in predicting dengue outbreak and the results demonstrated that LS-SVM produce the best result compare to NN in terms of predicting accuracy and computational time. Harrison and Kennedy (2005) developed and optimized NN model for diagnosis of *acute coronary syndrome* and concluded that NN is a suitable model for developing diagnosis algorithms for chest patients and this model calibrated and performed well on unseen data from different centers. Although some of them uses standalone model in their studies, they recommend applying a hybrid model to achieve the optimal parameter, thus improve the predicting accuracy.

These studies indicated that, hybrid model results outperform neural network, regression and even statistical methods using hybrid can provide better accuracy than standalone model because these methods prevent the NN from being stuck in the local minima (Shanti *et al.*, 2009). Although the hybrid between GA and NN has delivered comparable results against other standalone approach however from our observation, there is rarely any study on the application of hybrid model for dengue outbreak predicting.

Table 1. The previous studies review on disease prediction model

Author	Research Domain	Method used
Yusof (2011)	Predict <i>dengue outbreak</i> based on dengue cases data, rainfall data and proximity location data	Least Square Support Vector Machines (LS-SVM) and Neural Network Model
Aburas <i>et al.</i> (2010)	Predict <i>dengue cases</i> on mean temperature, mean relative humidity, total rainfall and total number of dengue cases	Neural network
Cen and Wang (2008)	Predict the <i>lung cancer</i> by introduces Multi-species Co-evolution Genetic Algorithm and Simulated Annealing algorithm.	Hybrid Genetic Algorithm-BP Neural Networks
Rachata <i>et al.</i> (2008)	Propose an automatic prediction system of <i>Dengue Hemorrhagic Fever (DHF)</i> outbreak on weather condition and DHF cases data	Neural network and Entropy technique
Harrison and Kennedy (2005)	To develop and optimize NN for diagnosis of <i>acute coronary syndrome</i> , to test the model on data collected prospectively from different centers.	Neural network and logistic regression.

Designing Model

In the context of this study, the hybridization between neural network and genetic is attempted to overcome the weaknesses of one with the strengths of the other. Basically gradient descent learning is used in order to obtain their weights in standalone back propagation neural network. However there always remains the problem of the network becoming stuck in local minimum.

Therefore, this hybrid model which uses genetic algorithm to determine the weights is expected to achieve adequate result. Figure 1 illustrates the design of hybrid GANN model that applied in this study. The determinations of weight for hybrid GANN are performed by several procedures which consist of coding, weight extraction, fitness function and preproduction. The procedures of determinations of weight for hybrid GANN will be explained in next section.

Implementation of Hybrid GANN Model

To develop a hybrid model, the main procedure is to train the neural network to obtain the weights for a given set of input data and to infer output given a testing set

using the weights obtained after training. This procedure involves two main files which are *training file* for train neural network and *testing file* for inferring the output of the given testing set respectively. This main procedure (*training*) will control all the procedures that must be executed in an exact order. All the main procedures will declare about seven sub procedures and these are: Weight extraction, produce genetic, convert weight, mating pool and fitness, cross sites and reproduce procedure.

The first procedure is to generate a population of random weights for the input-hidden and hidden-output layer of the models structure. At this stage we start with random number of generation. These models whose network configuration is $l-m-n$ where l is input nodes, m is hidden nodes and n is output nodes. The numbers of weights to be determined are $(l+n)m$ with each weight were being a real number. We refer the number of gene length in the weight as d . Here $d = 5$ is used for the number of digit allocated for each gene in the chromosome. The string $L = (l+n)md$ represents the number of population size. An initial number of population sizes were generated randomly. The output of weight is made available at the specific file.

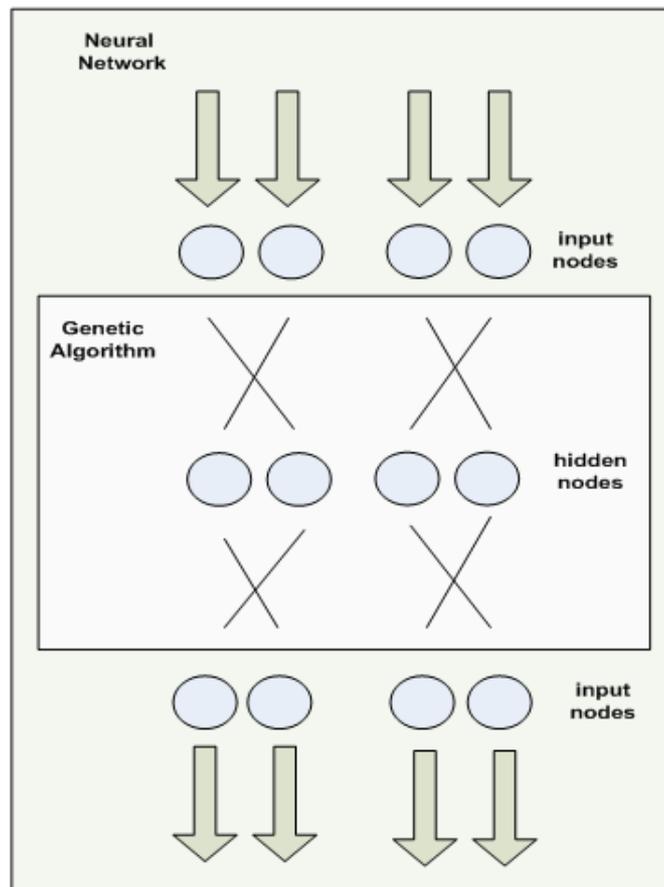


Fig. 1. The design of hybrid GANN model

Table 2. The models structure with specific parameters

Nodes	Struc 1	Struc 2	Struct 3	Struc 4
Input (l)	4	8	8	12
Output (n)	4	4	4	4
Hidden nodes (m)	4,9	8, 17	8,17	12, 25
Hidden layer	1	1	1	1
Weight (w)	16, 32	48, 96	48, 96	96, 192
Population size (L)	80, 160	240, 480	240, 480	480, 960

The number of input, hidden, output node and population size is determined based on configuration of the models structure as shown in Table 2.

For Structure 1, the input nodes is $l = 4$ and output is $n = 4$ nodes, hidden nodes are $m = 4$ and 9 with one hidden layer. Meanwhile the weight w are 16 and 32; and the population size are $L = 80$ and 160. Structure 2, the input is $l = 8$ and output is $n = 4$ nodes, hidden nodes are $m = 8$ and 17 with one hidden layer. Meanwhile the weight w are 48 and 96; and the population size is $L = 240$ and 480. For Structure 3, the input nodes is $l = 8$ and output is $n = 4$ nodes, hidden nodes are $m = 8$ and 17 with one hidden layer. Meanwhile the weight w are 48 and 96; and the population size are $L = 240$ and 480. Structure 4, the input nodes is $l = 12$ and output is $n = 4$ nodes, hidden nodes are $m = 12$ and 25 with one hidden layer. Meanwhile the weight w are 96 and 192; and the population size are $L = 480$ and 960.

The next process is to produce the genetic that represents the population of chromosomes. This procedure will merge input-hidden weight and hidden-output weight files and compute as a genetic. Weight for each member of the population will emerge on a single line to favor the application of genetic operations.

To define the values of fitness for each chromosome then we need to extract weights from each of the chromosomes. The next procedure is to extract weight procedure that proposes to convert the weights to be in the range of -5 to +5, existing in genetic file. This procedure will produce the final weight in other files. The number of population m , weight w and gene length d are determined based on configuration of the models structure as mention before. The actual weight w_k is given by this Equation 1 (Rajasekaran and Vijayalakshmi Pai, 2008):

$$w_k = \begin{cases} \frac{x_{kd+2}10^{d-2} + x_{kd+3}10^{d-3} + \dots x_{(k+1)d}}{10^{d-2}}, & \text{if } 5 \leq x_{kd+1} \leq 9 \\ -\frac{x_{kd+2}10^{d-2} + x_{kd+3}10^{d-3} + \dots x_{(k+1)d}}{10^{d-2}}, & \text{if } 0 \leq x_{kd+1} < 5 \end{cases} \quad (1)$$

Where:

- $x_1, x_2, \dots, x_d, \dots, x_L$ = Represent a chromosome
- $x_{kd+1}, x_{kd+2}, \dots, x_{(k+1)d}$ = Represent the k th gene ($k \geq 0$) in the chromosome.

The fitness function must be formulated for each problem to be solved. In the first place, we need to check the fitness file to ensure convergence of fitness. If the convergence has not been attained, run the *training procedure* and call for the rebuild option. Rebuild begins generation from the last formed population of chromosomes and their acquired fitness. Meanwhile, if the convergence has been attained, *testing procedure* must be performed to infer the output of the testing data set.

Mating pool made by eliminating that chromosome C_l with the least fitness F_{min} and change with duplicate copy of the chromosome C_k that reporting the highest fitness F^{max} . The parents are selected in pair at random. The offspring from the current population have their fitness calculated over again. In this study, the initial population of chromosome P_0 generated earlier with F_i , where $i = 1, 2 \dots n$, as their fitness values.

Results and Discussion

Result of Hybrid GANN Model

Table 3 illustrated the results of error and recognition rate for model structure 1, 2, 3 and 4 by using Hybrid GANN Model. As we can see at this table model structure 3 that comprise between the cases of dengue data with dengue cases at nearby location significantly show the lowest error compared with others model structure produce in significant.

At Hulu Langat district, the best MSE error with is 0.077524 with recognition rate 92.25% with the combination of hidden nodes eight, hidden layer one, eight input and four output. At Hulu Selangor the combination of hidden nodes eight, hidden layer one, eight input and four outputs produce the best MSE error with 0.060782 with recognition rate 93.92%. At Klang, the combination of hidden nodes 17, hidden layer one, eight input and four output produce the best MSE error with 0.065243 with recognition rate 93.48%.

Meanwhile, at Kuala Selangor the best MSE error is 0.065112 with recognition rate of 93.49% with the combination of hidden nodes eight, hidden layer one, eight input and four output. At Sepang, the combination of hidden nodes eight, hidden layer one, eight input and four output produce the best MSE error with 0.068891 with recognition rate 93.11%.

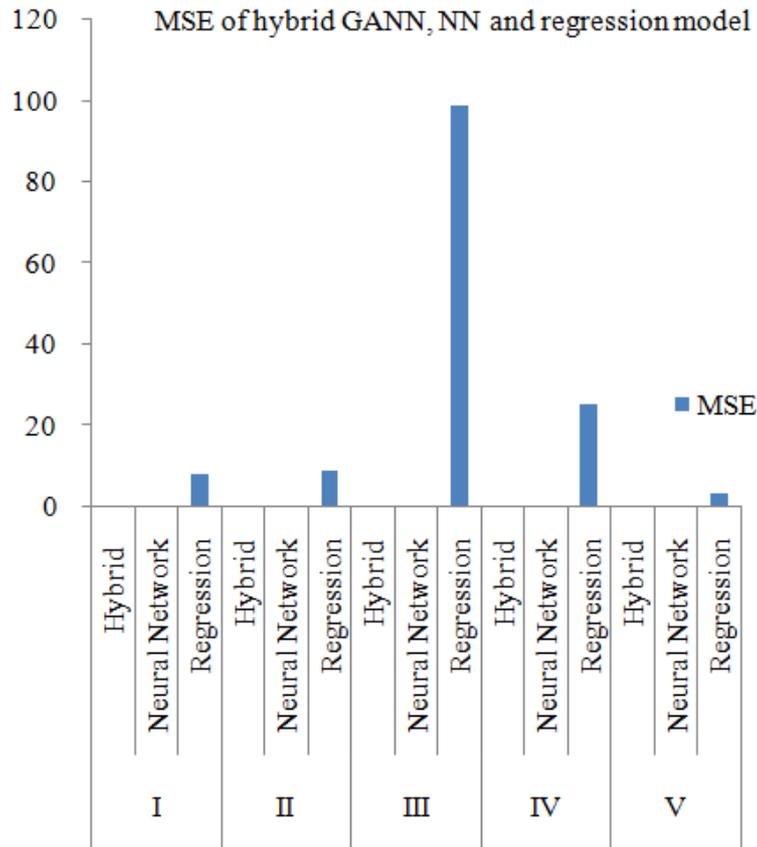


Fig. 2. Comparison result of MSE by using hybrid GANN, NN and regression model

Table 3. Results of recognition rate for model structure 1, 2, 3 and 4 at all district using hybrid GANN model

Structure	District	Error	Rate
1	I	0.111665	88.33
	II	0.06268	93.73
	III	0.06589	93.41
	IV	0.085603	91.44
	V	0.07681	92.32
2	I	0.072206	92.78
	II	0.06498	93.5
	III	0.07062	92.94
	IV	0.070689	92.93
	V	0.077099	92.29
3	I	0.068891	93.11
	II	0.06078	93.92
	III	0.07752	92.25
	IV	0.065243	93.48
	V	0.065112	93.49
4	I	0.078962	92.1
	II	0.06511	93.49
	III	0.07506	92.49
	IV	0.066073	93.39
	V	0.088945	91.11

I = Sepang, II = Hulu Selangor, III = Hulu Langat, IV = Klang and V = Kuala Selangor

Table 4. Result of hybrid GANN, neural network and regression

District	Model	Error
I	Hybrid	0.068891
	Neural Network	0.073620
	Regression	7.779000
II	Hybrid	0.060780
	Neural Network	0.064100
	Regression	8.667000
III	Hybrid	0.077520
	Neural Network	0.081930
	Regression	98.54900
IV	Hybrid	0.065243
	Neural Network	0.073900
	Regression	25.14700
V	Hybrid	0.065112
	Neural Network	0.069770
	Regression	3.247000

I = Sepang, II = Hulu Selangor, III = Hulu Langat, IV = Klang and V = Kuala Selangor

The Comparison Result of Hybrid GANN, NN and Regression Model

The main objective of this research is to determine which model structure will produce the best prediction

of the spread of dengue outbreak; which Structure 1 involved only the cases of dengue data, Structure 2 that comprise of dengue cases and rainfall data, Structure 3 that comprise of dengue cases and nearby location of dengue cases data and Structure 4 that include all the variables. Table 4 below illustrated the results between Hybrid GANN, NN and Regression Model in Kuala Selangor, Hulu Selangor, Sepang, Klang and Hulu Langat district based on the experimentation. From this table we can conclude that the Hybrid GANN model outperforms other standalone models. Figure 2 obviously shows that the regression model produce the highest error at all location.

Conclusion

This paper conduct experiments for finding which model structure will be the best for predicting the spread of dengue outbreak by comparing the performance of a hybrid model with conventional models namely neural network and regression model. The result from this experiment showed that the hybrid GANN model provides more accurate and efficient prediction compared to other standalone models. The result also showed that the model structure 3 that comprise with dengue cases and data on neighboring location was very effective in predicting dengue outbreak.

Moreover, this study has shown that the relative capability of hybrid model in dengue outbreak prediction and assisted to recognize the suitable variable and parameter for this experiment and their effect on predicting result. Hence, this study can be used as an indication for future research, especially in the context of prediction by comparing it with the performance of other models.

Future Work

Further research will be carried out to improve our dengue by observing other two parameters that persuaded the hybrid model performance and also explore different time frame. Besides, since the epidemic can be caused by different factors, we suggest exploring other factors that are not included in this research.

Acknowledgement

Much appreciation goes to the Department of Meteorological and Health State of Selangor, Malaysia and for providing information and relevant data for this study.

Author's Contribution

Nor Azura Husin: Collected, interpreted and analyzed the data and also implemented the model using C++ programming language. Mainly wrote the first draft of the manuscript.

Norwati Mustapha: Designed the main concepts and planned the steps of the proposed model. Final approval of the revision to published.

Md. Nasir Sulaiman and Razali Yaacob: Designed the main concepts and planned the steps of the proposed model.

Hazlina Hamdan and Masnida Hussin: Reviewed and analytically evaluated the manuscript for submission.

Ethics

This article is original and contains unpublished material. The corresponding author approved the manuscript and confirms that no ethical issues involved.

References

- Aburas, H.M., B.G. Cetiner and M. Sari, 2010. Dengue confirmed-cases prediction: A neural network model. *J. Expert Syst. Applic.*, 37: 4256-4260.
DOI: 10.1016/j.eswa.2009.11.077
- Cen, L. and M. Wang, 2008. Application of hybrid genetic algorithm-BP neural networks to diagnosis of lung cancer. *Proceedings of the International Conference on Computer Science and Software Engineering*, Dec. 12-14, IEEE Xplore Press, Wuhan, Hubei, pp: 36-39.
DOI: 10.1109/CSSE.2008.524
- Gray, A. and R. Kilgour, 1997. *Hybrid system FAQ*. Version Draft 1.00.
- Harrison, R.F. and R.L. Kennedy, 2005. Artificial neural network models for prediction of acute coronary syndromes using clinical data from the time of presentation. *Ann. Emergency Med.*, 46: 143-439.
DOI: 10.1016/j.annemergmed.2004.09.012
- Mahiran, M. and B.K. Ho, 2011. Clinical practice guideline on management of dengue infection in adult. *J. Malaysian Family Phys.*
- MOH, 2013. Malaysia calls for more concerted global R&D effort against dengue at 66th WHO Assembly, Switzerland. Ministry of Health Malaysia, Official Press Statement by the Deputy Director General.
- MOH, 2012. Health fact 2012, Health of Informatics Centre Planning and Development Division MOH/S/RAN/31.12(TR).
- Muhuidin, H. and T. Jamie, 2015. Variables that may affect the transmission of dengue-a case study for health management in Asia. *Topics in Public Health*.

- Rachata, N., P. Charoenkwan, T. Yooyativong, K. Chamnongthai and C. Lursinsap *et al.*, 2008. Automatic prediction system of dengue haemorrhagic-fever outbreak risk by using entropy and artificial neural network. proceedings of the International Symposium on Communications and Information Technologies, Oct. 21-23, IEEE Xplore Press, Lao, pp: 210-214.
DOI: 10.1109/ISCIT.2008.4700184
- Rajasekaran, S. and G.A. Vijayalakshmi Pai, 2008. Neural networks, fuzzy logic and genetic algorithms: Synthesis and applications. PHI Learning Private Limited.
- Shanti, D., G. Sahoo and N. Saravanan, 2009. Evolving connection weights of artificial neural networks using genetic algorithm with application to the prediction of stroke disease. *Int. J. Soft Comput.*, 4: 95-102.
- Yusof, H., 2011. Dengue outbreak prediction: A least squares support vector machines approach. *Int. J. Comput. Theory Eng.*, 3: 489-493.
DOI: 10.7763/IJCTE.2011.V3.355