

# Prediction of Dengue Outbreaks in Sri Lanka using Artificial Neural Networks

P.H.M.Nishanthi Herath  
Advanced Technological  
Institute  
Kandy, Sri Lanka

A.A.I. Perera  
Department of Mathematics,  
Faculty of Science  
University of Peradeniya  
Peradeniya, Sri Lanka

H.P.Wijekoon  
Department of Industrial  
Management,  
Faculty of Science  
University of Kelaniya,  
Kelaniya, Sri Lanka

## ABSTRACT

With nearly 30,000 cases reported annually all over the island, Dengue fever has become a major health hazard in Sri Lanka over the past few years. This research attempts to develop an Artificial Neural Network (ANN) to predict Dengue outbreaks. The study investigates the effects of weather variables and previous Dengue cases on the current Dengue cases. The weather variables, Average Temperature, Average Relative Humidity, Rainy Days per Week, Total Rainfall and the Previous Cases are identified with a time lag as input parameters to the ANN. The parameters and the specific time lags are defined by a correlation analysis between each individual variable with current Dengue cases. The ANN developed as an outcome of this research is capable of predicting Dengue outbreaks in Kandy district in Sri Lanka with fairly good accuracy.

## General Terms

Artificial Neural Networks, Predicting Dengue Outbreaks

## Keywords

Artificial Neural Networks, Multi Layer Perceptron Networks, Predicting Dengue Outbreaks, Effect of weather variables on Dengue

## 1. INTRODUCTION

The Dengue Fever (DF), Dengue Hemorrhage Fever (DHF) and Dengue Shock Syndrome (DSS) have been in high profile discussion, because of the dramatic increase of Dengue cases reported throughout the world [3]. The disease is a viral infection. The virus completes its life cycle partially in the mosquito – *Aedis sp* and partially in humans [5]. DF is a tropical infection [8]. Sri Lanka, being a tropical island has had significant experiences with the DF and DHF during the recent past. The annual number of DF cases reported amounts to be around 30,000. During the year 2012, nearly 40,000 cases were reported. Despite the number of medical researches done on the subject, little is accomplished on predicting DF. Proactive measures to prevent the incidence of disease in Sri Lanka still have room to improve. Only a handful of studies have been done so far in this regard [4]. However, there are a number of researches done in many other countries.

A research done in Malaysia [9] has attempted to develop an Artificial Neural Network Model (ANN) that predicts potential number of Dengue cases of a given geographical area based on the number of Dengue cases reported within a past period and the average rainfall received by the same geographical area. This research further compares the ANN model with Least Square – Support Vector Machines (LS –

SVM) model and states that LS – SVM model is better than the ANN model in terms of learning speed, but it is poor in performance with unseen data. However the author suggests that the model developed by this research can further be improved by considering the other factors affecting DF.

Another study done in Singapore [1] tries to predict the possible number of Dengue cases, also using an ANN model. This model predicts based on the number of cases reported during the previous period of time, with three weather parameters – mean temperature, mean rainfall and mean relative humidity of the relevant area during the time period considered. It points out that the accuracy of the model was increased by including the DF cases reported on the previous week in to the input layer. Results of this study reveal that ANN gave encouraging results in predicting Dengue in Singapore. The correlation between the predicted and actual values in the test set is 91%. The author suggests further research with more variables.

Among the few researches done in Sri Lanka, a study to develop an early warning system [4] attempts to develop an ANN to predict Dengue using the records of Western and Southern provinces. The data includes epidemiological records on DF cases reported, meteorological data on weather parameters and census data. It focuses on identifying the risk factors for DF outbreak using a regression analysis, predicting DF cases using an ANN and simulating the results using a GIS application. The accuracy of the model for the test set is stated as 67% and 62% for the two locations considered. The study further reveals that other than the weather variables, the rest has no significant effect over DF.

An ANN model is a software based model that can be used in predicting certain parameters based on the past data provided. The model learns through the past data and produces a forecast for the future. It has many advantages over the other types of computing models and statistical models, its ability to learn from the data itself, being one remarkable advantage [2].

Most of the research studies done on Dengue, involves weather parameters, considering the significance of weather variables in spreading DF. Not only the Dengue fever, but also the breeding and propagation of the vector mosquito population is positively and negatively affected by the climatic variables such as temperature, rainfall, relative humidity, cloudiness, etc., [3] A research study done in Singapore investigates the delayed effect of weather variables on incidence of DF in Singapore from year 2000 to 2010 [7]. This study establishes the effect of weather incidents such as rainfall or temperature that occurred at a time back in the previous week or month, on the DF cases reported in the

current week. The findings of the research includes that there is a delayed effect of weather variables for the rapid spreading of the disease.

Considering all the above research studies, this research intends to investigate the impact of weather variables and previously reported DF cases on current DF cases and to develop an ANN model to predict DF outbreaks in Sri Lanka. The similar research indicated as Reference [4] only incorporates weather data for the prediction [4]. The effect of previous dengue cases on the prediction is not considered in their research. Further, prediction considers weather data corresponding to the current week only. Thus, their model lacks the early prediction capacity. This research attempts to overcome aforesaid issues. Moreover, a comprehensive accuracy test is performed after developing the ANN to predict DF cases.

Rest of this paper elaborates on the methodology adapted in carrying out the study, the final results and conclusions drawn from those results.

## 2. METHODOLOGY

### 2.1 Data Collection

The data used in this study to predict sourced mainly from two locations. The data used for weather variables were extracted from the weather records maintained by the Meteorological Department of Sri Lanka. Readings recorded at the Katugastota observatory, for daily Average Minimum Temperature, Average Maximum Temperature, daily Total Rainfall, and daily Average Relative Humidity were obtained. The number of Dengue cases reported weekly was extracted from the records maintained at the Regional Epidemiology Unit at the Regional Director of Health Services Office in Kandy, Sri Lanka. The data of 9 Medical Office of Health (MOH) divisions in Kandy district for a period of 3 years (2010 – 2012) was used in the study. Since the prediction is generated for a period of one week, the weather data was also summarized into weekly figures. The total number of data points used in the study is 129, after removing the anomalies.

### 2.2 Selecting Input Parameters

The literature available on medical research on Dengue suggests that the communication of the disease as well as the propagation of the vector mosquito, are positively and negatively affected by certain weather variables [3]. On the other hand, delayed effects of weather variables on Dengue were studied by a one of the previous research [7]. Based on the facts established by those researches, a correlation analysis between each individual weather variable with the Dengue cases reported was carried out using the weather variables with time lags. The variables holding the highest correlation to the Dengue cases are selected as the input parameters to the ANN, with their corresponding time lag.

### 2.3 Designing the ANN

The ANN used to predict Dengue cases is designed as a Multi Layer Perceptron (MLP) network with three initial layers, namely – Input layer, Output layer and one hidden layer. The input layer consists of the input parameters identified in the previous step. The hidden layer is designed with one neuron and to be increased as the training process progresses on demand. Output layer consists of a single parameter which is the prediction of number of Dengue cases for the week ahead. The Log Sigmoid and Tan Sigmoid functions are alternatively used as the activation function for the hidden layer, with Levenburg – Marquardt algorithm as the learning algorithm. It

has shown successful results with feed forward neural networks according to previous research studies [6]. One of the researchers further suggests that Levenberg – Marquardt improves the speed and reliability of back propagation algorithms [6].

### 2.4 Training the ANN

The complete data set is divided into three portions, training set, validation set and the test set as shown in Table 1. The training set, which is used to train the ANN, consist of 70% of the data set. Training process is carried out using the training set, alternatively using Log Sigmoid and Tan Sigmoid functions as the activation function for the hidden layer. For each activation function, the number of hidden layer neurons is increased from 1 neuron up to 30 neurons and trained with re-initialized weights. The ANNs with minimum value of Mean Standard Error (MSE) for the results generated for both training and test sets were extracted to select the ANN most suitable and fit for the purpose.

Table 1. Proportions of data in each data set

Data set	Proportion
Training set	70%
Validation set	15%
Test set	15%

### 2.5 Testing and Validating the ANN

15% of unseen data from the entire data set is used for testing and validating the ANN, respectively. The validation set is particularly used to prevent over fitting the data, and thus using local minima in generating the result. The test set is the data which generates the real prediction.

The ANN is validated using the validation tools, Mean Magnitude of Relative Error (MMRE) and Pred(m).

### 2.6 Selecting the Optimum ANN to Predict Dengue Cases

The set of ANNs extracted during the training process is analyzed under a number of criteria to determine the optimum ANN that generates the optimum prediction. Initially, the ANNs with minimal MSE for both training and test sets were selected. Simultaneously, the correlation coefficient (R) between the predicted and actual values of each selected ANN is also calculated. For the set of ANNs with minimal MSE and R, the validation tools MMRE and Pred (m) are applied. For Pred (m), 25% is used hence the generated forecast lies within 25% tolerance of the actual values.

## 3. RESULTS AND DISCUSSION

### 3.1 Findings of the Correlation Analysis

The results of the correlation analysis done between individual weather variables with the weekly number of reported Dengue cases, revealed the highest correlations for Average Temperature and Average Relative Humidity with 1 week lag, and Total Rainfall and Number of Rainy days per week with 4 weeks lag. In addition, Average Temperature and Average Relative Humidity with 4 weeks lag also displayed higher correlations with the current Dengue cases. Taking the effect of temperature and humidity on blood feeding behaviors and other life cycle events of the vector mosquito, Average temperature and Average Relative Humidity with 4

weeks lag is also introduced as input parameters. Considering the effect of previously reported Dengue cases on the current number of cases reported, the Dengue cases reported on the previous week was also included as an input parameter [3]. Altogether, seven (07) input parameters were fed in to the ANN. Table 2 shows the input parameters.

**Table 2. Input parameters**

Input Parameters
Average Temperature with 1 week lag
Average Temperature with 4 weeks lag
Average Relative Humidity with 1 week lag
Average Relative Humidity with 4 weeks lag
Total Rainfall with 4 weeks lag
Number of Rainy Days with 4 weeks lag
Number of Dengue cases reported in the previous week

The ANN was trained using the above input parameters. The set of ANNs extracted through the training process is displayed in Table 3.

**Table 3. Set of candidate neural networks extracted from the training process**

Net ID	Activation Function	Number of Hidden Neurons
1	Tan sigmoid	3
2	Tan sigmoid	15
3	Tan sigmoid	13
4	Log sigmoid	1
5	Log sigmoid	4
6	Log sigmoid	15
7	Log sigmoid	23
8	Log sigmoid	28

### 3.2 Selecting the optimal ANN

From the training process, 8 ANNs with minimum training and test set MSEs and high correlation coefficients with the actually reported cases were extracted. A summary of the extracted ANNs are displayed in Table 3.

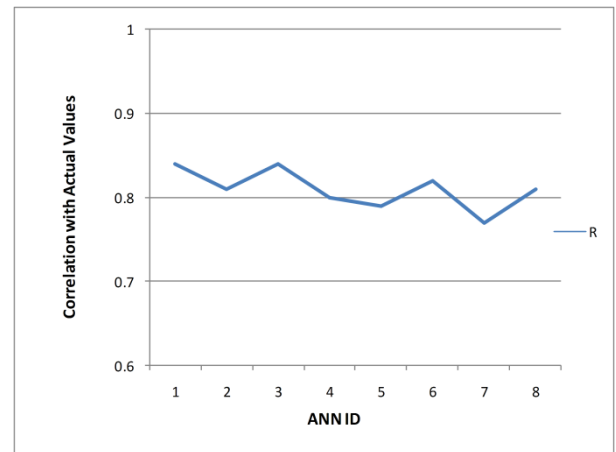
From the candidate ANNs tagged as Net 1 to Net 8, the correlation (R) between the predicted and actual values and the MSE obtained for both training and test set is considered in picking up an optimal ANN to be used in predicting.

Values obtained for R and training and test set MSEs is summarized in Table 4.

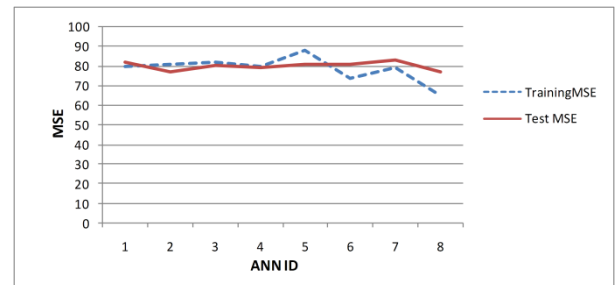
**Table 4. R and MSE values of the candidate ANNs**

Net ID	R	Training set MSE	Test set MSE
1	0.84	80	82
2	0.81	81	77
3	0.84	82	80
4	0.80	80	79
5	0.79	88	81
6	0.82	74	81
7	0.77	79	83
8	0.81	65	77

The variation in correlation coefficients and the MSE for both training and test sets are plotted in Figure 1 and Figure 2 respectively.



**Figure 1. Variation in R**



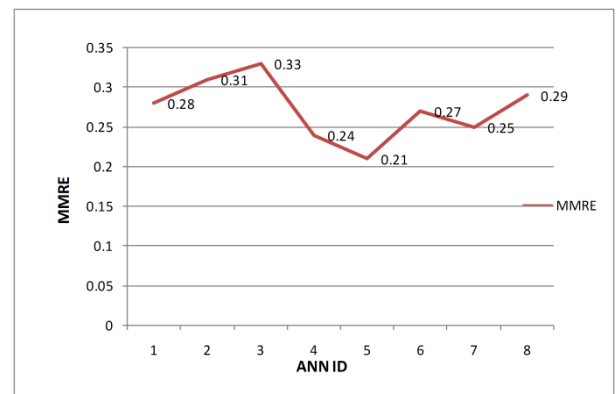
**Figure 2. Variations in training and test set MSE**

The values obtained for R ranges from 0.77 to 0.84, with a mean of 0.81.

The values for training MSE ranges from 65 to 88, with a mean of 78.625 while the test set MSE ranges from 77 to 83 with a mean of 80.

Since the initial selection criteria R and MSE fall within a very narrow range, the validation tools MMRE and Pred (25) is also used in selecting the optimal ANN to predict.

Figure 3 plots the variation obtained for MMRE and Pred (25).



**Figure 3. Variation in MMRE values**

The range of values is from 0.21 to 0.33, the mean being 0.25. The MMRE value displays a fair range and distribution with the minimum value recorded for the ANN tagged as Net 5.

The Pred (25) values are plotted in Figure 4.

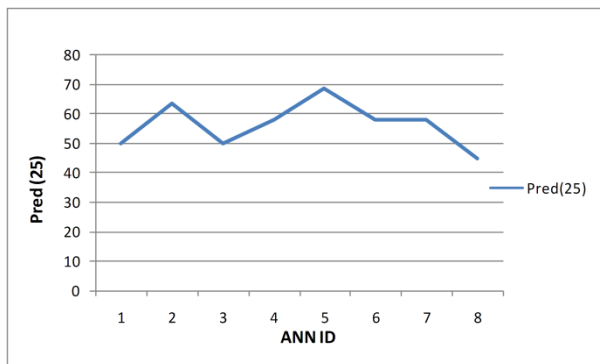


Figure 4. Variation in Pred (25) values

Pred (25) values lie within the range from 50 to 68.42 with a mean of 56.25. The highest value recorded is 68.42, which belongs to the ANN tagged as Net 5.

Thus, in terms of the accuracy of the forecast, both tools suggest the ANN, Net 5 is more suitable. The training and test set MSE being 88 and 81 respectively with a correlation of 0.79 with actual values; hence Net 5 is selected as the optimal ANN to predict Dengue outbreaks.

### 3.3 The Optimal ANN

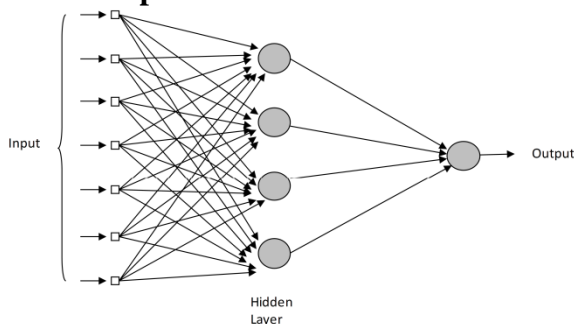


Figure 5. Selected ANN

The architecture of the selected optimal ANN is illustrated in Figure 5.

The selected ANN is a MLP network with seven inputs, one hidden layer with 4 neurons and an output layer consisting of a single output. The activation function of the hidden layer is Log sigmoid function with Levenburg – Marquardt as the learning algorithm. Pure Linear function is used as the output function.

The accuracy indicators of the selected optimal ANN (Net 5) is displayed in Table 5

Table 5. Summary of accuracy indicators of the selected ANN

Net 5	Training set	Test set
R	0.78	0.79
MSE	88.45	80.98
Pred (25)	81.57	68.42
MMRE	0.607	0.21

The output generated for the unseen data represented in the test set is plotted in Figure 6.

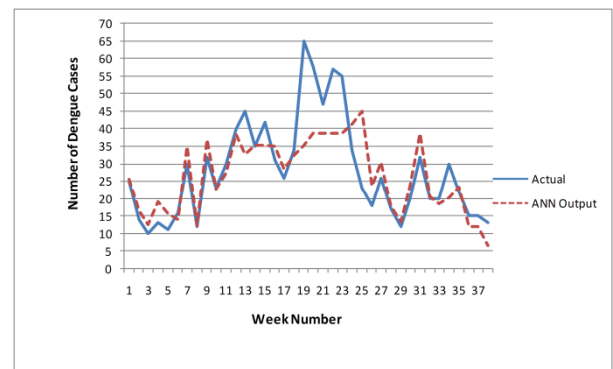


Figure 6. Test set output

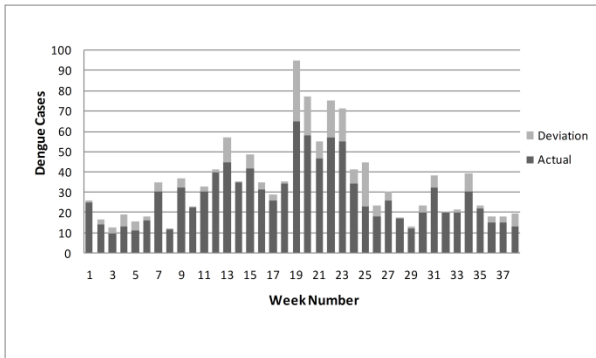
The ANN has performed fairly well with both training and test sets, except for the extremely high peak visible in the test data. However, the presence of this high peak may be attributed by the effect of some other variables which are out of the scope of this study. The fairly high values of the validity indicators Pred (25) and the low value for MMRE confirm the fact. The prediction is generated for a period of one week ahead.

## 4. CONCLUSION AND RECOMMENDATIONS

The results of the correlation analysis between the weather variables and currently reported Dengue cases reveals that average temperature with 1 week lag, average relative humidity with 1 week lag, total weekly rainfall with 4 weeks lag and number of rainy days with 4 week lag and the previously reported Dengue cases with 1 week lag has the highest correlation with the Dengue cases reported in the current week. On the other hand medical and entomological literature establishes effects of temperature and relative humidity on the blood feeding behaviors and development of life cycle stages of the vector mosquito. Thus average temperature with 4 weeks lag and average relative humidity with 4 weeks lag is also included as parameters to the input layer.

The ANN developed with the above data set as the input variables, is a Multi Layer Perceptron Model with three layers. There is one (01) hidden layer which has four (04) neurons. The activation function of the hidden layer the Log Sigmoid function and the Pure Linear function is used as the output function. The learning algorithms used for this ANN is Levenberg – Marquardt algorithm.

The ANN generates predictions with an accuracy of 68.5% according to Pred (25), which means 68.5% of the generated prediction lies within a 25% tolerance of the actually reported cases. On the other hand, the accuracy indicator MMRE records a value of 0.21. The MSE of testing output of this model is 80. The ANN output shows a regression coefficient of 0.79 with the corresponding actual number of Dengue cases. Thus in terms of accuracy, this model outperforms from the previous researches [1], [4] and [9] done on the area. Figure 7 illustrates the actual cases and the deviation in the form of a stacked bar chart.



**Figure 7. Deviations in the prediction**

The data for Average Temperature, Total Rainfall and Relative Humidity used in this research was measured in the Climatic Data Collection Centre in Katugasthota. Therefore this ANN can be used for predicting Dengue outbreaks in Kandy District, at the present. But the selected ANN can be used to generate predictions for any other district of Sri Lanka if the necessary input data is provided.

Due to the unavailability of sufficient data for the vector population density indices, the indices were not used as input parameters in this model. The accuracy of the prediction may be further improved by considering the indices – House Index (HI), Container Index (CI) and Breteau Index (BI) as input parameters. The reason for not predicting some of the peaks may be the effect of variables like above which have not been considered in this study. Therefore in future studies these variables also can be added as input parameters to the ANN and improve the prediction accuracy.

## 5. ACKNOWLEDGMENTS

Our sincere gratitude goes to Dr. (Mrs) Anoma de Silva, the Regional Epidemiologist for Central Province in Sri Lanka, for giving her extensive support in gathering data on Dengue cases reported at MOH offices.

The Information Technology division of the Meteorological Department of Sri Lanka is also recalled with thankfulness for the kind assistance given in obtaining the weather data.

## 6. REFERENCES

- [1] Hani M. Aburas B. Gultekin Cetiner, Murat Sari. 2009. Dengue confirmed-cases prediction: A neural network model' Expert Systems with Applications 37 (2010) 4256–4260.
- [2] Lippman R.P. 1987. An Introduction to Computing with Neural Nets, IEEE ASP Magazine vol 3 No. 4 pp 4-22.
- [3] Medical Research Institute and Dengue Coordination Unit of Ministry of Healthcare and Nutrition of Sri Lanka. 2011. The Practical Manual and Guidelines for Dengue Vector Surveillance.
- [4] Munasinghe A., Premaratne H.L., Fernando M.G.N.A.S. 2013. Towards an Early Warning System to Combat Dengue. International Journal of Computer Science and Electronics Engineering (IJCSEE) Volume 1, Issue 2 (2013).
- [5] Rudnick A., Lim T.W. 1986. Dengue fever studies in Malaysia. Bulletin from the Institute for Medical Research in Malaysia.
- [6] Suratgar A.A., Thawaloki N.B., Hosienibadi M. 1994. Modifying Levenburg Marquardt algorithm for neural network. World Academy of Science, Engineering and Technology, 2005.
- [7] Wang Z., Chan H.M., Zhang T., Kim A.P., Hibbered M.L., Lee G.K.K.L. 2012. Delayed effects of weather variables on incidence of Dengue fever in Singapore from 2000 to 2010. International Journal of Environmental Science and Development, volume3, No 4.
- [8] World Health Organization. Dengue strategic plan for the Asia Pacific Region (2008-2015).
- [9] Yuhani Y., Zuriani M. 2011. Dengue Outbreak Prediction: A Least Squares Support Vector Machines Approach. International Journal of Computer Theory and Engineering, Vol. 3, No. 4, August.