Developing a decision support testing algorithm to detect severity level of dengue.

**ABSTRACT**

Dengue is a vector borne disease that has become a global threat. In order to reduce the mortality rate early detection of dengue severity level is crucial. This study is an extension of the decision models developed individually for inflammatory mediators and immune parameters. The objective of this study is to improve the individual models by considering their combined effect and to improve the decision making at 96 hours from onset of illness. In order to combine these, three approaches are attempted including, combining together the individual full models on inflammatory mediators and immune parameters, combining the immune parameters based model with decision tree informed cytokines and implementing a decision tree informed model with immune parameters and inflammatory mediators. The decision tree algorithm that is used in model development is Improved ID3 algorithm. The decision tree based model is a two-step decision system with the initial decision being made using the parameters TNF-α, IL-10, dengue NS1 antigen and dengue IgG antibody and, the operator values above 0.4413, are then subjected to the second test including platelet and Platelet Activating Factor. The decision tree based model performed well with an accuracy of 76.19% and 82.3% of DHF patients were correctly classified. Sensitivity analysis indicated the model to be robust. Overestimation of the risk of the DF patients is a drawback of the model.

**INTRODUCTION**

Dengue is one of the most concerned mosquito borne viral diseases in the world with the reported dengue cases increasing worldwide from 2.2 million in 2010 to 3.2 million in 2015 (World Health Organization, 2016). In 2013, estimated 9000 deaths occurred globally with a majority being reported from low and middle income countries (Stanaway, et al., 2016). Dengue virus which is transmitted by *Aedes Aegypti* and *Aedes Albopictus* mosquitoes can result in asymptomatic infection, dengue fever (DF) or severe forms of dengue haemorrhagic fever (DHF) and life threatening dengue shock syndrome (DSS). Although a commercial vaccine against dengue is being licensed in several countries, it is still not in use in many parts of the world which suffer from consistent dengue outbreaks (World Health Organization, 2016). Therefore, early detection and careful body fluid management remains important in treating against dengue so as to prevent a patient from moving into critical phase.

Cytokines are believed to be associated with increased vascular permeability that may lead to severe forms of dengue (Appanna, et al., 2012). In this analysis Platelet activating factor (PAF), sphingosine 1- phosphate (S1P), Tumor Necrosis Factor -α (TNF-α) and Interleukin -10 (IL-10) are used as parameters. Previous studies have shown the impact that cytokines and inflammatory mediators can have on determining dengue disease severity. Elevated levels are observed for IL-1β, IFN-γ, IL-4, IL-6, IL-13 and IL-7 in DHF patients than in DF patients and TNF-α is shown to be associated with thrombocytopenia (Bozza, et al., 2008). Also, compared with DF patients, DHF patients have shown to have significantly lower S1P levels throughout the course of illness than DF patients (Gomes, et al., 2014).

Immune parameters used in the study are dengue NS1 antigen levels, dengue IgG antibody levels, platelet counts and lymphocyte counts. NS1 positivity is found to be associated with severe dengue, especially on day 5-6 after illness. Also, it is found to be inversely correlated with lymphocyte count (Paranavitane, et al., 2014). During a secondary dengue infection IgG is present in high values even in the acute phase and thus IgG/IgM antibody ratios are used to distinguish between primary and secondary infection (World Health Organization, 2009). Thrombocytopenia is believed to be a resultant of bone marrow suppression and peripheral destruction of platelets (de Azeredo, et al., 2015).

This study is an extension of the studies in (Jayasundara, et al., 2016) and (Premaratne, et al., 2016) in which fuzzy based decision models are developed separately for inflammatory mediators (Jayasundara, et al., 2016) and immune parameters (Premaratne, et al., 2016). The objective of this study is to develop a decision tree informed fuzzy decision system that can detect dengue severity at an early stage. Thus, this model focuses on detection at 96 hours from onset of illness rather than at 108 or 120 hours from onset of illness as was the case in the previous models. This model is a combination of the previous two models and is targeted at determining disease severity at an early stage. The combined interaction of inflammatory mediators together with immune parameters is of interest to study as, previous studies have shown that S1P levels are significantly correlated with platelet counts in DHF patients (Green, et al., 1999) and IL-10 levels are significantly and inversely correlated with lymphocyte counts (Malavige, et al., 2013). Furthermore, this model targets at reducing the number of parameters that are required for decision making.

Fuzzy logic is commonly used in models involving medicine as it has the ability to handle the imprecision and uncertainty associated in medical decision making. Decision trees are commonly used to handle biological problems and in this study, Improved ID3 (IID3) algorithm is used to determine the effect that inflammatory mediators and immune parameters have on detecting disease severity.

**DEVELOPMENT OF DECISION SUPPORT MODEL**

***Data***

Data was obtained from 36 adult patients who were admitted to the Colombo South Teaching Hospital, Sri Lanka. According to 2011 WHO guidelines, out of these, 11 patients are classified as DF while 25 are classified as DHF. These patients are admitted to the hospital at varying time points ranging from 72-144 hours from onset of illness. However, our analysis is limited to only 96 hours from onset of illness as the aim is for early detection and sufficient data did not exist for earlier time points. In the data set, there are 17 DHF and 4 DF patients who are admitted at 96 hours from onset of illness.

***Decision Tree***

The decision tree algorithm that is used in this analysis is Improved ID3 algorithm (IID3). ID3 algorithm is one of the most widely used algorithms in decision trees. This uses information gain to determine the most suitable property for each node and the attribute with the highest information gain is selected as the attribute for that particular node. However, as ID3 algorithm tend to be biased towards selecting the attribute with many values, this is modified using an association function to overcome this drawback and the modified IID3 algorithm is developed (Jin, et al., 2009).

***Theoretical Framework***

***Fuzzy set***

Elements of a fuzzy set are mapped using a membership function which measures the degree to which an element belongs to the set (Zadeh, 1965).

***Hamacher Operator***

(1)

where *,*  are the membership function values of the fuzzy sets A and B respectively.

***Ordered Weighted Aggregation (OWA) Operator***

OWA operator is defined as,

where is the jth largest element of the collection of aggregated objects and and .

A way to determine the weights of OWA operator is suggested by (Yager, 1993) and (Yager, 1988) as,

and the non-decreasing quantifier Q is defined in (Zadeh, 1983) as,

where.

‘Orness measure’ which measures the degree to which the aggregation operation is like an ‘OR’ operation is defined by (Yager, 1988) as

which lies in the interval [0,1].

***Model Development***

The fuzzy membership functions are developed according to the previous studies (Jayasundara, et al., 2016) and (Premaratne, et al., 2016).

The membership functions for IL-10, TNF- α, PAF and S1P are given by the equations (3), (4), (5) and (6) respectively.

The membership functions for platelet, dengue NS1 antigen, lymphocyte and dengue IgG antibody levels are given in equations (7), (8), (9) and (10) respectively.

Three approaches are attempted to combine the individual models in (Jayasundara, et al., 2016) and (Premaratne, et al., 2016). For future reference, the cytokines and inflammatory mediators based model of (Jayasundara, et al., 2016) is defined as Model A and immune parameters based model of (Premaratne, et al., 2016) is defined as Model B.

1st approach: Combining together the individual models, Model A and Model B.

2nd approach: Combining Model B with only the decision tree informed cytokines in Model A.

3rd approach: Decision tree informed model consisting of immune parameters and inflammatory mediators.

***Decision Tree Output***

IID3 algorithm based decision trees are implemented for 96 and 108 hours from onset of illness for the parameters in Model A and Model B.

*96 hours from onset of illness.*

The decision tree for inflammatory mediators and immune parameters evaluated at 96 hours from onset of illness is shown in Fig.1 and Fig.2 respectively. In the figures, y=1 refers to DHF patients and y=2 refers to DF patients and ‘n’ refers to the number of patients classified under that particular decision making.

Fig.1: Decision tree at 96 hours from onset of illness for cytokines.

As it can be seen from Fig.1, at 96 hours from onset of illness decisions are made using only the parameters TNF- α and IL-10 and the parameters PAF and S1P are not used in the decision process. At this time point, 12 out of 17 (70.59%) of DHF patients are classified based only on the IL-10 concentration. All four DF patients are categorized by the decision criteria IL-10 >= 31.81(pg/ml) and TNF <8.16 (pg/ml).

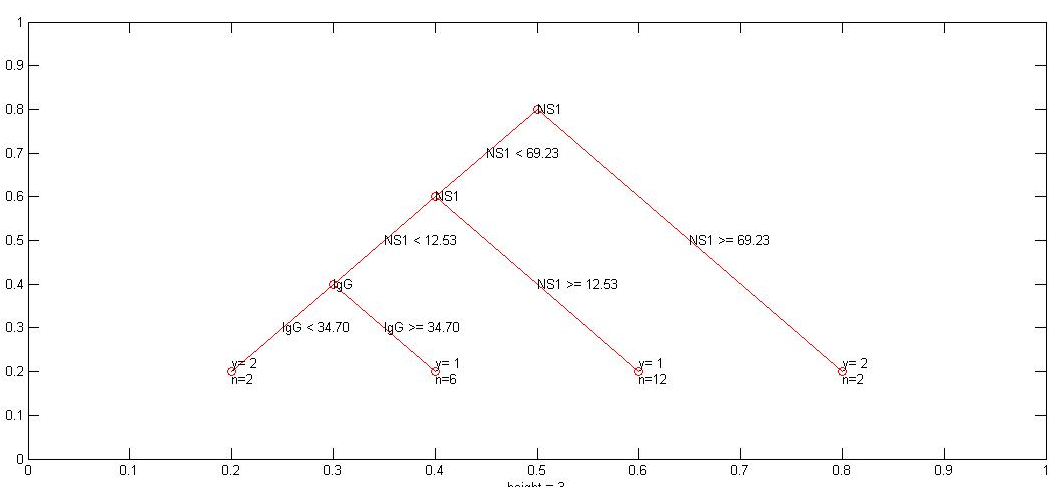
Also, according to Fig.1, if the IL-10 concentration is less than 31.81 (pg/ml) the patients are directly categorized as DHF and thus indicates that IL-10 is an important parameter in making decisions as it was in (Brasier, et al., 2012). However, according to (Perez, et al., 2004), (Green, et al., 1999) IL-10 levels have shown to be higher in DHF patients than in DF patients and according to (Chen, et al., 2006) DHF patients showed a median IL-10 level of 110.8 pg/ml and DF patients a median of 15.5 pg/ml. Therefore, further analysis is required to determine if the low levels of IL-10 in DHF patients is significant or whether it is specific to this data set alone.

Fig.2: Decision tree at 96 hours from onset of illness for immune parameters.

From Fig.2 it can be seen that at 96 hours from onset of illness decisions are made using only the parameters dengue NS1 antigen and dengue IgG antibody, indicating these parameters to be informative at 96 hours from onset of illness. The parameters platelet and lymphocyte counts are not used in the decision process.

*108 hours from onset of illness.*

The decision tree for inflammatory mediators and immune parameters evaluated at 108 hours from onset of illness is shown in Fig.3 and Fig.4 respectively. In the figures, y=1 refers to DHF patients and y=2 refers to DF patients and ‘n’ refers to the number of patients classified under that particular decision making

Fig.3: Decision tree at 108 hours from onset of illness for inflammatory mediators.

As it can be seen from Fig.3 at 108 hours from onset of illness since TNF-α has the highest information gain, decision making is started with TNF-α. Decisions are made using only the parameters TNF- α and PAF. 13 out of 17 (70.59%) of DHF patients are classified based only on TNF-α values.

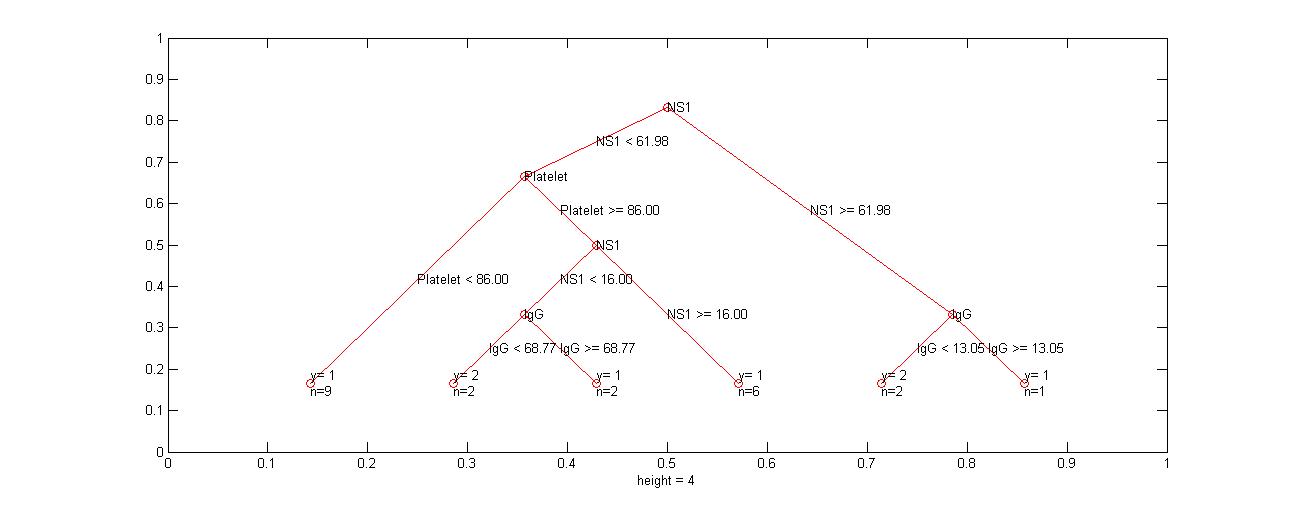


Fig.4: Decision tree at 108 hours from onset of illness for immune parameters.

As it can be seen from Fig.4, at 108 hours from onset of illness since NS1 has the highest information gain, decision making is started with it. Decisions are made using only the parameters NS1, IgG and platelet count but, the lymphocyte count is not taken for consideration.

**RESULTS AND DISCUSSION**

***1st approach: Combining together the individual models, Model A and Model B.***

Model A and Model B are combined together by combining the individual model output values of the two models through the Hamacher operator. The accuracy of the model is calculated as,

At 96 hours the combined model performed with an accuracy of 58.82%. However, still 42.85% of the DHF patients’ model results fell in between 0.3151-0.5264, which is the ambiguous region defined in Model B and the DF patients’ results did not improve. However, none of the DHF patients fell into the non-severe region.

***2nd approach: Combining Model B with only the decision tree informed cytokines in Model A.***

Since the immune parameters in Model B are easy to measure and the inflammatory mediators are costly, it is decided to use the full Model B and improve its results by incorporating the cytokines which are indicated as informative by the decision tree. As the objective is for early detection, the decision tree results evaluated at 96 hours from onset of illness are used to modify the Model B. As it can be seen from Fig.1, since the decision tree results are based on the cytokines TNF-α and IL-10, their combined effect is considered along with the results from Model B.

In the Model B, 43.75% of the DHF patients fell into the non-severe region and 12.5% of DHF patients fell into the ambiguous region. Therefore, to improve the outcome of the misclassification in Model B, the patients whose operator value is above the lower limit of the ambiguous region (0.3151), were further subjected to the combined effect given in equation (11).

where are the membership values of TNF-α and IL-10 respectively and both are concentrated by 1.1 as in (Jayasundara, et al., 2016). H1 produces the Hamacher product between TNF-α and IL-10.

H1 is combined with the Model B results as given in equation (12).

where *OWAa* refers to the final model output value given by model B.

This implementation of the decision system performed with an accuracy of 64.7%. However, 28.57% of the DHF patients are categorized as non-severe and the DF patients’ results did not improve. The misclassification of the DHF patients into non-severe region is a concern and hence further development of the model is required.

***3rd approach: Decision tree informed model consisting of immune parameters and inflammatory mediators.***

IID3 decision tree algorithm is evaluated for 96 hours from onset of illness for the two sets of parameters, one including cytokines (Fig.1) and the other including immune parameters (Fig.2).

Since, at 96 hours from onset of illness, the decision tree based on inflammatory mediators outputs the results based only on the parameters TNF-α and IL-10 (Fig.1) and the immune parameters based decision tree outputs the results based only on the parameters NS1 and IgG (Fig.2), the initial decision model is developed using only these four parameters. In order to account for the combined effect from the parameters, the Hamacher operator is performed as described in equations (11) and (13). The fuzzy concentration and dilution of the parameters is performed as in Model A and Model B.

where are the membership values of NS1 and IgG respectively and NS1 is diluted by 0.2 and IgG is concentrated by 1.1 as in Model B.

The Hamacher results in (11) and (13) are then combined using the Ordered Weighted Aggregation Operator (OWA) as in (14).

The weights of the OWA operator are determined from Model A.

The ambiguous region of the model is constructed using the concepts in Model A and Model B and the ambiguous region is found to lie in between the values 0.4413 and 0.5727. Thus, values above 0.5727 indicate the region with DF patients (non-severe) whereas the region with DHF patients (severe) is shown for values less than 0.4413.

When the model is validated at 96 hours from onset of illness, it could be seen that 58.8% of the DHF patients are correctly classified into severe region. Although this model has a much lower percentage of misclassifications than the Model A and Model B, still the model misclassifies 41.2% of DHF patients into ambiguous and non-severe (DF) regions. This is a costly error as, by classifying a patient as DF who indeed should be classified as a DHF patient, the patient may not receive proper attention and thus the patient can easily proceed into critical stage without being detected.

Therefore, the existing model is further improved by using information from the decision trees evaluated at 108 hours from onset of illness (Fig.3 and Fig.4). Since, Fig.3 uses PAF and Fig.4 uses platelet count in the decision process, the model is further improved to include the effect from the parameters PAF and Platelet count. Since the model in previous stage misclassifies DHF patients into non-severe region, the patients with initial model operator values above 0.4413 are subjected to second decision making which includes PAF and Platelet count as given in equation (14).

where are the membership values of Platelet and PAF respectively and both are concentrated by 4.

For these patients, after going through the second decision criteria, the final model output result is obtained through equation (15).

When this improved model which includes second testing with PAF and platelet count is validated at 96 hours from onset of illness, it can be seen that the model output results have improved for the DHF patients than it was in the previous stage, as now 82.3% of DHF patients are correctly classified. Without testing for PAF and Platelet counts five DHF patients (29.4%) were misclassified and two (11.76%) were in the ambiguous region. However, after considering their PAF and platelet counts there were only two misclassifications of DHF patients and only one DHF patient fell into the ambiguous region.

At 96 hours from onset of illness, the final model performed with an accuracy of 76.19%. This improved model, implemented using decision tree results performs with much higher accuracy than the individual models developed in (Jayasundara, et al., 2016) and (Premaratne, et al., 2016). The final validation results of the three models are shown in Table 1. From Table 1, it can be seen that the decision tree informed model performs the best with an accuracy of 76.19%.

Table 1: Accuracy of the models evaluated at 96 hours from onset of illness.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model B | Model A | Model in 1st approach | Model in 2nd approach | Model in 3rd approach |
| 53.00% | 71.43% | 58.82% | 64.7% | 76.19% |

As the decision tree informed model (3rd approach) performed with an accuracy of 76.19% at 96 hours from onset of illness, and resulted in 82.3% of correct DHF classification, this model can be used for early decision making. This model uses a reduced number of parameters than the previous studies (Jayasundara, et al., 2016) and (Premaratne, et al., 2016). The cytokines IL-1β and S1P and the immune parameter lymphocyte count are not used in this model. This makes it a relatively cost effective model as well. The accuracy of this model at 96 hours from onset of illness is highly improved than Models A and B. This indicates that the model with combined interactions between inflammatory mediators and immune parameters performs well than when they are used individually to measure dengue severity and that decision trees play a vital role in making decisions about parameter selection for the decision model.

***Sensitivity analysis***

Sensitivity analysis is performed to see how the decision tree based model outcomes would be affected when the degrees of fuzziness are changed. Fig. 5 denotes the change in the ambiguous region when the lower and upper boundary of the membership functions are changed over a small range. It can be seen that as the boundary values of the membership functions are changed the ambiguous region does not change in a way that could affect the classification of the patients.





(c)

(b)

(a)



(d)

(e)



(f)

Fig.5: Behaviour of the ambiguous region when the lower cut off value and the upper cut off value of the membership function is changed for the parameters TNF (a, b), NS1 (c, d), IL-10 (e, f), IgG (g, h).

Fig.6 denotes the change in the ambiguous region when the amounts of concentration and dilution are changed.

(a)

(b)

(d)



(c)

Fig.6: Behaviour of ambiguous region for a change in concentration weights for TNF (a), IL-10 (b), IgG (c), NS1 (d).

It can be seen that when the parameter concentration and dilutions are varied over a small range the model outcomes remain unchanged as the ambiguous region remains relatively unchanged. Therefore, Fig. 5 and Fig.6 indicate that the model is robust.

**CONCLUSIONS**

This study is an attempt to develop a decision model to predict the severity level of dengue patients and is an improved extension of the models in (Jayasundara, et al., 2016) and (Premaratne, et al., 2016) . Three variations are carried out including, combining together the individual full models in (Jayasundara, et al., 2016) and (Premaratne, et al., 2016), combining the model in (Premaratne, et al., 2016) with decision tree informed cytokines and implementing a decision tree informed model with immune parameters and inflammatory mediators. In the decision tree based model, IID3 decision trees are used to determine which inflammatory mediators and immune parameters should be used in constructing the decision model.

It could be seen that the decision tree based model, performs with the highest accuracy and that the other two models succumb to the more serious error of misclassifying DHF patients in to the non-severe region.

The decision tree based model makes decisions mainly based on the parameters NS1, IgG, TNF-α and IL-10. Second decision criteria is made using PAF and platelet counts. Unlike in (Jayasundara, et al., 2016) this model uses only three inflammatory mediators for the decision making and this can be economically beneficial. This model can be considered as an improved decision maker than the individual models in (Jayasundara, et al., 2016) and (Premaratne, et al., 2016). Although the model tends to overestimate the risk of DF patients, this model reduces the more serious risk of incorrectly classifying DHF patients as DF patients. The model validation results for patients at 96 hours from onset of illness achieved an accuracy of 76.19% and only 2 DHF patients (out of 17) were misclassified. The sensitivity analysis indicated the model to be robust. Therefore this model will be useful as an early indicator to detect dengue severity.

The drawback of this model is it tend to overestimate the risk of DF patients. In the final decision tree informed model 3 out of 4 DF patients were misclassified into severe region. However, more data on DF patients is needed to accurately measure their behaviour with the model and the data limitation on DF patients contributes to the model being biased towards severe level. Also, all the patients considered in this sample are adult patients and it is important to test the decision rules on samples which include children as well, as severe dengue and death is common among children (World Health Organization, 2016).

# REFERENCES

Appanna, R. et al., 2012. Cytokine factors present in dengue patient sera induces alterations of junctional proteins in human endothelial cells. *Am J Trop Med Hyg,* Volume 87, pp. 936-942.

Bozza, F. A. et al., 2008. Multiplex cytokine profile from dengue patients: MIP-1 beta and IFN-gamma as predictive factors for severity. *BMC Infectious Diseases,* 8(86).

Brasier, A. R. et al., 2012. A three component biomarker panel for prediction of dengue hemorrhagic fever. *Am J Trop Med,* 86(2), pp. 341-348.

Chen, L. et al., 2006. Correlation of serum levels of macrophage migration inhibitory factor with disease severity and clinical outcome in dengue patients.. *Am J Trop Med Hyg. ,* Volume 74, pp. 142-147.

de Azeredo, E. L., Monteiro, R. Q. & de-Oliveira Pinto, L. M., 2015. Thrombocytopenia in Dengue: Interrelationship between Virus and the Imbalance between Coagulation and Fibrinolysis and Inflammatory Mediators. *Mediators of Inflammation,* Volume 2015.

Gomes, L. et al., 2014. Sphingosine 1-Phosphate in Acute Dengue Infection. *PloS ONE,* 9(11).

Green, S. et al., 1999. Elevated plasma interleukin-10 levels in acute dengue correlate with disease severity. *J Med Virol,* Volume 59, pp. 329-334.

Jayasundara, P., Perera, S. S. N., Malavige, G. N. & Jayasinghe, S., 2016. Mathematical modelling and a systems science approach to describe the role of cytokines in the evolution of severe dengue. *Manuscript submitted for publication.*

Jin, C., De-lin, L. & Fen-xiang, M., 2009. *An Improved ID3 Decision Tree Algorithm.*

Kittugul, L., Temprom, W., Sujirara, D. & Kittugul, C., 2000. Determination of tumor necrosis factor- alpha in dengue virus infected patients by sensitive biotin-streptravidin enzyme-linked immunosorbent assay. *J Virol Methods.,* Volume 90, pp. 51-57.

Malavige, G. N. et al., 2013. Serum IL-10 as a marker of severe dengue infection. *BMC Infect Dis.,* 13(341).

Paranavitane, S. A. et al., 2014. Dengue NS1 antigen as a marker of severe clinical disease. *BMC infectious diseases,* 14(570).

Perez, A. et al., 2004. IL-10 levels in Dengue patients: Some findings from the exceptional epidemiological conditions in Cuba. *J Med Virol,* Volume 59, pp. 230-234.

Potts, J. A. et al., 2010. Prediction of Dengue Disease Severity among Pediatric Thai patients using early clinical laboratory indicators. *PLoS Negl Trop Dis,* 4(8).

Premaratne, M. K., Perera, S. S. N., Malavige, G. N. & Jayasinghe, S., 2016. Mathematical Modelling of Immune Parameters in the Evolution of Severe Dengue. *Manuscript submitted for publication.*

Stanaway, J. D. et al., 2016. The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. *Lancet Infect Dis.,* 16(6), pp. 712-723.

World Health Organization, 2009. *Dengue Guidelines for Diagnosis, Treatment, Prevention and Control,* France.

World Health Organization, 2016. *Dengue and Severe Dengue.* [Online]   
Available at: http://www.who.int/mediacentre/factsheets/fs117/en/

World Health Organization, 2016. *Weekly Epidemiological Record,* Geneva: World Health Organization.

Yager, R. R., 1988. On ordered weigted averaging aggregation operators in multicriteria decision making.. *IEEE Trans. Syst., Man, Cybern. ,* Volume 18, pp. 183-190.

Yager, R. R., 1993. Families of OWA operators. *Fuzzy Set Syst,* Volume 59, pp. 125-48.

Zadeh, L. A., 1983. A computational approach to fuzzy quantifiers in natural languages. *Comput Math Appl.,* Volume 9, pp. 149-184.

Zadeh, L. A., 1965. Fuzzy Sets. *Information and Control,* Volume 8, pp. 338-353.