

## Research Article

# Fuzzy Multidimensional Model to Cluster Dengue Risk in Sri Lanka

I. T. S. Piyatilake <sup>1</sup> and S. S. N. Perera<sup>2</sup>

<sup>1</sup>Department of Computational Mathematics, University of Moratuwa, Sri Lanka

<sup>2</sup>Research & Development Center for Mathematical Modeling, Department of Mathematics, Faculty of Science, University of Colombo, Sri Lanka

Correspondence should be addressed to I. T. S. Piyatilake; [thilinispi@uom.lk](mailto:thilinispi@uom.lk)

Received 8 May 2020; Revised 31 July 2020; Accepted 17 September 2020; Published 5 November 2020

Academic Editor: Adriana Carr

Copyright © 2020 I. T. S. Piyatilake and S. S. N. Perera. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Dengue is the world's rapidly transmitting mosquito-borne viral disease. It is mostly found in subtropical countries in the world. The annual number of global deaths caused by dengue fever is about 25,000. The Sri Lanka dengue situation is also not different to other countries. In the year 2019, dengue fever caused 120 deaths in Sri Lanka. Most of these deaths were reported from the main administrative district Colombo. Health authorities have to pay their attention to control this new situation. Therefore, identifying the hot spots in the country and implementing necessary actions to control the disease is an important task. This study aims to develop a clustering technique to identify the dengue hot spots in Sri Lanka. Suitable risk factors are identified using expert ideas and reviewing available literature. The weights are derived using Chang's extent method. These weights are used to prioritize the factors associated with dengue. Using the geometric mean, the interaction between the triggering variable and other variables is calculated. According to the interaction matrices, five dengue risk clusters are identified. It is found that high population movement in the area plays a dominant role to transmit the disease to other areas. Most of the districts in Sri Lanka will reach to moderate risk cluster in the year 2022.

## 1. Introduction

The World Health Organization (WHO) considered *dengue* as one of the world's top ten global health hazards in 2019 [1]. It is a mosquito-borne viral disease, and this virus belongs to the *Flaviviridae* family of viruses. Dengue virus appears as four different, but closely associated serotypes of viruses, namely, DENV-1, DENV-2, DENV-3, and DENV-4 [2]. Infected female mosquitoes of the categories *Aedes aegypti* and *Aedes albopictus* are found to be spreading the dengue virus. These viruses are getting to healthy human bodies by infected mosquito bites. If a human being recovers from one serotype of dengue, he will not have that type of dengue in future, but it does not grant the cross protection to other serotypes. A consecutive contamination by different serotypes is a risk factor for having severe forms of dengue, namely, Dengue Hemorrhagic Fever (DHF) and Dengue Shock Syndrome (DSS), and these two forms are considered as life-threatening stages

of dengue [3]. High fever, headache, pain behind the eye, nausea, vomiting, muscle and joint aches, different types of rashes, bleeding from the nose and gums, and bruises and facial flushing are the symptoms of dengue fever. In addition, widespread bleeding, low blood pressure, and body organ such as liver and kidney failure can occur in DHF and DSS stages. Things such as used tires and food containers which can collect and store water and poor collecting and disposing of garbage in urban areas are some of the factors helping the breeding of the mosquitoes. Also, some rarely used water collecting methods such as water tanks, wells, and poor access to sanitation facilities help the breeding of the mosquitoes. Above all these, helpful climate effects for mosquito breeding process is significant.

WHO statistics indicate the risk of exposure to dengue is nearly half the world human population [1]. As an average, 390 million dengue cases were reported worldwide in every past year [1]. The annual number of global deaths caused by severe forms of dengue is about 20,000 [1]. Among the infected

people of dengue, 90% of them are children and most of them are from Asian and Latin American countries. The global dengue treatment expenditure is about 8 billion dollars [4].

The situation in Sri Lanka is also not different to other countries. Colombo was the capital city and was the busiest industrial area of Sri Lanka in 1965 to 1966. The first dengue outbreak was reported in Colombo in the same period. Until 1989, the dengue virus progressed at a very slow rate and was identified as a sporadic disease. During the period 1989 to 1990, the number of recorded DHF cases was 206 and that is the first epidemic of dengue fever in Sri Lanka [5]. Thereafter, the numbers of dengue cases and deaths have abruptly increased. The year 2017 reported the largest outbreak of dengue in Sri Lanka with 186,101 cases and 440 deaths [6, 7]. While the Colombo district reported 18,186 cases, the Gampaha district reported 12,121 cases [8]. These two figures were the highest in the country. In recent decades, dengue becomes a leading public health problem in Sri Lanka, because the annual number of patients was increased by this disease. Not only that, it spreads to new areas due to infected travelers. For instance, in 1965, the dengue cases are initially found mostly in and around Colombo district, but they progressively spread to other districts and reached outbreak proportions in several main districts such as Gampaha, Kurunegala, Kalutara, Batticaloa, Kegalle, Ratnapura, Kandy, Matara, Galle, and Trincomalee districts in 2017 [9].

Still, vaccine has not been developed by scientist to prevent dengue attacks. Destroying mosquito breeding sites is of the utmost importance to control spreading dengue. Therefore, it is beneficial to identify the clusters with a similar risk of dengue transmission. This would facilitate the relevant authorities to implement the same control strategies in similar risk areas effectively. It also helps to allocate funds and resources efficiently based on the intensity of dengue risk. Therefore, this study is aimed at developing a multidimensional risk model to cluster the dengue hot spots in urban areas of Sri Lanka based on risk factors and predict the possible future hot spots by identifying the variation of appropriate factors.

Several studies focused to develop clustering algorithms to identify the area with a similar risk of dengue transmission. The study developed by Lowe et al. [10] proposed a Bayesian inference-based model which is used to identify the intensity of dengue risk in Brazil during the 2014 FIFA World Cup season. The factors altitude, population density, temperature, and rainfall were used to find the risk of dengue. The results of their study revealed that the risk of dengue was low in the selected cities for the football matches. A statistical analysis was performed by Vicente et al. [11] to detect the spatial variation of dengue temporal data in Brazil. They considered the relative risk of dengue, house index, population density, and income for their analysis. According to their study, there were 11 clusters in Vitória, Brazil. In the Sri Lankan context, there were a very limited number of studies carried out in this regard. Sumanasinghe et al. [12] developed a geostatistical risk model to cluster the high-risk dengue areas in Sri Lanka. The effect of rainfall and population density were considered for their study. They identified five risk areas using geographically weighted regression. Combining past dengue records with climate data, a new two cluster model

TABLE 1: Linguistic variables used in the model construction and their triangular fuzzy number representation.

Linguistic term	Triangular fuzzy scale	Fuzzy scale for $\delta = 0.5$
Absolutely more important	$(3 - \delta, 3, 3 + \delta)$	$(5/2, 3, 7/2)$
Very strongly more important	$(5/2 - \delta, 5/2, 5/2 + \delta)$	$(2, 5/2, 3)$
Strongly more important	$(2 - \delta, 2, 2 + \delta)$	$(3/2, 2, 5/2)$
Weakly more important	$(3/2 - \delta, 3/2, 3/2 + \delta)$	$(1, 3/2, 2)$
Equally important	$(1 - \delta, 1, 1 + \delta)$	$(1/2, 1, 3/2)$
Just equal	$(1, 1, 1)$	$(1, 1, 1)$

was proposed by Sun et al. [13]. They used spatial-temporal clustering method, to investigate the dengue clusters in Sri Lanka from 2012 to 2016. All these Sri Lankan studies were based on only climate factors and population density. Therefore, it is worthwhile to build a multifactorial model considering social and climate factors to cluster the dengue risk areas. These social and climate factors are highly uncertain, and so, it is not possible to define the exact boundaries of those factors. In the present study, the combined effect of multifactors is used in an uncertain environment to find the dengue risk of a given area which was not discussed in the previous studies.

Defining multifactorial risk models is quite a sophisticated task due to the uncertainty nature of such factors. Fuzzy mathematics which was invented in 1965 by Zadeh [14] is one path to model the uncertainty situations. Therefore, fuzzy mathematical concepts are used to develop the problem. The model is constructed as a hierarchical process by including factors and their risk categories in two different levels. The analytical hierarchical process (AHP) is a mathematical tool which was invented by Saaty in 1970 to handle hierarchical processes. Using AHP, we can decompose the given decision problem into different subproblems which can easily handle. Due to the uncertainty of the factors in our problem, fuzzy analytical hierarchical process (FAHP) is used to construct the model. Finally, the dengue risk categories are constructed combining the results of FAHP and Haddon matrices.

The remaining sections of this paper are organized as follows. Section 2 introduces mathematical approaches, including basic definitions of fuzzy theory, the FAHP technique, the concept of Haddon matrix, and geometric mean calculation which we used to build the risk model. The model construction and factor selection processes are presented in Section 3. The clustering algorithm, risk categories, and obtained results with discussion are presented in Section 4. Finally, the conclusion and remarks can be found.

## 2. Theoretical Background

In this section, we will look at some mathematical theories including commonly used definitions in fuzzy theory, Chang's extent analysis method, geometric mean, and Haddon matrix. These theories are gathered from the references [15–21].

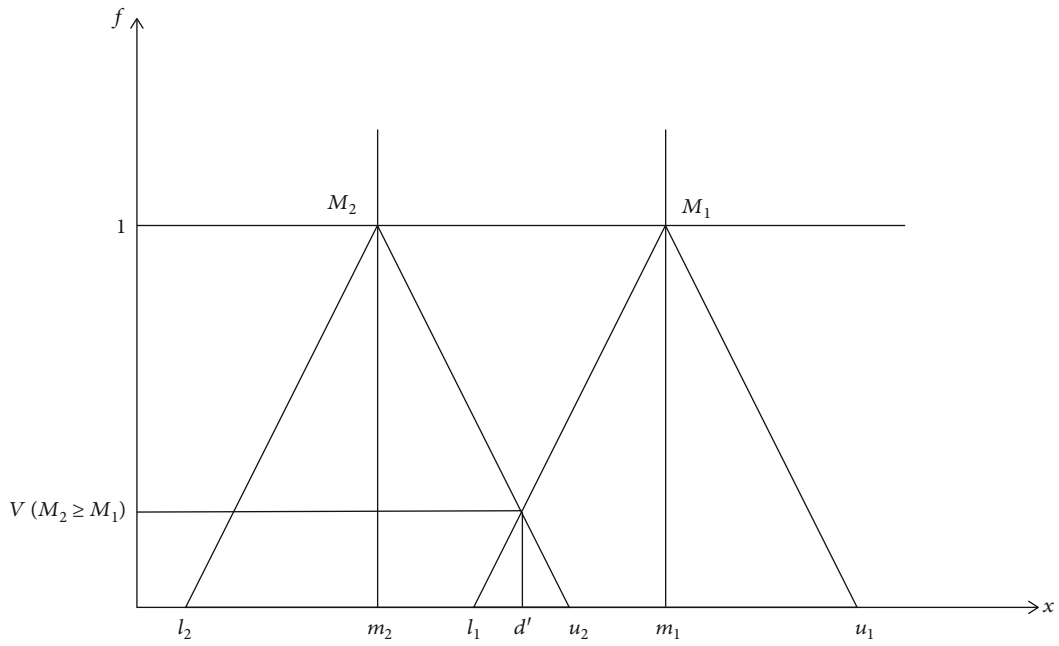


FIGURE 1: The intersection of  $M_1$  and  $M_2$ .

TABLE 2: Selected risk factors with their subcategories.

Risk factor	High	Risk level Moderate	Low
T	$30^{\circ}\text{C} < T < 34^{\circ}\text{C}$	$16^{\circ}\text{C} \leq T \leq 30^{\circ}\text{C}$ or $34^{\circ}\text{C} \leq T \leq 37^{\circ}\text{C}$	$T < 16^{\circ}\text{C}$ or $T > 37^{\circ}\text{C}$
R	$10\text{ mm} < R < 30\text{ mm}$	$5\text{ mm} \leq R \leq 10\text{ mm}$ or $30\text{ mm} \leq R \leq 55\text{ mm}$	$R < 5\text{ mm}$ or $R > 55\text{ mm}$
D	$D > 4$	$2 \leq D \leq 4$	$D < 2$
H	$H > 75$	$50 \leq H \leq 75$	$H < 50$
G	$G > 100$	$50 \leq G \leq 100$	$G < 50$
P	$P > 300$	$100 \leq P \leq 300$	$P < 100$
U	$U > 40$	$20 \leq U \leq 40$	$U < 20$
M	$M > 200,000$	$100,000 \leq M \leq 200,000$	$M < 100,000$

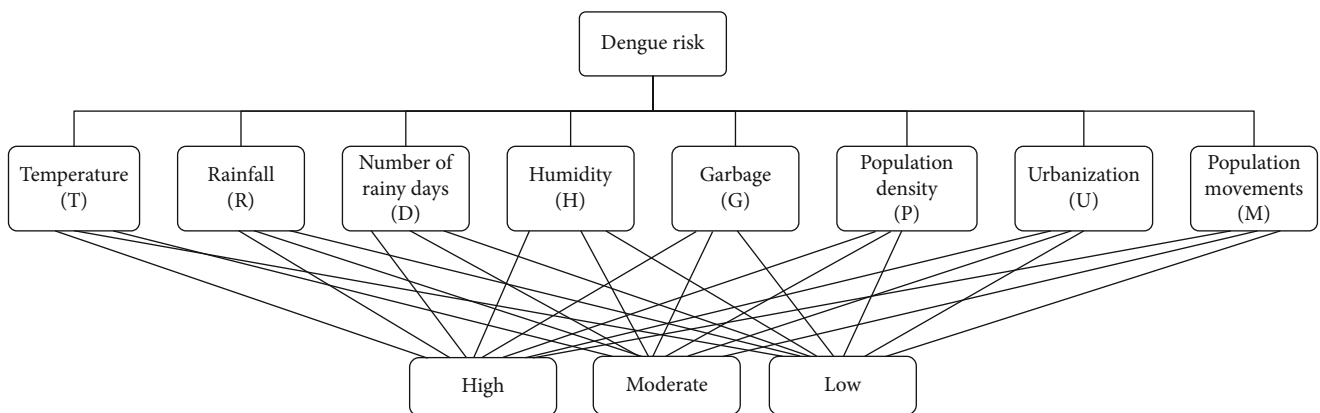


FIGURE 2: The proposed hierarchical model.

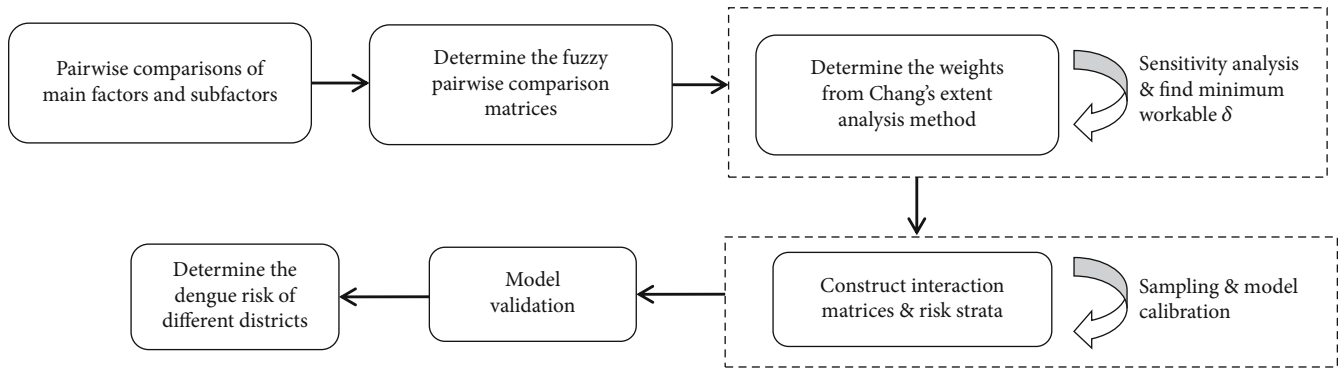


FIGURE 3: The framework for dengue risk model development.

TABLE 3: Matrix to find the interaction between the trigger variable with the other variables [21].

Risk level of variables	Risk level of trigger variable ( $\Omega$ )		
	High	Moderate	Low
High	$\Omega_{\text{High}} \times \Phi_{\text{High}}$	$\Omega_{\text{Moderate}} \times \Phi_{\text{High}}$	$\Omega_{\text{Low}} \times \Phi_{\text{High}}$
Moderate	$\Omega_{\text{High}} \times \Phi_{\text{Moderate}}$	$\Omega_{\text{Moderate}} \times \Phi_{\text{Moderate}}$	$\Omega_{\text{Low}} \times \Phi_{\text{Moderate}}$
Low	$\Omega_{\text{High}} \times \Phi_{\text{Low}}$	$\Omega_{\text{Moderate}} \times \Phi_{\text{Low}}$	$\Omega_{\text{Low}} \times \Phi_{\text{Low}}$

TABLE 4: Construct risk strata using GM values [21].

$\Phi$	$\Omega$ high	$\Omega$ moderate	$\Omega$ low
High	$R_{H,H} = \sqrt[7]{\prod_{i=1}^7 \Omega_{\text{High}} \times \Phi_{\text{High}}}$	$R_{H,M} = \sqrt[7]{\prod_{i=1}^7 \Omega_{\text{Moderate}} \times \Phi_{\text{High}}}$	$R_{H,L} = \sqrt[7]{\prod_{i=1}^7 \Omega_{\text{Low}} \times \Phi_{\text{High}}}$
Moderate	$R_{M,H} = \sqrt[7]{\prod_{i=1}^7 \Omega_{\text{High}} \times \Phi_{\text{Moderate}}}$	$R_{M,M} = \sqrt[7]{\prod_{i=1}^7 \Omega_{\text{Moderate}} \times \Phi_{\text{Moderate}}}$	$R_{M,L} = \sqrt[7]{\prod_{i=1}^7 \Omega_{\text{Low}} \times \Phi_{\text{Moderate}}}$
Low	$R_{L,H} = \sqrt[7]{\prod_{i=1}^7 \Omega_{\text{High}} \times \Phi_{\text{Low}}}$	$R_{L,M} = \sqrt[7]{\prod_{i=1}^7 \Omega_{\text{Moderate}} \times \Phi_{\text{Low}}}$	$R_{L,L} = \sqrt[7]{\prod_{i=1}^7 \Omega_{\text{Low}} \times \Phi_{\text{Low}}}$

TABLE 5: The derived weights of sublevels of factors.

Factor	Risk level		
	High	Moderate	Low
T	0.5584	0.3446	0.0970
R	0.5905	0.2120	0.1975
D	0.4632	0.3905	0.1734
H	0.3694	0.3307	0.3000
G	0.5301	0.4480	0.0219
P	0.3694	0.3307	0.3000
U	0.5301	0.4480	0.0219
M	0.5905	0.2120	0.1975

2.1. Basic Definitions in Fuzzy Theory

Definition 1. A fuzzy set,  $A$ , in a universe of discourse,  $U$ , is a function of the form

$$f_A : U \rightarrow [0, 1]. \tag{1}$$

The function  $f_A$  is called the membership function of  $A$ , and for any  $x$  in  $U$ ,  $f_A(x)$  in  $[0, 1]$  represents the grade of membership of  $x$  in  $A$ .

Definition 2 (triangular fuzzy number). A triangular fuzzy number (TFN),  $A$ , can be defined by a triplet  $(l, m, u)$ , where  $l, m, u$  represent the smallest possible value, the most promising value, and the largest possible value of an event, respectively. This representation is interpreted using the membership function as follows:

$$f_A(x) = \begin{cases} \frac{x-l}{m-l}, & \text{if } x \in [l, m] \\ \frac{u-x}{u-m}, & \text{if } x \in [m, u] \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

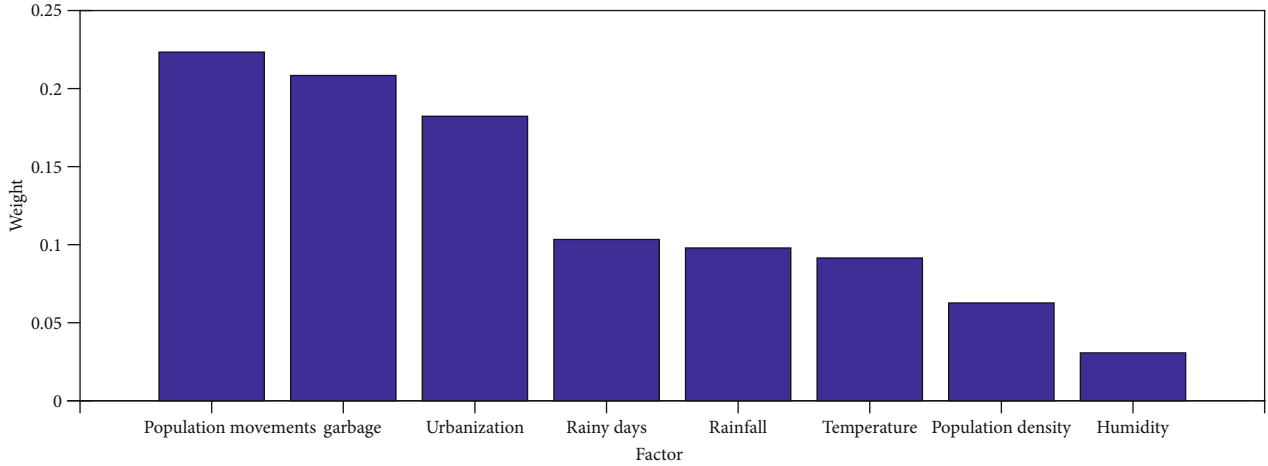


FIGURE 4: Prioritized weights of the predictive factors.

*Definition 3* (basic operations). Assume that  $A$  and  $B$  are two triangular fuzzy numbers with  $A = (l_1, m_1, u_1)$  and  $B = (l_2, m_2, u_2)$ . The basic operations are:

- (1) Addition

$$A \oplus B = (l_1 + l_2, m_1 + m_2, u_1 + u_2). \tag{3}$$

- (2) Multiplication

$$A \otimes B = (l_1 l_2, m_1 m_2, u_1 u_2). \tag{4}$$

- (3) Inverse

$$A^{-1} \approx \left( \frac{1}{u_1}, \frac{1}{m_1}, \frac{1}{l_1} \right). \tag{5}$$

*2.2. Fuzzy Pair Wise Comparison Matrix.* Let  $\tilde{A}$  represent a fuzzified reciprocal  $n \times n$  judgment matrix containing all pairwise comparisons  $\tilde{a}_{ij}$  between elements  $i$  and  $j$  for all  $i, j \in 1, 2, 3, \dots, n$ .

$$\tilde{A} = \begin{matrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{matrix} \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ (1, 1, 1) & \tilde{a}_{12} & \cdots & \tilde{a}_{1n} \\ \tilde{a}_{21} & (1, 1, 1) & \cdots & \tilde{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{a}_{n1} & \tilde{a}_{n2} & \cdots & (1, 1, 1) \end{bmatrix}, \tag{6}$$

where  $\tilde{a}_{ij} = (1, 1, 1) : \forall i = j, \tilde{a}_{ji} = \tilde{a}_{ij}^{-1}$ ,  $n$  is the criteria number to be evaluated,  $c_i$  is the  $i$ th criteria,  $\tilde{a}_{ij}$  is the importance of the  $i$ th criteria according to the  $j$ th criteria, and all  $\tilde{a}_{ij}$  are triangular fuzzy numbers  $\tilde{a}_{ij} = (l_{ij}, m_{ij}, u_{ij})$ .

*2.3. Linguistic Variable*

*Definition 4* (linguistic variable). A linguistic variable [14] is characterized by a quintuple  $(x, T(x), U, G, \tilde{M})$ , in which  $x$  is the name of the variable,  $T(x)$  denotes the term set of  $x$ , that is, the set of names of linguistic values of  $x$ . Each of these values is a fuzzy variable, denoted generically by  $X$  and ranging over a universe of discourse  $U$ , which is associated with the base variable  $u$ ;  $G$  is a syntactic rule (which usually has the form of a grammar) for generating the name,  $X$ , of values of  $x$ .  $M$  is a semantic rule for associating with each  $X$  its meaning. Here,  $\tilde{M}(X)$  is a fuzzy subset of  $U$ . A particular  $X$ , that is, a name generated by  $G$ , is called a term.

For the analysis purposes, the expert judgments regarding variables are processed using linguistic variables. Traditionally, we represent these linguistic variables using numerical scales. Commonly used numerical scales are 1 to 3, 1 to 5, 1 to 7, and 1 to 9. As we know, in the real scenario, these expert judgments are uncertain. Therefore, in this study, these uncertain ideas are analyzed using triangular fuzzy numbers. The linguistic scale and their triangular fuzzy representation are shown in Table 1.

*2.4. Chang's Extent Analysis Method.* Let  $X = \{x_1, x_2, x_3, \dots, x_n\}$  be an object set and  $G = \{g_1, g_2, g_3, \dots, g_n\}$  be a goal set. Then, each object is taken and the extent analysis for each goal is performed, respectively. There are  $m$  extent analysis values for each object, and the values are  $M_{g_i}^1, M_{g_i}^2, \dots, M_{g_i}^m, i = 1, 2, \dots, n$  where  $M_{g_i}^j (j = 1, 2, \dots, m)$  are TFNs. The steps of Chang's extent analysis method are as follows:

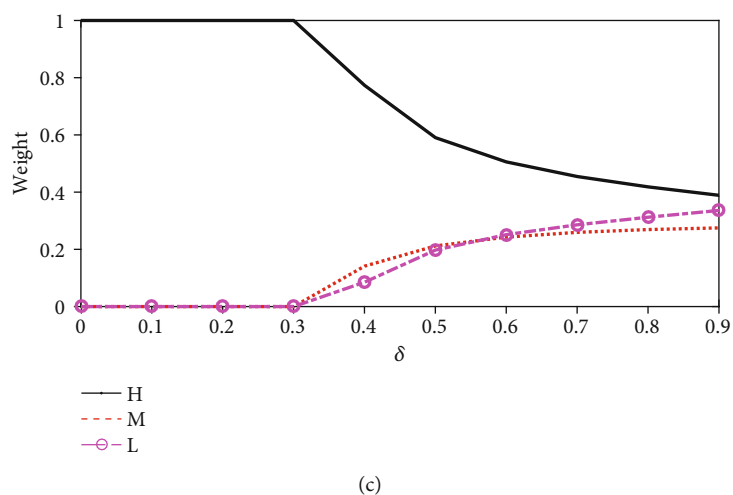
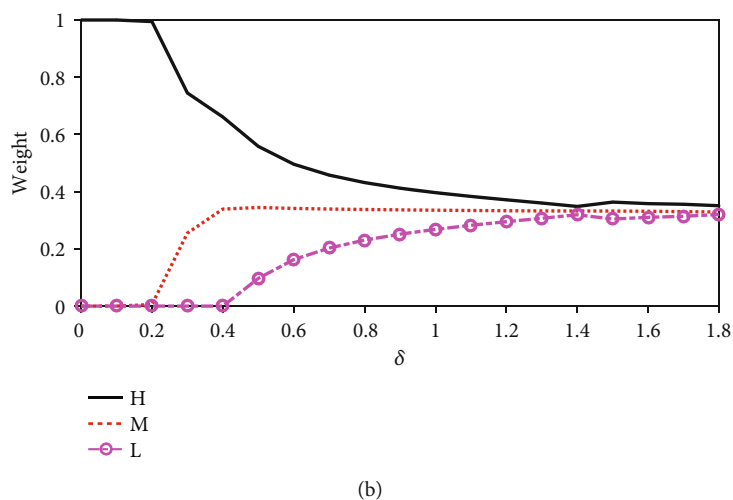
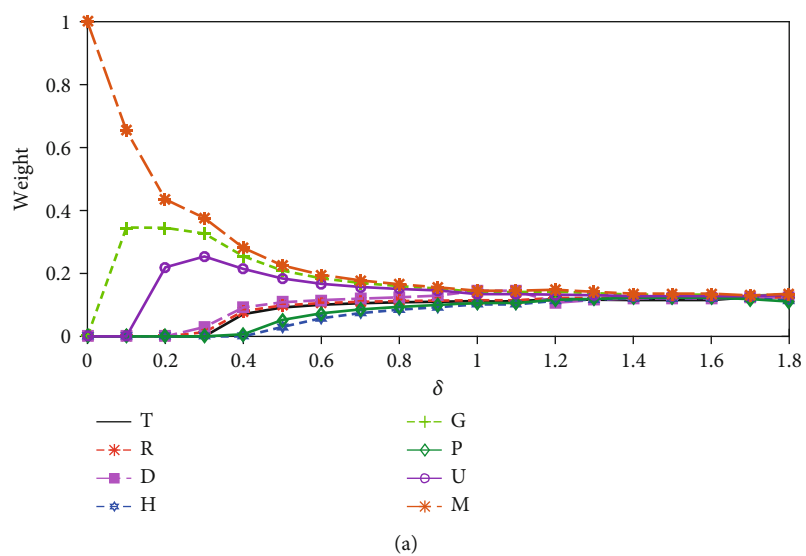
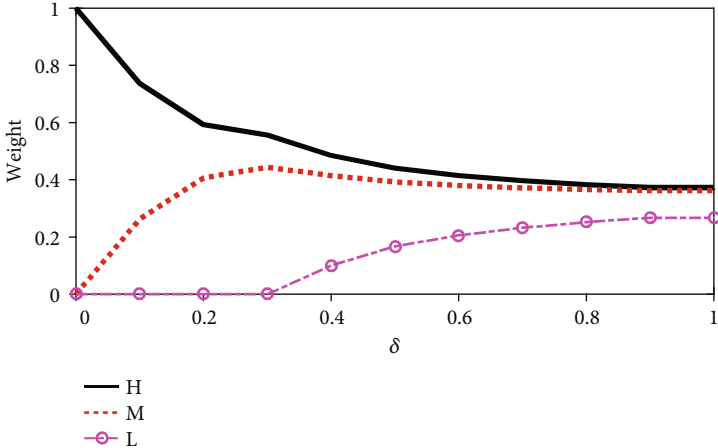
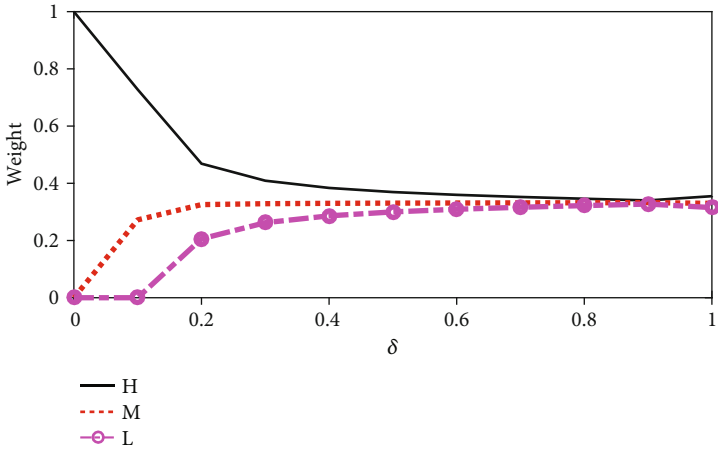


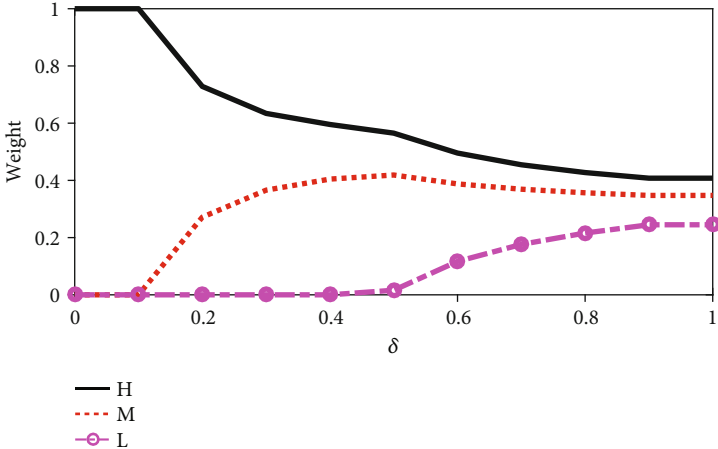
FIGURE 5: Continued.



(d)

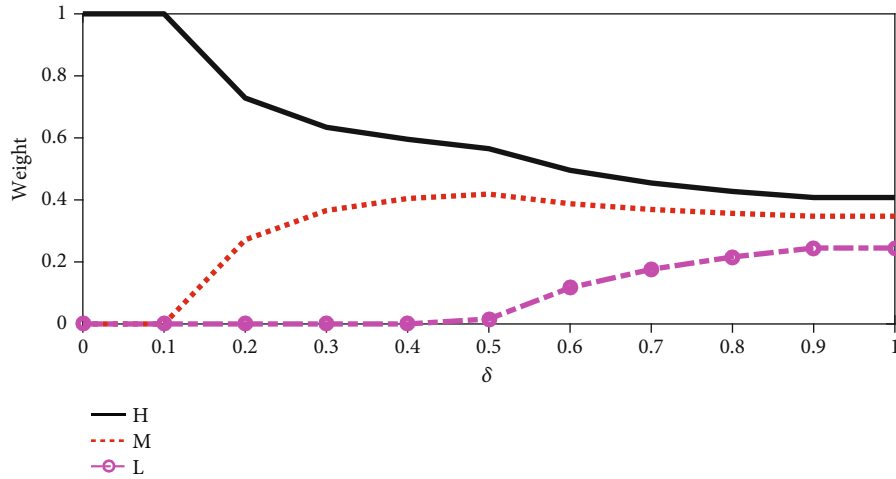


(e)

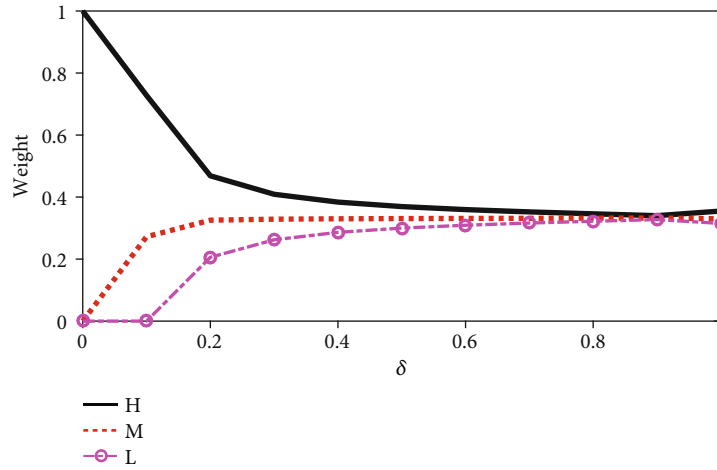


(f)

FIGURE 5: Continued.



(g)



(h)

FIGURE 5: Sensitivity of the weights of factors for different values of  $\delta$ : (a) all predictive factors; (b) temperature; (c) rainfall; (d) number of rainy days; (e) humidity; (f) garbage (g); population density; (h) urbanization.

Step 1. The value of fuzzy synthetic extent with respect to the  $i$ th object is

$$s_i = \sum_{j=1}^m M_{g_i}^j \otimes \left[ \sum_{i=1}^n \sum_{j=1}^m M_{g_i}^j \right]^{-1} \quad (7)$$

The value of  $\sum_{j=1}^m M_{g_i}^j$  is obtained using the “fuzzy addition operation” of  $m$  extent analysis values for a particular matrix, and the value  $[\sum_{i=1}^n \sum_{j=1}^m M_{g_i}^j]^{-1}$  is obtained using the “fuzzy addition operation” of  $M_{g_i}^j (j = 1, 2, \dots, m)$  values.

Step 2. The degree of possibility of two triangular fuzzy numbers  $M_2 = (l_2, m_2, u_2) \geq M_1 = (l_1, m_1, u_1)$  can be expressed as

follows:

$$V(M_2 \geq M_1) = \begin{cases} 1 & \text{if } m_2 \geq m_1 \\ 0 & \text{if } l_1 \geq u_2 \\ \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)}, & \text{otherwise.} \end{cases} \quad (8)$$

According to the values of  $V(M_1 \geq M_2)$  and  $V(M_2 \geq M_1)$ , two triangular fuzzy numbers can be compared.

Step 3. Assume that

$$d'(A_i) = \min V(S_i \geq S_k), \quad (9)$$

for  $k = 1, 2, \dots, n; k \neq i$ , where  $d'$  is the ordinate of the highest intersection point  $D$  between  $\mu_{M_1}$  and  $\mu_{M_2}$ , and it is shown in Figure 1.



TABLE 6: GM of different risk factor levels.

Cluster	High	Moderate	Low
High	0.2833	0.1017	0.0948
Moderate	0.2059	0.0739	0.0689
Low	0.0622	0.0223	0.0208

TABLE 7: Proposed risk clusters with their color scheme.

Risk cluster	GM	Color
Very high	0.2833-1.0000	Maroon
High	0.2059-0.2833	Red
Moderate	0.0689-0.2059	Yellow
Low	0.0208-0.0689	Light green
Very low	0.0000-0.0208	Green

Step 4. Then, the weight vector is given by

$$W' = \left( d'(A_1), d'(A_2), \dots, d'(A_n) \right)^T, \quad (10)$$

where  $A_i, (i = 1, 2, \dots, n)$  are  $n$  elements.

Step 5. Normalizing (10) can obtain the normalized weight vector

$$W = (d(A_1), d(A_2), \dots, d(A_n))^T, \quad (11)$$

where  $d(A_i) = (d'(A_i)) / (\sum_{i=1}^n d'(A_i))$ . Now,  $W$  is a non-fuzzy number.

2.5. *Haddon Matrix.* The concept of the Haddon matrix [22] was invented by William Haddon in 1970. This was first introduced as a tool to recognize the risk factors associated with injury occurrences. This conceptual framework facilitates users to think about injury prevention activities and their association with factors in three stages: before injury, injury, and after injury. The cells of this matrix provide users different approaches to prevent the injury.

2.6. *Geometric Mean.* Geometric mean (GM) [23] is used to measure the central location of a given data set. For a given set of positive random variables  $\{x_i\}_{i=1}^N$ , we can define GM value as follows:

$$GM = \sqrt[N]{\prod_{i=1}^N x_i} = \left( \prod_{i=1}^N x_i \right)^{\frac{1}{N}}, \quad (12)$$

where  $N$  is the population (or sample) size. For the zero and negative random variables, we can define GM value considering the following three cases.

Case 1. Given  $N$  is an odd number and all the random variables  $\{x_i\}_{i=1}^N$  are less than 0, in this situation, GM value is given by

$$GM = \sqrt[N]{\prod_{i=1}^N (-x_i)} = -\sqrt[N]{\prod_{i=1}^N |x_i|}. \quad (13)$$

Case 2. Given  $N_1$  is the number of a negative random variable in the given set of data and contains both positive and negative values, then the GM value is given by

$$GM = \sqrt[N]{\prod_{i=1}^N x_i} = \sqrt[N]{\prod_{i=1}^{N_1} x_i^- \prod_{i=N_1+1}^N x_i^+} = \sqrt[N]{\prod_{i=1}^{N_1} x_i^-} \sqrt[N]{\prod_{i=N_1+1}^N x_i^+}. \quad (14)$$

According to (14), the total GM value is the multiplication of GM of negative values and GM of positive values. Therefore, this total GM value can be called as bigeometrical. In this case, the total GM can be calculated as a weighted average, and it is given by,

$$GM = W_1 G_- + W_2 G_+, \quad (15)$$

where  $G_- = \sqrt[N]{\prod_{i=1}^{N_1} x_i^-}$ ,  $G_+ = \sqrt[N]{\prod_{i=N_1+1}^N x_i^+}$ ,  $W_1 = N_1/N$ , and  $W_2 = N_2/N$ . Here,  $N_2$  is the number of positive values in the set.

Case 3. Given some of the random variable values are equal to 0, in this situation, we should consider about three different GM values to compute the overall GM. Therefore, this GM value is called as trigeometrical. In here, the overall GM value is given by

$$GM = W_1 G_- + W_2 G_+ + W_3 G_0, \quad (16)$$

where  $G_- = \sqrt[N]{\prod_{i=1}^{N_1} x_i^-}$ ,  $G_+ = \sqrt[N]{\prod_{i=N_1+1}^{N_2} x_i^+}$ ,  $G_0 = 0$ ,  $W_1 = N_1/N$ ,  $W_2 = N_2/N$ , and  $W_3 = N_3/N$ . Here,  $N_2$  is the number of positive values in the set,  $N_3$  is the number of zero values in the set, and  $N = N_1 + N_2 + N_3$ .

### 3. Methodology

3.1. *Selection of Predictive Factors.* After reviewing the facts of the national and international sources regarding the dengue transmission, eight different predictive factors are identified. They are the average value of daily temperature ( $^{\circ}\text{C}$ ), average value of daily rainfall (mm), average number of rainy days per week, average value of daily humidity (%), average daily collection of garbage (Mt/Day),

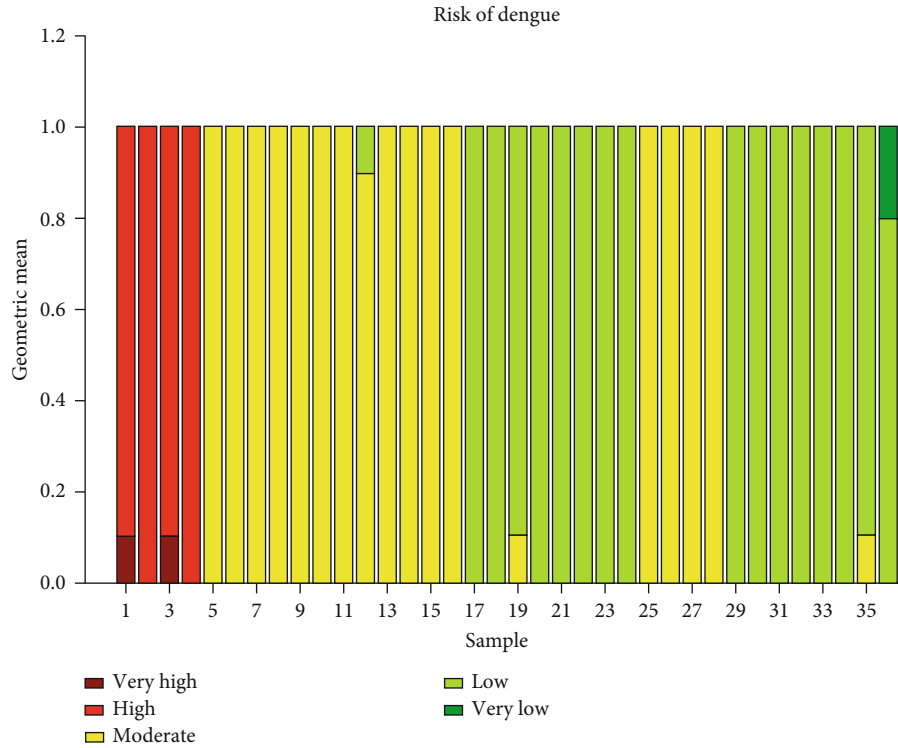


FIGURE 6: Model calibration with different samples.

population density (persons/km<sup>2</sup>), percentage of urbanization, and number of population movements to the target area. By considering the impact of these factors on dengue severity, they are further subdivided into three risk levels, namely, high, moderate, and low. Expert ideas and literature are used for this purpose [21, 24–26]. The selected factors with their subcategories are given in Table 2. Using all the above facts, the proposed hierarchical model is presented in Figure 2.

**3.2. Constructing Multidimensional Dengue Risk Clusters.** We pursue the steps which are presented in Figure 3 to construct the risk clusters. Using expert opinions in the field, the subfactors are compared and constructed the fuzzy pair-wise comparisons matrices. The weights of the factors are derived using Chang's extent analysis method. According to the derived weights, the factor with the highest numerical value is defined as the trigger variable ( $\Omega$ ) in the context of dengue transmission. Then, the interaction between the variable  $\Omega$  and the remaining seven factors ( $\Phi; i = 1, 2, \dots, 7$ ) are calculated. These calculated interaction values are stored in seven  $3 \times 3$  matrices as in Table 3 [21, 24]. Finally, these seven matrices are combined using the GM values as in Table 4.

According to the output of Table 4 and using the concept of the Haddon matrices, three risk clusters can be defined as follows:

(1) *Low-risk cluster.* The value of  $R_{L,L}$  is the lower limit, and the value of  $R_{M,L}$  is the upper limit of this cluster.

(2) *Moderate risk cluster.* The value of  $R_{M,H}$  is the lower limit, and the value of  $R_{M,L}$  is the upper limit of this cluster.

(3) *High-risk cluster.* The value of  $R_{M,H}$  is the lower limit, and the value of  $R_{H,H}$  is the upper limit of this cluster.

To find the risk clusters of districts membership values of predictive factors are aggregated using GM. MATLAB and R statistical softwares are used to carry out the simulations.

## 4. Results

### 4.1. Algorithm

*Step 1.* Compare the predictive factors using expert ideas.

With the aid of expert ideas, pairwise comparison matrices as in Table S1 to Table S9 (see Supplementary material S1) are created.

*Step 2.* Prioritize the predictive factors.

Synthetic extent value of each factor is derived using (7) and the results in Table S1 (see Supplementary material S1). These derived values are given as follows:

$$\begin{aligned}
 S_T &= (0.0644, 0.1062, 0.1713) & S_R &= (0.0577, 0.1045, 0.1798) \\
 S_D &= (0.0601, 0.1058, 0.1854) & S_H &= (0.0500, 0.0794, 0.1320) \\
 S_G &= (0.1023, 0.1749, 0.2949) & S_P &= (0.0527, 0.0859, 0.1545) \\
 S_U &= (0.0953, 0.1554, 0.2584) & S_M &= (0.1148, 0.1878, 0.2978)
 \end{aligned}$$



FIGURE 7: A map of the districts in Sri Lanka.

Then, to find the degrees of possibilities above, the values are compared according to the condition given in (8). The results are given as follows:

$V(S_T \geq S_R) = 1$	$V(S_R \geq S_T) = 0.9852$	$V(S_D \geq S_T) = 0.9964$	$V(S_H \geq S_T) = 0.7163$	$V(S_G \geq S_T) = 1$	$V(S_P \geq S_T) = 0.8161$
$V(S_T \geq S_D) = 1$	$V(S_R \geq S_D) = 0.9893$	$V(S_D \geq S_R) = 1$	$V(S_H \geq S_R) = 0.7479$	$V(S_G \geq S_R) = 1$	$V(S_P \geq S_R) = 0.8390$
$V(S_T \geq S_H) = 1$	$V(S_R \geq S_H) = 1$	$V(S_D \geq S_H) = 1$	$V(S_H \geq S_D) = 0.7320$	$V(S_G \geq S_D) = 1$	$V(S_P \geq S_D) = 0.8262$
$V(S_T \geq S_G) = 0.5013$	$V(S_R \geq S_G) = 0.5238$	$V(S_D \geq S_G) = 0.5459$	$V(S_H \geq S_G) = 0.2372$	$V(S_G \geq S_H) = 1$	$V(S_P \geq S_H) = 1$
$V(S_T \geq S_P) = 1$	$V(S_R \geq S_P) = 1$	$V(S_D \geq S_P) = 1$	$V(S_H \geq S_P) = 0.9245$	$V(S_G \geq S_P) = 1$	$V(S_P \geq S_G) = 0.3696$
$V(S_T \geq S_U) = 0.6071$	$V(S_R \geq S_U) = 0.6238$	$V(S_D \geq S_U) = 0.6447$	$V(S_H \geq S_U) = 0.3258$	$V(S_G \geq S_U) = 1$	$V(S_P \geq S_U) = 0.4599$
$V(S_T \geq S_M) = 0.4095$	$V(S_R \geq S_M) = 0.4382$	$V(S_D \geq S_M) = 0.4626$	$V(S_H \geq S_M) = 0.1374$	$V(S_G \geq S_M) = 0.9329$	$V(S_P \geq S_M) = 0.2805$
			$V(S_U \geq S_T) = 1$	$V(S_M \geq S_T) = 1$	
			$V(S_U \geq S_R) = 1$	$V(S_M \geq S_R) = 1$	

$$\begin{aligned}
V(S_U \geq S_D) &= 1 & V(S_M \geq S_D) &= 1 \\
V(S_U \geq S_H) &= 1 & V(S_M \geq S_H) &= 1 \\
V(S_U \geq S_G) &= 0.8893 & V(S_M \geq S_G) &= 1 \\
V(S_U \geq S_P) &= 1 & V(S_M \geq S_P) &= 1 \\
V(S_U \geq S_M) &= 0.8161 & V(S_U \geq S_T) &= 1
\end{aligned}$$

Then, using (10), the minimum degree of possibilities is calculated. Normalizing these values according to (11), the final weights of the factors are derived. Then, the final weight vector of the predictive factors is

$$W = (0.0915, 0.0979, 0.1033, 0.0307, 0.2084, 0.0627, 0.1823, 0.2234). \quad (17)$$

Similarly, we can derive the weights of the sublevel of the factors, and these values are given in Table 5. Finally, the derived weights are used to prioritize the predictive factors as shown in Figure 4. According to the output of Figure 4, the population movements in the area have the highest weight. Therefore, the variable population movements in the area are selected as  $\Omega$ .

#### Step 3. Justification of selection of $\delta$ .

In this step, we carry out a sensitivity analysis. This is to identify the impact of  $\delta$  to the weights of the factors. In order to identify this, we carry out simulations for different values of  $\delta$ , and these results are shown in Figure 5. According to the simulated results, we can observe that some of the weights are equal to 0 for some values of  $\delta$ . Therefore, we must select a value for  $\delta$  which gives a weight value for all the factors. For this study, we select this  $\delta$  value as 0.5.

#### Step 4. Define dengue risk clusters.

The multiplicative interaction between each predictive factor with the subcategories of population movements is calculated using the concept explained in Table 3. These results are shown in Table S10 to Table S16 (see Supplementary material S2).

As explained in Table 4, the cells of Table S10 to Table S16 (see Supplementary material S2) combined using GM. The generated results are in Table 6. The proposed risk clusters and their boundaries based on levels of dengue risk with the proposed color scheme are shown in Table 7.

#### Step 5. Model calibration.

The random sampling technique is used to calibrate the proposed model [21, 24]. For each sublevel of the trigger variable, 36 random samples are generated. To generate these samples, other variables are kept in three different probability levels. These probability levels are 80%, 50%, and 20%. For example, think that the trigger variable is in its high-risk sublevel. Then, one variable is selected with 80% probability and the remaining six variables selected with its complementary probability of 20%. Likewise, 10 repeated samples are generated randomly for this selection. Then, using the GM value, the risk cluster of each sample is generated. Then, the number of clusters with similar results are counted. Similarly, random samples are generated for other sublevels of the trigger variable, and the results are given in Figure 6.

TABLE 8: Ranks of districts in Sri Lanka and their dengue categories.

District	GM	Risk cluster	Relative dengue cases (%)
Ampara	0.0707	Moderate	0.53
Anuradhapura	0.0422	Low	1.66
Badulla	0.0488	Low	2.14
Batticaloa	0.0696	Moderate	3.20
Colombo	0.2230	High	19.57
Galle	0.0508	Low	3.58
Gampaha	0.1414	Moderate	18.07
Jaffna	0.0774	Moderate	3.47
Kalutara	0.1392	Moderate	6.26
Kandy	0.1414	Moderate	8.23
Kegalle	0.0488	Low	5.46
Kilinochchi	0.0793	Moderate	0.30
Kurunegala	0.1392	Moderate	6.43
Hambantota	0.0511	Low	2.04
Mannar	0.0711	Moderate	0.31
Matarara	0.0462	Low	3.62
Mullaitivu	0.0700	Moderate	0.22
Nuwara Eliya	0.0462	Low	0.51
Puttalam	0.0488	Low	4.48
Ratnapura	0.0503	Low	6.45
Trincomalee	0.0619	Low	2.86
Vavuniya	0.0679	Low	0.61

As per Figure 6, we can identify the following three population movement clusters.

#### (1) Cluster with high population movements

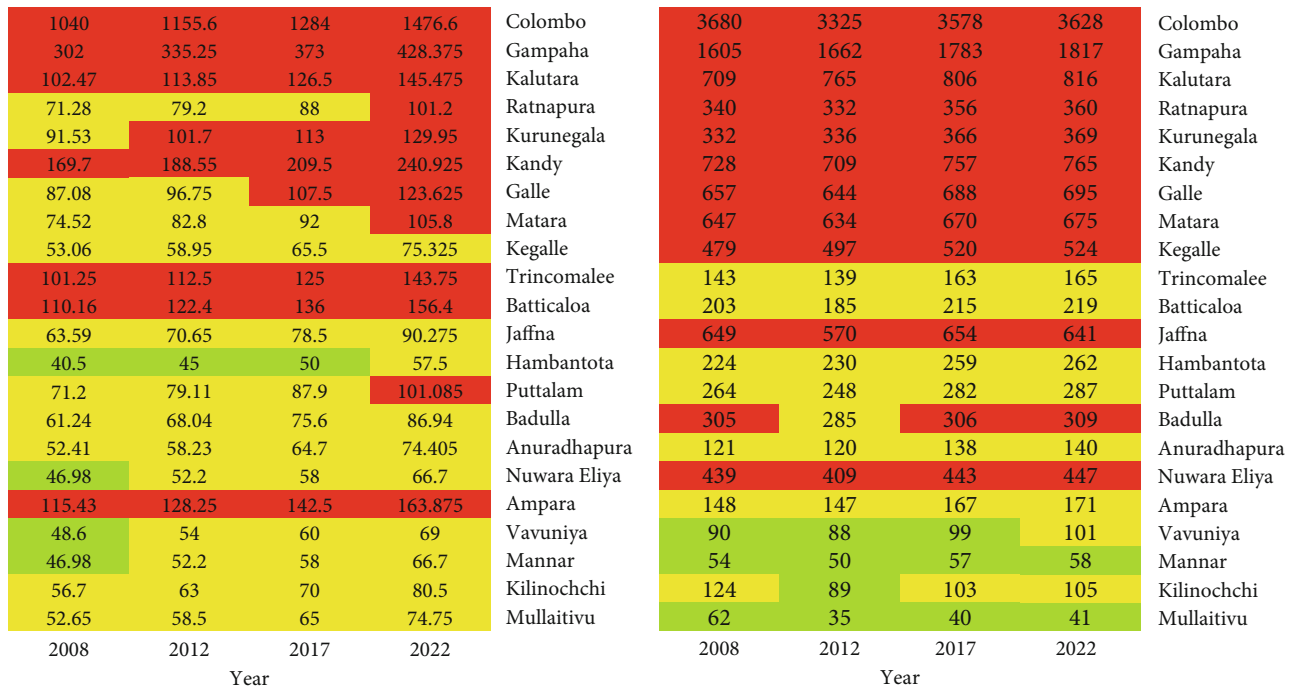
In Figure 6, the first twelve samples are generated with high population movements. According to Figure 6, we can see maroon, red, and yellow bars appear in this region. That means, when there are high population movements, the dengue risk belong to very high, high, and moderate dengue clusters.

#### (2) Cluster with moderate population movement

In Figure 6, the second set of twelve samples are generated with moderate population movements. According to Figure 6, we can see yellow and light green bars appear in this region. That means, when there are moderate population movements, the dengue risk belongs to the moderate and low dengue clusters.

#### (3) Cluster with low population movement

In Figure 6, the last set of twelve samples are generated with low population movements. According to Figure 6, we can see yellow, light green, and green bars appear in this region. That means, when there are low population movements, the dengue risk belongs to moderate, low, and very low dengue clusters.

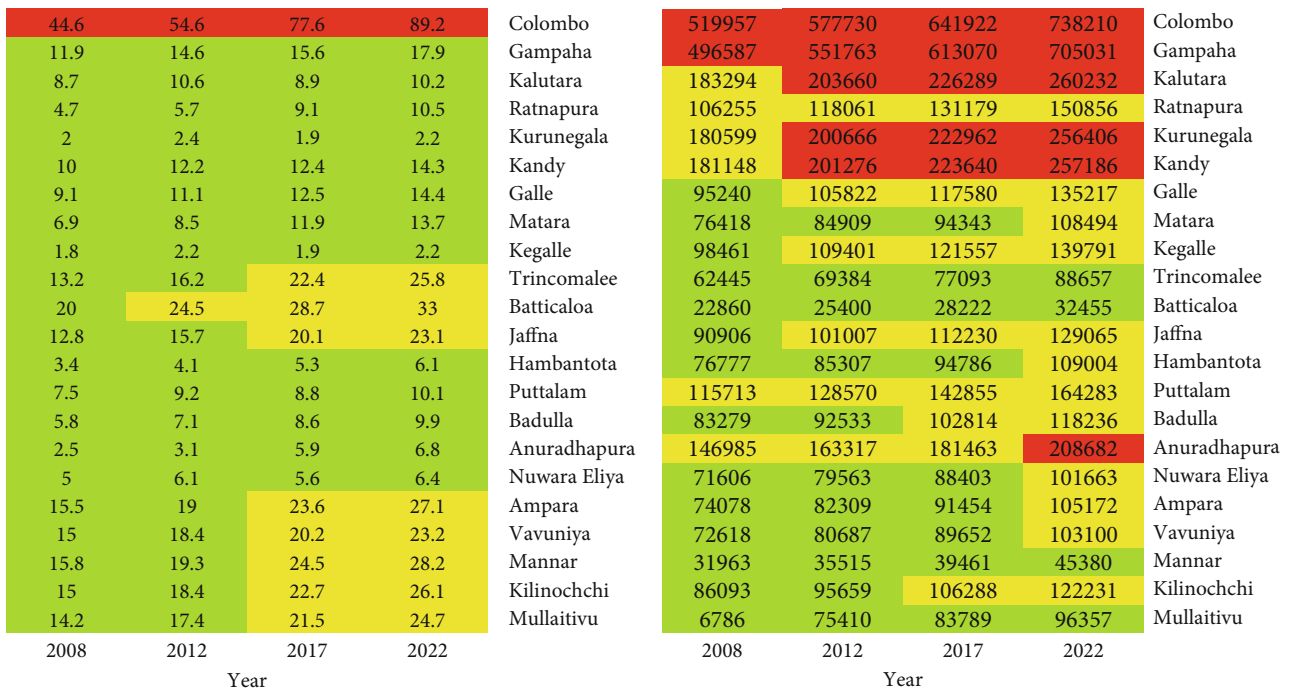


■ High  
 ■ Moderate  
 ■ Low

■ High  
 ■ Moderate  
 ■ Low

(a) Garbage collection

(b) Population density



■ High  
 ■ Moderate  
 ■ Low

■ High  
 ■ Moderate  
 ■ Low

(c) Urbanization

(d) Population movements

FIGURE 8: Heat maps of predictive factors.

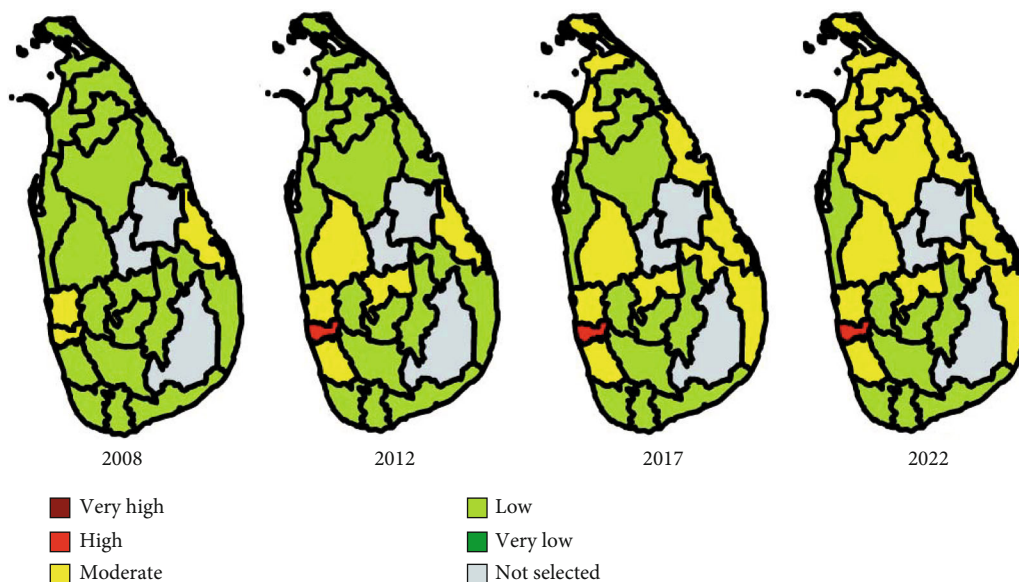


FIGURE 9: Dengue risk maps for the years 2008, 2012, 2017, and 2022.

**4.2. Model Validation.** Twenty-two districts in Sri Lanka are selected for the model validation. The selected districts are Ampara, Anuradhapura, Badulla, Batticaloa, Colombo, Galle, Gampaha, Jaffna, Kalutara, Kandy, Kegalle, Kilinochchi, Kurunegala, Hambantota, Mannar, Matara, Mullaitivu, Nuwara Eliya, Puttalam, Ratnapura, Trincomalee, and Vavuniya. These districts are shown in Figure 7. Three districts, namely, Matale, Moneragala, and Polonnaruwa, are neglected from the study due to the unavailability of data. The risk factor details of these districts are collected from the Department of Census and Statistics, Epidemiology Unit of Health Ministry, Department of Meteorology, and Central Bank of Sri Lanka websites.

Using the proposed model, GM value of each district is generated. The obtained results and risk categories of districts are given in Table 8. Then, the generated risk categories are compared with the percentage relative value of dengue cases considering the year 2017 and the risk clusters announced by the International Federation of Red Cross and Red Crescent Societies [9]. According to the results, the model accuracy level is 72.73%.

The results revealed that the Colombo districts have the highest risk in terms of dengue in the year 2017, and it belongs to the high-risk cluster. Since Colombo is the main administrative district in Sri Lanka, main industries, schools, and universities are situated in this area. It also a highly dense area in Sri Lanka. Due to these reasons, there is a high dengue transmission in the Colombo district. Gampaha, Kalutara, Kurunegala, Kandy, Batticaloa, Jaffna, Ampara, Mannar, Kilinochchi, and Mullaitivu belong to moderate-risk cluster in the context of dengue transmission. The districts in the low-risk cluster are Anuradhapura, Badulla, Galle, Hambantota, Kegalle, Matara, Nuwara Eliya, Puttalam, Ratnapura, Trincomalee, and Vavuniya.

**4.3. Capture the Future Risk Clusters of Dengue in Sri Lanka.** It is important to predict the future dengue risk cluster in dis-

tricts. In order to recognize the increasing trend of dengue clusters, 2008, 2012, and 2017 epidemic years are considered. The dengue risk prediction is carried out considering the year 2022. In order to simulate the results, we assumed that there is no change in the climate factors, the garbage increasing rate is 10%, the urbanization rate is 10%, and the population movements increase in selected areas by 10%. The annual population growth rates of the selected districts are obtained from the Department of Census and Statistics [27]. Figure 8 shows the heat maps generated for dengue risk factors. Figure 9 shows the heat map generated to visualize the past and future risk clusters of different Sri Lankan districts.

## 5. Discussion

This study presents the multifactorial model to determine the dengue risk of cities in an uncertain environment. Eight different social and climate factors are considered for the model development process. The developed mathematical model was applied to Sri Lankan districts. According to the results of the present study, we identified the population movements in the area as a significant factor to increase the dengue risk. This result agrees with the findings of the literature [28, 29]. The model predicted the dengue risk in selected districts with 72.73% accuracy level. Out of 22 districts, 16 district risk clusters have been predicted accurately by the developed model. Correctly predicted districts are Anuradhapura, Badulla, Batticaloa, Colombo, Gampaha, Hambantota, Jaffna, Kalutara, Kandy, Kilinochchi, Kurunegala, Mannar, Mullaitivu, Nuwara Eliya, Puttalam, and Vavuniya.

Variables such as garbage collection, urbanization percentage, and population movements considered in this research are novel factors for dengue risk prediction models because the literature contains the results considering only the climate factors and population density. Most of the models in the literature were developed considering the statistical concepts like logistic regression [30, 31].

However, when we are dealing with more factors, there is an uncertainty of the model. In order to handle this uncertainty, a fuzzy theory-based mathematical model is developed. Also, in the proposed model, dengue risk clusters are defined by intervals. Hence, small changes in factors do not show a drastic change in the results. Further, this study considers the effect of multifactors and their interaction to predict the dengue risk. This also helps to increase the predictive effect of the present study compared to the existing studies.

## 6. Conclusions

In this research, a multidimensional risk-based model was introduced to cluster the dengue hot spots in Sri Lanka. The model incorporates the social and climate factors including the eight different variables which can be easily assessed and analyzed. Since these variables are highly uncertain, FAHP was used to prioritize the variables. The interactions of the variables were calculated using GM. Considering the concept of Haddon matrices, five dengue risk clusters were identified. High population movement to target district plays a dominant role to transmit the disease.

Due to the limited resources, we selected only a few risk factors of dengue for model construction. Therefore, it is necessary to include social variables such as school density, factory density, household density, conditions of houses, and infrastructure facilities in the area to the model, considering more number of factors will help to recognize the dengue transmission risk in districts more accurately. Here, we measure the interaction between two variables. But there may be multidimensional interaction between the dengue risk factors. It would be necessary to incorporate these multifactor interactions to the model in future studies. In this study, the obtained weights do not depend on time. However, in the real scenario, the selected factors such as rainfall, temperature, and humidity heavily depend on time. Hence, in the future studies, we can consider the time-dependent FAHP model to prioritize the variable. Note that the developed risk clustering algorithm can be used to identify future dengue hotspots in the country. This will facilitate the government and relevant authorities to develop control strategies. It also provides information to people to make correct decisions when they travel places with a high risk of dengue.

## Data Availability

The datasets used during the current study are obtained from the Department of Census and Statistics, Epidemiology Unit of Health Ministry, Department of Meteorology, and Central Bank of Sri Lanka websites.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

## Supplementary Materials

Supplementary material S1. Fuzzy pairwise comparison matrices. Supplementary material S2. Multiplicative interaction matrices. Supplementary material S3. MATLAB codes to find sensitivity of the weights. Supplementary material S4. R code to generate heat maps. (*Supplementary Materials*)

## References

- [1] WHO, *Ten threats to global health in 2019*, 2019, May 2019, <https://www.who.int/emergencies/ten-threats-to-global-health-in-2019>.
- [2] *Dengue Virus Net, Dengue virus*, 2019, May 2019, <http://www.denguevirusnet.com>.
- [3] *Centers for Disease Control and Prevention, Dengue and dengue hemorrhagic fever*, 2009, May 2019, <https://www.breakdengue.org/Dengue/resources>.
- [4] *Break Dengue, Uncovering the hidden economic cost of dengue*, 2017, May 2019, <https://www.breakdengue.org/economic-cost-of-dengue/>.
- [5] W. B. Messer, U. T. Vitarana, K. Sivananthan et al., "Epidemiology of dengue in Sri Lanka before and after the emergence of epidemic dengue hemorrhagic fever," *The American Journal of Tropical Medicine and Hygiene*, vol. 66, no. 6, pp. 765–773, 2002.
- [6] *Epidemiology Unit, Dengue update*, 2019, May 2019, <http://www.epid.gov.lk/>.
- [7] The Ministry of Health, *Nutrition and Indigenous Medicine, Sri Lanka, Annual performance report*, 2017, May 2019, <http://www.health.gov.lk/>.
- [8] WHO, *Dengue fever Sri Lanka*, 2017, May 2019, <https://www.who.int/csr/don/19-july-2017-dengue-sri-lanka/en/>.
- [9] International Federation of Red Cross and Red Crescent Societies, *Dengue drefinal report Sri Lanka: Dengue*, Sri Lanka Red Cross Society, Sri Lanka, 2018.
- [10] R. Lowe, C. Barcellos, C. A. S. Coelho et al., "Dengue outlook for the World Cup in Brazil: an early warning model framework driven by real-time seasonal climate forecasts," *The Lancet Infectious Diseases*, vol. 14, no. 7, pp. 619–626, 2014.
- [11] C. R. Vicente, K. H. Herbing, C. Cerutti Junior, C. Malta Romano, A. de Souza Areias Cabidelle, and G. Fröschl, "Determination of clusters and factors associated with dengue dispersion during the first epidemic related to Dengue virus serotype 4 in Vitória, Brazil," *PLoS One*, vol. 12, no. 4, pp. e0175432–e0175626, 2017.
- [12] N. Sumanasinghe, A. Mikler, C. Tiwari, and J. Muthukudage, "Geo-statistical dengue risk model using GIS techniques to identify the risk prone areas by linking rainfall and population density factors in Sri Lanka," *Ceylon Journal of Science*, vol. 45, no. 3, pp. 39–46, 2016.
- [13] W. Sun, L. Xue, and X. Xie, "Spatial-temporal distribution of dengue and climate characteristics for two clusters in Sri Lanka from 2012 to 2016," *Scientific Reports*, vol. 7, no. 1, article 12884, 2017.
- [14] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Journal of Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.
- [15] C. G. E. Boender, J. G. de Grann, and F. A. Lootsma, "Multi-criteria decision analysis with fuzzy pairwise comparisons," *Fuzzy Sets and Systems*, vol. 29, no. 2, pp. 133–143, 1989.

- [16] N. A. Cinar, "Decision support model for bank branch location selection," *International Journal of Mechanical, Industrial Science and Engineering*, vol. 3, pp. 126–131, 2009.
- [17] O. Meixner, "Fuzzy AHP group decision analysis and its applications for the evaluation of energy sources," in *10th International Symposium on the Analytic Hierarchy Process/Network Process*, Pittsburgh, PA, USA, July 2009.
- [18] Y. C. Tang and T. W. Lin, "Application of the fuzzy analytic hierarchy process to the lead-free equipment selection decision," *International Journal of Business and Systems Research*, vol. 5, no. 1, pp. 35–56, 2011.
- [19] I. T. S. Piyatilake, S. K. Boralugoda, and S. S. N. Perera, "Developing a decision support index to quantify air quality," *Journal of Basic Applied Research International*, vol. 13, pp. 120–136, 2016.
- [20] I. T. S. Piyatilake and S. S. N. Perera, "Mathematical model to quantify air pollution in cities," in *Recent Advances in Applications of Computational and Fuzzy Mathematics*, S. Chakraverty and S. S. N. Perera, Eds., pp. 147–178, Springer, Singapore, 2018.
- [21] I. T. S. Piyatilake and S. S. N. Perera, "Developing a multiparametric risk index for dengue transmission," in *Mathematical Modelling and Analysis of Infectious Diseases*, K. Hattaf and H. Dutta, Eds., pp. 213–244, Springer International Publishing, Switzerland, 2020.
- [22] A. F. Williams, "The Haddon matrix: its contribution to injury prevention and control," in *Third National Conference on Injury Prevention and Control*, R. McClure, Ed., pp. 15–16, Queensland University of Technology, Brisbane, Queensland, 1999.
- [23] E. A. E. Habib, "Geometric mean for negative and zero values," *Journal of Research and Reviews in Applied Sciences*, vol. 11, pp. 419–432, 2012.
- [24] M. Ranjbar, A. Shoghli, G. Kolifarhood, S. M. Tabatabaei, M. Amlashi, and M. Mohammadi, "Predicting factors for malaria re-introduction: an applied model in an elimination setting to prevent malaria outbreaks," *Malaria Journal*, vol. 15, no. 1, p. 138, 2016.
- [25] W. P. T. M. Wickramaarachchi and S. S. N. Perera, "Developing a two dimensional climate risk model for dengue disease transmission in Urban Colombo," *Journal of Basic and Applied Research International*, vol. 20, pp. 168–177, 2017.
- [26] W. P. T. M. Wickramaarachchi and S. S. N. Perera, "A mathematical model with control to analyse the dynamics of dengue disease transmission in urban Colombo," *Journal of National Science Foundation Sri Lanka*, vol. 46, no. 1, pp. 41–49, 2018.
- [27] *Department of Census and Statistics, Population and housing*, 2019, June 2019, <http://www.statistics.gov.lk/>.
- [28] A. Wesolowski, T. Qureshi, M. F. Boni et al., "Impact of human mobility on the emergence of dengue epidemics in Pakistan," *National Academy of Sciences of the United States of America*, vol. 112, no. 38, pp. 11887–11892, 2015.
- [29] Public Library of Science, *Population movement can be critical factor in dengue's spread*, 2009, July 2020, <http://www.sciencedaily.com/releases/2009/11/091110065920>.
- [30] L. Chandrakantha, "Risk prediction model for dengue transmission based on climate data: logistic regression approach," *Stats*, vol. 2, no. 2, pp. 272–283, 2019.
- [31] D. T. T. Toan, W. Hu, P. Q. Thai, L. N. Hoat, P. Wright, and P. Martens, "Hot spot detection and spatio-temporal dispersion of dengue fever in Hanoi, Vietnam," *Global Health Action*, vol. 6, pp. 1–9, 2013.