

## Research Article

# Predictive Models for the Medical Diagnosis of Dengue: A Case Study in Paraguay

Jorge D. Mello-Román <sup>1</sup>, Julio C. Mello-Román,<sup>1</sup> Santiago Gómez-Guerrero,<sup>2</sup>  
and Miguel García-Torres <sup>3</sup>

<sup>1</sup>Universidad Nacional de Concepción, Concepción 8700, Paraguay

<sup>2</sup>Universidad Nacional de Asunción, San Lorenzo 2111, Paraguay

<sup>3</sup>Universidad Pablo de Olavide, Sevilla 41013, Spain

Correspondence should be addressed to Jorge D. Mello-Román; mellojorgepy@gmail.com

Received 30 December 2018; Accepted 26 May 2019; Published 29 July 2019

Academic Editor: Maria E. Fantacci

Copyright © 2019 Jorge D. Mello-Román et al. 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.

Early diagnosis of dengue continues to be a concern for public health in countries with a high incidence of this disease. In this work, we compared two machine learning techniques: artificial neural networks (ANN) and support vector machines (SVM) as assistance tools for medical diagnosis. The performance of classification models was evaluated in a real dataset of patients with a previous diagnosis of dengue extracted from the public health system of Paraguay during the period 2012–2016. The ANN multilayer perceptron achieved better results with an average of 96% accuracy, 96% sensitivity, and 97% specificity, with low variation in thirty different partitions of the dataset. In comparison, SVM polynomial obtained results above 90% for accuracy, sensitivity, and specificity.

## 1. Introduction

According to the World Health Organization, the incidence and prevalence of dengue have been increasing in endemic areas of the tropical and subtropical regions. Based on mathematical estimates of the model, approximately 50 million infections occur each year [1, 2], and historically, the Southern Cone (Argentina, Brazil, Chile, Paraguay, and Uruguay) is the subregion that contributes between 50–60% of the dengue cases in the Americans [3, 4].

Early diagnosis of dengue disease is a recurrent need in the world public health system, and machine learning techniques can help doctors diagnose and predict diseases at an early stage, helping not only in improving the accuracy performance of classification but also in saving diagnostics time, cost, and the pain accompanying pathology tests [5, 6]. Raval et al. [7] point out that the machine learning techniques have been used in medical diagnosis allowing the disease to be analysed based on clinical and laboratory

symptoms, providing an accurate result. They also point out that artificial neural network (ANN) is one of the main techniques used to solve medical diagnostic problems, and also the support vector machines (SVM) provide accurate results when evaluating a single disease.

ANN has been applied in several studies of dengue forecast models. In [8], Ibrahim et al. use clinical and epidemiological data from Malaysia. In another work, Cetiner et al. [9] apply ANN to weather data from the Singaporean National Environment Agency (SNEA). A weather dataset was also used by Rachata et al. [10], and feature selection algorithms were applied to predict the incidence of dengue in Thailand.

SVM is another technique quite popular to address this problem. Wu et al. [11] apply SVM to Singapore weather dataset to predict dengue fever. Gomes et al. [12] use gene expression data and apply a radial basis function (RBF) kernel on a small sample from Brazil. Support vector regression (SVR) is also applied by Guo et al. [13]. In this case,

they compare several machine learning approaches to data collected from the Guangdong region (China). In a more recent work, Carvajal et al. [14] study the incidence of dengue in Philippines using meteorological factors. They compare several machine learning techniques such as general additive modelling, seasonal autoregressive integrated moving average with exogenous variables, random forest, and gradient boosting.

In this work, the objective is to compare the performance of ANN and SVM classification models to predict the presence of dengue disease, in terms of accuracy, specificity, and sensitivity [15, 16]. The models were trained in a real dataset of patients admitted because of fever in health centres of northern Paraguay, prediagnosed with dengue, and subsequently confirmed or discarded by means of laboratory criteria during the period 2012–2016.

## 2. Classification Models

**2.1. Artificial Neural Network (ANN).** Artificial neural networks are, as their name indicates, computational networks which attempt to simulate, in a gross manner, the decision process in networks of nerve cell (neurons) of the biological (human or animal) central nervous system [17, 18].

The ANN structure consists of three main layers as shown in Figure 1. First, an input layer, which connects the input signal ( $X_j$ ) to the neuron via a set of weights ( $W_{kj}$ ). Next, a hidden layer, which summarizes the bias values ( $b_k$ ) and the input signals, is weighted by the respective weight values of the neuron. Finally, an output layer is used for limiting the amplitude of the output of the neuron using the activation transfer function. In addition, a bias is added to the neuron to increase or decrease the net output of the neuron [19].

The mathematical structure of a neuron  $k$  in ANN is represented as follows [20, 21]:

$$\begin{aligned} U_k &= \sum_{j=1}^n (W_{kj} X_j), \\ Y_k &= f(U_k + b_k), \end{aligned} \quad (1)$$

where  $X_j$  are the input signals,  $W_{kj}$  are the weights for neuron  $k$ ,  $b_k$  is the bias value,  $U_k$  is the linear combiner,  $f(\cdot)$  is the activation transfer function, and  $Y_k$  is the output signal of the neuron. Two neural networks widely used as supervised training methods are the multilayer perceptron (MLP) and radial basis function (RBF) networks [22, 23].

MLP is a network formed by an input layer, at least one hidden layer and an output layer, trained with the use of the backpropagation learning algorithm, which consists of using the error generated by the network and propagating it backwards. RBF is a neural network that uses radial basis functions as activation functions. There are different types of radial basis functions, but the most widely used type is the Gaussian function. The architecture used by the RBF is very similar to that of the multilayer perceptron,

with the characteristic that the RBF always uses three layers: an entry layer, a hidden layer, and an exit layer, while MLPs may have more [24, 25]. A description in some more detail of the different types of ANN systems can be found in [26].

**2.2. Support Vector Machine (SVM).** SVM is a classification technique that has given satisfactory results in many practical applications. It also works very well with high dimensional data. SVM looks for a separator hyperplane with the highest margin [27, 28]. An example of nonlinear SVM is shown in Figure 2.

Consider the nonlinear transformation  $\Phi : R^m \rightarrow \mathcal{H}$  that allows to project the input vectors from the original coordinate space  $x$  to a transformed space  $\Phi(x) \in \mathcal{H}$ , provided with scalar product. The scalar product between two given vectors in the transformed space enunciated in terms of a function of similarity in the original space  $\Phi(u) \cdot \Phi(v) = K(u, v)$  is known as the kernel function [30, 31].

Some well-known kernel functions are as follows:

- (i) Linear  $K(u, v) = u \cdot v$
- (ii) Polynomial  $K(u, v) = (u \cdot v + b)^p$
- (iii) Gaussian  $K(u, v) = \exp(-\sigma \|u - v\|^2)$

Consider the problem of binary classification consisting of  $N$  examples of training. Each example is indicated by a tuple  $(X_i, y_i)$ , where  $X$  corresponds to the set of attributes for example  $i$  and the class denomination is indicated by  $y_i \in \{1, -1\}$ . The learning task with SVM can be formalized as the following constrained optimization problem [32, 33]:

$$\text{maximise } L = \sum_i^N \lambda_i + \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j K(X_i, X_j), \quad (2)$$

such that  $\sum_i^N \lambda_i y_i = 0$  and  $\lambda_i \geq 0$  for all  $i$ .

A test case  $Z$  can be classified using the equation  $f(z) = \text{sign}(\sum_i^N \lambda_i y_i K(X_i, Z) + b)$ , where  $\lambda_i$  is a Lagrange multiplier,  $b$  is a parameter, and  $K$  is a kernel function.

**2.3. Evaluation of Classification Models.** In the medical sciences, there are many measures used to calculate indices related to diagnostic accuracy: sensitivity, specificity, and others [34]. In this paper, to evaluate the performance of classification models ANN and SVM, the accuracy, sensitivity, and specificity of the confusion matrices obtained for the test datasets will be calculated [15] and the variation thereof in several test sets obtained from randomised partitions of the dataset.

The accuracy of a test is its ability to differentiate the patient and healthy cases correctly, the sensitivity is its ability to determine the patient cases correctly, and the specificity is its ability to determine the healthy cases correctly. Mathematically, this can be stated as [35]

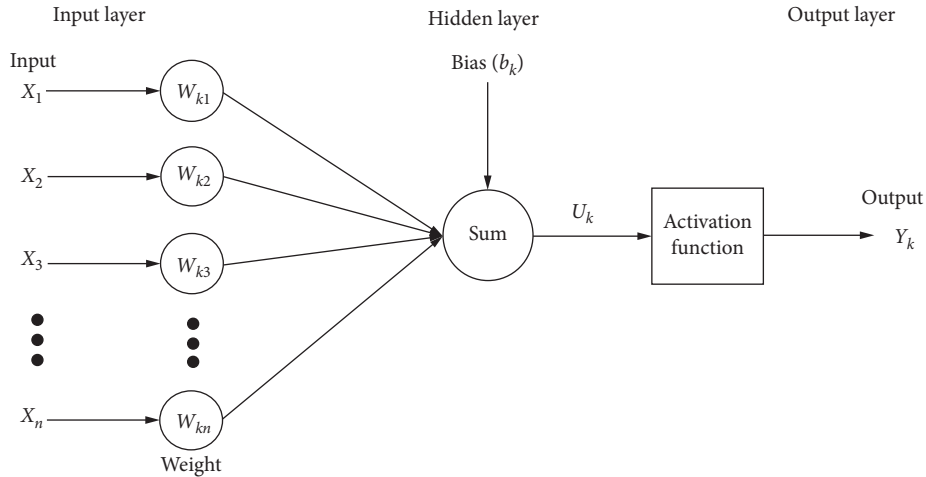


FIGURE 1: Example of a feedforward architecture with one hidden layer.

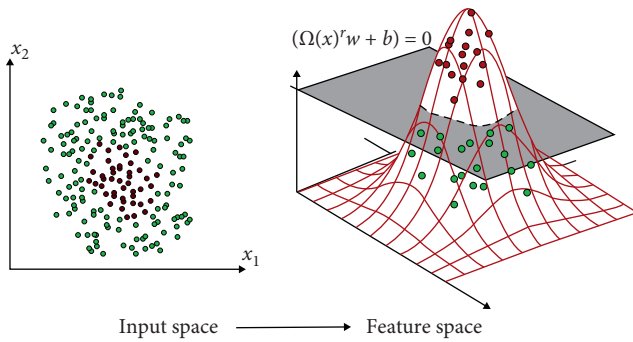


FIGURE 2: Example of nonlinear SVM [29].

$$\begin{aligned} \text{accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \\ \text{sensitivity} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}}, \end{aligned} \quad (3)$$

where TP is the number of cases correctly identified as patient, FP is the number of cases incorrectly identified as patient, TN is the number of cases correctly identified as healthy, and FN is the number of cases incorrectly identified as healthy.

### 3. Dataset

The original dataset is composed of cases registered by the public health system of Paraguay, such as patients admitted for fever and initially diagnosed with dengue, in various health centres of the Department of Concepción, between the years 2012 to 2016. Concepción is located to the northwest of the eastern region of the country, bordering Brazil to the north, as shown in Figure 3.

There were 4332 cases registered, of which 53% corresponded to women and 47% to men. The age of the patients is normally distributed, with the age group between 20 and

39 being the most frequent, 83% of the patients stated that they reside in urban areas, and 55% said that similar cases were occurring simultaneously in their surroundings. Of all the cases, 82% were treated as outpatients.

The cases occurred mainly between the months of December and May over the five years. Figure 4 shows the evolution of the number of patients diagnosed with dengue by week of onset of fever. The datasets related to patients registered in health centres generally have many difficulties throughout the world, characterised by being incomplete, incorrect, scarce, and inaccurate [36, 37]. This dataset is not the exception and leads to the need for an exhaustive pre-processing in order to determine which are the cases and variables that provide useful information for processing.

**3.1. Dataset Preprocessing.** The original dataset was constituted by dengue cases evaluated by one of the laboratory criteria for diagnosis (dengue IgM serological tests or virological tests such as viral isolation or RT-PCR) or by the criterion of the epidemiological link. The epidemiological link is to confirm the probable cases of dengue from laboratory-confirmed cases using the association of person, time, and space [38–40]. Only cases confirmed or discarded with laboratory criteria were included in the study.

In order to ensure minimal loss of information, we proceeded to treat missing values in three stages. First, variables with more than 20% of missing values were excluded. According to [41, 42], it is not recommended to impute data in situations in which the omission in one or more variables reaches percentages higher than an established threshold, because it puts statistical reliability at risk. As a second stage, cases with a nonresponse rate higher than 80% were eliminated [43, 44].

For the selection of the appropriate imputation of missing values, there are no specific rules. It depends on the type of the dataset, file size, nonresponse type, pattern of loss of response, of the research objectives, specific characteristics of the population, general characteristics of the



FIGURE 3: Department of Concepción, Paraguay.

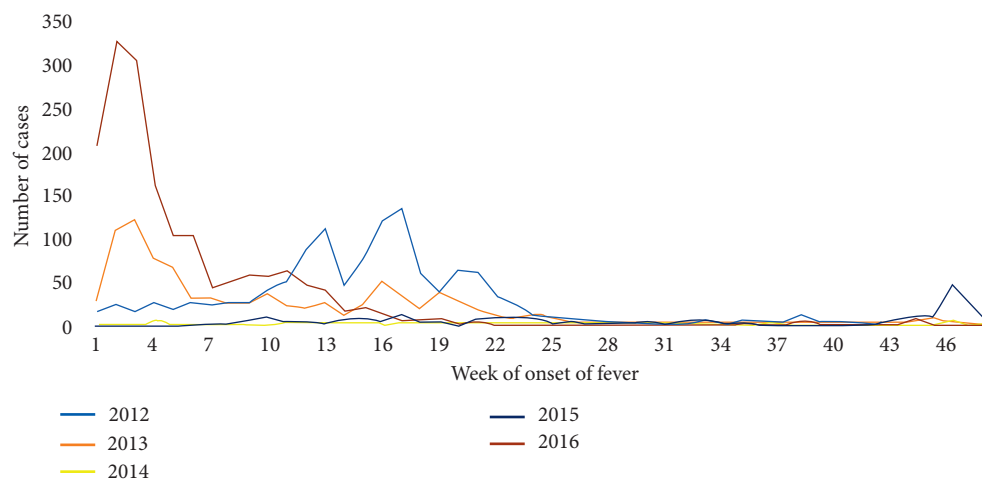


FIGURE 4: Number of patients diagnosed with dengue by week of onset of fever.

organization of the study, or available software [45]. As the third stage and given the characteristics of the data input process in the dataset and its epidemiological nature, it was considered most appropriate to impute the mean of the valid adjacent data to the missing values.

After treatment of missing values, the final dataset is comprised of 668 cases and the variables included are described in Table 1.

#### 4. Implementation and Results

The tests were carried out with the support of the IBM SPSS Modeler software [46]. Firstly, the classification models have been configured in such a way that random partitions of the dataset are carried out by 90% for the training dataset and 10% for the test dataset. For the ANN classifier, the objective

has been to increase the accuracy of the model, and the automatic calculation of the number of hidden layers has been arranged for two models of the neuronal network: the multilayer perceptron (MLP) and radial basis function (RBF).

For the SVM classifier, the performance of three kernel functions has been evaluated: linear, Gaussian, and polynomial. Previously, several tests have been carried out to determine the best parameters of the kernel functions and the optimal penalty parameters, based on the accuracy of the classification in the test datasets. The results are described in Table 2.

To evaluate the performance of the ANN and SVM classifiers as predictive models, thirty (30) random partitions of the dataset were performed in the percentages indicated for training and test datasets. The performance of the ANN

TABLE 1: Variables included.

Variables	Description	Measure	Values
Week	Week declared by patient when fever started	Scale	$[1, 48] \in \mathbb{N}$
Year	Registration year	Scale	$[1, 5] \in \mathbb{N}$
Age	Age of the patient at the time of registration	Scale	$[0, 99] \in \mathbb{N}$
Sex	Patient's sex	Nominal	1 = male, 0 = female
District	Geographic area of the Department of Concepción where the patient resides	Nominal	$[1, 9] \in \mathbb{N}$
Hospitalised	Indicates whether the patient was hospitalised or treated as an outpatient	Nominal	1 = yes, 0 = no
Headache	Presence of the symptom	Nominal	1 = yes, 0 = no
Myalgia	Presence of the symptom	Nominal	1 = yes, 0 = no
Arthralgia	Presence of the symptom	Nominal	1 = yes, 0 = no
Retro-ocular pain	Presence of the symptom	Nominal	1 = yes, 0 = no
Pruritus	Presence of the symptom	Nominal	1 = yes, 0 = no
Sickness	Presence of the symptom	Nominal	1 = yes, 0 = no
Cough and dyspnea	Presence of the symptom	Nominal	1 = yes, 0 = no
Oligoanuria	Presence of the symptom	Nominal	1 = yes, 0 = no
Epistaxis	Presence of the symptom	Nominal	1 = yes, 0 = no
Gingivorrhagia	Presence of the symptom	Nominal	1 = yes, 0 = no
Hemoptysis	Presence of the symptom	Nominal	1 = yes, 0 = no
Melena	Presence of the symptom	Nominal	1 = yes, 0 = no
Black vomiting	Presence of the symptom	Nominal	1 = yes, 0 = no
Exanthema	Presence of the symptom	Nominal	1 = yes, 0 = no
Shock	Presence of the symptom	Nominal	1 = yes, 0 = no
Conjunctival injection	Presence of the symptom	Nominal	1 = yes, 0 = no
Bipalpebral edema	Presence of the symptom	Nominal	1 = yes, 0 = no
Tachycardia	Presence of the symptom	Nominal	1 = yes, 0 = no
Hepatomegaly	Presence of the symptom	Nominal	1 = yes, 0 = no
Splenomegaly	Presence of the symptom	Nominal	1 = yes, 0 = no
Sensory alteration	Presence of the symptom	Nominal	1 = yes, 0 = no
Stiff neck	Presence of the symptom	Nominal	1 = yes, 0 = no
Petechia	Presence of the symptom	Nominal	1 = yes, 0 = no
Purpura	Presence of the symptom	Nominal	1 = yes, 0 = no
Jaundice	Presence of the symptom	Nominal	1 = yes, 0 = no
Other symptoms	Presence of other symptoms	Nominal	1 = yes, 0 = no
Travel	Patient's statement if he/she travelled the last 15 days	Nominal	1 = yes, 0 = no
Camping	Patient's statement if he/she was camping the last 15 days	Nominal	1 = yes, 0 = no
Similar health condition	Patient's statement if he/she had a similar health condition before	Nominal	1 = yes, 0 = no
Live alone	Patient's statement if he/she lives alone	Nominal	1 = yes, 0 = no
Poverty	Doctor's impression of whether the patient is in poverty	Nominal	1 = yes, 0 = no
Final classification	Laboratory diagnosis for confirming dengue virus infection	Nominal	$C = \text{confirmed}, D = \text{discarded}$

TABLE 2: Selection of best parameters of the SVM classifier.

Kernel	Penalty	Kernel parameters
Linear	$C = 10$	—
Gaussian	$C = 11$	$\sigma = 0.01$
Polynomial	$C = 10$	$p = 3; b = 0$

and SVM classifiers was evaluated exclusively on the test datasets; Table 3 shows the sample mean and the coefficient of variation (the ratio between the sample standard deviation and the sample mean) of the accuracy, sensitivity, and specificity.

It is important to highlight that a 96% sensitivity of the ANN-MLP classifier represents a 4% probability of

TABLE 3: The performance of the ANN and SVM classifiers on 30 test datasets.

Classifiers	Accuracy (%)		Sensitivity (%)		Specificity (%)	
	Mean	Variation	Mean	Variation	Mean	Variation
ANN-MLP	96	2	96	4	97	3
ANN-RBF	55	11	58	10	52	24
SVM linear	64	8	56	18	71	10
SVM Gaussian	86	9	84	11	89	10
SVM polynomial	92	14	93	12	92	16

The ANN-MLP classifier achieved better results with an average of 96% accuracy, 96% sensitivity, and 97% specificity, with low variation in thirty different partitions of the dataset. Also, the SVM polynomial classifier obtained results above 90% for the three indicators with acceptable variations.

TABLE 4: ANN-MLP confusion matrix.

Laboratory diagnosis of dengue virus infection	Predicted	
	Confirmed	Discarded
Confirmed	36	0
Discarded	1	43

TABLE 5: SVM polynomial confusion matrix.

Laboratory diagnosis of dengue virus infection	Predicted	
	Confirmed	Discarded
Confirmed	45	0
Discarded	1	33

committing a false-negative type error, which in the medical case is equivalent to diagnosing sick individuals as healthy.

Thirty confusion matrices were obtained for each classifier evaluated, one matrix for each partition of the dataset. The confusion matrices of ANN-MPL and SVM polynomial classifiers, with higher accuracy, are described in Tables 4 and 5.

## 5. Conclusions

We compared the results of two classifiers, ANN and SVM, frequently used in medical sciences as machine learning techniques to assist in medical diagnosis. ANN-MLP and SVM polynomials have shown that they can perform as classifiers in the diagnosis of dengue disease with high averages of accuracy, sensitivity, and specificity, within the spatial and temporal context determined by the dataset.

An important methodological step of this paper is the preprocessing of the dataset considering the particularity of its origin in the public health system and the problems that this entails in the loading of the data. The dataset is validated for the construction of machine learning classifiers.

The results obtained by both classifiers are encouraging for its use in experimental stage and lead to propose more exhaustive investigations with controlled data capturing systems. Additionally, the predictive models generated can be integrated into computer systems to assist in the diagnosis of dengue disease in Paraguay.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This research was developed as part of the research project PINV15-706 “COMIDENCO—Construction of an Dengue

Incidence Model Applied to Communities of Paraguay,” funded within the framework of the PROCENCIA program of the National Council of Sciences and Technology of Paraguay (CONACYT).

## References

- [1] World Health Organization, “Strengthening implementation of the global strategy for dengue fever/dengue haemorrhagic fever prevention and control,” Report of the Informal Consultation, WHO HQ, Geneva, Switzerland, 1999.
- [2] M. Guzman and G. Kouri, “Dengue and dengue hemorrhagic fever in the Americas: lessons and challenges,” *Journal of Clinical Virology*, vol. 27, no. 1, pp. 1–13, 2003.
- [3] J. L. San Martín, J. O. Solórzano, M. G. Guzmán et al., “The epidemiology of dengue in the Americas over the last three decades: a worrisome reality,” *American Journal of Tropical Medicine and Hygiene*, vol. 82, no. 1, pp. 128–135, 2010.
- [4] D. S. Shepard, E. A. Undurraga, M. Betancourt-Cravioto et al., “Approaches to refining estimates of global burden and economics of dengue,” *PLoS Neglected Tropical Diseases*, vol. 8, no. 11, article e3306, 2014.
- [5] A. Jain, “Machine learning techniques for medical diagnosis: a review,” DU, Conference Center, New Delhi, India, 2015.
- [6] I. Kononenko, “Machine learning for medical diagnosis: history, state of the art and perspective,” *Artificial Intelligence in Medicine*, vol. 23, no. 1, pp. 89–109, 2001.
- [7] D. Raval, D. Bhatt, M. K. Kumhar, V. Parikh, and D. Vyas, “Medical diagnosis system using machine learning,” *International Journal of Computer Science & Communication*, vol. 7, no. 1, pp. 177–182, 2016.
- [8] F. Ibrahim, M. N. Taib, W. A. B. W. Abas, C. C. Guan, and S. Sulaiman, “A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using artificial neural network (ANN),” *Computer Methods and Programs in Biomedicine*, vol. 79, no. 3, pp. 273–281, 2005.
- [9] B. G. Cetiner, M. Sari, and H. M. Aburas, “Recognition of dengue disease patterns using artificial neural networks,” in *Proceedings of the 5th International Advanced Technologies Symposium (IATS’09)*, pp. 359–362, Karabuk, Turkey, May 2009.
- [10] N. Rachata, P. Charoenkwan, T. Yooyativong, K. Chamnongthai, C. Lursinsap, and K. Higuchi, “Automatic prediction system of dengue haemorrhagic-fever outbreak risk by using entropy and artificial neural network,” in *Proceedings of the International Symposium on Communications and Information Technologies*, IEEE, Lao, China, October 2008.
- [11] Y. Wu, G. Lee, X. Fu, and T. Hung, “Detect climatic factors contributing to dengue outbreak based on wavelet, support vector machines and genetic algorithm,” in *Proceedings of the World Congress on Engineering*, vol. 1, London, UK, July 2008.
- [12] A. L. V. Gomes, L. J. K. Wee, A. M. Khan et al., “Classification of dengue fever patients based on gene expression data using support vector machines,” *PLoS One*, vol. 5, no. 6, Article ID e11267, 2010.
- [13] P. Guo, T. Liu, Q. Zhang et al., “Developing a dengue forecast model using machine learning: a case study in China,” *PLoS Neglected Tropical Diseases*, vol. 11, no. 10, article e0005973, 2017.
- [14] T. M. Carvajal, K. M. Viacrusis, L. F. T. Hernandez, H. T. Ho, D. M. Amalin, and K. Watanabe, “Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in Metropolitan Manila, Philippines,” *BMC Infectious Diseases*, vol. 18, no. 1, p. 183, 2018.

- [15] S. Fathima and N. Hundewale, "Comparison of classification techniques-SVM and naives bayes to predict the arboviral disease-Dengue," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)*, IEEE, Atlanta, GA, USA, December 2011.
- [16] K. Kesorn, P. Ongruk, J. Chomposri et al., "Morbidity rate prediction of dengue hemorrhagic fever (DHF) using the support vector machine and the *Aedes aegypti* infection rate in similar climates and geographical areas," *PLoS One*, vol. 10, no. 5, Article ID e0125049, 2015.
- [17] D. Graupe, *Principles of Artificial Neural Networks*, vol. 7, World Scientific, Singapore, 2013.
- [18] G. Li, X. Zhou, J. Liu et al., "Comparison of three data mining models for prediction of advanced schistosomiasis prognosis in the Hubei province," *PLoS Neglected Tropical Diseases*, vol. 12, p. 2, article 0006262, 2018.
- [19] T. Kerdphol, K. Fuji, Y. Mitani, M. Watanabe, and Y. Qudaih, "Optimization of a battery energy storage system using particle swarm optimization for stand-alone microgrids," *International Journal of Electrical Power & Energy Systems*, vol. 81, pp. 32–39, 2016.
- [20] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1994.
- [21] E. Singhal and S. Tiwari, "Skin cancer detection using artificial neural network," *International Journal of Advanced Research in Computer Science*, vol. 6, p. 1, 2015.
- [22] M. Kayri, "An intelligent approach to educational data: performance comparison of the multilayer perceptron and the radial basis function artificial neural networks," *Educational Sciences: Theory and Practice*, vol. 15, no. 5, pp. 1247–1255, 2015.
- [23] A. K. Dwivedi and U. Chouhan, "Multilayer perceptron and evolutionary radial basis function neural network models for discrimination of HIV-1 genomes," *Current Science*, vol. 115, no. 11, p. 2063, 2018.
- [24] M. A. Ghorbani, H. A. Zadeh, M. Isazadeh, and O. Terzi, "A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction," *Environmental Earth Sciences*, vol. 75, no. 6, p. 476, 2016.
- [25] D. Mercado Polo, L. Pedraza Caballero, and E. Martínez Gómez, "Comparación de redes neuronales aplicadas a la predicción de series de tiempo," *Prospectiva*, vol. 13, no. 2, pp. 88–95, 2015.
- [26] P. Lorrentz, *Artificial Neural Systems: Principle and Practice*, Bentham Science Publishers, Sharjah, UAE, 2015.
- [27] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Pearson, London, UK, 1st edition, 2005.
- [28] M. Chu, X. Liu, R. Gong, and J. Zhao, "Support vector machine with quantile hyper-spheres for pattern classification," *PLoS One*, vol. 14, no. 2, Article ID e0212361, 2019.
- [29] P. Davidson and A. M. Waas, "Probabilistic defect analysis of fiber reinforced composites using kriging and support vector machine based surrogates," *Composite Structures*, vol. 195, pp. 186–198, 2018.
- [30] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, Cambridge, UK, 2004.
- [31] L. C. Padierna, M. Carpio, A. Rojas-Domínguez, H. Puga, and H. Fraire, "A novel formulation of orthogonal polynomial kernel functions for SVM classifiers: the Gegenbauer family," *Pattern Recognition*, vol. 84, pp. 211–225, 2018.
- [32] J. Wei, L. Jiake, W. Xuan, and S. Rongrong, "A novel hybrid approach of KPCA and SVM for crop quality classification," in *Proceedings of the International Conference on Computer Science and Software Engineering*, vol. 4, IEEE, Hubei, China, December 2008.
- [33] F. Ye, X. Y. Lou, and L. F. Sun, "An improved chaotic fruit fly optimization based on a mutation strategy for simultaneous feature selection and parameter optimization for SVM and its applications," *PLoS One*, vol. 12, no. 4, Article ID e0173516, 2017.
- [34] P. Mishra, C. Pandey, U. Singh, S. Yadav, and V. Sharma, "Evaluation and application of diagnostic accuracy in clinical decision-making," *International Journal of Medical Science and Public Health*, vol. 5, no. 10, pp. 2190–2193, 2016.
- [35] A. Baratloo and S. Safari, "Evidence based medicine; simple definition and calculation of accuracy, sensitivity, and specificity," *Iranian Journal of Emergency Medicine*, vol. 2, pp. 105–107, 2015.
- [36] G. D. Magoulas and P. Andriana, "Machine learning in medical applications," in *Advanced Course on Artificial Intelligence*, Springer, Berlin, Heidelberg, 1999.
- [37] D. B. Neill, "Using artificial intelligence to improve hospital inpatient care," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 92–95, 2013.
- [38] World Health Organization, *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control*, World Health Organization, Geneva, Switzerland, 2009.
- [39] J. S. Torres-Roman, C. Díaz-Vélez, J. Bazalar-Palacios, and L. M. Helguero-Santin, "Hospital management in patients with dengue: what challenges do we face in Latin America?," *Le Infezioni in Medicina: Rivista Periodica di Eziologia, Epidemiologia, Diagnostica, Clinica e Terapia Delle Patologie Infettive*, vol. 24, no. 4, pp. 359–360, 2016.
- [40] M. Khurshed, U. R. Khan, K. Ejaz, J. Fayyaz, I. Qamar, and J. A. Razzak, "A comparison of WHO guidelines issued in 1997 and 2009 for dengue fever-single centre experience," *Journal of Pakistan Medical Association*, vol. 63, no. 6, pp. 670–674, 2013.
- [41] F. Medina and M. Galván, "Estudios estadísticos y prospectivos," in *Imputación de Datos: Teoría y Práctica. División Estadística y Proyecciones Económicas Naciones Unidas*, CEPAL, Santiago, Chile, 2007.
- [42] D. P. Jannat-Khah, M. Unterbrink, M. McNairy et al., "Treating loss-to-follow-up as a missing data problem: a case study using a longitudinal cohort of HIV-infected patients in Haiti," *BMC Public Health*, vol. 18, no. 1, p. 1269, 2018.
- [43] A. P. Goicoechea, *Imputación Basada en Árboles de Clasificación*, EUSTAT, San Sebastian, Spain, 2002, <http://www.eustat.es/documentos/datos/ct%204>.
- [44] C. D. Nyström, J. D. Barnes, and M. S. Tremblay, "An exploratory analysis of missing data from the Royal Bank of Canada (RBC) Learn to Play–Canadian Assessment of Physical Literacy (CAPL) project," *BMC Public Health*, vol. 18, no. S2, p. 1046, 2018.
- [45] U. Castro, L. María, D. Maria, and M. Ávila, *Una Introducción a la Imputación de Valores Perdidos: Terra Nueva Etapa*, vol. 22, Universidad Central de Venezuela, Caracas, Venezuela, 2006.
- [46] J. Salcedo and K. McCormick, *IBM SPSS Modeler Essentials: Effective Techniques for Building Powerful Data Mining and Predictive Analytics Solutions*, Packt Publishing Ltd., Birmingham, UK, 2017.