



Contents lists available at ScienceDirect

## New Horizons in Translational Medicine

journal homepage: [www.elsevier.com/locate/nhtm](http://www.elsevier.com/locate/nhtm)

Research article

## PSO-ANN based diagnostic model for the early detection of dengue disease

Shalini Gambhir<sup>a</sup>, Sanjay Kumar Malik<sup>a</sup>, Yugal Kumar<sup>b,\*</sup><sup>a</sup> Department of Computer Science and Engineering, SRM University, Haryana, India<sup>b</sup> Department of Computer Science and Engineering, JUIT, Himachal Pradesh, India

## ARTICLE INFO

## Keywords:

Classification  
Neural network  
Decision tree  
Naive Bayes  
CART  
PSO

## ABSTRACT

Large numbers of machine learning approaches have been developed for analysis of medical data in recent years. These approaches have also proved their significance through accurate and earlier diagnosis of diseases. The objective of this work is to develop a diagnostic model for earlier diagnosis of dengue disease. Dengue fever is spread through the bite of the female mosquito (*Aedes aegypti*). The symptoms of this fever are similar to other fever such as that of Viral influenza, Chikungunya, Zika fever, and so on. However, in this fever, human life can be at risk due to severe depletion of blood platelets. Therefore, early diagnosis of dengue disease can help in protecting human lives by making a preventive move before it turns into an infectious disease. In this work, an effort is made to develop a PSO-ANN based diagnostic model for earlier diagnosis of dengue fever. In the proposed model, PSO technique is applied to optimize the weight and bias parameters of ANN method. Further, PSO optimized ANN approach is used to detect dengue patients. The effectiveness of the proposed model is evaluated based on accuracy, sensitivity, specificity, error rate and AUC parameters. The results of the proposed model have been compared with other existing approaches like ANN, DT, NB, and PSO. It is observed that the proposed diagnostic model is a proficient and powerful model for more accurate and earlier detection of dengue fever.

## 1. Introduction

Dengue is a mosquito-based viral disease that can quickly spread in favorable climatic condition. This communicates through female mosquito named 'Aedes aegypti'. The main reason behind the widespread prevalence of dengue disease over the tropics is due to variations in rainfall, temperature, and unplanned rapid urbanization. In recent years, dengue cases have grown up rapidly around the world, however, the actual numbers of dengue cases are either never reported or sometimes classified inaccurately. According to WHO report, every year, 390 million dengue infections are reported in the entire world, out of this 96 million are clinically reported with the severity of disease [1]. The other study, on the occurrence of dengue disease, indicates that dengue viruses can infect 3.9 billion people in 128 countries [2]. The number of cases registered for dengue is increased from 2.2 million (in 2010) to 3.2 million (in 2015). Dengue is one of most fatal and widespread viral infection in the world today. It is an increasingly prevalent tropical virus infection with significant morbidity and fatality rate [3]. Dengue infection has been recognized to be endemic in India for over two centuries as a benign and self-limited disease. In recent years, the

disease has shifted its course manifesting in the severe form of DHF and with increasing frequency of outbreaks [4]. Dengue infection in a previously non-immune host produces a principal response of antibodies characterized by a slow and low-titer antibody response. IgM antibody is the first immunoglobulin isotype to appear. In a suspected case of dengue, the presence of anti-dengue IgM antibody suggests recent infection. Anti-dengue IgM detection using enzyme-linked immunosorbent assay (ELISA) represents one of the most important advances and has become an invaluable instrument for routine dengue diagnosis [5].

In recent years, various decision support systems and diagnostic models have been developed for improving experiences and abilities of physicians to accurate detection and diagnosis of diseases. It is observed from the recent research trends that artificial neural networks have been widely used in the field of medical data mining and number of decision support systems have been developed with the help of ANN due to its ability of prediction, parallel operation and Adaptivity [6–13]. The multilayer neural networks (MLNNs) have been successfully used in replacing conventional pattern recognition methods for the disease diagnosis systems and it can be back-recognized as a

*Abbreviations:* ANN, Artificial Neural Network; ANFIS, Adaptive Neuro-Fuzzy Inference System; AUC, Area Under the Curve; BPSO, Binary Particle Swarm Optimization; CART, Classification and Regression Tree; DHF, Dengue Hemorrhagic Fever; DT, Decision Tree; GA, Genetic Algorithm; MLNN, Multilayer Neural Network; MLP, Multilayer Perceptron; NB, Naive Bayes; PSO, Particle Swarm Optimization; SVM, Support Vector Machine; WHO, World Health Organization

\* Corresponding author at: Department of Computer Science and Engineering, JUIT, Himachal Pradesh, India.

E-mail addresses: [shalinigambhir.21@gmail.com](mailto:shalinigambhir.21@gmail.com) (S. Gambhir), [skmalik9876@gmail.com](mailto:skmalik9876@gmail.com) (S.K. Malik), [yugalkumar.14@gmail.com](mailto:yugalkumar.14@gmail.com) (Y. Kumar).

<http://dx.doi.org/10.1016/j.nhtm.2017.10.001>

Received 19 July 2017; Received in revised form 12 September 2017; Accepted 8 October 2017

2307-5023/ © 2017 European Society for Translational Medicine. Published by Elsevier Ltd. All rights reserved.

powerful tool for training of the MLNNs [6–10]. In this work, a PSO-ANN based diagnostic model is proposed for earlier detection of dengue disease. In the proposed model, PSO method is employed to optimize the parameters of ANN approach. Further, the optimized ANN is applied for the detection of dengue patients. The paper is organized as follows: Section 2 summarizes the related works in the field of disease diagnosis. Section 3 illustrates dengue disease data set and the attribute information. In Section 4, artificial neural network approach is discussed. The proposed diagnostic model is explained in Section 5. The results of the study are presented in Section 6. Finally, the entire work is concluded in Section 7.

## 2. Related work

For automatic detection of normal and Coronary Artery Disease (CAD) conditions, Giri et al. [14], have used heart signals to design a methodology. The heart signals, are associated with several frequency sub bands using Discrete Wavelet Transform (DWT). Hence, to retrieve the heart signal from DWT and also to reduce the dimensions of data, three statistical methods are applied to the set of DWT coefficients. Further, the normalized information is passed through a set of four classifiers such as SVM, Probabilistic Neural Network, Gaussian Mixture Model, and K-Nearest Neighbor for decision-making. The results indicate that the GMM model coupled with ICA provides higher accuracy rate. Further, to improve the classification performance of the CAD, Babaoglu et al. [15] have developed a BPSO, GA, and SVM based classification model based on stress testing data. In the proposed model, BPSO and GA techniques are applied for obtaining the relevant set of features to predict CAD. Whereas, SVM is adopted as a classifier system to classify patients with CAD. For enhancing the diagnostic rate, 10-fold cross-validation techniques are incorporated in SVM classifier. Results indicate that the proposed classification model achieves higher accuracy rate and have less complexity. Patil et al. [16] have developed a hybrid prediction model to investigate the characteristics of diabetic patients and also to predict type-2 diabetic effectively. In the proposed prediction model k-means clustering algorithm is applied to validate the class labels of diabetic data set and the classical C4.5 method is selected to build the final classifier using k-fold cross-validation method. From results, it is stated that the proposed system attains higher sensitivity and specificity rate than the other methods being compared. An intelligent hepatitis diagnostic system for caring and treatment of hepatitis patients is reported in [17]. Another application of SVM technique is also reported for effective diagnosis of cancer affected patients [18]. Ucar et al. [19], have proposed a new hybrid machine learning method for identification of tuberculosis disease. The proposed machine learning method is the combination of adaptive neuro-fuzzy inference system and rough sets. The results indicate that the proposed method provides more viable results in comparison to other algorithms being compared. To predict the heart valve disorder, Uguz has developed an adaptive neuro-fuzzy based system [20]. In the proposed system, three layers are presented for feature extraction, feature selection and classification. Feature extraction is done by DWT, whereas, the Shannon entropy algorithm is adopted for feature selection. Finally, all selected features are classified using ANFIS classifier. The proposed system obtains 98.3% classification accuracy. For the effective diagnosis of CAD, Muthukaruppan et al. [21], have presented particle swarm optimization based fuzzy expert system. In the proposed system, the initial decision tree is implemented to find the best features for better prediction of CAD. Further, these features are converted into fuzzy if-then rules and make a database of fuzzy rules. Prior to convert the crisp set into fuzzy set, a fuzzy membership function is applied in fuzzy set theory and this function has significant impact on the fuzzy output. Further, the authors have adopted PSO algorithm to tune the value of the fuzzy membership function. From the results, it's clear that the fuzzy system is more capable of diagnosis of CAD in comparison to other methods being compared. Seera and Lim have described a fuzzy

min-max neural network for the medical data classification task in [22]. It consists of Fuzzy Min–Max neural network, regression tree, and the random forest algorithms. The efficacy of the system tests for breast cancer, diabetes, and liver disorders diseases and provides significant results. To achieve better accuracy and diagnostic rate of breast cancer affected patients, Ubeyli et al. [23], suggested an adaptive neuro-fuzzy based inference system for breast cancer detection. The first-order Sugeno model is applied with two fuzzy if-then rules and further, this model integrates with neural network architecture for effective detection of breast cancer disease. It is found that the ANFIS achieves 99.08% accuracy rate. Kumar et al., have developed a rule-based classification model to predict the different types of liver diseases and claimed that decision tree based classification model gives higher accuracy [24]. In continuation of their work, five machine learning approaches are applied to evaluate the seminal quality and it is observed that PSO-SVM approach provides better results than MLP, DT, NB, and SVM approaches [25]. Yadav et al. [26], have applied three popular machine learning approaches Tree, Statistical, and SVM classifiers to discover a Parkinson's illness influenced patients. It is seen that the performance of logistic regression classifier is better than other classifiers.

## 3. Dataset description

The details of dengue disease data set have been described in this section. The actual data have been collected from the different hospitals located in Delhi during the months May 2016 to October 2016. A detailed questionnaire is prepared to acquire the relevant information about the patients which is given in Annexure-1. The discretion of dengue dataset is being summarized in Table 1. This data set consists of 110 data instances with sixteen attributes and having two classes such dengue positive and negative. The attribute of dengue dataset are Age (years), Sex (M/F), Temperature, Pulse, Platelet count, Fever, Vomit, Abdominal pain, Chills, Bodyache, Headache, Weakness, Dengue Antigen (NS1), IgM, IgG, Dengue NS1 Antigen (Elisa). The objective of this data set is the early diagnosis of Dengue disease. It is a two-class problem with class label dengue positive and negative. The dengue positive class describes a patient with dengue disease, whereas, dengue negative class specifies a person without the dengue disease. There are 85 samples of class dengue positive and 25 samples of class dengue negative. The purpose of the study is to investigate the relationship between the dengue diagnostic result, environmental, and physiological parameters. The normalized values of all attributes of dengue data set are presented in Table 2. Moreover, Fig. 1 shows the visual representation of the attributes of dengue data set.

**Table 1**  
Normalized values of different attributes of dengue data set.

Sr. No.	Attribute	Values
1	Age (years)	Continuous
2	Sex (M/F)	1 = Male, 0 = Female
3	Temperature	Continuous
4	Pulse	Continuous value in Beats Per Minute
5	Platelet count	Continuous value
6	Fever	1 = Yes, 0 = No
7	Vomit	1 = Yes, 0 = No
8	Abdominal pain	1 = Yes, 0 = No
9	Chills	1 = Yes, 0 = No
10	Bodyache	1 = Yes, 0 = No
11	Headache	1 = Yes, 0 = No
12	Weakness	1 = Yes, 0 = No
13	Dengue Antigen (NS1)	1 and 0
14	IgM	1 and 0
15	IgG	1 and 0
16	Dengue NS1 Antigen (Elisa)	Continuous value

**Table 2**  
Description of dengue dataset.

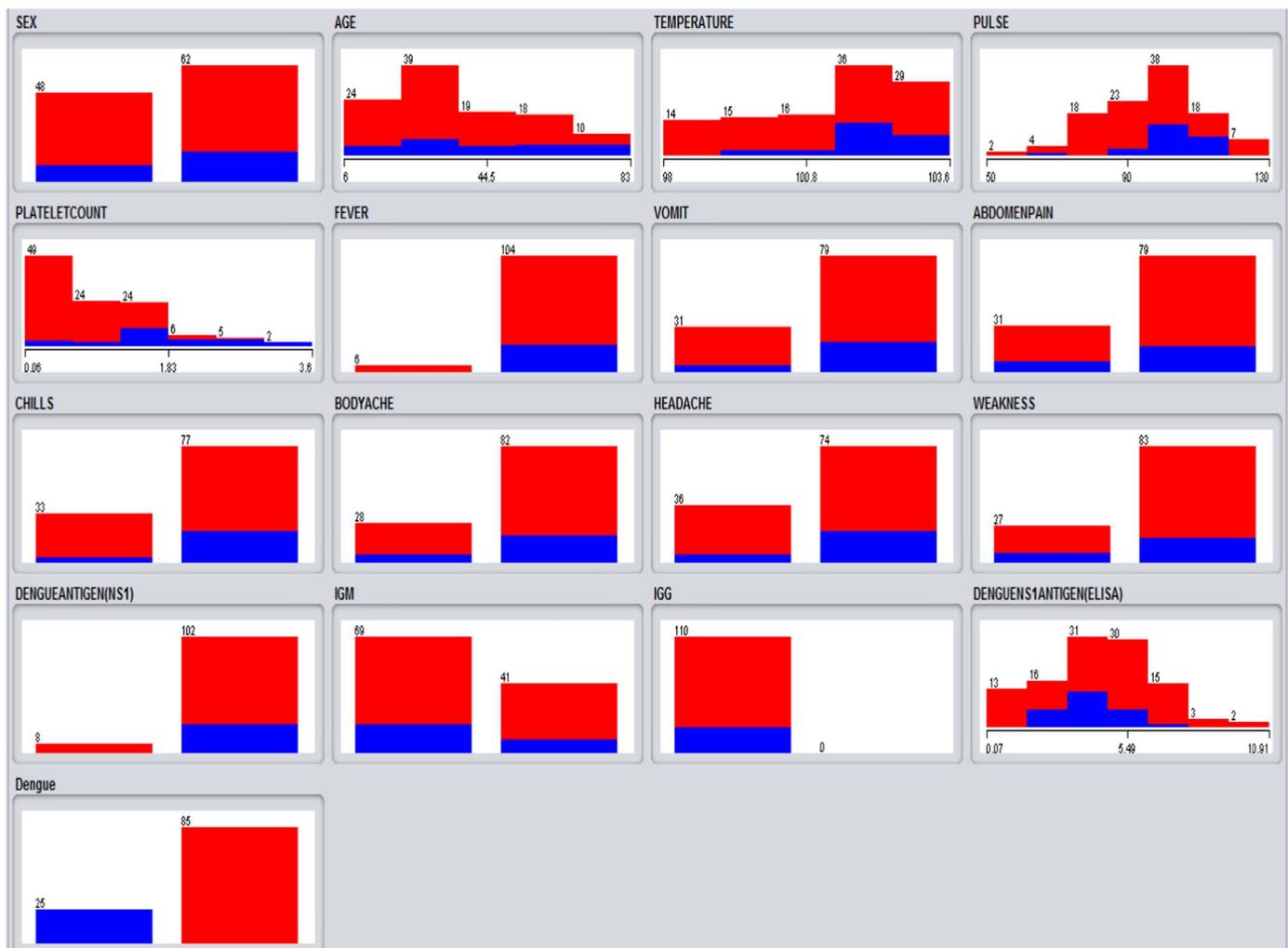
Data set	No. of samples	Input Attribute	Output Attribute	Output Classes	Total No. of Attributes	Missing attributes status	Noisy attributes status
Real life Hospital data of Dengue patients	110	16	1	2	16	No	No

#### 4. Artificial neural networks

An artificial neural network (ANN) is a soft computing technique that can be widely used to process the information [13,14]. It is inspired by biological nervous systems, such as the working mechanism of the brain. It is represented in terms of weighted directed graphs in which nodes act as artificial neurons and directed edges between neurons defined weights. It can be divided into two categories: Feedforward and Recurrent networks. The Feedforward networks are a static while, recurrent networks are dynamic in nature. They produce only one set of output values rather than a sequence of values from a given input. The Feedforward networks are memory-less and independent of the state of the previous network. When a new input pattern is presented, the neuron outputs are computed. On the basis of the feedback paths, the inputs to each neuron are then modified, which leads the network to enter a new state. In the literature, it is observed that the multilayer neural networks have been successfully implemented in decision support systems for disease diagnosis systems [27,28]. In

Multilayer Perceptron (MLP), [29–32] multiple layers of neurons are present, but the minimum layers are three i.e. input layer, one hidden layer, and an output layer which is responsible to generate results. It is also observed that the backpropagation (BP) algorithm [10] is widely used to train the network. Apart from the input layer, every neuron in the other layers acts as a computational element with a nonlinear activation function. The principle of the neural network is that when data is presented at the input layer, the network neurons run calculations in the consecutive layers until an output value is obtained at each of the output neurons. The output

of neural network specifies the appropriate class for the input data. Each neuron in the input and hidden layers is connected to all other neurons of the next layer by some weight values. The neurons of hidden layers are responsible to compute the weighted sums of their inputs and add a threshold. Fig. 2 shows the structure of multilayer perceptron using an input layer, a hidden layer, and an output layer. The input layer represents the attributes of datasets. The working of hidden layer represents the attributes of datasets which are not linearly separated



**Fig. 1.** Visual representation of dengue data set attributes.

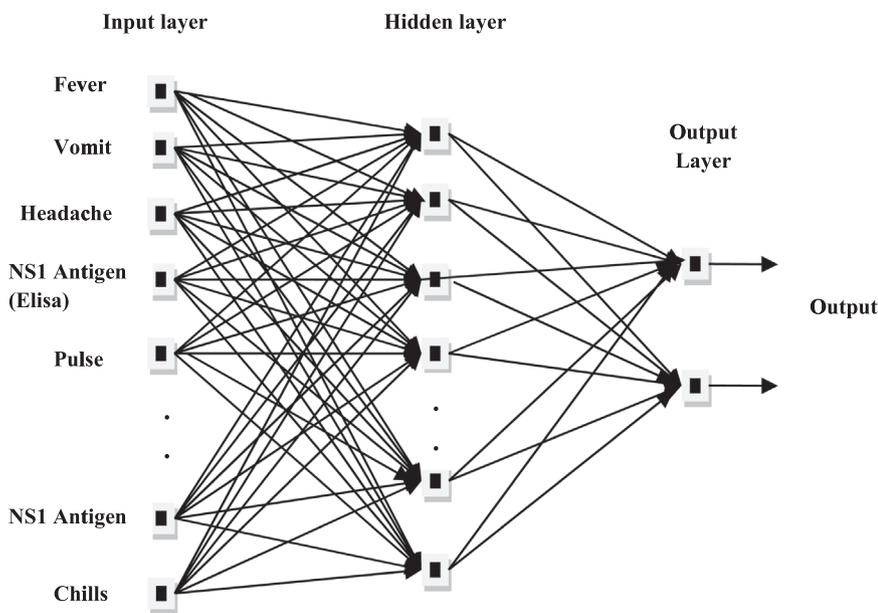


Fig. 2. Three-layer Feedforward neural network: one input, one hidden layer and one output layer for predicting Dengue disease.

and output layer provides the desired results. A threshold node is also added in input layer which specifies the weight function. The resulting sums are used to obtain the activity of the neurons by applying a sigmoid activation function. This process is defined using Eq. (1).

$$p_j = \sum_{i=1}^n w_{ji}x_i + \theta_j, m_j = f_j(p_j) \quad (1)$$

Where  $p_j$  is the linear combination of inputs  $x_1, x_2, \dots, x_n$ , and the threshold  $\theta_j$ ,  $w_{ji}$  is the connection weight between the input  $x_i$ , and the neuron  $j$ , and  $f_j$  is the activation function of the  $j$ th neuron, and  $m_j$  is the output. A sigmoid function is a common choice of activation function and described in Eq. (2).

$$f(t) = \frac{1}{1+e^{-t}} \quad (2)$$

To train the MLP, the backpropagation learning method has been used [32] which is a gradient descent method for the adaptation of the weights. All the weight vectors ( $w$ ) are initialized with small random values from a pseudorandom sequence generator. However, this process can take too many steps to train the network, and the adjusted weights are computed at each step. To overcome the above-mentioned problems, a particle swarm optimization based approach is utilized to compute the optimal value of the weight and threshold functions, because PSO have the capability to determine weight parallel and finding the optimal solutions.

## 5. Proposed approach

This section describes the proposed model for the early detection and diagnosis of dengue disease. In this work, the PSO optimized neural network is applied for effective diagnosis of dengue. The Subsection 5.1 describes the parameter optimization of ANN approach while Subsection 5.2 illustrates the proposed diagnostic model.

### 5.1. Parameter optimization of ANN using PSO

It is noticed that weight ( $w$ ) and bias ( $b$ ) parameters have great influence on the performance of ANN. In this work, the PSO method is employed to optimize the parameters of ANN. The PSO is a populated search method, which derives from the research for the movement of organisms in a bird flocking or fish schooling [33]. The method is easy to implement and has few parameters to adjust. PSO performs searches

using a population (called swarm) of individuals (called particles) that are updated with each iteration. To discover the optimal solution, each particle moves in the direction of its previous best position ( $p_{best}$ ) and its best global position ( $g_{best}$ ). The velocity and position of particles can be updated using Eqs. (3) and (4).

$$V_{ij}(t+1) = W*V_{ij}(t) + c_1*r_1*(X_{pbest}(t) - X_{ij}(t)) + c_2*r_2*(X_{gbest}(t) - X_{ij}(t)) \quad (3)$$

$$X_{ij}(t+1) = X_{ij}(t) + V_{ij}(t+1) \quad (4)$$

Where “ $t$ ” denotes the iteration counter,  $v_{ij}$  is the velocity of the particle “ $i$ ” on the  $j$ th dimension, whose value is limited to the range of  $[v_{min}; v_{max}]$ ;  $p_{ij}$  is the position of particle “ $i$ ” on the  $j$ th dimension, whose value is limited to the range  $[X_{min}; X_{max}]$ ;  $X_{pbest}$  is the  $p_{best}$  position of the particle “ $i$ ” on the  $j$ th dimension, and  $X_{gbest}$  is the  $g_{best}$  position of the swarm on the  $j$ th dimension. The inertia weight “ $w$ ” is used to balance the global exploration and local exploitation. The  $r_1$  and  $r_2$  are the random function in the range  $[0, 1]$ ,  $b$  is constraint factor used to control the velocity weight, whose value is usually set to 1. Positive constant  $c_1$  and  $c_2$  are personal and social learning factors, whose values are usually set to 2. Here, the particle is composed of the parameters  $w$  and  $b$ . Fig. 3 presents the process of optimizing the ANN parameters with PSO, which is described below:

The main steps of PSO based parameter optimization process are summarized as:

#### Step 1: Initialization

In this step, different parameters of PSO are initialized with a population of random particles and velocities.

#### Step 2: Train the ANN model and evaluate the fitness function

ANN model is trained with the parameters  $c$  and  $r$  included in the current particle. The 10-fold cross-validation technique is applied to evaluate fitness function. In 10-fold cross-validation, the training data set is randomly divided into 10 mutually exclusive subsets of approximately equal size, in which 9 subsets are used to train the data and the last subset is used to test the data. The above-mentioned procedure is repeated 10 times so that each subset is used once for testing. The fitness function is defined as the  $1-CA_{validation}$  of the 10-fold cross-validation method in the training data set, which is shown in Eqs. (5) and (6). Moreover, a solution with a bigger  $CA_{validation}$  has a smaller fitness

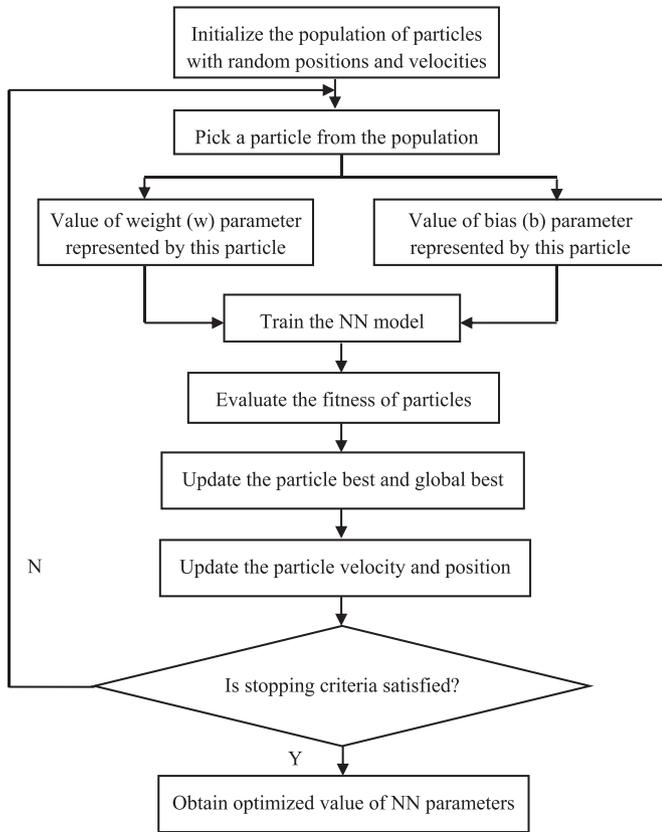


Fig. 3. PSO based Parameter optimization process of ANN.

value.

$$\text{Fitness} = 1 - \text{CA}_{\text{validation}} \quad (5)$$

$$\text{CA}_{\text{validation}} = 1 - \frac{1}{10} \sum_{i=1}^{10} \left| \frac{y_c}{y_c + y_f} \right| \times 100 \quad (6)$$

Where,  $y_c$  and  $y_f$  represent the number of true and false classifications respectively.

Step 3: Update the global and personal best positions

In this step, the global best and personal best positions of particles are updated according to the fitness function values.

Step 4: Update the velocity and position

The position and velocity of each particle are updated using the Eqs. (3 and 4) and obtained the new positions of particles for further iterations.

Step 5: Termination Condition

Repeat steps 2–4 until termination conditions are not satisfied.

### 5.2. Proposed diagnostic model

The proposed diagnostic model for the early detection and prediction of dengue disease has been prescribed in this section. A PSO-ANN based classifier is applied to accurately detect the dengue. In the proposed classifier, PSO is applied to optimize the weight and bias parameters of ANN technique. In the proposed model, 10-fold cross-validation technique is implemented which divides the dengue data into ten equal parts. The detailed description of this technique is given in the

next section. In the proposed model, sixteen input neurons are defined in input layer corresponding to sixteen attributes of dengue data set. The hidden layer consists of seventeen neurons, whereas output layer consists of only two neurons corresponding of class labels i.e. dengue positive and dengue negative. The weight of each neuron is computed using PSO technique and the optimized weight is used to train the neural network. The classification accuracy and confusion matrix are used to evaluate the performance of the proposed diagnostic model. Fig. 4 shows the proposed diagnostic model for the early detection of dengue disease.

## 6. Experimental results

This section summarizes the results of proposed PSO-ANN based approach in diagnosing and predicting dengue effectively. The parameter setting of all approaches is illustrated in Table 3. The k-fold cross-validation method is also used to validate the results of the proposed approach. Further, a confusion matrix is designed to evaluate the performance of PSO-ANN approach. Confusion matrix defines the relationship between value predicted by the model and actual value. The upper left cell denotes the number of data classified as true while they were true (i.e., True Positive) and the upper right cell denotes the number of data classified as true while they actually were false (i.e., False Positive). The lower left cell denotes the number of data classified as false while they actually were true (i.e., False Negative) and lower right cell denotes the number of data classified as false while they were actually false (i.e., True Negative). The performance of the PSO-ANN diagnostic model is measured using accuracy, sensitivity, specificity, error rate and area under the curve (AUC) parameters. The results of the proposed model are also compared with other approaches like Naive Bayes, Decision Tree, ANN, and PSO.

### 6.1. Performance measures

**Accuracy:** Accuracy of a model is defined as the total positive instances of the model divided by the total number of instances. Accuracy parameter provides the percentage of correctly classified instances. The accuracy of the model is defined using Eq. (7).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (7)$$

**Sensitivity:** This parameter is used to determine the degree of the attributes to correctly classify the person with diseases and can be computed using Eq. (8).

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

**Specificity:** This parameter is used to determine the degree of the attributes to correctly classify the person without diseases and evaluated using Eq. (9).

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (9)$$

**Area Under the Curve:** Area under the curve (AUC) is the important parameter which is used to assess the performance of diagnostic tests as well as to identify the prevalence of a disease. It is two-dimensional plots between the sensitivity and specificity and measures the validity of medical tests.

### 6.2. K-fold cross-validation technique

The k-fold cross-validation technique is widely used to validate the performance of data mining models as well as statistical analysis of data sets [34,35]. In the k-fold cross-validation technique, k is defined as the number of folds in the data set in which k-1 folds are used as training

Fig. 4. Proposed PSO-ANN based diagnostic model.

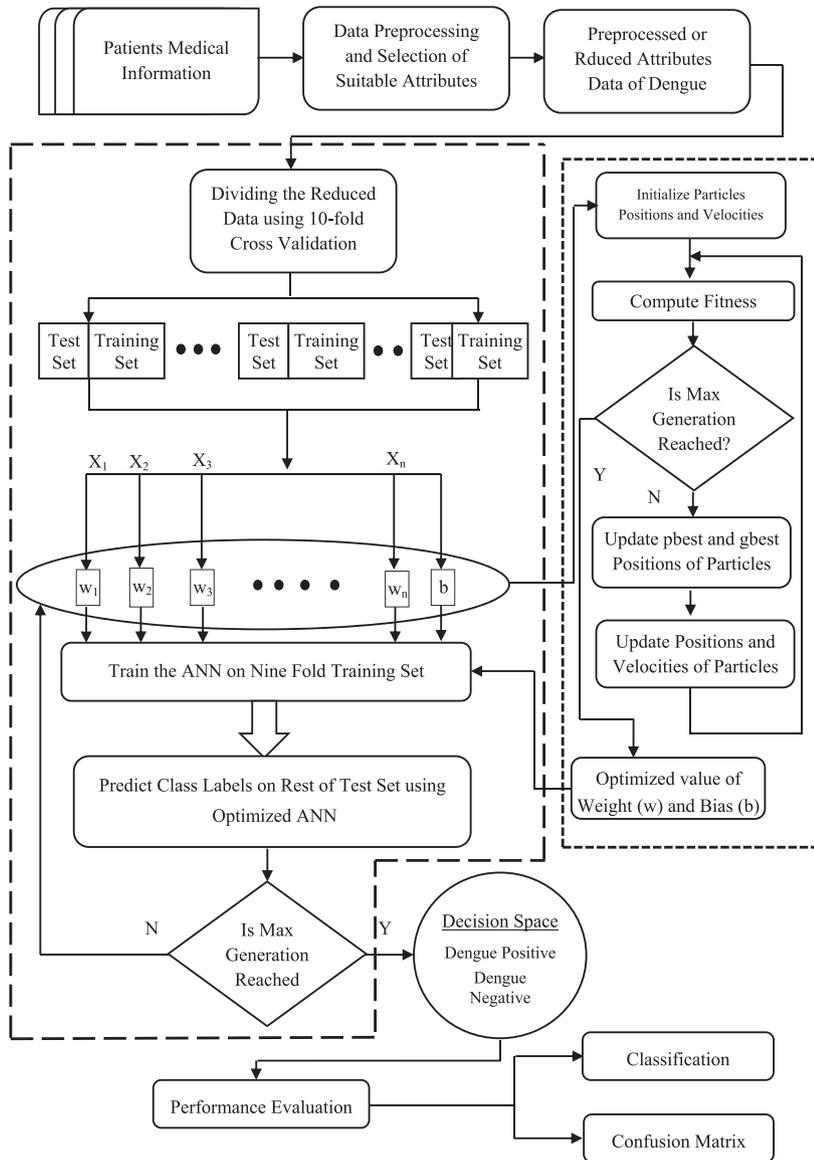


Table 3  
Parameters setting of DT, ANN, NB, and PSO approaches.

DT		ANN		NB		PSO	
Parameters	Values	Parameters	Values	Parameters	Values	Parameter	Value
Criterion	Information Gain	Network	MLP	Estimation Method	Greedy	Number of particles	20
Size of split	2	Hidden layers	3	Minimum Bandwidth	0.01	$c_1 = c_2$	2
Minimal leaf size	2	Training Cycles	600	Number of Kernels	10	$\omega_{min}$	0.5
Minimal Gain	0.01	Learning Rate	0.3			$\omega_{max}$	1
Maximal Depth	20	Momentum	0.25			Velocity	0.9
Confidence	0.5						

instances and  $k$ th fold is used as test instance. In this paper, the numbers of folds are 10 i.e. the data set is divided into 10 parts such that  $D = \{a_1, a_2, a_3, \dots, a_{10}\}$ . However, 9 folds out of 10 folds are used as training instances and the 10th fold is used as test instance that gives

the accuracy of the model. In the  $k$ -fold cross-validation technique, the test instance is used to predict the class labels. A system diagram of 10-fold cross-validation technique is given in Fig. 5.

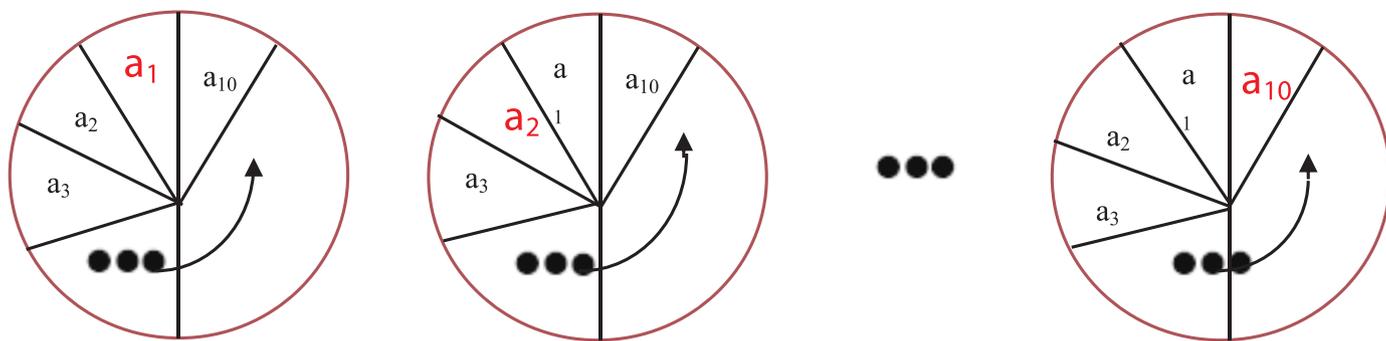


Fig. 5. The system diagram of 10-fold cross-validation technique in which bold red alphabet portion of the wheel act as a test instance while others act as training instances. This process is executed up to 10 iterations and each iteration consists of different test instances.

Table 4  
Confusion Matrix for DT, ANN, NB, PSO, and PSO-ANN diagnostic model.

Confusion Matrix	DT PREDICTED	ANN		NB		PSO		PSO-ANN			
		Dengue (-)	Dengue (+)								
A C T U A L	Dengue (-)	10	15	10	15	11	14	12	13	17	8
	Dengue (+)	14	71	8	77	11	74	13	72	6	79

6.3. Result analysis and discussion

This subsection describes the results of our study. Table 4 illustrates the confusion matrix obtained from dengue data set using ANN, DT, NB, PSO, and PSO-ANN approaches. The confusion matrix is used to derive the results of performance measures used in this study. This matrix contains actual and predicted values of all data instances of dengue disease. The experimental results of all above-mentioned approaches are described in Table 5. It is seen that PSO-ANN based approach provides better results in comparison to other approaches. It is also noticed that PSO-ANN approach based diagnostic model obtains higher accuracy, sensitivity and specificity rate. Further, it is also observed that DT approach provides the worst performance among all. Also, the performance of ANN approach is significantly improved by optimizing its weight and bias parameters through PSO method. Moreover, the ranking of each approach is also computed using performance measures and it is noticed that PSO-ANN approach obtains the first rank for all the parameters and DT got the worst rank.

Table 6 summarizes the results of AUC parameter of all different approaches. The final AUC is determined using AUC of dengue negative and positive. This parameter is widely used to validate the diagnostic accuracy. It is seen that PSO-ANN approach has the highest AUC rate. This parameter also validates that the PSO-ANN based diagnostic model provides more accurate results and can be effectively applied to the early diagnosis of dengue disease.

Table 5  
Performance comparison of DT, ANN, NB, PSO, and PSO-ANN diagnostic model.

Parameters	Techniques				
	NB	DT	ANN	PSO	PSO-ANN
Accuracy	77.2	73.63	79.09	76.3	87.27
Rank	3	5	2	4	1
Sensitivity	44	41.66	55.55	48	68
Rank	4	5	2	3	1
Specificity	84.88	82.55	88.5	87.05	92.94
Rank	4	5	2	3	1
Error Rate	22.8	26.33	20.9	23.7	12.73
Rank	3	5	2	4	1
Overall Rank	3.5	5	2	3.5	1

Table 6  
AUC results of different classifiers and proposed diagnostic model.

	NB	DT	ANN	PSO	PSO-ANN
Dengue Negative	0.752	0.723	0.779	0.736	0.826
Dengue Positive	0.742	0.729	0.776	0.747	0.819
Total	0.747	0.726	0.778	0.742	0.823

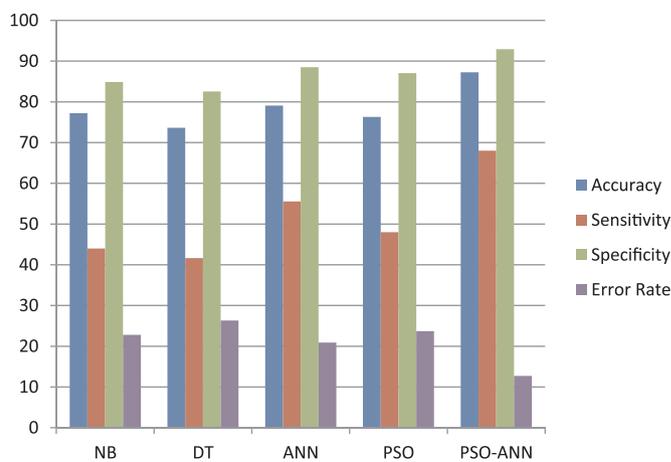


Fig. 6. Performance comparison of accuracy, sensitivity, specificity, and error rate of different approaches.

Fig. 6 shows the performances of different approaches using accuracy, specificity, sensitivity, and error rate parameters graphically. It is observed that PSO-ANN approach obtained higher accuracy, sensitivity, specificity rate and low error rate for dengue while, DT approach presented worst performance. Based on the experimental study, the following point can be highlighted:

- The higher accuracy rate can be achieved using PSO-ANN approach. It can be possible by optimizing weight (w) and bias (b) parameters of ANN using PSO technique. From the PSO algorithm, an initially optimized value of the weight and bias parameter is determined and after that, the optimized values pass in ANN parameters.

- It is observed that slight variations are occurring between the performances of NB and PSO approaches, especially for sensitivity and specificity parameters. The average rank of both the techniques is same, i.e. 3.5. This indicates that both of the techniques have similar performance for diagnosis of dengue disease. But, the AUC performance measure indicates that NB approach is better than PSO due to higher AUC rate.
- To validate the diagnostic accuracy, AUC performance measure is also used. The result of this performance measure also confirms that the PSO-ANN based diagnostic model is more effective and efficient model.

## 7. Conclusion

Dengue is a life-threatening disease. A PSO-ANN based diagnostic model has been implemented and validated for the early detection and accurate prediction of dengue disease. In this model, at first PSO method is adopted to optimize the parameters, and then the ANN approach has been applied to identify the dengue influenced patients. Further, the 10-fold cross-validation method has been employed to validate the results. It is observed that the proposed model provides better outcomes in comparison to other approaches like ANN, DT, NB, and PSO. In addition, the parameter optimization of ANN has a significant impact in accomplishing an improved performance. The PSO-ANN approach has a large computational time than other methodologies. The research presented in this paper depicts that the PSO-ANN based diagnostic model is capable to diagnose dengue in its early stage with high accuracy.

## Conflict of interest

In this work, there is no conflict of interest. We are not receiving funding for any agencies. This article does not contain any studies with human participants or animals performed by any of the authors.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.nhtm.2017.10.001>.

## References

- [1] S. Bhatt, P.W. Gething, O.J. Brady, J.P. Messina, A.W. Farlow, C.L. Moyes, et al., The global distribution and burden of dengue, *Nature* 496 (2016) 504–507.
- [2] O.J. Brady, P.W. Gething, S. Bhatt, J.P. Messina, J.S. Brownstein, A.G. Hoen, et al., Refining the global spatial limits of dengue virus transmission by evidence-based consensus, *PLoS Negl. Trop. Dis.* 6 (2012) 1760, <http://dx.doi.org/10.1371/journal.pntd.0001760>.
- [3] D.J. Gubler, Dengue and dengue hemorrhagic fever, *Clin. Microbiol. Rev.* 11 (3) (1998) 480–496.
- [4] E. Gupta, L. Dar, G. Kapoor, S. Broor, The changing epidemiology of dengue in Delhi, India, *Virology* 3 (2006) 92.
- [5] A.K. Hati, Studies on dengue and dengue hemorrhagic fever (DHF) in West Bengal state, India, *J. Commun. Dis.* 38 (2) (2006) 124–129.
- [6] A. Sengur, An expert system based on linear discriminant analysis and adaptive neuro-fuzzy inference system to diagnosis heart valve diseases, *Exp. Syst. Appl.* 35 (1–2) (2008) 214–222.
- [7] A. Sengur, An expert system based on principal component analysis, artificial immune system and fuzzy k-NN for diagnosis of valvular heart diseases, *Comput. Biol. Med.* 38 (3) (2008) 329–338.
- [8] C.-C. Hsu, C.-S. Ho, A new hybrid case-based architecture for medical diagnosis, *Inform. Sci.* 166 (2004) 231–247.
- [9] R. Das, I. Turkoglu, A. Sengur, Effective diagnosis of heart disease through neural networks ensembles, *Exp. Syst. Appl.* 36 (2009) 7675–7680.
- [10] H. Yan, Y. Jiang, J. Zheng, ChenglinPeng and Li Q., A multilayer perceptron-based medical decision support system for heart disease diagnosis, *Exp. Syst. Appl.* 30 (2006) 272–281.
- [11] OrhanEr, N. Yumusak, F. Temurtas, Diagnosis of chest diseases using artificial immune system, *Exp. Syst. Appl.* 39 (2) (2012) 1862–1868.
- [12] M. Karabatak, M. CevdetInce, An expert system for detection of breast cancer based on association rules and neural network, *Exp. Syst. Appl.* 36 (2009) 3465–3469.
- [13] OrhanEr, F. Temurtas, A. CetinTanrikulu, Tuberculosis disease diagnosis using artificial neural networks, *J. Med. Syst.* 34 (2010) 299–302.
- [14] D. Giri, U.R. Acharya, R.J. Martis, S.V. Sree, T.C. Lim, T. Ahamed, J.S. Suri, Automated diagnosis of coronary artery disease affected patients using LDA, PCA, ICA and discrete wavelet transform, *Knowl.-Based Syst.* 37 (2013) 274–282.
- [15] İ. Babaoglu, O. Findik, E. Ulker, A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine, *Exp. Syst. Appl.* 37 (4) (2010) 3177–3183.
- [16] B.M. Patil, R.C. Joshi, D. Toshniwal, Hybrid prediction model for type-2 diabetic patients, *Exp. Syst. Appl.* 37 (12) (2010) 8102–8108.
- [17] D. Çalişir, E. Dogantekin, A new intelligent hepatitis diagnosis system: PCA–LSSVM, *Exp. Syst. Appl.* 38 (8) (2011) 10705–10708.
- [18] B. Zheng, S.W. Yoon, S.S. Lam, Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms, *Exp. Syst. Appl.* 41 (4) (2014) 1476–1482.
- [19] T. Ucar, A. Karahoca, D. Karahoca, Tuberculosis disease diagnosis by using adaptive neuro fuzzy inference system and rough sets, *Neural Comput. Appl.* 23 (2) (2013) 471–483.
- [20] H. Uguz, Adaptive neuro-fuzzy inference system for diagnosis of the heart valve diseases using wavelet transform with entropy, *Neural Comput. Appl.* 21 (7) (2012) 1617–1628.
- [21] S. Muthukaruppan, M.J. Er, A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease, *Exp. Syst. Appl.* 39 (14) (2012) 11657–11665.
- [22] M. Seera, C.P. Lim, A hybrid intelligent system for medical data classification, *Exp. Syst. Appl.* 41 (5) (2014) 2239–2249.
- [23] E.D. Ubeyli, Implementing automated diagnostic systems for breast cancer detection, *Exp. Syst. Appl.* 33 (4) (2007) 1054–1062.
- [24] Y. Kumar, G. Sahoo, Prediction of different types of liver diseases using rule based classification model, *Technol. Health Care* 21 (5) (2013) 417–432.
- [25] Sahoo, J. Anoop, Y. Kumar, Seminal quality prediction using data mining methods, *Technol. Health Care* 22 (4) (2014) 531–545.
- [26] G. Yadav, Y. Kumar, G. Sahoo, Predication of Parkinson's disease using data mining methods: A comparative analysis of tree, statistical and support vector machine classifiers. in: *Proceedings of National Conference on Computing and Communication Systems (NCCCS)*, IEEE pp. 1–8.
- [27] F. Amato, A. López, E.M. Peña-Méndez, P. Vañhara, A. Hampl, J. Havel, Artificial neural networks in medical diagnosis, *J. Appl. Biomed.* 11 (2013) 47–58.
- [28] J. Nahar, T. Imam, K.S. Tickle, Y.P.P. Chen, Computational intelligence for heart disease diagnosis: a medical knowledge driven approach, *Exp. Syst. Appl.* 40 (1) (2013) 96–104.
- [29] B.D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.
- [30] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Englewoods Cliffs, Prentice-Hall, NJ, 1998.
- [31] C.M. Bishop, *Neural Networks for Pattern Recognition*, Oxford Univ Press, 2005.
- [32] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, *Nature* 323 (1985) 533–536.
- [33] J. Kennedy, *Particle Swarm Optimization*, Encyclopedia of Machine Learning, Springer, US, 2011, pp. 760–766.
- [34] C. Schaffer, Selecting a classification method by cross-validation”, *Mach. Learn.* 13 (1993) 135–143.
- [35] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *Proceedings of International Joint Conference on AI*, 1995, pp. 1137–1145.