



## A NOVEL APPROACH TO PREDICT DENGUE DISEASES IN PATIENTS USING COATI OPTIMIZATION-BASED SUPPORT VECTOR MACHINE

Rama Krishna K<sup>1\*</sup>, Dr. Mohan K G<sup>2</sup>, Dr. Mahalakshmi R<sup>3</sup>

### Abstract

The risk prediction in dengue patients and risk criteria definition for dengue disease are included in the range of challenges that have been addressed. For life-saving measures, surveillance and clinical care, it is essential to effectively and accurately predict the risk stage for dengue disease. The existing studies suggested several techniques but it met a few of the shortcomings in terms of computational and time complexities with prediction accuracy. Hence, this study proposed a novel approach by applying machine-learning algorithms for dengue disease prediction. Initially, the patient's data like Age, Sex, Body Temp, Platelets Level Count, Electrolytes Count, Red Blood Cell (RBC) Count and White Blood Cell (WBC) count are gathered from the Government Medical College, Bellary. After that, this study proposed a novel Coati Optimization algorithm-based Support Vector Machine (CO-SVM) model to predict the presence or absence of dengue diseases based on the symptoms of Fever, Headache, Rash skin, Severe Muscle and joint pain, Severe pain behind the eyes, Swollen Glands and Exhaustion. For this study, the patient's data were collected from Government Medical College, Bellary during the time period of 2019 to 2022. The total of 1480 patients is admitted in hospital. Out of this, 978 patients are dengue positive cases and the rest of 502 cases are dengue negative classes. Statistical parameters like prediction accuracy (A), Mean Squared Error (MSE), Mean Absolute Error (MAE) Root Mean Square Error (RMSE) measures are to validate the effectiveness of the proposed CO-SVM method for dengue disease prediction and the proposed work outcomes are compared to other existing state-of-art methods.

**Keywords:** Dengue disease, Fever, Joint pain, Coati Optimization and Support Vector Machine.

---

<sup>1\*</sup> Assistant Professor, Computer Science and Engineering, Presidency University, Bangalore.  
Ramakrishna@presidencyuniversity.in

<sup>2</sup> Professor, Computer Science and Engineering, Gitam University, Bangalore, mkabadi@gitam.edu

<sup>3</sup> Professor, Computer Science and Engineering, Presidency University, Bangalore,  
mahalakshmi@presidencyuniversity.in

**\*Corresponding Author:** Rama Krishna K

\*Assistant Professor, Computer Science and Engineering, Presidency University, Bangalore.  
Ramakrishna@presidencyuniversity.in

## **1. Introduction:**

Dengue [1] is an arbovirus sickness that occurs among four main infectious dengue strains of the virus that are transmitted through infection. It is currently the most common infectious disease spread by mosquitoes that affect people and pose a severe threat to global security. An expected Three billion humans worldwide are in danger of getting sick, and each year, there are claimed to be ailment experiences, 20,000 mortality, and illness occurrences. Now, approximately 100 nations have confirmed it, with Southeast Asia accounting for nearly 70% of global zika illness and death, the Caribbean, Southeast Asia, and Latin America [2] are all affected. Nevertheless, as more medically significant occurrences are indeed going unreported in the neighborhood by the maternity ward security measures often employed in collecting ailment investigations, the true prevalence of the condition is believed to have several orders of magnitude larger.

The major contribution of this study is summarized;

- ❖ The patient's health data such as Age, Sex, Body Temp, Platelets Level Count, Electrolytes Count, RBC and WBC are gathered from the hospital.
- ❖ The data noise or null blocks are removed and also the unimportant data for disease predictions are removed during the process of data pre-processing.
- ❖ Dengue disease is predicted using the CO-SVM model based on the symptoms of Fever, Headache, Rash skin, Severe Muscle and joint pain, severe pain behind the eyes, Swollen Glands and Exhaustion.

The rest of the article is summarized as; Section 2 reviews the related works and the proposed methodology is described in Section 3. Section 4 discusses the result and the paper is concluded in section 5.

## **2. Related Works:**

Amin et al. [9] have described machine learning techniques for detecting dengue using only messages shared on social media. With the whole material, users then extrapolate important information including doing various analyses that can enhance overall living conditions. The post may contain details about a specific individual, including if they are ill with infectious diseases specifically or just in general. It is effective to

analyze comment content for disease diagnosis. However, the quantity of hashtags is not a reliable predictor of outbreaks.

Umar et al. [10] have presented a genetic algorithm (GA) that provides the parameters to resolve the dynamic framework of dengue fever. This method uses a genetic variation to handle simultaneously confined as well as unregulated domains. It is a graphics rendering worldwide search method. The process of optimum is continuously used to address the numerous inflexible rather than complex structures and correct the specific demographic findings. It is examined by consistently gathering the findings that overlap at a valid and acceptable level of precision for resolving the dengue disease. Hence, it is analyzed statistically to resolve differential modelling.

Latif et al. [11] have determined machine learning methods installation and Usage of Health Diagnosis Systems for Online Health Records. The nodes are denoted as terminal and non-terminal nodes. Every non-terminal node represents a constraint or assessment of a data element. This method is frequently used to classify entries, which is useful when problems involve association and prediction. It enhances the prediction outcomes by utilizing fresh methodologies in various serious diseases. Furthermore, these methods lack the intelligence to determine whether certain illnesses are consistent or extend past the set of guidelines.

Gupta et al. [12] have evaluated machine learning (ML) text classification approaches employed by monitoring systems to collect information from social networking sites in the medical industry.

The investigation identifies the various media or internet platforms utilized for observation in the medical sector, as well as the medical topic(s), predicated on the collection of 148 citations. The capacity to anticipate diseases significantly increased also with the incorporation of bandwidth in security measures. Hence, it is inadequate to produce accurate and relevant epidemic diagnoses.

## **3. Proposed Framework:**

The proposed framework introduced a novel coati optimization-based support vector machine model to predict dengue disease based on different

health data of the patients. Figure 1 describes the overall architecture of the proposed methodology.

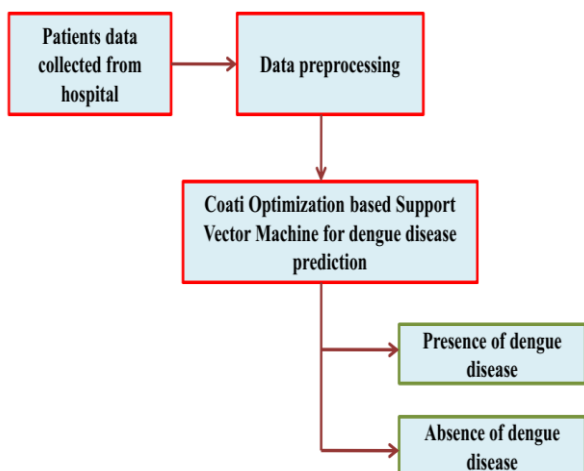


Fig 1: Overall proposed framework

### 3.1 Data Pre-processing:

The dengue patient's data collected from the hospital with respect to the patient's health data such as Age, Sex, Body Temp, Platelets Level Count, Electrolytes Count, RBC and WBC. The data noise or null blocks are removed and also the unimportant data for disease predictions are removed during the process of data pre-processing [17]. The missing data is handled by identifying via the approach of mean calculation. The median values are imputed with missing and values are ignored. Finally, the data is converted into the required form for further processing.

### 3.2 Dengue Disease Prediction:

Next to data pre-processing, the data is fed for dengue disease prediction. This section describes the novel Coati Optimization algorithm-based Support Vector Machine (CO-SVM) model to predict dengue disease prediction and its steps are described in the below sub-section.

#### 3.2.1 Support Vector Machine:

To determine dengue disease, the SVM model is applied to make training samples error. The SVM regression function is  $X_j \in \mathfrak{R}^m, Y_j \in \mathfrak{R}, K=1,2,\dots,m$  for the given sample sets  $(X_j, Y_j)$ . The below equation illustrates the SVM regression function ( $fun(X)$ ) [18].

$$fun(X) = \chi\lambda(X) + B \quad (1)$$

The non-linear mapping, offset degree and weights of the SVM are  $\lambda(x)$ ,  $B$  and  $\chi$ . The structural risk idea minimization and loss function combination transforms to SVM is gives as;

$$\begin{cases} \min \frac{1}{2} \|\chi\|^2 + CP \sum_j^n (\delta_j + \delta_j^*) \\ s.t \quad Y_p - \chi\lambda(X_j) - B \leq \beta + \delta_j \\ \quad \quad Y_p + \chi\lambda(X_j) + B \leq \beta + \delta_j^* \\ \quad \quad \delta_j \geq 0, \delta_j^* \geq 0 \end{cases} \quad (2)$$

The relaxation variables are  $\delta_j$  and  $\delta_j^*$ . The penalty parameter of SVM is  $CP$ . The below equation defines the kernel function.

$$ker\ nel(Y_j, Y_k) = \lambda(X_j)\lambda(X_k) \quad (3)$$

The below method adopts the standardized function of RBF kernel.

$$ker\ nel(Y_j, Y_k) = \exp\left(-G\|Y_j - Y_k\|^2\right) \quad (4)$$

The kernel function with respect to variance is  $G$ .

#### 3.2.2 Coati Optimization algorithm:

The population members consider the population-based metaheuristics called the Coati Optimization algorithm (COA).

For the decision variables, the values are determined with every coati position in the search space. The problem of the candidate solution is represented by the COA.

$$Y_j : Y_{j,k} = BL_k + R \cdot (BU_k - BL_k), \quad j=1,2,\dots,M, \quad k=1,2,\dots,n \quad (5)$$

According to the search space, the  $j^{th}$  coati position is  $Y_j$ . In the  $j^{th}$  decision variable, the value is  $Y_{j,k}$  and a random real number is  $R$  tends to the interval  $[0, 1]$ . Based on  $k^{th}$  decision variables, the lower and upper bounds are  $BL_k$  and  $BU_k$ . The population matrix  $Y$  represents the coat population [19].

$$Y = \begin{bmatrix} Y_1 \\ M \\ Y_j \\ M \\ Y_M \end{bmatrix} = \begin{bmatrix} Y_{1,1} & \Lambda & Y_{1,k} & \Lambda & Y_{1,n} \\ M & O & M & N & M \\ Y_{j,1} & \Lambda & Y_{j,k} & \Lambda & Y_{j,n} \\ & N & M & O & M \\ Y_{M,1} & \Lambda & Y_{M,k} & \Lambda & Y_{M,n} \end{bmatrix}_{M \times n} \quad (6)$$

The problem objective function vector ( $fun$ ) is expressed as;

$$fun = \begin{bmatrix} fun_1 \\ M \\ fun_j \\ M \\ fun_M \end{bmatrix}_{M \times 1} = \begin{bmatrix} fun(Y_1) \\ M \\ fun(Y_j) \\ M \\ fun(Y_M) \end{bmatrix}_{M \times 1} \quad (7)$$

An objective function is a candidate solution quality measure in COA.

**(i) Exploration stage or Attacking and Hunting mechanism:**

While attacking the iguana, the simulating strategy model the population of coatis' in the search space. The iguana position assumes the best population member position in COA design. Equation (8) mathematically simulates the coatis rising position from the tree.

$$Y_j^{P1} : Y_{j,k}^{P1} = Y_{j,k} + R \cdot (Iguana - J \cdot Y_{j,k}) \quad (8)$$

For  $j = 1, 2, \dots, \lfloor \frac{M}{2} \rfloor$  and  $k = 1, 2, \dots, n$

The random position in a search space placed after the iguana falls to the ground. In the search space, equation (9) and (10) simulates the coatis on the ground moving depending upon the random position.

$$Iguana^g : Iguana_k^g = BL_k + R \cdot (BU_k - BL_k), \quad k = 1, 2, \dots, n \quad (9)$$

$$Y_j^{P1} : Y_{j,k}^{P1} = \begin{cases} Y_{j,k} + R \cdot (Iguana - J \cdot Y_k^g), & fun_{Iguana}^g < fun_j \\ Y_{j,k} + R \cdot (Y_{j,k} - Iguana_k^g), & Otherwise \end{cases} \quad (10)$$

For  $j = \lfloor \frac{1}{2} \rfloor + 1, \lfloor \frac{1}{2} \rfloor + 2, \dots, M$  and  $k = 1, 2, 3, \dots, n$

While enhancing the objective function, calculate the new position is to every coati. Equation (11) simulates the updated condition.

$$Y_j = \begin{cases} Y_j^{P1}, & fun_j^{P1} < fun_j \\ Y_j, & Otherwise \end{cases} \quad (11)$$

For the  $j^{th}$  coati, update the new position  $Y_j^{P1}$  and the objective function  $fun_j^{P1}$ . The iguana's position on the ground is  $Iguana^g$ .

**(ii) Exploitation phase or process of escaping from predators:**

Equation (12) and (13) locates each coati and the position generates the random position to simulate the behavior.

$$BL_k^{local} = \frac{BL_k}{T}, \quad BU_j^{local} = \frac{BU_j}{T}, \quad \text{where } T = 1, 2, \dots, t \quad (12)$$

$$Y_j^{P2} : Y_{j,k}^{P2} = Y_{j,k} + (1 - 2R) \cdot (BU_k^{local} + R \cdot (BU_k^{local} - BL_k^{local}))$$

$j = 1, 2, \dots, M$  and  $k = 1, 2, \dots, n$

(13)

The objective function is improved and acceptable is a new position calculated.

$$Y_j = \begin{cases} Y_j^{P2}, & fun_j^{P2} < fun_j \\ Y_j, & otherwise \end{cases} \quad (14)$$

The new position is  $Y_j^{P2}$  and the iteration is T. Complete the COA iteration based on the exploration and exploitation strategy then updates all coatis' position. The COA iteration is repeated till the population is updated based on equations (8-14).

**3.2.3 Proposed CO-SVM model for dengue disease prediction:**

Next to data pre-processing, the dengue disease patients are predicted via the Coati Optimization algorithm-based Support Vector Machine (CO-SVM) model. The hyper-parameter in SVM is tuned by employing the Coati Optimization algorithm. Figure 2 illustrates the CO-SVM model to detect dengue disease patients. There are two processes included in the proposed framework like (i) evaluation of outer classification and (ii) parameter optimization using the CO algorithm. The optimal solution from the CO algorithm dynamically adjusts the parameters of the SVM such as kernel function and weights in which the optimized SVM provide the classification of dengue disease [20]. The fitness function of the proposed work is classification accuracy with respect to CO-SVM.

$$Fitness = \left( \sum_{j=1}^k Accuracy / k \right) \quad (15)$$

The 5-fold cross-validation based CO-SVM achieves an average accuracy thereby obtaining superior prediction accuracy for dengue disease based on the symptoms of Fever, Headache, Rash

skin, Severe Muscle and joint pain, Severe pain behind eyes, Swollen Glands and Exhaustion.

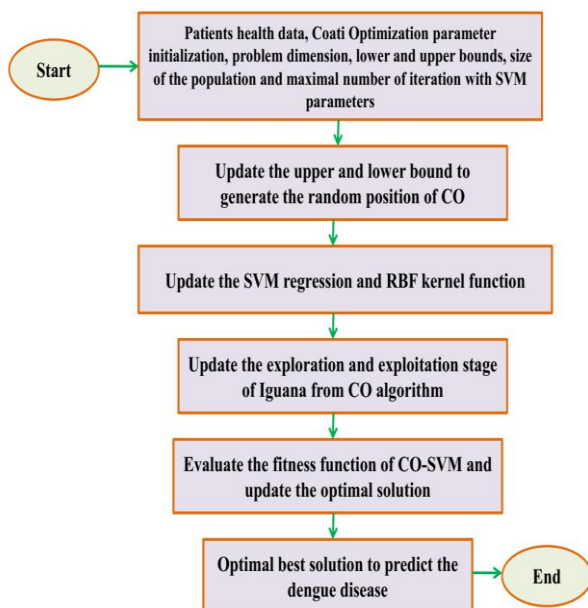


Fig 2: Representation of flowchart for dengue disease prediction using CO-SVM

#### 4. Result and Discussion:

This section discusses the experimental study and the proposed model implemented in the JAVA platform. By using the implementation JAVA platform, we have achieved the performance analysis of state-of-art study results with respect to various statistical parameters are as shown in the below sections. Patients' data were collected from Government Medical College in Bellary for this study between the year of 2019 and 2022. A total of 1480 patients have been admitted to the hospital. 978 of these patients are dengue positive, while the remaining 502 are dengue negative. The patient's health data such as Age, Sex, Body Temp, Platelets Level Count, Electrolytes Count, RBC and WBC are gathered from Government Medical College Bellary. The symptoms of Fever, Headache, Rash skin, Severe Muscle and joint pain, Swollen Glands, severe pain behind the eyes, and Exhaustion are important symptoms to predict dengue disease.

##### 4.1 Statistical Parameters:

Statistical measures such as prediction accuracy (A), Mean Squared Error (MSE), Mean Absolute Error (MAE) Root Mean Square Error (RMSE) and etc are used to analyze the efficiency of proposed CO-SVM for dengue disease prediction.

$$\text{Prediction Accuracy} = \frac{TN+TP}{TP+FP+TN+TP} \quad (16)$$

This is simple to calculate and make precision decisions on that. It represents the proportion of accurate evidence in the patient's data.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (17)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (18)$$

The number of truly positive (521patients) and true negative (457 patients) classes are *TP* and *TN* with the number of false positive (380 patients) and false negative (122 patients) classes are *FP* and *FN*.

It is defined as the measure of prediction error that occurs while computing the comparative study and it can be expressed as,

$$\text{MAE} = \frac{\sum_{i=1}^m |X_i - H_i|}{m} \quad (19)$$

It is the measure of the difference between the predicted and expected value and predicting the errors. It can be expressed as,

$$\text{RMSE} = \left[ \frac{\sum_{i=1}^m |X_i - H_i|}{m} \right]^{1/2} \quad (20)$$

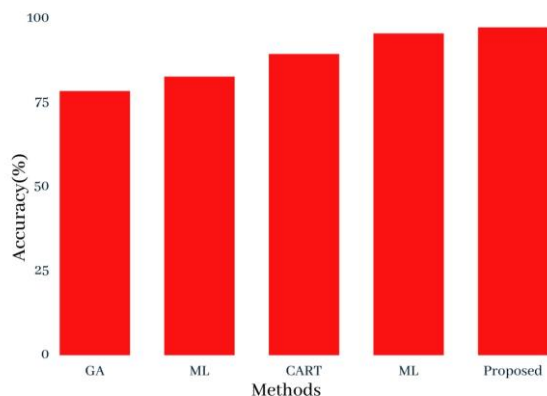
It is defined as the average difference between the expected and predicted value while conducting the dengue disease prediction. It can be expressed as outlined below,

$$\text{MSE} = \left[ \frac{\sum_{i=1}^m |X_i - H_i|^2}{m} \right] \quad (21)$$

Here,  $X_i$  is the expected output and the predicted output is  $H_i$ . The number of samples is denoted as  $m$ .

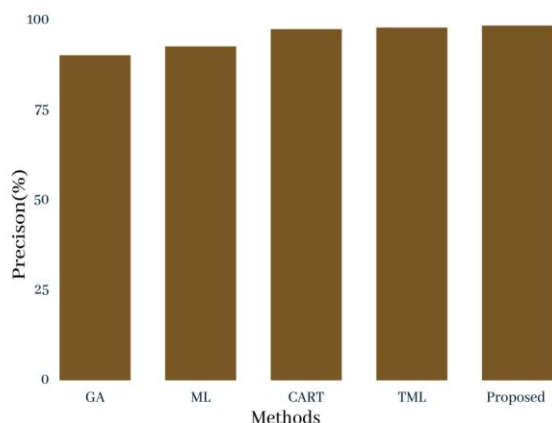
##### 4.2 Performance Analysis:

Figure 3 explains the comprehensive study of dengue disease prediction with different approaches such as GA [10], ML [12], CART [15], TML [13], and the proposed approach in terms of accuracy. From the visualization, it is clear that the proposed approach's accuracy that 97.56% and other approaches.



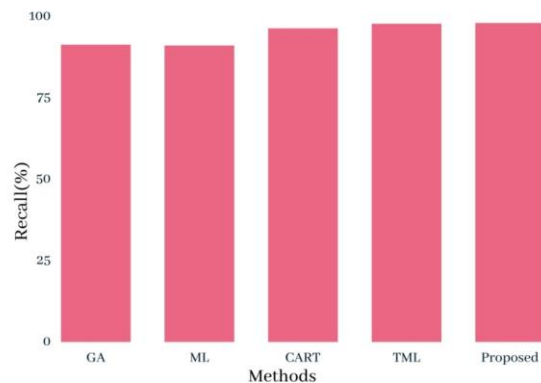
**Fig 3:** Comprehensive study based on the accuracy of the prediction

The comprehensive study of the GA [10], ML [12], CART [15], TML [13], and proposed approach in terms of precision for dengue disease prediction is explained in figure 4. Since our approach utilizes novel approach the precision of our approach is higher at 98.45%. The GA approach provides lesser precision of 90.21%.

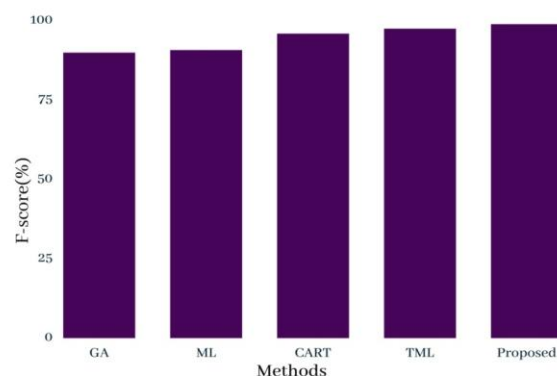


**Fig 4:** Comprehensive study based on the precision of the prediction

Comparative study based on recall is illustrated in figure 5 with different approaches such as GA [10], ML [12], CART [15], TML [13], and proposed approach. The visual displays that the recall of proposed approach is 98.01% and GA is achieved 91.34%, which is lesser of all approaches. The comparative study based on the F-score is illustrated in figure 6 with various approaches such as GA [10], ML [12], CART [15], TML [13], and proposed approach. Since novel approach is utilized our work produces higher F-score and others show fewer results.

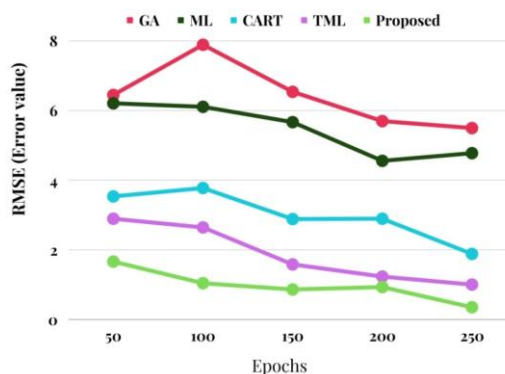


**Fig 5:** Comprehensive study based on the recall of the prediction



**Fig 6:** Comprehensive study based on the F-score of the prediction

The RMSE of proposed approach is small for our approach from the figure 7. This represents that the predicted output is more or less same as the expected output as shown and others show little higher RMSE. The MAE value is illustrated in figure 8 for different approaches such as GA [10], ML [12], CART [15], TML [13], and proposed approach. The MAE of our approach is small and reduces with the number of epochs as shown.



**Fig 7:** Comparative outcome based on RMSE for different epochs

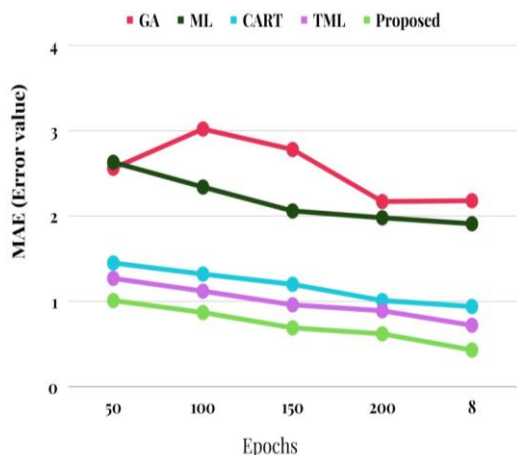


Fig 8: Comparative outcome in terms of MAE

The methods such as GA [10], ML [12], CART [15], TML [13], and proposed approach are the state-of-art method to investigate the performance of dengue disease. Comparative study based on the MSE is visualized in figure 9. The proposed approach shows lower MSE with increasing epochs and other approaches show higher MSE.

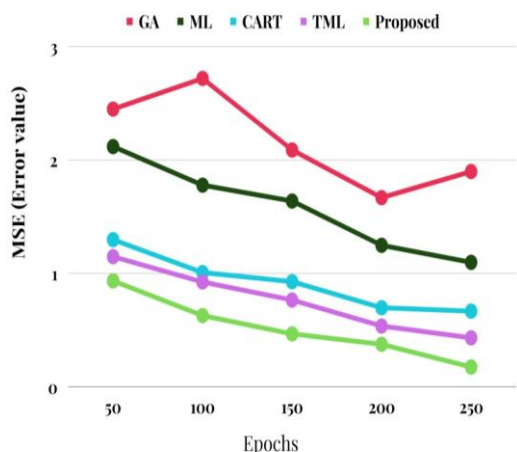


Fig 9: Comparative outcome in terms of MSE

### 5. Conclusion:

This study presented a novel dengue disease prediction using the Coati Optimization algorithm based Support Vector Machine (CO-SVM) model with respect to the symptoms of Fever, Headache, Rash skin, Severe Muscle and joint pain, Severe pain behind eyes, Swollen Glands and Exhaustion and patient's dataset consists of Age, Sex, Body Temp, Platelets Level Count, Electrolytes Count, RBC and WBC. In the proposed CO-SVM, effectiveness is demonstrated based on statistical parameters such as accuracy (A), precision (P), recall (R), Mean Squared Error (MSE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). When compared to the GA, ML, CART and TML methods, the proposed work

outperformed superior performances of accuracy precision and recall with minimum MSE, MAE and RMSE results. The proposed approach's accuracy is 97.56% which is more than other approaches.

### References:

1. Appice, A., Gel, Y.R., Iliev, I., Lyubchich, V. and Malerba, D., 2020. A multi-stage machine learning approach to predict dengue incidence: a case study in Mexico. *Ieee Access*, 8, pp.52713-52725.
2. Pravin, A., Jacob, T.P. and Nagarajan, G., 2020. An intelligent and secure healthcare framework for the prediction and prevention of Dengue virus outbreak using fog computing. *Health and Technology*, 10, pp.303-311.
3. Dey, L., Chakraborty, S. and Mukhopadhyay, A., 2020. Machine learning techniques for sequence-based prediction of viral-host interactions between SARS-CoV-2 and human proteins. *Biomedical journal*, 43(5), pp.438-450.
4. Hedayatipour, A. and Mcfarlane, N., 2020. Wearables for the next pandemic. *Ieee Access*, 8, pp.184457-184474.
5. Shaukat, K., Luo, S., Varadharajan, V., Hameed, I.A. and Xu, M., 2020. A survey on machine learning techniques for cyber security in the last decade. *IEEE Access*, 8, pp.222310-222354.
6. Jelodar, H., Wang, Y., Orji, R. and Huang, S., 2020. Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. *IEEE Journal of Biomedical and Health Informatics*, 24(10), pp.2733-2742.
7. Fernandes, M.S., Cordeiro, W. and Recamonde-Mendoza, M., 2021. Detecting Aedes aegypti mosquitoes through audio classification with convolutional neural networks. *Computers in Biology and Medicine*, 129, p.104152.
8. Riad, M.M. and Eldamak, A.R., 2020. Coplanar waveguide based sensor using paper superstrate for non-invasive sweat monitoring. *IEEE Access*, 8, pp.177757-177766.
9. Amin, S., Uddin, M.I., Hassan, S., Khan, A., Nasser, N., Alharbi, A. and Alyami, H., 2020. Recurrent neural networks with TF-IDF embedding technique for detection and classification in tweets of dengue disease. *Ieee Access*, 8, pp.131522-131533.

10. Umar, M., Sabir, Z., Raja, M.A.Z. and Sánchez, Y.G., 2020. A stochastic numerical computing heuristic of SIR nonlinear model based on dengue fever. *Results in Physics*, 19, p.103585.
11. Latif, J., Xiao, C., Tu, S., Rehman, S.U., Imran, A. and Bilal, A., 2020. Implementation and use of disease diagnosis systems for electronic medical records based on machine learning: A complete review. *IEEE Access*, 8, pp.150489-150513.
12. Gupta, A. and Katarya, R., 2020. Social media based surveillance systems for healthcare using machine learning: a systematic review. *Journal of biomedical informatics*, 108, p.103500.
13. Gangula, R., Thirupathi, L., Parupati, R., Sreeveda, K. and Gattoju, S., 2021. Ensemble machine learning based prediction of dengue disease with performance and accuracy elevation patterns. *Materials Today: Proceedings*.
14. Khan, S., Sajjad, M., Hussain, T., Ullah, A. and Imran, A.S., 2020. A review on traditional machine learning and deep learning models for WBCs classification in blood smear images. *Ieee Access*, 9, pp.10657-10673.
15. Phakhounthong, K., Chaovalit, P., Jittamala, P., Blacksell, S.D., Carter, M.J., Turner, P., Chheng, K., Sona, S., Kumar, V., Day, N.P. and White, L.J., 2018. Predicting the severity of dengue fever in children on admission based on clinical features and laboratory indicators: application of classification tree analysis. *BMC pediatrics*, 18, pp.1-9.
16. Chatterjee, S., Dey, N., Shi, F., Ashour, A.S., Fong, S.J. and Sen, S., 2018. Clinical application of modified bag-of-features coupled with hybrid neural-based classifier in dengue fever classification using gene expression data. *Medical & biological engineering & computing*, 56, pp.709-720.
17. Mishra, P., Biancolillo, A., Roger, J.M., Marini, F. and Rutledge, D.N., 2020. New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends in Analytical Chemistry*, 132, p.116045.
18. Cherkassky, V. and Ma, Y., 2004. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks*, 17(1), pp.113-126.
19. Dehghani, M., Montazeri, Z., Trojovská, E. and Trojovský, P., 2023. Coati Optimization Algorithm: A new bio-inspired metaheuristic algorithm for solving optimization problems. *Knowledge-Based Systems*, 259, p.110011.
20. Ali, L., Wajahat, I., Amiri Golilarz, N., Keshtkar, F. and Bukhari, S.A.C., 2021. LDA-GA-SVM: improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine. *Neural Computing and Applications*, 33, pp.2783-2792.