

Revolutionizing Public Health Surveillance: Computational Solution for Dengue Prediction

Dr. Rajesh Keshavrao Deshmukh¹, Mohit Shrivastav²

¹Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India

²Research Scholar, Department of CS & IT, Kalinga University, Raipur, India

KEYWORDS

Dengue prediction,
risk assessment,
public health,
climate, illness

ABSTRACT

Dengue fever is a rapidly growing vector-borne viral illness that is posing a threat to an increasing number of areas worldwide. Numerous scientists have focused on various strategies to stop and limit the spread of illness. Additionally, the development of a variety of methods for ascertaining and predictive modeling through quantifiable, numerical analysis of machine learning (ML) is investigated. This research introduces a novel cubic-kernelized support vector machine with a bee optimizer (CSVM-BO) approach for dengue identification. The climate data is initially collected and preprocessed utilizing the decimal scale normalization method. In addition, we provide a feature extraction approach for linear discriminant analysis (LDA), which attempts to extract essential information for early identification and risk assessment. The findings demonstrate that the framework suggested in this research has considerable benefits in public health for dengue prediction and that our proposed CSVM-BO technique performs optimally in terms of latency (5 s), time complexity (0.62 s), and accuracy (98.57%). The problems identified by this thorough study offer a helpful foundation for epidemiology and public health research.

1. Introduction

Among the most common vector-borne diseases in the world, dengue fever is frequently recognized as such. Humans get Dengue Virus (DeV) from mosquito bites, which is the source of the illness. The main mosquito species that transmit DeV to people are female *Aedes aegypti* or *Aedes albopictis* [1]. A serious illness, dengue fever can strike anyone of any age, including adults, teenagers, children, and newborns. Human disease transmission is caused by the *Aedes aegypti* mosquito and occurs more frequently in the rainy season [2]. Early illness detection lessens the need for needless hospital stays, lessens the societal burden associated with the disease, and stops it from spreading [3]. To combat the dengue virus, people's immune systems need to have adaptable and natural responses [9]. Even though there is no specific therapy for dengue, it is nevertheless important to diagnose and treat the illness as soon as possible [11]. Dengue is characterized by a complex epidemiological nature, difficult monitoring, and costly and inefficient tracking methods [4]. The spectrum of symptoms linked to dengue infection encompasses skin rashes, high fever, headaches, joint pain and muscle pain, and retro-ocular discomfort, in addition to severe hemorrhagic fever and shock syndrome. An innovative CSVM-BO method for dengue identification is presented in this study.

The article's subsequent sections: portion 2 provides of literature work; portion 3 outlines the technique; and portion 4 shows the simulation. The study is concluded in portion 5.

Literature work

Sharma et al. [5] proposed a global framework to integrate important elements utilizing many inputs, including epidemiological data, health infrastructure data, and remote sensing, to predict dengue sickness. The primary climate variables linked to dengue epidemics were determined by Yavari and Varathan [12]. These characteristics were identified by correlation analysis, and ML models employed these factors as input parameters [14]. Rajendran et al. [7] stated that the task was to predict future incidences of disease by utilizing a recurrent neural network (RNN), and a deep learning (DL) model. The model would help hospitals handle the growing demand for medical care and the government takes preventive action [6]. Corthis et al. [8] used the coati optimization algorithm (COA) in conjunction with the attention-based long short-term memory (ALSTM) method to decrease misdiagnosis error. The purpose of the study by Baker et al. [13] was to estimate the number of disease cases that occur in Iquitos and San Juan each week by analyzing a weather-related dataset using several ML regression techniques.

2. Methodology

Materials

This suggested work analyzes public health for dengue prediction utilizing DL classification methods. The main stages of this suggested research are as follows: gathering of information, pre-processing of the data, feature selection, and optimal data prediction method. Figure 1 illustrates the suggested approach's public health implication for dengue prediction and categorization.

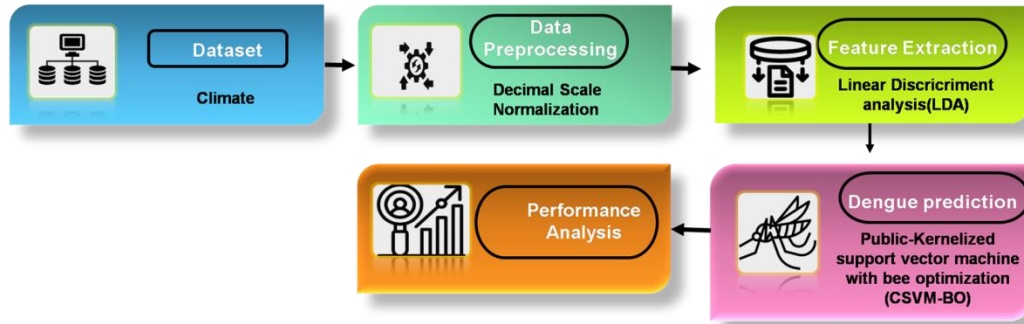


Figure 1 Flow of proposed method

Dataset

This dataset shows every anticipated number about dengue cases in a certain area based on local meteorological data. The dataset is gathered from (<https://www.kaggle.com/datasets/siddhvr/dengue-prediction/data>).

Data preprocessing

The feature values are scaled by a power of ten in the public health for dengue prediction using decimal scaling normalization. Decimal scaling is the method that generates the range between -1 and 1. Therefore, using the decimal scaling method (equation (1)).

$$v^j = \frac{v}{10^i} \quad (1)$$

Where, v^j -the scaled values, v -is the value range, and i -is the least significant number $Max(|v^j|) < 1$.

Feature extraction using linear discriminant analysis (LDA)

The supervised dimensionality reduction and classification method known as LDA is used in statistics and ML. It works by locating linear feature combinations in a dataset that optimize the separation between various classes. A different approach minimizes the average cost function (ECM) to provide the ranking rule. The class response variable is examined in LDA with several independent factors. A discriminant linear function that traverses an average of each group's center is employed to separate the two groups in this case. Classification needs $k - 1$, where k is several classes when there are numerous prediction groups. Consider two groups.

The fisher discriminant function separates as follows if \bar{w}_1 and \bar{w}_2 is the corresponding means for a combined variance-covariance matrix, T represents the first and second groups. If T_j fits into group one, Equation (2).

$$z = (\bar{w}_1 - \bar{w}_2)'T^{-1}T_j \geq \frac{1}{2}(\bar{w}_1 - \bar{w}_2)'T^{-1}(t_1 - t_2) \quad (2)$$

If T_j is a part of the second group, then Equation (3):

$$z = (\bar{w}_1 - \bar{w}_1)'T^{-1}T_j < \frac{1}{2}(\bar{w}_1 - \bar{w}_1)'T^{-1}(t_1 - t_2) \quad (3)$$

The assumptions of LDA include the fact that it demonstrated that a linear discriminant function is resistant to the deviation from the multivariate normality caused by a homogeneous variance-

covariance matrix, contains outlier data, violates multivariate normalization, is linear and lacks multicollinearity amongst independent variables.

Dengue prediction using cubic-kernelized support vector machine with bee optimizer (CSVM-BO)

In this section, we create the CSVM and BO, which were introduced with the suggested approach of the cubic-kernelized support vector machine with the bee optimizer (CSVM-BO) algorithm, which is explained as follows.

Bee optimization (BO)

Finding an optimal solution requires less iteration overall when the queen's speed-reducing factor within BO is appropriately chosen, balancing the world's exploration and extraction. Conversely, selecting the speed reduction factor of the queen accurately depicts the bees cooperative behavior, resulting in a more precise and strong global optimum solution. The goal of employing this strategy is to enhance global search during the first phases of optimization and encourage particle convergence toward the global optimal state after the process (see equation (4)).

$$\alpha(s) = (N - n(s))/N \quad (4)$$

Where $n(s)$ in Equation (4) the entire is amount of drones chosen for coupling during the initial s transitions, and N is the size of the spermatheca.

Cubic-kernelized support vector machine (CSVM)

Structural risk minimization and statistical learning theory provide the foundation of the SVM approach. By projecting input values into a high-dimensional region, the SVM seeks to determine the border between the two classes. $(y_j, x_j) 1 \leq j \leq Q$ represents the input values, while y_j and $y_j \in S^u$ display their size. The output value is found in the x_j values (equation (5)).

$$z_j(x \times w_j + a) > 0, j = 1, \dots, M \quad (5)$$

A hyper plane (5) indicates that a linear separation of the dataset is possible. In this instance, the Equation (6) is derived by constantly resizing x and a .

$$z_{1 \leq j \leq M} z_j(x \times w_j + a) \geq 1, j = 1, \dots, M \quad (6)$$

The distance $1/|x|$ represents the separation between the nearest point and the hyperplane. Equation (1) is therefore changed to Equation (7).

$$z_j(x \times w_j + a) > 1 \quad (7)$$

$$x(\alpha) = \sum_{j=1}^M \alpha_j - \frac{1}{2} \sum_{j,i=1}^M \alpha_j \alpha_i z_j z_i w_j \times w_i \quad (8)$$

Following the discovery of vector $\alpha^0 = (\alpha_1^0, \dots, \alpha_M^0)$ in Equation (8) as a solution to the maximizing issue, OSH $(x_0 a_0)$ possesses Equation (9).

$$x_0 = \sum_{j=0}^M \alpha_j^0 z_j w_j \quad (9)$$

Equation (7) is obtained from the support vectors. The positions where $\alpha_j^0 > 0$ is found are the support vectors. The hyperplane decision function is demonstrated by Equation (10) by considering the expansion Equation (9) of x_0 .

$$e(w) = \text{sgn}(\sum_{j=0}^M \alpha_j^0 z_j w_j \times w + a_0) \quad (10)$$

These kernel functions is classified as radial basis function (RBF), polynomial, linear, and non-linear sigmoid. The definition of the kernel is Equation (11).

$$l(\bar{y}) = \begin{cases} 1 & \text{if } |\bar{y}| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

Algorithm 1: Cubic-kernelized support vector machine with bee optimizer (CSVM-BO)

Input: Determine the various climate evaluation data

Output: predicted class X

Initialize candidates $U = \{\text{most similar pair from the opposing class}\}$

While several points are violated **do**/ Find a violator

applicant SV-BO=applicant SV U_{violator}

if any $\alpha p < 0$ due to the addition of c to S then

applicant SV-BO=applicant SV-BO\p

repeat/**end if**

3. Results and Discussion

This study uses MATLAB 2017b for testing, using a system setup of an i5 CPU, 2GB of GPU RAM, and 8GB of RAM. The effectiveness and accuracy of a suggested strategy CSVM-BO, are contrasted with those of existing techniques such as Deep reinforcement learning (DRL), Deep neural network (DNN), and Adaptive teaching learning optimized convolutional neural network-LSTM (ATLBO optimized CNN-LSTM) to show their effectiveness [10]. The outcome of the given technique establishes the estimated latency, time complexity, and accuracy. The numerical results of classification techniques are shown in Table 1.

Table 1 Numerical outcome of classification methods

Methods	Latency (s)	Time complexity (s)	Accuracy (%)
ATLBO optimized CNN-LSTM	0.075	10	96.9
DNN	0.093	28	93
DRL	0.079	25	92
Proposed	0.062	5	98.57

Latency is the length of time a data packet takes to travel from a source to its destination. Regarding performance optimization, it's critical to evaluate site performance by simulating high latency and optimize for consumers with reduced connections by reducing delay-causing factors. Although each elementary action has a predetermined execution time, the primary manner to estimate the amount of time required is to count the number of fundamental procedures the algorithm completes.

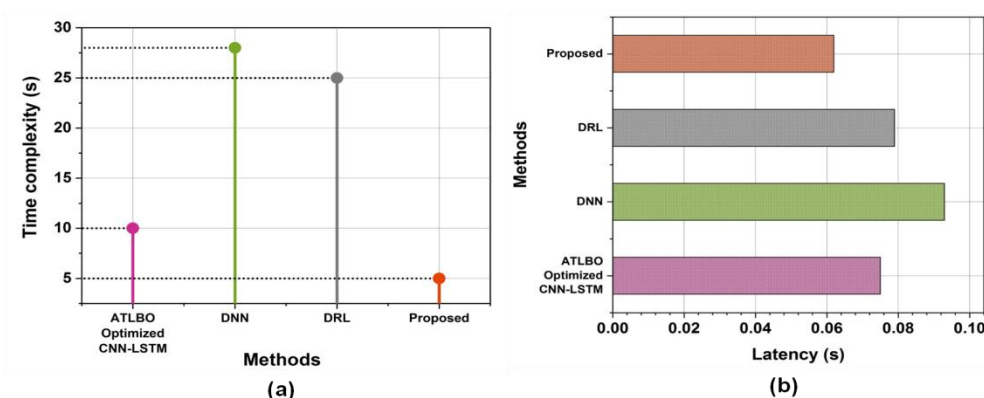


Figure 2 Comparison of the (a) latency, (b) time complexity

Figures 2 (a) and (b) show the suggested system's latency and time complexity. These include the recommended approach and the present systems' latency and temporal complexity of consumption forecasting. While the advanced system achieves the proposed latency in 0.062 s, DRL has gained

0.079 s, ATLO-optimized CNN-LSTM is gathered at 0.075 s, and DNN has attained 0.093 s. The recommended method CSVM-BO achieves 5 s time complexity compared to 28 s for DNN, ATLBO optimized CNN-LSTM is obtained 10 s, and 25 s for DRL. It shows that the recommended action method is more effective than the existing method.

Accuracy indicates that accuracy is a fundamental metric for correctness and precision across many fields. It illustrates how closely a measured value matches the real deal regarding data analysis. Accuracy evaluates each proportion of the correct predictions a model produces, giving information about its efficiency.

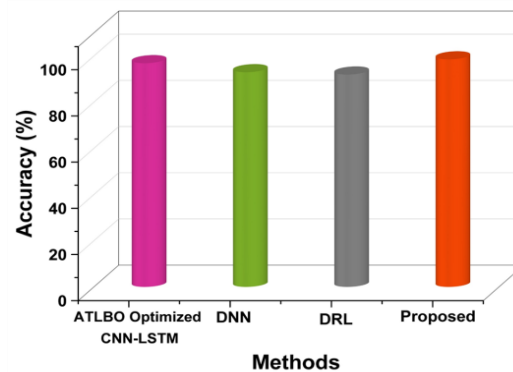


Figure 3 Comparison of the accuracy

The accuracy of the existing and suggested systems is displayed in Figure 3. The advanced system achieves 98.57% of the existing methods for DRL with 92% accuracy, DNN has gained 93%, and ATLBO-optimized CNN-LSTM has attained 96.9%. It shows that the recommended method is more effective than the existing methods.

4. Conclusion and future scope

In this paper, public health for dengue prediction is predicted using a cubic-kernelized support vector machine with bee optimizer (CSVM-BO). Dengue vector numbers and seasonal patterns of transmission are strongly influenced by weather. The results showed that public health for dengue prediction, the effective approach, such as CSVM-BO, should be evaluated thoroughly using accuracy (98.57%), latency (0.062 s), and time complexity (5 s). The results indicated a considerable reduction in the risk of dengue prediction by applying prompt ML case evaluation and efficient control strategies. The use of ML approaches in public health for dengue prediction has the potential to guide the creation of policies and support efforts to eradicate the illness.

Reference

- [1] S.K.Sood, V. Sood, I.Mahajan, and Sahil, "An intelligent healthcare system for predicting and preventing dengue virus infection", *Computing*, 105(3), pp.617-655, 2023. <https://doi.org/10.1007/s00607-020-00877-8>
- [2] G. Gupta, S. Khan, V. Guleria, A. Almjally, B.I. Abdullah, T.Siddiqui, B.M.Albahlal, S.A.Alajlan, and M.Al-Subaie, "DDPM: A dengue disease prediction and diagnosis model using sentiment analysis and machine learning algorithms", *Diagnostics*, 13(6), p.1093, 2023. <https://doi.org/10.3390/diagnostics13061093>
- [3] S. Neelima, Manoj Govindaraj, Dr.K. Subramani, Ahmed ALkhayyat, & Dr. Chippy Mohan. (2024). Factors Influencing Data Utilization and Performance of Health Management Information Systems: A Case Study. *Indian Journal of Information Sources and Services*, 14(2), 146–152. <https://doi.org/10.51983/ijiss-2024.14.2.21>
- [4] S. Amin, M.I. Uddin, M.A. Zeb, A.A. Alarood, M. Mahmoud, and M.H. Alkinani, "Detecting Information on the Spread of Dengue on Twitter Using Artificial Neural Networks" *Computers, Materials & Continua*, 67(1), 2021. <https://doi.org/10.32604/cmc.2021.014733>
- [5] V. Sharma, S.K. Ghosh, and S. Khare, "A proposed framework for surveillance of dengue disease and prediction" *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, pp.317-323, 2023. <https://doi.org/10.5194/isprs-archives-XLVIII-M-1-2023-317-2023>

- [6] Kodric, Z., Vrhovec, S., & Jelovcan, L. (2021). Securing edge-enabled smart healthcare systems with blockchain: A systematic literature review. *Journal of Internet Services and Information Security*, 11(4), 19-32.
- [7] N.M. Rajendran, M. Karthikeyan, B. Karthik Raja, K. Pragadishwaran, E.A. Gopalakrishnan, and V. Sowmya, "Communicable Disease Prediction Using Machine Learning and Deep Learning Algorithms", In *International Conference on Information, Communication and Computing Technology* (pp. 979-992). Singapore: Springer Nature Singapore, 2023, May. https://doi.org/10.1007/978-981-99-5166-6_66
- [8] P.B. Corthis, G.P. Ramesh, and A.B. Jayachandra, "A Meta heuristic-based deep learning classifier for effective dengue disease prediction in IoT-Fog system", *Expert Systems*, p.e13605, 2024. <https://doi.org/10.1111/exsy.13605>
- [9] Malathi, K., Shruthi, S.N., Madhumitha, N., Sreelakshmi, S., Sathya, U., & Sangeetha, P.M. (2024). Medical Data Integration and Interoperability through Remote Monitoring of Healthcare Devices. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, 15(2), 60-72. <https://doi.org/10.58346/JOWUA.2024.I2.005>
- [10] S.N. Manoharan, K.M. Kumar, and N. Vadivelan, "A novel CNN-TLSTM approach for dengue disease identification and prevention using IoT-fog cloud architecture", *Neural Processing Letters*, 55(2), pp.1951-1973, 2023. <https://doi.org/10.1007/s11063-022-10971-x>
- [11] S. Saturn, "Development of prediction and forecasting model for Dengue disease using machine learning algorithms" In *2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)* (pp. 6-11). IEEE, 2020, October. <https://doi.org/10.1109/DISCOVER50404.2020.9278079>
- [12] F. Yavari Nejad, and Varathan, K.D. "Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction" *BMC Medical Informatics and Decision Making*, 21(1), p.141, 2021. <https://doi.org/10.1186/s12911-021-01493-y>
- [13] Q.B. Baker, D. Faraj, and A. Alguzo, "Forecasting dengue fever using machine learning regression techniques", In *2021 12th International Conference on Information and Communication Systems (ICICS)* (pp. 157-163). IEEE, 2021, May. <https://doi.org/10.1109/ICICS52457.2021.9464619>
- [14] Bobir, A.O., Askariy, M., Otabek, Y.Y., Nodir, R.K., Rakhima, A., Zukhra, Z.Y., Sherzod, A.A. (2024). Utilizing Deep Learning and the Internet of Things to Monitor the Health of Aquatic Ecosystems to Conserve Biodiversity. *Natural and Engineering Sciences*, 9(1), 72-83.