Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 2 (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

An Efficient and Hybrid Antlion Optimization-Based Genetic Algorithm Assisted Machine Learning Technique: A Metaheuristic Approach for Dengue Prediction

Mrs. G. Ruba[1], Dr.S. Rizwana[2]

[1]Research Scholar, Department of Computer Science, Erode Arts and Science College, Erode.

Email ID: ruba.mca@gmail.com

[2]Research Supervisor, Assistant Professor and Head, Department of Computer Science (SF), Erode Arts and Science College, Erode.

Email ID: rizureesh@gmail.com

Cite this paper as: Mrs. G. Ruba, Dr.S. Rizwana, (2025) An Efficient and Hybrid Antlion Optimization-Based Genetic Algorithm Assisted Machine Learning Technique: A Metaheuristic Approach for Dengue Prediction. *Journal of Neonatal Surgery*, 14 (2), 198-210

ABSTRACT

Dengue Prediction is a very valuable task in public health, particularly in areas that are most at risk of dengue fever. In this research, an improved the Antlion Optimization Algorithm (ALO) with a Genetic Algorithm (GA) are proposed and applied to improve the predictive accuracy of dengue cases. The precondition for input data preparatory is an elaborate pre-processing pipeline which include handling missing values using mode imputation, removing noise using outlier detection methods and normalizing data. Categorical variables are then encoded While class imbalance problems are addressed either by Synthetic Minority Over-sampling Technique (SMOTE). After that feature selection and extraction processes are fine-tuned using the hybrid ALO-GA approach; adopting ALO for feature selection and GA for tuning the parameters, the proposed technique minimizes overfitting and improve the accuracy of the classifier. The proposed hybrid model is compared with other classifiers under the machine learning domain and good performance improvements are found in areas such as sensitivity, specificity, and F1-score. This metaheuristic approach allows for choosing the right epidemiological and environmental variables for the model making it more accurate.

Keywords: *Optimization, feature selection, dengue fever, over-sampling, synthetic sampling, and accuracy.*

1. INTRODUCTION

Dengue fever being a viral disease attributed to the virus dengue transmitted mostly by the Aedes aegypti and Aedes albopictus mosquitoes is one of the most vital threats to health in the tropical and subtropical world [1]. Symptoms of dengue can be mild like normal flu symptoms to severe like dengue hemorrhagic fever and dengue shock syndrome which are also fatal when not well managed. The World Health Organization states that the flavivirus every year infects millions of people and incidence tends to increase because of factors like urbanization, changes in climate and population growth [2]. Because dengue is perceived to affect many people and spreads quickly, accurate and timely prediction of the disease has emerged as a critical concern. When these agencies can predict such incidences, they can use data to proactively control the spread of mosquitos, begin treatment, and spread information to discourage the virus [3, 4].

The dengue is dynamic in its interaction with climatic factors, epidemiological and demographic ones making prediction inefficient and difficult. Dengue fever forecast models are mainly dependent on the past record of dengue cases and symptoms, demographic factors and characteristics, and climate factors namely temperature, rainfall, and humidity. These models are expected to detect factors that relate to high risk of dengue transmission in the data [5]. Statistical regression models are more frequently used in dengue prediction in accordance with traditional epidemiological models; however, they are also deficient in providing adequate information about non-linear connections and interactions between variables. In turn, Machine learning (ML) methods are more permissive in terms of volume of input data and capability of finding irregularities that cannot be observed with classical methods [6]. However, the performance of ML models highly relies on the process of data pre-processing, feature selection and hyperparameters optimization all these steps must use optimization techniques to achieve the most accurate outcomes [7].

Machine learning has been used in a very promising way for improving the dengue prediction because it offers models that could be trained out of past data and used for future prediction. Decision Trees, Support Vector Machines (SVM) as well as

Neural Networks have been used in the ML approach to predict and classify dengue fever to reasonable accuracy. However, with high dimensional and complicated dengue related data, the application of these models often demands fine-tuning of the strategies for feature extraction, data sampling and parameter optimization for the best results [8].

GA, PSO and ALO optimization algorithms have been used in dengue prediction models to enhance the accuracy of their prediction. Specifically, the use of several algorithms in which the strengths of various algorithms are combined holds considerable potential. In this study, we formulate a compound optimization strategy that acknowledges the ALO combined with GA to assist in ML-based dengue prediction [9]. ALO is outstandingly good at exploration and exploitation when it comes to selecting features from high-dimensional data. By integrating ALO with GA, which effectively search the solution space for the best hyperparameters, research can obtain a new model that increases the predictive accuracy, minimizes overfitting, and increases the model's ability to generalize. This makes the two improvements help in the selection of features and the optimization of model parameters as seen below as they complement each other since the individual methods have their shortcomings [10].

There is a rising caseload of dengue and currently, the demand for better predictions that would help prevent the disease is also high [11, 12]. The existing models of epidemiology have some drawbacks since they presuppose rather strict assumptions and do not include certain parameters in the situation and demographics. Some methods coming from the machine learning area only, are flexible, but they have problems with feature selection, with the parameters setting and with computational complexity [13, 14].

The integration of metaheuristic optimization methods with machine learning can be an interesting solution in this context as it will enhance model accuracy and interpretability [15]. Because ALO can search spaces optimally and GA can further adjust the parameters of the model, we think that combining both ALO and GA will improve dengue prognosis enormously. We plan to integrate these algorithms to create a model that will overcome the problem observed with feature dimensionality, data imbalance and overfitting to ensure that an accurate, reliable, and efficient framework for dengue prediction is developed.

In this paper, an ALO-GA is proposed to develop a more effective model for accurate dengue prediction. The introduction focuses on the importance of making accurate forecasts, the related works describe the existing approach and reveal the current gaps. This paper outlines the proposed methodology focusing on data pre-processing and optimization technique know as ALO-GA. Results and discussion present findings of model evaluation by various criteria including accuracy and sensitivity and demonstrate how ALO-GA works. The conclusion gives an overview of research outcomes, their relevance to public health and future research recommendations.

2. RELATED WORKS

Dengue fever is a highly endemic mosquito-borne disease and is causing a burdensome to global health, most especially in the tropical and subtropical world regions. Due to interactions between various environmental and epidemiological factors such as increased rate of urbanization, climatic change, and inefficient control of the mosquito vector, it becomes difficult to forecast dengue outbreaks. Conventional approaches for predicting the incidence of dengue cases have mainly relied on extrapolation methods by means of statistical time series analysis like the ARIMA. However, these methods do not allow for the analysis of detailed, nonlinear dynamics that characterize the dengue disease transmission. In the recent past, the application of machine learning (ML) and deep learning (DL) has been considered due to their capabilities in modeling complex data set structures which will aid in making more accurate predictive models of outbreaks, thereby enhancing operation of health interventions. To improve the accuracy of the prognosis, the recent investigations have investigated the integration of ML and DL with augmented optimization and data elements. These advancements point to further improved multi-disciplinary methods as a means of achieving a more effective means of predicting dengue. The described papers illustrate how different methods work with the focus on certain aspects of dengue prediction aiming at enhancing the models' accuracy, promptness, and stalwartness. The comprehensive analysis of dengue fever prediction is given in Table 1.

Table 1. The Comprehensive Analysis of Dengue Fever Prediction

Author Name and Year	Methodology Used	Inference	Significance
Akter, T., Islam, M. T., Hossain, M. F., & Ullah, M. S. (2024) [16]	Neural Networks and ARIMA time series analysis	Neural networks outperform ARIMA in predicting dengue fever with the lowest MAPE.	Machine learning, specifically neural networks, is more effective for dengue prediction in Dhaka, providing better forecasting performance than traditional time series analysis.

Mumtaz, Z., Rashid, Z., Saif, R., & Yousaf, M. Z. (2024) [17]	Deep Learning (LSTM and FNN)	The DL-DVE model effectively predicts evolving DENV serotypes with high accuracy and classification performance.	The model demonstrates the importance of deep learning in forecasting emerging DENV serotypes, aiding in better understanding and controlling dengue outbreaks.
Kuo, C. Y., Yang, W. W., & Su, E. C. Y. (2024) [18]	Random Forests with Feature Selection	Incorporating AQIs improves dengue fever prediction in Taiwan.	The study introduces the novel use of air quality indices in dengue prediction, providing a new dimension to public health interventions.
Gupta, G., Khan, S., Guleria, V., Almjally, A., Alabduallah, B. I., Siddiqui, T., & Al-Subaie, M. (2023) [19]	Sentiment Analysis and Machine Learning	Proposed a machine learning method for early-stage dengue fever prediction using sentiment analysis.	The model offers an innovative approach by integrating sentiment analysis with machine learning for early dengue diagnosis, improving the accuracy of early-stage prediction.
Manoharan, S. N., Kumar, K. M., & Vadivelan, N. (2023) [20]	Hybrid CNN-TLSTM with IoT-Fog Cloud Architecture	The ATLBO optimized CNN-TLSTM model effectively diagnoses dengue with high accuracy and precision.	The approach integrates IoT and fog-cloud architecture for timely disease diagnosis and emergency prevention, enhancing public health response capabilities.
Majeed, M. A., Shafri, H. Z. M., Zulkafli, Z., & Wayayok, A. (2023) [21]	LSTM with Spatial Attention	SSA-LSTM model outperforms traditional models in predicting dengue cases in Malaysia.	The use of spatial attention in LSTM models significantly improves dengue prediction, offering a robust tool for managing dengue outbreaks in diverse geographic regions.

Thus, by focusing on the most recent investigations of dengue prediction, important aspects of ML and DL approaches remain under-optimized. Still, as in Kuo et al. (2024), using feature selection enhances the performance of the models, and at the same time, most of the proposed approaches disregard the use of the concurrent environmental data including AQI and other related indices of different regions. The studies together imply the use of neural networks, feature selection, spatial attention and real time environmental conditions for more efficient and scalable dengue prediction models.

3. PROPOSED METHODOLOGY - HYBRID ANTLION OPTIMIZATION-BASED GENETIC ALGORITHM ASSISTED MACHINE LEARNING TECHNIQUE

The research proposes the integration of a Machine learning model with two algorithms namely the Antlion Optimization Algorithm (ALO) and the Genetic Algorithm (GA) to enhance accuracy of dengue disease prediction. The following steps are taken in the methodological approach: data pre-processing, feature selection, parameter optimization, model building.

Data pre-processing is crucial for producing good quality inputs that in turn have a direct relation to model efficiency. In both training and test datasets, the dataset consists of missing values in different features. Null values are handled through mode imputation where missing values are replaced by the common value of the feature in an efficient way without much distortions. The optional parameters as reanalysis_tdtr_k, reanalysis_specific_humidity_g_per_kg and station_avg_temp_c contain up to 43 missing values, while the ndvi_ne contains 194 missing values. The test set also has missing values with 12 and 43 missing values in station_avg_temp_c and ndvi_ne, respectively.

Given a dataset $D = \{(X_i, y_i)\}_{i=1}^{n}$ where X_i represents the feature vector of the i-th instance and y_i denotes its corresponding dengue occurrence label (binary or continuous value, depending on the target), the goal is to develop a predictive model f such that:

$$f(X) \approx y$$

Our aim is to maximize the predictive accuracy of f by selecting the optimal features $X^{opt} \subset$ X and tuning parameters θ that define the model structure.

In the process of z-score analysis, outliers that can hinder the appraisal of the model is eliminated. This technique aids in the control of the quality of the data by leaving out such outliers which actually are an exaggeration of the trends. In addition, all features are scaled by normalizing the continuous variables in order to make sure that all of them would lie in the similar ranges and make the model converge well. Some of the procedures like Min-Max scaling and Z-score normalization are used to scale the features to equal range. Let $X_{miss} \subset X$ represent features with missing values. We apply mode imputation on each feature $x_j \in X_{miss}$ as:

$$x_j = mode(x_j) \qquad if\ x_j\ is\ missing$$

Using **Z-score determine** outliers as values x_{ij} that satisfy:

$$outlier(x_{ij}) = \begin{cases} x_{ij} > \mu_j + k\mu_j \\ x_{ij} < \mu_j - k\sigma_j \end{cases}$$

where μ_j and μ_j are the mean and standard deviation of feature x_{ij}, respectively, and k is a threshold (typically 3).

Apply Min-Max scaling or Z-score normalization for each feature x_i:

$$x_j^{norm} = \frac{x_j - \min(x_j)}{\max(x_j) - \min(x_j)}$$

Categorical features are encoded using one hot encoding or label encoding which helps to convert the non numerical data into machine learning acceptable format. Thus, the encoded variables contain all unique categories while the data is ready for model training. Dengue datasets include variance in terms of the numbers of infected and unaffected populations. To overcome with this Synthetic Minority Over-sampling Technique (SMOTE) is used to balance the dataset artificially. These techniques produce samples in the minority class for better generalization of the model without prejudice to other classes. Using SMOTE and ADASYN, synthetic samples for minority class $y_{minority}$ are generated to balance class distribution:

$$X_{balanced} = SMOTE(X, y)$$

The feature selection is useful to prevent the model from considering irrelevant features, as well as to improve the efficiency and accuracy of the resulting model. The hybrid feature selection and extraction model, ALO-GA, is proposed for the feature selection and extraction. ALO is a metaheuristic optimization algorithm been derived from the hunting techniques of the antlions. It shows great exploration and exploitation properties which makes it suitable for feature selection tasks. In this methodology, ALO finds out which features are more important when a feature subset effects on the prediction model in terms of the performance measure of a prediction task. This step of the process greatly diminishes the feature space with only the most important features being chosen which serves to minimize computational cost and enhance results of the model.

Determine the feature selection process as an optimization problem with $f_{fitness}$ measuring the model's predictive performance (e.g., accuracy) on a given subset of features. For each iteration t, update the feature vector $X^{(t)}$ by modeling the position update of antlions in the search space:

$$X^{(t+1)} = X^t + r \times (X_{antlion}^{(t)} - X_{prey}^{(t)})$$

where r is a random walk variable influenced by the antlion position $X_{antlion}$ and prey position X_{prey}.

The generalization of Adapative optimization algorithm, GA is introduced to further tune the parameters of the machine learning model. In this hybrid framework, GA works next to ALO for improving deep learning such as learning rate, numbers of layer, and batch size. This procedure of selection, crossover and mutation operations GA search for configuration of parameters that improve the performance of the model. This framework is even better than ALO or GA alone because the overfitting issue is solved plus the generalization ability is enhanced when generalized from one dataset to another.

In this step, the Genetic Algorithm (GA) optimizes the parameters θ of the predictive model to achieve high predictive accuracy. GA evolves a population of candidate solutions (parameter sets) by iteratively applying selection, crossover, and mutation processes. A population P of N parameter sets $\{\theta_1, \theta_2, \ldots, \theta_N\}$ is initialized randomly. Each θ_i is a vector of parameters, such as the learning rate, the number of layers, or other hyperparameters that define the model structure. Each parameter set θ_i is evaluated using a fitness function $f_{fitness}(\theta i)$, which measures the model's performance, often based on validation accuracy or error. Formally, for each parameter set θ_i :

$$f_{fitness}(\theta_i) = validation\ accuracy\ \theta_i$$

The objective is to maximize $f_{fitness}(\theta_i)$, which represents how well the parameter set θ_i enables the model to predict the target variable accurately.

The selection process involves choosing the best-performing parameter sets from the population based on their fitness scores. A common method is **tournament selection** or **roulette wheel selection**. For simplicity, in this context, let's assume the top M parameter sets with the highest fitness scores are selected to form a mating pool:

$$Mating\ Pool = \{\theta_{(1)}, \theta_{(2)}, \dots, \theta_{(N)}\}$$

where $\theta_{(1)}$ to $\theta_{(M)}$ are the top M parameter sets ordered by their fitness values such that $f_{fitness}(\theta_{(1)}) \geq f_{fitness}(\theta_{(2)}) \dots \dots > f_{fitness}(\theta_{(M)})$

Crossover combines two parameter sets (parents) to produce new parameter sets (offspring), encouraging exploration of the parameter space. For example, for two parent parameter sets θ_p and θ_q from the mating pool, crossover produces a new parameter set θ' as:

$$\theta' = \alpha\theta_p + (1-\alpha)\theta_q$$

where α is a crossover coefficient (e.g., α=0.5) that determines the contribution of each parent to the offspring. Multiple crossovers generate a new population, preserving diversity and increasing the chance of finding optimal parameters.

Mutation introduces small random changes to parameters to maintain diversity and avoid premature convergence. For a parameter θ_j in θ', mutation is applied as:

$$\theta_j' = \theta_j + \delta$$

where $\delta \sim N(0, \sigma^2)$ is a small random noise sampled from a Gaussian distribution with mean 0 and variance σ^2. This process explores neighboring solutions in the parameter space, which can lead to better solutions.

The GA process of selection, crossover, and mutation is repeated for a set number of generations (iterations), with each generation producing a new population that ideally has higher fitness on average. After sufficient iterations, the algorithm converges to an optimal parameter set θ_{opt}. At the end of the GA process, the best-performing parameter set θ_{opt} is selected:

$$\theta^{opt} = arg \max_{\theta \epsilon P} f_{fitness}(\theta)$$

When some features are chosen and the parameters are adjusted, the SVM and NN model are built on the given dataset. These algorithms, designed to work with time series data and interactions between the features, are well suited for dengue prediction considering that dengue relies heavily on environmental and time series data patterns. The final model is developed in two stages, using feature selection in ALO-GA and training on the processed data set. In addition to evaluating the internal consistency, the final model is checked for generalization performance through validation and testing.

Train the predictive model f using the optimized features and parameters:

$$f(X^{opt}, \theta^{opt}) \rightarrow y$$

with f being a machine learning model (e.g., SVM, NN) trained to minimize the error:

$$E = |y - f(X^{opt}, \theta^{opt})|$$

The final objective is to minimize prediction error, formalized as:

$$\min_{\theta, X^{opt}} E = \min_{\theta, X^{opt}} \frac{1}{n} \sum_{i=1}^{n} |y_i - f(X_i^{opt}, \theta)|$$

subject to the constraints defined by ALO-GA for feature and parameter selection.

Like the training data, the test dataset also has missing values and has similar features as the training data, pre-processing of the test data is done in exactly the same manner as that of the training data. This unseen data gives an insight into the stability and efficiency in real life situations of the constructed model. The methodology to be proposed, ALO-GA assisted analysis, combines a systematic preprocessing step with a complex hybrid optimization scheme suitable for high-dimensional and imbalanced dengue datasets. This has been done effectively by using ALO for feature selection thus; GA for parameter tuning thus enhancing prediction and minimizing over fitting. This pipeline offers all-in-one, large-scale dengue prediction, which, in turn, could be supportive of the overall dengue prevention in case of accurate outbreak prediction.

4. RESULT AND DISCUSSION

The dataset in this study has 1872 samples and 25 predictors and includes meteorological and geographical features. Looking at the training dataset, missing values are most common in `ndvi_ne` with 194, `station_avg_temp_c` with 43, and `ndvi_nw` with 52 missing values; other features such as `precipitation_amt_mm` and `station_min_temp_c` are missing 10 to 22 values. Similar to the main dataset, the test dataset also contains missing values, where there are 43 missing values in 'ndvi_ne' & 12 missing values in 'station_diur_temp_rng_c'. Even in the data pre-processing stage, null values are dealt with through imputation and outliers to make data good for modeling. The feature correlation is illustrated in Figure 1.

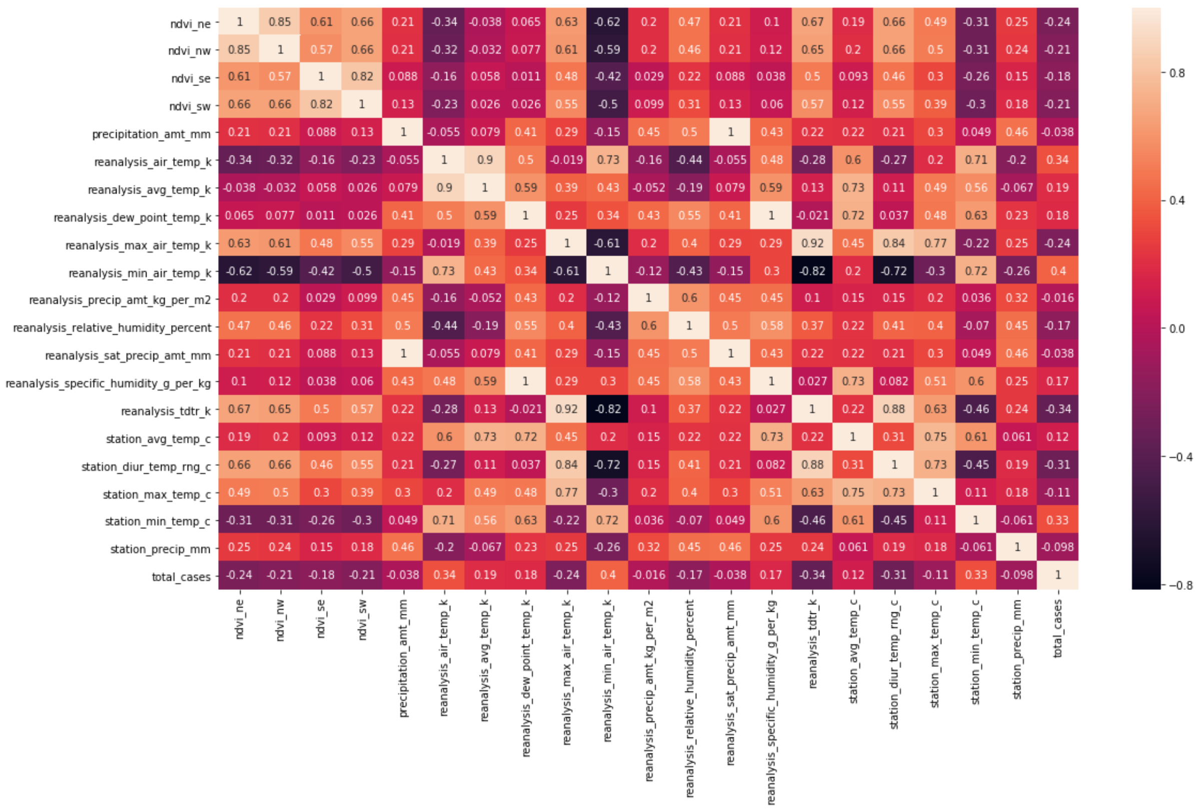

Figure 1. Correlation Matrix

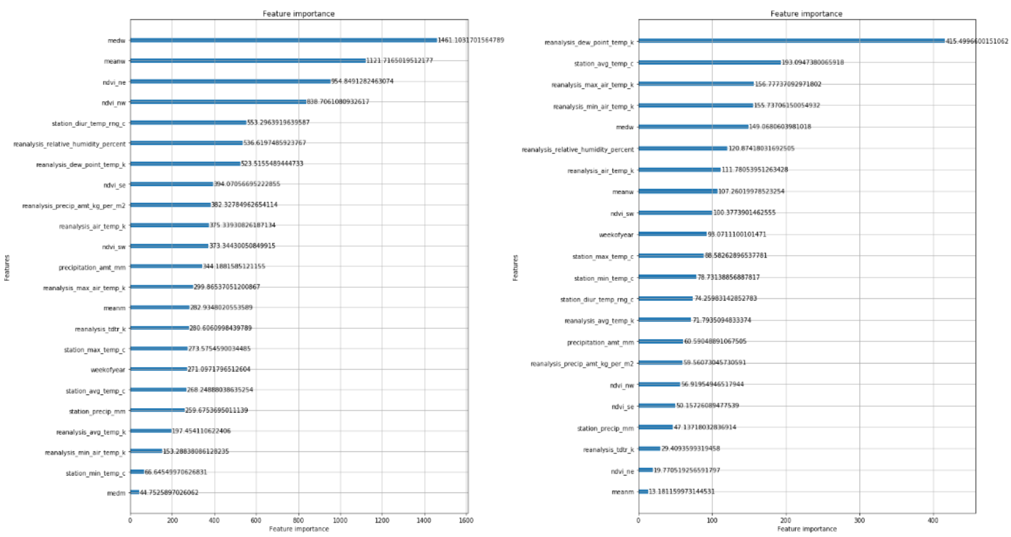

Figure 2. Feature Importance

Performance measures are crucial ingredients in the assessment of accuracy of the models used in the prediction of dengue fever. Classification problems use basic measures such as accuracy, sensitivity/recall, specificity, and F1-score. Accuracy measures the percent of all instance to which the model assigned the correct class. Recall also known as sensitivity refers to the ability of the model to identify the actual number of positive cases it was trained on. Specificity reflects the percentage of false negatives, within the total number of negative instances. F1-score is the average of precision and recall, and therefore is the harmonic mean of the two. These metrics provide a good assessment of the model, more so in cases where the data it is tested on is unbalanced.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Specificty = \frac{TN}{TN + FP}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Table 2. Comparison of Accuracy

Iteration	SVM	ALO-GA-SVM	NN	ALO-GA-NN
200	87	91	88	92
400	87.5	91.5	88.5	93.1
600	88	91.8	89	93.8
800	89	92	89.5	94.9
1000	89.5	92.5	90	94

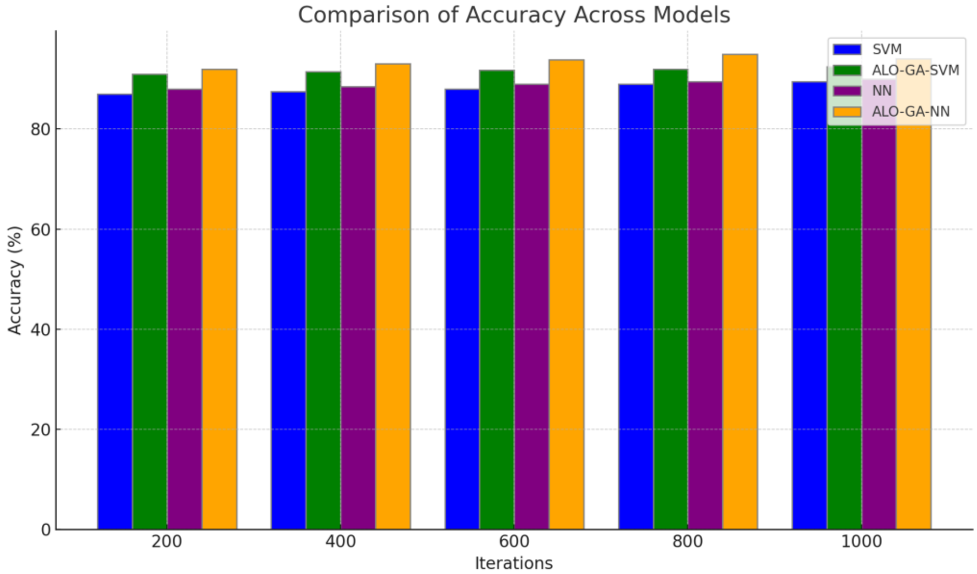

Figure 3. Comparison of Accuracy

The accuracy comparison between the standard machine learning models (SVM and NN) and the optimized models (ALO-GA-SVM and ALO-GA-NN) demonstrates a significant improvement with the incorporation of the hybrid ALO-GA optimization approach. For SVM, accuracy increases from 87% at 200 iterations to 89.5% at 1000 iterations. ALO-GA-SVM shows a consistent increase, achieving 92.5% accuracy at 1000 iterations. Similarly, for neural networks (NN), accuracy improves from 88% at 200 iterations to 90% at 1000 iterations, while ALO-GA-NN sees a remarkable increase from 92% at 200 iterations to 94% at 1000 iterations. This indicates that the ALO-GA-based optimization significantly enhances both SVM and NN models, improving their predictive performance and robustness over time.

Table 3. Comparison of Recall

Iteration	SVM	ALO-GA-SVM	NN	ALO-GA-NN
200	81.5	87	82	89
400	82	87.2	82.7	88.7
600	82.6	87.5	83	90

800	82.9	88	83.5	90.1
1000	83	88.8	84	90.3

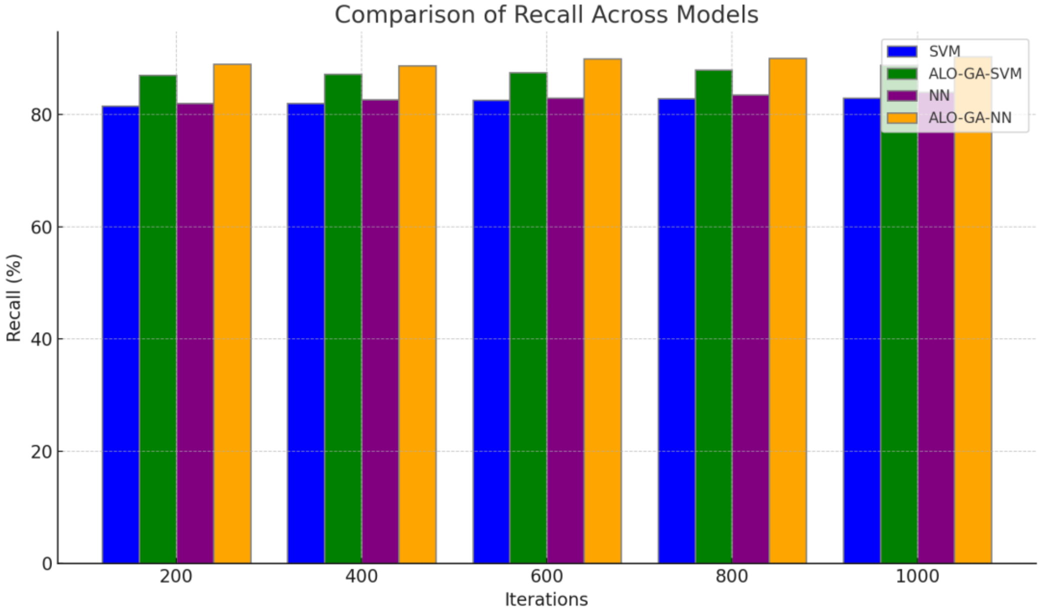

Figure 4. Comparison of Recall

Recall, also known as sensitivity or true positive rate, measures the ability of a model to correctly identify positive instances. For SVM, recall increases steadily from 81.5% at 200 iterations to 83% at 1000 iterations. ALO-GA-SVM improves recall significantly, starting at 87% at 200 iterations and reaching 88.8% at 1000 iterations. In the case of neural networks (NN), recall improves from 82% at 200 iterations to 84% at 1000 iterations, while ALO-GA-NN achieves an impressive increase from 89% at 200 iterations to 90.3% at 1000 iterations. This improvement in recall for the optimized models highlights their ability to better detect positive instances, which is crucial in applications like dengue prediction, where identifying true positive cases is essential for timely intervention.

Table 4. Comparison of Specificity

Iteration	SVM	ALO-GA-SVM	NN	ALO-GA-NN
200	82.1	87.3	82	89.5
400	82.4	87.8	83	89.7
600	82.9	87.9	83.2	89.9
800	83	88.5	83.6	90
1000	83.4	88.9	83.9	90.5

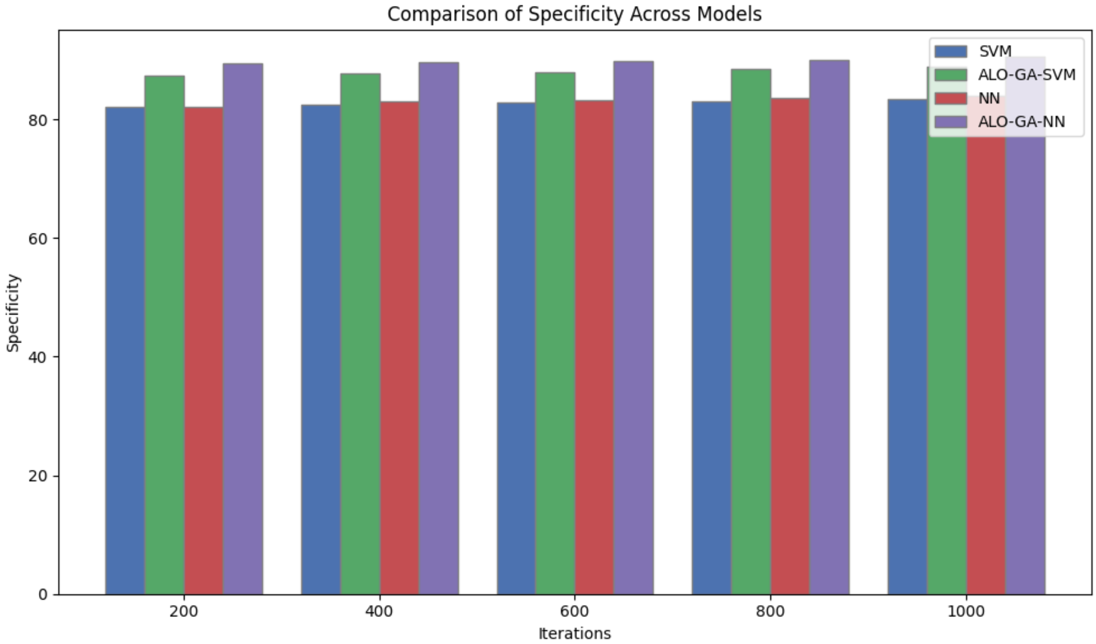

Figure 5. Comparison of Specificity

Specificity measures the proportion of true negatives correctly identified by the model. The comparison of specificity across models shows that ALO-GA-based optimization techniques enhance specificity in both SVM and NN models. For SVM, specificity increases from 82.1% at 200 iterations to 83.4% at 1000 iterations, whereas ALO-GA-SVM improves from 87.3% to 88.9% over the same iterations. Neural networks (NN) show an increase in specificity from 82% at 200 iterations to 83.9% at 1000 iterations, with ALO-GA-NN reaching 90.5% at 1000 iterations, compared to 89.5% at 200 iterations. This shows that the hybrid ALO-GA approach improves the model's ability to correctly identify negative instances, reducing false positives, and enhancing the model's precision in the predictive tasks.

Table 5. Comparison of F1-Score

Iteration	SVM	ALO-GA-SVM	NN	ALO-GA-NN
200	0.456	0.71	0.51	0.75
400	0.562	0.745	0.59	0.81
600	0.67	0.763	0.63	0.86
800	0.68	0.789	0.69	0.91
1000	0.69	0.82	0.71	0.93

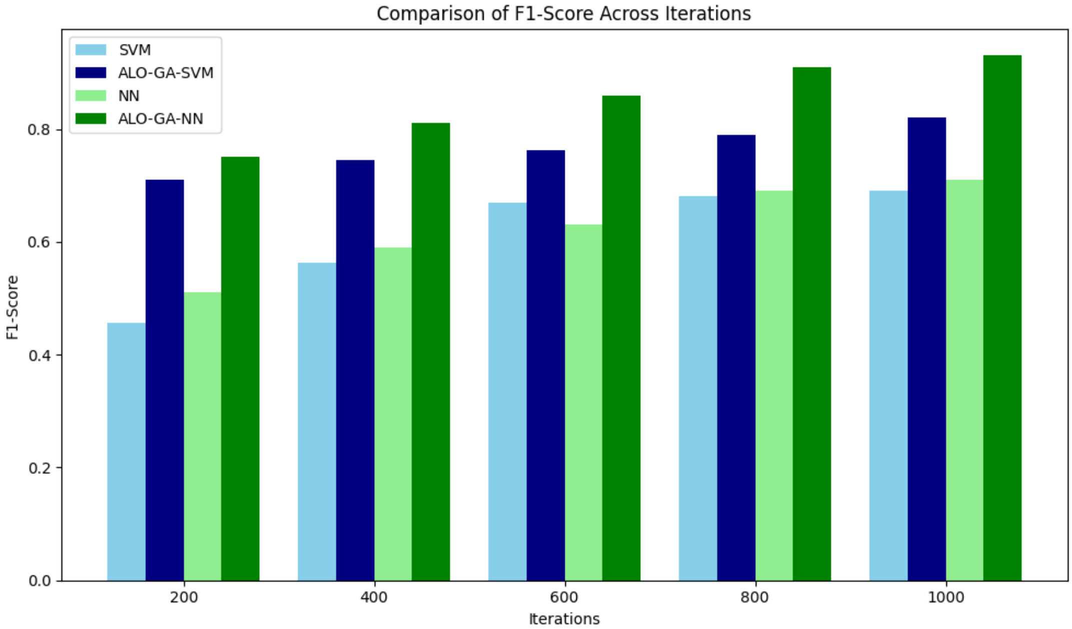

Figure 6. Comparison of F1-Score

The F1-Score, which is the harmonic mean of precision and recall, serves as an overall measure of a model's predictive performance. For the SVM model, the F1-score starts at 0.456 at 200 iterations and increases to 0.69 at 1000 iterations. ALO-GA-SVM demonstrates a steady improvement, with the F1-score rising from 0.71 at 200 iterations to 0.82 at 1000 iterations. Neural networks (NN) show a lower increase, from 0.51 at 200 iterations to 0.71 at 1000 iterations. However, ALO-GA-NN exhibits a significant improvement, with the F1-score increasing from 0.75 at 200 iterations to 0.93 at 1000 iterations. The F1-score analysis highlights that the hybrid ALO-GA optimization significantly enhances the balance between precision and recall, offering a more robust and reliable prediction, especially in the context of dengue fever prediction where accurate and balanced classification is crucial.

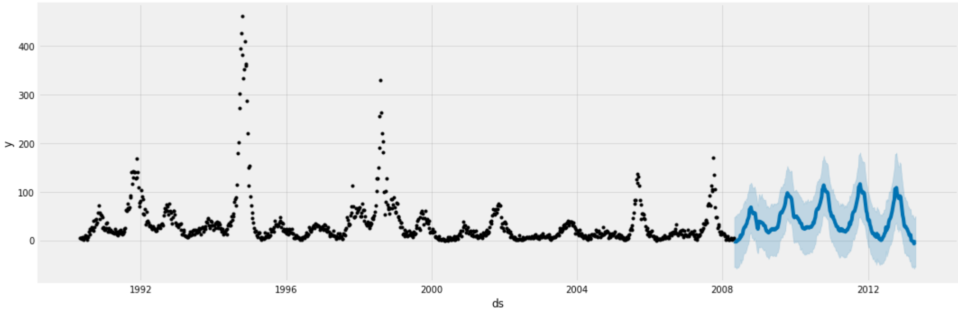

Figure 7. Dengue Prediction

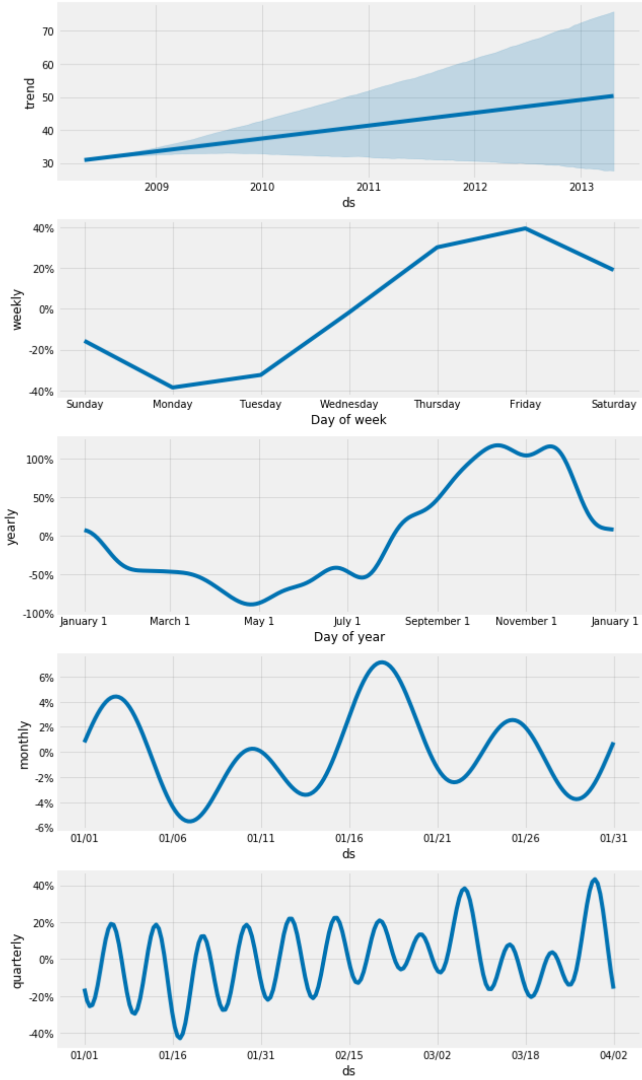

Figure 8. Diverse Period Dengue Prediction

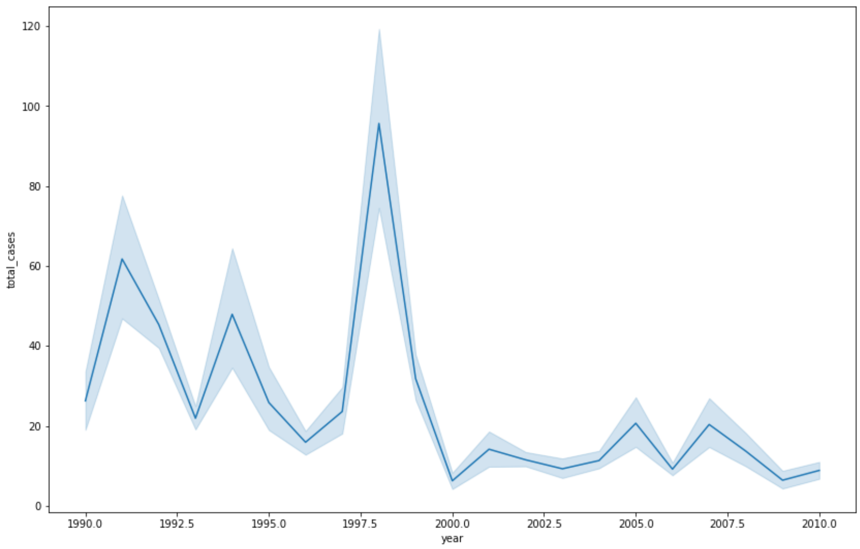

Figure 9. Distribution of Total Cases

5. CONCLUSION

The ALO-GA hybrid model used in this research for feature selection and parameter optimization for the prediction of dengue fever is proposed. Data pre-processing to handle missing data, noise and class imbalance is the initial phase of the proposed methodology followed by feature optimization, parameters using ALO & GA. The ALO-GA approach also minimises over fitting that is commonly encountered in most high-dimensional databases and increases the predictive power since the right features are selected and right parameters tuned. The proposed technique optimizes dengue prediction models and thus an asset in dengue monitoring and control. The ability to perform well on large, noisy datasets, and the capacity to predict the spread of a disease is a benefit to disease prediction and serves as a framework for future work in other epidemiological fields and vector-borne diseases.

REFERENCES

[1] Setyawan, A. F., Ariyanto, A. D. P., & Fikriah, F. K. (2024). CLASSIFICATION OF DENGUE FEVER DISEASE USING A MACHINE LEARNING-BASED RANDOM FOREST ALGORITHM. JIKO (Jurnal Informatika dan Komputer), 7(2), 88-92.

[2] Van Duong, C., Tran, U. T. P., Van Nguyen, V., & Bae, Y. J. (2021). Predator selection and predator-prey interactions for the biological control of mosquito dengue vectors in northern Vietnam. Journal of Vector Ecology, 46(2), 163-172.

[3] Jitonnom, J., Meelua, W., Tue-Nguen, P., Saparpakorn, P., Hannongbua, S., & Chotpatiwetchkul, W. (2024). 3D-QSAR and molecular docking studies of peptide-hybrids as dengue virus NS2B/NS3 protease inhibitors. Chemico-Biological Interactions, 396, 111040.

[4] Sarma, D., Hossain, S., Mittra, T., Bhuiya, M. A. M., Saha, I., & Chakma, R. (2020, December). Dengue prediction using machine learning algorithms. In 2020 IEEE 8th R10 humanitarian technology conference (R10-HTC) (pp. 1-6). IEEE.

[5] Dourjoy, S. M. K., Rafi, A. M. G. R., Tumpa, Z. N., & Saifuzzaman, M. (2021). A comparative study on prediction of dengue fever using machine learning algorithm. In Advances in Distributed Computing and Machine Learning: Proceedings of ICADCML 2020 (pp. 501-510). Springer Singapore.

[6] Salim, N. A. M., Wah, Y. B., Reeves, C., Smith, M., Yaacob, W. F. W., Mudin, R. N., ... & Haque, U. (2021). Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. Scientific reports, 11(1), 939.

[7] Majeed, M. A., Shafri, H. Z. M., Zulkafli, Z., & Wayayok, A. (2023). A deep learning approach for dengue fever prediction in Malaysia using LSTM with spatial attention. International journal of environmental research and public health, 20(5), 4130.

[8] Gupta, G., Khan, S., Guleria, V., Almjally, A., Alabduallah, B. I., Siddiqui, T., ... & Al-Subaie, M. (2023). DDPM: A dengue disease prediction and diagnosis model using sentiment analysis and machine learning algorithms. Diagnostics, 13(6), 1093.

[9] Nguyen, V. H., Tuyet-Hanh, T. T., Mulhall, J., Minh, H. V., Duong, T. Q., Chien, N. V., ... & Son, M. T. (2022). Deep learning models for forecasting dengue fever based on climate data in Vietnam. PLoS Neglected Tropical Diseases, 16(6), e0010509.

[10] Cabrera, M., Leake, J., Naranjo-Torres, J., Valero, N., Cabrera, J. C., & Rodríguez-Morales, A. J. (2022). Dengue prediction in Latin America using machine learning and the one health perspective: a literature review. Tropical Medicine and Infectious Disease, 7(10), 322.

[11] Gangula, R., Thirupathi, L., Parupati, R., Sreeveda, K., & Gattoju, S. (2023). Ensemble machine learning based prediction of dengue disease with performance and accuracy elevation patterns. Materials Today: Proceedings, 80, 3458-3463.

[12] Roster, K., Connaughton, C., & Rodrigues, F. A. (2022). Machine-learning–based forecasting of dengue fever in Brazilian cities using epidemiologic and meteorological variables. American Journal of Epidemiology, 191(10), 1803-1812.

[13] Yavari Nejad, F., & Varathan, K. D. (2021). Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction. BMC Medical Informatics and Decision Making, 21(1), 141.

[14] Jeevika Tharini, V., & Vijayarani, S. (2020). Bio-inspired High-Utility Item Framework based Particle Swarm Optimization Tree Algorithms for Mining High Utility Itemset. In Advances in Computational Intelligence and Informatics: Proceedings of ICACII 2019 (pp. 265-276). Springer Singapore.

[15] Dey, S. K., Rahman, M. M., Howlader, A., Siddiqi, U. R., Uddin, K. M. M., Borhan, R., & Rahman, E. U. (2022). Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in

Bangladesh: A machine learning approach. PloS one, 17(7), e0270933.

[16] Akter, T., Islam, M. T., Hossain, M. F., & Ullah, M. S. (2024). A Comparative Study between Time Series and Machine Learning Technique to Predict Dengue Fever in Dhaka City. Discrete Dynamics in Nature and Society, 2024(1), 2757381.

[17] Mumtaz, Z., Rashid, Z., Saif, R., & Yousaf, M. Z. (2024). Deep learning guided prediction modeling of dengue virus evolving serotype. Heliyon.

[18] Kuo, C. Y., Yang, W. W., & Su, E. C. Y. (2024). Improving dengue fever predictions in Taiwan based on feature selection and random forests. BMC Infectious Diseases, 24(Suppl 2), 334.

[19] Gupta, G., Khan, S., Guleria, V., Almjally, A., Alabduallah, B. I., Siddiqui, T., ... & Al-Subaie, M. (2023). DDPM: A dengue disease prediction and diagnosis model using sentiment analysis and machine learning algorithms. Diagnostics, 13(6), 1093.

[20] Manoharan, S. N., Kumar, K. M., & Vadivelan, N. (2023). A novel CNN-TLSTM approach for dengue disease identification and prevention using IoT-fog cloud architecture. Neural Processing Letters, 55(2), 1951-1973.

[21] Majeed, M. A., Shafri, H. Z. M., Zulkafli, Z., & Wayayok, A. (2023). A deep learning approach for dengue fever prediction in Malaysia using LSTM with spatial attention. International journal of environmental research and public health, 20(5), 4130.