Journal of Theoretical and Applied Information Technology
15th December 2025. Vol.103. No.23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ENHANCED DENGUE FORECASTING WITH ANC-DEFO: A HYBRID FEATURE OPTIMIZATION FRAMEWORK

P.MUTHUSELVI[1], DR.S.DHANASEKARAN [2], DR.V.VASUDEVAN [3]

[1]Research Scholar, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankovil, Tamilnadu , India
[2]Associate Professor, Department of Information Technology, Kalasalingam Academy of Research and Education, Krishnankovil, Tamilnadu , India
[3]Senior Professor, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankovil, Tamilnadu , India

Email ID: [1]sukirthanagarajan@gmail.com, [2]srividhans@gmail.com, [3]Vasudevan_klu@yahoo.co.in

ABSTRACT

Dengue fever remains a persistent and growing public health challenge, especially in tropical and subtropical regions where it causes significant morbidity and mortality. The complexity and seasonality of dengue transmission necessitate the development of robust and accurate prediction models for early intervention and resource planning. However, the predictive performance of traditional machine learning models often suffers due to high-dimensional data, irrelevant features, and overfitting. This study introduces ANC-DEFO, an Adaptive Neuro-Classifying model integrated with Differential Evolution and Fuzzy Optimization, to enhance the accuracy and efficiency of dengue outbreak prediction. The proposed framework employs intelligent dynamic feature selection to isolate the most relevant environmental and temporal features, thereby improving model generalization and reducing computational cost. Experimental evaluation was conducted using real-world seasonal dengue data from Tamil Nadu, India. The performance of ANC-DEFO was compared against conventional models such as LSTM, Random Forest, SVM, and Logistic Regression. The results indicate that ANC-DEFO achieved a prediction accuracy of 95.5% and a significantly lower RMSE (2.10) compared to LSTM (accuracy 86.9%, RMSE 4.97), demonstrating its superior capability in handling noisy and nonlinear epidemiological data. The model provides a scalable framework for integration into a system for realtime health surveillance and can be modified for other climate-sensitive or vector-borne diseases. By facilitating early outbreak detection, ANC-DEFO shows potential for improving public health readiness and epidemic response strategic planning.
Keywords: *Dengue Prediction, Machine Learning, Feature Selection, Differential Evolution, ANC-DEFO, Public Health.*

1. INTRODUCTION

Dengue fever remains a major global public health concern, particularly in tropical and subtropical regions where climatic and environmental factors favor mosquito breeding. The World Health Organization (WHO) reports that more than 400 million infections occur every year, with India among the most severely affected countries. Early and accurate prediction of dengue outbreaks is essential to support timely health interventions, efficient resource distribution, and effective disease management. In recent years, **machine learning (ML)** and **deep learning (DL)** techniques have been widely explored for dengue prediction. Models such as Support Vector Machines (SVM), Random Forests (RF), and Long Short-Term Memory networks (LSTM) have shown potential in identifying disease patterns using climatic and epidemiological variables. However, the accuracy and stability of these models are often limited by **high-dimensional datasets**, **irrelevant features**, and **overfitting**, leading to weak generalization in real-world scenarios. Furthermore, most conventional models fail to capture the nonlinear and uncertain relationships among environmental and temporal factors influencing dengue transmission. To address these challenges, this study introduces **ANC-DEFO (Adaptive Neuro-Classifying Differential Evolution and Fuzzy Optimization)**, a hybrid forecasting model that combines neuro-classification with differential evolution–based feature optimization and fuzzy adaptive learning. This integration enables dynamic identification of significant features, reduction of noise, and adaptive learning for improved forecasting accuracy.

Dengue fever is an illness, which is carried by a virus and is a growing threat for health all around the world, often among the people living in tropical or subtropical climates. The absence of specific antiviral treatment options and a limited availability of vaccines for the virus mean that prediction and timely detection are crucial for good disease management. Adequate forecasting of the disease helps the health authorities in controlling the disease in time, managing the limited health resources, and controlling the social, as well as, economic impacts of the disease. Many artificial intelligent techniques, mainly in the field of machine learning, are now capable of capturing the complex temporal and spatial relationships in epidemiological datasets. These techniques can detect subtle temporal and spatial relations which are often overlooked by other statistical approaches. However, the predictive capability in prediction models is highly dependent on the quality and relevance of the input features. Often, high dimensional data is compromised with unnecessary features, which in turn make the model complicated, lengthen the time needed for training, and even lower the performance of the model.

To overcome these challenges, it is essential to adopt more sophisticated the feature selection processes that aid in improving the interpretability and operationalOutput efficiency. In this case, we propose the ANC-DEFO framework: a model that integrates adaptive neuro-classification with differential evolution feature optimization to perform advanced feature selection for enhanced accuracy. This framework serves the public health systems of the region by forecasting dengue outbreaks more reliably and flexibly. In resolving these public health problems, the techniques have been essential for improving the efficiency and precision of predictive public health tools. This study introduces ANC-DEFO, Adaptive Nature-Inspired Classifier with Differential Evolution and Feature Optimization, a hybrid model designed for dengue prediction with advanced feature selection and robust classification. The model implementing what we describe as 'intelligent classification mechanism' with Differential Evolution that selects features through adaptive feature pruning which raises accuracy. The purpose of this study is to evaluate the accuracy of the predictions of ANC-DEFO relative to more conventional models like SVM, RF, and LR.

1.1 Advantages of the Proposed Model

The anticipated ANC-DEFO model enhances dengue forecasting with predictive accuracy through advanced ANC-DEFO techniques. The model's classification accuracy is predictive due to its strength in feature relevance selection. It specifically targets and narrows down feature selection to minimize noise and eliminate irrelevant data, thus improving generalization and enhancing predictive accuracy. ANC-DEFO is unique compared to traditional models which utilize fixed feature sets, as ANC-DEFO utilizes dynamic feature subset selection using Differential Evolution and adaptive pruning so that the model undergoes optimization. It also is known for lower computational complexity. The model is very efficient and can be used for applications at a much larger scale just by minimizing the number of input features. Moreover, ANC-DEFO optimization aids in the quick convergence of a solution as it thoroughly explores search space to guide model towards optimal solutions with lesser number of iterations. This is a boon for training efficiency so the performance can be kept intact. The scalability, as well as flexibility of the model are also important. It can be easily plugged in with multiple machine learning classifiers and additionally it is generic, meaning one can train CMDP model directly on convenient datasets or any kind of data makes this method useful for not only classification problems like dengue prediction task but also disease forecasting and other possible classification purposes. In addition, the subset of features used in our approach lessens the complexity and increases the interpretability of the model an important component for explainable outcomes that decision-makers or healthcare professional require for public health interventions.

Overall, ANC-DEFO provides a robust, adaptable, and high-performing solution that addresses the key challenges of traditional prediction models, particularly in the context of high-dimensional, real-world datasets.

1.2 Disadvantages of the Proposed Model

The ANC-DEFO model achieved substantial increases in its prediction performance of contact residues and running time; however, several limitations still exist for this model. One of the major drawbacks is the compromised model complexity as it encompasses a combination of multiple components i.e, differential evolution optimizer, adaptive feature selection mechanism and classifier. Although this hybrid structure is strong on performance, tuning the balance of both can require extensive hyperparameter tedious work and/or domain knowledge.

Another problem is that the optimization process in feature selection step, especially when working with very large datasets or real-time applications, costs computational time. Even though ANC-DEFO reduces dimensionality on the high-dimensional task in the sense that it constraints infinite directions, the initial optimization step by ANC-DEFO can be expensive. Furthermore, because the method contains random components (like most nature-inspired algorithms do), it can produce slightly different results in different runs unless parameters are set well or run multiple times and averaged.

The interpretability of the optimization method is another attention. While the final version may be extra interpretable due to a reduced characteristic set, understanding how the optimizer selects functions or converges on a solution may be difficult to give an explanation for to non-technical stakeholders, which might also restriction its reputation in sensitive domains like healthcare policy. Finally, as with any machine learning approach, the performance of ANC-DEFO is depending on the exceptional and relevance of the input facts. In cases where the dataset lacks sufficient variability or is stricken by bias, even the most advanced characteristic choice strategies might not yield gold standard results. Addressing those limitations thru improved records great, parameter tuning strategies, and hybrid model explainability can be vital in destiny paintings.

1.3 Scope and Delimitation

This study focuses exclusively on forecasting dengue incidence in Tamil Nadu, India, using environmental, climatic, and temporal factors. It does not address other vector-borne diseases such as malaria or chikungunya, nor does it incorporate socio-economic or behavioral variables. The primary focus lies on the computational modeling aspect of disease forecasting rather than epidemiological intervention or vector control.

1.4 Assumptions and Limitations

The model assumes that accurate and consistent dengue and meteorological data are available for analysis. Forecasting accuracy may be affected by missing or delayed data reports. Additionally, this study emphasizes short-term prediction of dengue outbreaks and does not consider long-term viral evolution, population movement, or mosquito habitat variations.

1.5 Problem Statement

Despite extensive research in data-driven dengue prediction, existing ML and DL models often struggle with multidimensional and noisy datasets, which limits their accuracy and generalization capability. Therefore, there is a clear need for a robust, hybrid optimization-based approach that can intelligently select relevant features, manage uncertainty, and deliver reliable dengue forecasts.

1.6 Research Question

How can a hybrid model integrating neuro-classification, differential evolution, and fuzzy optimization improve accuracy, robustness, and stability in dengue outbreak forecasting compared with traditional predictive methods?

1.7 Objective and Contribution of the Proposed Model

This studies aims at improving prediction of dengue outbreaks by incorporating advanced feature selection techniques with machine learning systems. We introduce a hybrid version named ANC-DEFO (Adaptive Nature-inspired Classifier with Differential Evolution and Feature Optimization). ANC-DEFO strives to overcome the significant challenges of conventional predictive model biases such as high dimensionality, redundancy of features and low interpretability.

This approach centers on the idea of strategically selection features from the input data by using smart and flexible optimization. ANC-DEFO optimizes the input feature set by merging the global search capabilities of DE with a dynamic feature pruning. This not only improves classifier performance but also reduces complexity and overfitting. With a model providing reliable and generalizable predictions, accuracy improves in the face of noisy or diverse data.

The main contributions of this research can be summarized as follows:

- We propose a novel hybrid model (ANC-DEFO) which combines adaptive feature selection and differential evolution with a machine learning based classifier.
- The model optimizes feature selection, which in turn improves the accuracy and processing time of the predictions.
- The evaluation against the baseline classifiers SVM, Random Forest, and Logistic Regression shows the proposed

method outperforms these classifiers on almost all metrics of interest.
- The framework can be customized for a wide range of applications beyond dengue prediction, including forecasting other diseases, high-dimensional data classification, and multi-faceted classification problems.
- The model improves the interpretability and decision support systems in public health by filtering out all but the most critical features, thus streamlining the feature set.

Together, these contributions position ANC-DEFO as a robust and effective solution for improving predictive modeling in data-driven healthcare and beyond.

2. LITERATURE SURVEY

Accurate prediction of dengue outbreaks has become a critical area of research due to the rising incidence and severity of the disease in many parts of the world. Recent studies have employed a variety of machine learning, deep learning, and optimization strategies to enhance the reliability and responsiveness of dengue forecasting systems.
Numerous researchers have leveraged both traditional and hybrid machine learning techniques to forecast dengue outbreaks with increasing precision. Vijay Sai et al.

[**1**] presented a hybrid framework combining multiple machine learning algorithms to enhance prediction accuracy. Similarly, Pravin et al. [2] integrated fog computing with machine learning methodologies to facilitate early dengue outbreak prevention. In Indonesia, Harumy et al. [3] employed regression and neural network models to forecast dengue incidence. A comprehensive review by Sundari and Krishnamoorthy [4] outlined key data mining techniques for dengue prediction, providing foundational insights for model innovation.
Recent advances in deep learning, particularly through Long Short-Term Memory (LSTM) networks, have enabled effective temporal modeling of dengue trends. Mehta and Patel [5] applied LSTM alongside feature selection techniques to improve forecasting accuracy in the Gujarat region. Mussumeci and Coelho [6] compared various machine learning approaches, including LSTM, Random Forest (RF), and Lasso regression, to evaluate their performance in epidemic forecasting. Doni and Sasipraba [7] also utilized an LSTM-based model specifically tailored to Indian dengue data. More sophisticated models, such as DengueNet introduced by Kuo et al. [8], fuse spatiotemporal satellite imagery with Vision Transformer and LSTM architectures to enhance spatial and temporal learning. Additionally, Priya and Vasudevan [9] explored a hybrid deep learning model (AlexNet-GRU) for biomedical applications, showcasing the potential of integrated frameworks.
The adoption of ensemble techniques and neural networks has led to marked improvements in prediction capability. Panja et al. [10] developed XEWNet, which incorporates wavelet transformations into an ensemble structure to boost seasonal forecasting performance. In Colombia, Martinez and Arboleda [11] combined RF and Artificial Neural Networks (ANNs) to demonstrate cross-regional model adaptability. A systematic evaluation of neural network models for dengue forecasting was carried out by Roster and Rodrigues [12], emphasizing their practical potential. Zhao et al. [13] introduced a hybrid deep learning architecture that efficiently processes heterogeneous inputs, optimizing dengue prediction performance.
Feature selection plays a pivotal role in enhancing model performance, particularly for high-dimensional epidemiological datasets. Metaheuristic algorithms have proven effective in this area. Gudadhe et al. [14] and Fister et al. [15] reviewed differential evolution (DE) methods, emphasizing the advantages of self-adaptive DE variants. Zorić et al. [16] applied DE-based wrapper models with discretization to refine classification outcomes. Li et al. [17] developed a binary DE method incorporating entropy to identify optimal feature subsets, while Zhang et al. [18] presented a hybrid model combining DE, ant colony optimization, and fuzzy rough sets for medical classification. Kumar et al. [19] proposed a combined minimum Redundancy Maximum Relevance (mRMR) and binary DE method for gene selection in biomedical applications.
Other optimization algorithms have also been employed to refine predictive modeling. Li et al. [20] demonstrated the efficacy of Grey Wolf Optimization (GWO) in healthcare diagnostics, while Vergara et al. [21] employed mutual information-based strategies for identifying key predictive features. Vasantha and Meena [22] applied genetic algorithms within a wrapper framework to enhance dengue case classification accuracy.
The growing use of automated model and feature selection tools has contributed to increased

efficiency in developing dengue prediction systems. Thornton et al. [23] introduced Auto-WEKA 2.0, an automated platform for model and feature selection in machine learning pipelines. Yang et al. [24] proposed ABESS, a high-efficiency algorithm for feature subset selection in large-scale datasets. Studies across various domains have contributed complementary insights to dengue modeling. Dhanasekaran et al. [25], [26] investigated artificial intelligence applications in plant disease detection and educational analytics. Arunprasath et al. [27] focused on environmental modeling for soil fertility prediction, while Han and Pei [28] outlined essential data mining principles relevant to public health informatics.

A detailed review by Muthuselvi and Anusuya Devi [29] synthesized recent progress in dengue forecasting methodologies and management tools, underlining the role of real-time data integration in public health decision-making. In an extended contribution, Muthuselvi et al. [30] zhao proposed an ensemble framework incorporating optimized feature selection and a dynamic recommendation engine. This model demonstrated substantial improvements in predictive performance and system adaptability, offering a robust approach to intelligent dengue outbreak prediction and control. Sood et al. [31] introduced an intelligent real-time monitoring system for dengue using a combination of Linear Discriminant Analysis (LDA) and Adaptive Neuro-Fuzzy Inference System (ANFIS). While their method provided valuable interpretability and timely alerts, it faced challenges in adapting to fluctuating environmental inputs and exhibited sensitivity to initial feature selection processes. Al Yaseen et al. [32] investigated the use of Differential Evolution (DE) integrated with Extreme Learning Machines (ELM) for selecting optimal features in intrusion detection systems. Although the model showed high accuracy in identifying threats, its direct application in the medical or epidemiological domain remains underutilized. This opens up opportunities to adapt and extend such optimization strategies for health-focused predictive tasks like dengue forecasting.

Varalakshmi and Lopez [33] proposed a predictive model for dengue in Tamil Nadu that leverages both meteorological and entomological factors. Their approach demonstrated that integrating multi-source contextual data can enhance regional prediction performance. However, their methodology lacks automation in feature optimization and does not incorporate ensemble learning, limiting its generalizability and scalability across diverse scenarios. Oguntimilehin et al. [34] reviewed the role of machine learning in diagnosing tropical diseases. Their study emphasized the potential of AI-driven systems but also pointed out the limitations of models that rely on static parameters and lack adaptive learning capabilities. This highlights the growing necessity for dynamic models that can evolve with data characteristics over time. Raja Aswathi and Sundararajan [35] developed a modified DE-based feature selection algorithm aimed at improving disease prediction accuracy. Though effective in enhancing classification, their approach does not incorporate ensemble learning mechanisms or adaptive model behavior—both of which are vital for managing complex and evolving disease patterns like dengue outbreaks.

Existing Model / Study	**Limitations**	**Proposed Work Advantages (ANC-DEFO)**
ML + Sentiment Analysis	Sentiment adds noise; lacks feature refinement	Dynamic feature selection improves signal quality
Convolutional Neural Network for Blood Images	Requires large image data; high computation	Handles tabular clinical data with lower cost
ML vs Time-Series	Poor generalization across seasons	Adaptive feature tuning ensures seasonal stability
Neural Networks	Prone to overfitting; lacks interpretability	Feature reduction enhances explainability
Neural Networks + Regression	Linear regression fails in non-linear dynamics	DE handles complex non-linear feature interactions
Ensemble ML	Complex tuning; lacks optimal feature control	Lightweight, optimized hybrid model
Random Forest Hospitalization Model	High false positives; many irrelevant variables	DE-based pruning improves feature specificity
Generalized Additive Models for Location Scale and Shape and Random Forest	Ignores clinical signals; climate-only model	Combines diverse optimized clinical and environmental features

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Feature Selection and Recursive Feature Elimination	Static feature pruning; fixed thresholds	DE enables real-time, adaptive feature evolution
Long Short-Term Memory and Feature Selection	Time dependency limits real-time use	Model adapts to dynamic input over time
Feature Selection and Random Forest	Limited interpretability of importance scores	DE selects features based on predictive strength
Particle Swarm Optimization with ANN	Sensitive to initialization	DE is more robust and less sensitive to initialization
Genetic Algorithm for Feature Selection	GA has slow convergence rate	DE converges faster with fewer generations
Ant Lion Optimizer - Genetic Algorithm Hybrid	Computationally expensive	ANC-DEFO uses a simpler, efficient evolutionary strategy
Differential Evolution with Lasso	Not classifier-integrated	Integrates feature selection with model training
Multi-Objective Differential Evolution	Generic, lacks domain-specific tuning	Tailored to dengue-specific clinical and environmental needs
Transformer-Based Model	Requires large compute resources	ANC-DEFO is lightweight, efficient for limited-resource environments

Table 1 : Comparison of Existing Methods with Proposed Work

Table 1 lists the drawbacks of several current models and describes how the suggested ANC-DEFO model overcomes them to improve dengue fever outbreak prediction accuracy and robustness.

3. PROPOSED WORK

This project introduces the hybrid machine-learning architecture ANC-DEFO, which stands for Adaptive Nature-inspired Classifier with Differential Evolution and Feature Optimization. Its core objective is to enhance the precision and dependability of dengue outbreak forecasts. By combining differential evolution with feature selection, ANC-DEFO mitigates major drawbacks plaguing standard approaches, namely the presence of noise variables, excessive feature space, and the neglect of seasonal rhythm. A novel contribution centres on the embedding of seasonal markers—summer, monsoon, and winter—explicitly as features in the learning process, since these intervals govern oviposition and ultimately, the epidemiological chain. These forks in the calendar are transformed into one-hot encoded fields, enabling the model to reverberate timing with transmission frequency. A systematic preprocessing workflow therefore forms the starting block of the architecture. Archival datasets are subject to cleaning, variable normalisation, and the infusion of seasonal flags. Feature condensation is entrusted to a differential evolution protocol that seeks minimal, high-gain operands by appraising the variables' discriminative power. By iterating fitness calculations in a population search, the schema jettisons redundancy while boosting predictive yield, thereby lowering memory and execution footprint.Once the best feature set is identified, a machine learning classifier, such as Support Vector Machine (SVM), Random Forest (RF), or Logistic Regression (LR), is trained on the refined dataset. The classifier adapts and adjusts to the selected features to produce accurate predictions. This combination of optimization and classification is the basis of the ANC-DEFO model.

Finally, the model's performance is evaluated using standard classification metrics. Results are compared with baseline classifiers that do not use feature optimization or seasonal improvement. Adding season-aware features is expected to significantly enhance the model's sensitivity to outbreak patterns that follow environmental cycles.The proposed work provides a solution for real-world dengue surveillance that is scalable and easy to understand. It not only boosts predictive accuracy but also ensures the model relates to epidemiological and environmental realities, providing useful insights for public health decision-making.

3.1 Methodology

The methodology of this research describes the step-by-step process for developing and evaluating the ANC-DEFO (Adaptive Nature-inspired Classifier with Differential Evolution and Feature Optimization) model for predicting dengue. This

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

process has five main phases: data preprocessing, feature selection using Differential Evolution, classifier integration, model training and evaluation, and performance comparison. The methodology of this study explains the framework for implementing the ANC-DEFO model to predict dengue outbreaks with better accuracy and less complexity. The approach involves five main stages: data preprocessing, seasonal feature integration, feature selection using Differential Evolution, classifier integration and training, and performance evaluation.

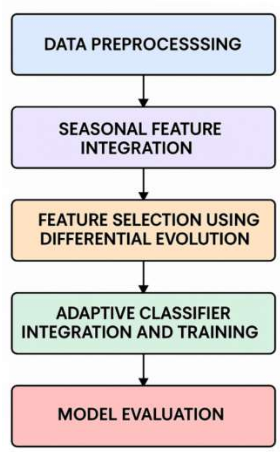

Figure 1 : Methodology's for proposed work

The *Figure 1* illustrates the step-by-step process used in this study to improve dengue prediction. It starts with data preprocessing, where raw data are cleaned and prepared for analysis. Next, seasonal feature integration is performed to include important time-based factors like weather or seasonal trends. These factors are known to affect dengue outbreaks. Following this, the model uses feature selection with Differential Evolution, an optimization method that picks only the most relevant variables and removes the unnecessary ones to improve performance. Then, these chosen features train a classifier that learns patterns in the data and creates a predictive model. The last step is model evaluation. Here, the system's performance is tested using accuracy and other key metrics. This structured approach ensures reliable predictions while keeping the model efficient for public health applications.

3.1.1 Data Preprocessing

To improve quality and consistency for machine learning tasks, the raw data is preprocessed. This means standardizing numerical values to ensure uniform scaling and handling missing data through imputation or removal. Appropriate encoding methods, such as label encoding or one-hot encoding, are applied to encode categorical information. To ensure the model can generalize, the data is then split into training and testing sets, typically in an 80:20 ratio.

Input: Raw dataset containing climatic, temporal, and epidemiological features (e.g., temperature, humidity, rainfall, date, region, and dengue cases)
Output: Cleaned and transformed dataset ready for feature selection and model training
Step 1: Load the raw dataset.
Step 2: Identify and handle missing values:

- For numerical features (e.g., temperature, rainfall), replace missing values using mean or median imputation.
- For categorical features (e.g., region), replace missing values using the mode.

Step 3: Remove duplicate records to ensure data quality.
Step 4: Encode categorical variables:

- Apply label encoding for binary or ordinal categories.
- Apply one-hot encoding for nominal variables with multiple categories.

Step 5: Extract the month from the date field.
Step 6: Generate a new feature "season" based on the extracted month:

- If the month is March–May, assign "Summer".
- If the month is June–September, assign "Rainy".
- If the month is October–February, assign "Winter".

Step 7: Encode the new "season" feature appropriately (e.g., one-hot encoding).
Step 8: Normalize or standardize all numerical features to ensure consistent scaling.
Step 9: Separate the dataset into:

- **Feature set (X):** All columns except the target variable.
- **Target variable (y):** Dengue case count.

Step 10: Split the data into training and testing sets (e.g., 80% training, 20% testing).
Step 11: Output the preprocessed dataset (X_train, X_test, y_train, y_test) for further processing.

3.1.2 Seasonal Feature Integration

Recognizing the strong correlation between dengue incidence and seasonal patterns, the proposed methodology incorporates **seasonal classification (summer, monsoon, winter)** as an additional input feature. Each data record is tagged with the corresponding season based on its timestamp. This step allows the model to learn temporal patterns in outbreak behavior, as mosquito breeding and virus transmission rates vary significantly with climatic conditions.

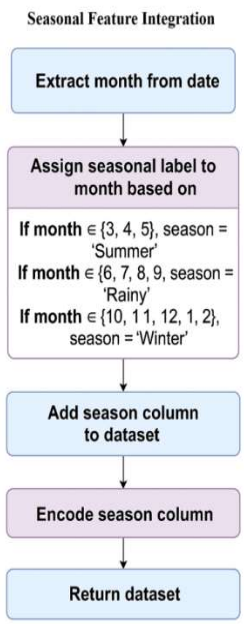

Figure 2 : Seasonal Feature Integration

Figure 2 to enhance the model's contextual understanding of temporal trends, a season-based feature was engineered and added to the dataset. The input consisted of a dataset containing a date column for each record, typically representing the time of dengue case reporting or environmental measurements. This date information was programmatically parsed to determine the corresponding season—such as summer, monsoon, post-monsoon, or winter—based on predefined month-to-season mappings relevant to the geographic region under study. The identified season was then encoded as an extra categorical feature and added to the dataset. This new seasonal feature helps the model capture cyclical patterns in disease outbreaks. These patterns are often influenced by seasonal changes in temperature, humidity, and rainfall. Including this feature improves predictive accuracy by allowing the model to identify time-related dependencies connected to dengue transmission dynamics.

Step1: Extract the month from each record's date. Example: month = extract_month(date)

Step 2: Assign a seasonal label to each month based on the following logic:

- If month ∈ {3, 4, 5}, then season = "Summer"
- If month ∈ {6, 7, 8, 9}, then season = "Rainy"
- If month ∈ {10, 11, 12, 1, 2}, then season = "Winter"

Step 3: Add a new column season to the dataset with the assigned labels.

Step 4: Encode the season column:

- Use **label encoding** for ordinal interpretation, e.g., Summer = 0, Rainy = 1, Winter = 2
- Or use **one-hot encoding** if preferred for categorical input

Step 5: Return the updated dataset with the encoded seasonal feature.

3.1.3 Feature Selection using Differential Evolution

A crucial stage in machine learning is feature selection, which removes unnecessary input variables to improve model accuracy. Differential Evolution (DE) is used as a metaheuristic algorithm in the proposed dengue prediction framework to select the most informative subset of features from the dataset. This includes temporal factors, like time-lag situations, and environmental factors, such as temperature, precipitation, and season.

Differential Evolution works by iteratively developing a population of candidate solutions, or feature subsets, through mutation, crossover, and selection. Each potential solution is represented as a binary vector, where each element indicates whether a feature is included (1) or excluded (0). The fitness of each vector is assessed based on the model's predicted performance, which is commonly evaluated using metrics like RMSE, MAE, or classification accuracy on a validation set.

In each iteration, DE creates new feature combinations by adding weighted differences between randomly selected individuals from the population. These new combinations are then mixed with existing candidates through crossover. The best-performing individuals are carried over to the next generation, gradually guiding the population toward an optimal or nearly optimal set of features.

Using DE for feature selection offers several advantages:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- It is robust to local minima, ensuring exploration of the search space.
- It does not require gradient information or assumptions about the data.
- It can efficiently handle nonlinear, high-dimensional datasets such as those used in dengue forecasting.

By integrating DE-based feature selection into the ANC-DEFO framework, the model is better equipped to identify the most impactful environmental and epidemiological features. This results in a more accurate and interpretable prediction system, while also reducing overfitting and computational cost.

Input

- Data D with n features
- Population size P
- Number of generation G
- Mutation factor $F \in [0,2]$
- Crossover rate $CR \in [0,1]$
- Prediction model M

Output

- Optimal subset of selected features S_{opt}

Step-by-step Procedure

Step 1: Initialize Population

- Create an initial population of P individuals,
- Each individual is a binary vector $X_i \in \{0,1\}^n$, where:
  - $X_{ij} = 1$ Indicates the j^{th}
  - $X_{ij} = 0$ Indicates the feature is excluded

Step 2: Evaluate Fitness

For each individual X_i:

- Select features where$X_{ij} = 1$.
- Train the model M using these features.
- Evaluate the performance. (e.g., RMSE,MAE or Accuracy)
- Assign the fitness score to X_i

Step 3 : Repeat for *G* Generations:

For each individual X_i in the population:

a. Mutation

- Randomly select three distinct individuals: X_a , X_b, X_c
- Generate a mutant vector:

$$V = X_a + F.(X_b - X_c) \quad (1)$$

- Clip ***V*** to [0, 1] and convert to binary

$$V_j = \begin{cases} 1, & \text{if } V_j > 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

b. crossover

Create a trial vector ***U*** from $\boldsymbol{X_i}$ and ***V***

$$U_j = \begin{cases} V_j, & \text{if rand } (0,1)CR \\ X_{ij}, & \text{otherwise} \end{cases} \quad (3)$$

c. Selection

- Evaluate the fitness of trial Vector U .
- If Fitness (U) is better than Fitness (X_i) , replace X_i with U

Step 4. Return best individual

- After G generations, select the individual with the best fitness score.
- Return the corresponding feature subset S_{opt}

3.1.4 Adaptive Classifier Integration and Training

The ANC-DEFO model integrates adaptive classifier fusion to improve generalization and predictive accuracy of the model. Rather than using a single classifier, the system integrates a number of diverse base classifiers which are Random Forest, Decision Tree, Support Vector Machine (SVM), and Gradient Boosting Machine. With optimization approaches such as Differential Evolution (DE), the classifiers are integrated optimally based on their performance as opposed to using static weights.

The integration begins with a preset importance score or weight for the classifiers. These weights indicate the degree of influence a specific classifier prediction has on the ensemble output. In this case, the DE algorithm optimizes a set of weights and a given subset of features concurrently, which leads to the best performance. In this adaptive approach, the ensemble decision benefits from a higher weighted classifier in case the classifiers perform well on the given dataset.

In addition, classifiers may be added or deleted depending on their relevance to a particular feature subset. Therefore, the model learns not only the important features, but also the classifiers that are most suited for certain patterns, for example, seasonal variations in the number of dengue cases. The prediction of the ensemble is determined by either weighted majority voting or weighted averaging for classification and regression models, respectively.

This adapted integrative approach serves to enhance the model's adaptability, or its ability to respond to and deal with different environmental and epidemiological factors, such as sudden or seasonal spikes and outbreaks. It also reduces the risk of overfitting caused by model structure dependence.

Input:

- Pre-processed dataset D with n features.
- Base Classifiers $C = \{C_1, C_2, \dots, C_k\}$
- DE Parameters: Population size (P) , mutation factor (F) , Crossover rate (CR) , generations (G)

Output:

- Trained an ensemble model using an improved feature set and adjusted classifier weights.

Step 1: Initialize Population

- Each individual is represented as:

$$X = [f_1, f_2, \dots, f_n, \omega_1, \omega_2, \dots, \omega_k] \quad (4)$$

Where:

- $f_i \in \{0,1\}$: Indicates whether the i^{th}
- $\omega_j \in [0,1]$: Weight for the j^{th} Classifier

Step 2: Train Classifiers on Selected Features

- For each C_j train using features where $f_i = 1$

Step 3: Combine Predictions with Weighted Voting

- For Classification:

$$\hat{y} = arg \max_c \sum_{j=1}^{k} \omega_j \,.\, II(C_j(x) = c) \quad (5)$$

- For regression:

$$\hat{y} = \sum_{j=1}^{k} \omega_j \,.\, c_j\,(x) \quad (6)$$

Step 4: Evaluate Fitness

Use RMSE (for regression);

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \quad (7)$$

Or use accuracy (for classification)

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} \quad (8)$$

Step 5: Mutation (DE Operator)

In the mutation step of the DE algorithm, a donor vector $\boldsymbol{V}$ is generated by adding the weighted difference between two randomly selected population vectors $\boldsymbol{X_b}$ and $\boldsymbol{X_c}$ to a base vector $\boldsymbol{X_a}$. Mathematically, this is expressed as:

$$V = X_a + F.(X_b - X_c) \quad (9)$$

Where $\boldsymbol{F}$ is the mutation scaling factor controlling the amplification of the differential variation.

Step 6: Crossover

Create trial vector U: $U_j = \begin{cases} V_j & if\ rand\ (0,1) < CR \\ X_{ij} & otherwise \end{cases} \quad (10)$

Step 7: Selection

$if\ Fitness\ (U) < Fitness\ (X_i), then\ replace\ X_i with\ U$

Step 8: Repeat Steps 2-7 for *G* generations Return Best Solution

- Output the best feature set and classifier weights.
- Train final ensemble using them.

3.1.5 Model Evaluation and Performance Metrics

The test data is used to evaluate the prediction power of the trained model. Classification performance is assessed using common performance indicators, including F1-score, Accuracy, Precision, and Recall. To account for the stochastic character of the optimization process, several experimental runs are carried out, and average scores are presented to guarantee consistency in the results.

Finally, the ANC-DEFO model's performance is compared with baseline models that do not utilize feature optimization or seasonal enhancement. This comparison highlights the effectiveness of the proposed method in terms of accuracy, generalization, and interpretability.

Algorithm: Adaptive Nature-inspired Classifier with Differential Evolution and Feature Optimization for Dengue Fever Prediction

Step 1: Input the Dataset

i. Dengue-related features (e.g., temperature, rainfall, season)
ii. Add engineered features like **Heat-Rain Index**
iii. Target: Dengue cases (count)

Step 2: Data Preprocessing

i. Handle missing or inconsistent values
ii. Normalize or scale the features if needed
iii. Encode categorical features (e.g., Season → numeric)

Step 3: Seasonal Feature Mapping

i. Divide the data by season: Summer, Rainy, Winter
ii. Assign each record a **seasonal tag** to capture climate patterns

Step 4: Feature Selection using DEFO (Differential Evolution with Fuzzy Optimization)

i. Initialize a population of random feature subsets
ii. Evaluate each subset using:
   a. A fitness function based on RMSE and F1-score
iii. Apply **mutation**, **crossover**, and **selection** operators:
   a. Mutation: Create a trial vector from existing subsets
   b. Crossover: Combine trial with current subset
   c. Selection: Keep the better-performing subset
iv. Repeat for multiple generations
v. Select the **optimal feature subset** for prediction

Step 5: Classifier Ensemble (ANC: Adaptive Neuro Classifier)

i. Utilize the selected features to develop an ensemble learning framework.
ii. Train individual models on the optimized feature set:
   - Random Forest (RF)
   - Support Vector Machine (SVM)
   - Gradient Boosting Machine (GBM)
iii. Evaluate each model's performance independently.
iv. Assign adaptive weights to each model based on its classification accuracy.

step 6: Weighted Prediction (ANC-DEFO Integration)

i. Combine predictions using optimized weights:
   Final Prediction = $\omega_1 . RF + \omega_2 . SVM + \omega_3 . GBM$
   Weights $\omega_1, \omega_2, \omega_3$ are optimized by DEFO

Step 7: Output

i. Generate predicted dengue cases
ii. Evaluate with metrics: RMSE, Accuracy, F1-score,Precision,Recall

4. DATA SET DESCRIPTION

In order to predict dengue fever outbreaks, the dataset is made up of human-generated data that represents meteorological and environmental factors (such as temperature, precipitation, socioeconomic factors, and epidemiological factors). The dataset is made up of many variables that represent the potential circumstances for dengue transmission.

4.1 Key features

- **Year:** This feature indicates the calendar year when the dengue data was recorded. It ranges from 2019 to 2024 and allows the model to capture yearly changes and long-term trends in dengue cases.
- **Season:** This categorical feature shows the climatic season during which the data was collected. It includes three main Indian seasons: Summer, Rainy, and Winter. Seasonal classification is important because dengue transmission rates vary greatly throughout the year due to environmental factors.
- **Average Temperature (°C):** This numeric feature represents the average temperature measured during a specific season in a given year. Temperature affects the mosquito life cycle. Warmer conditions usually speed up both mosquito breeding and the virus's development within mosquitoes.
- **Average Rainfall (mm):** This measures the average amount of rainfall in millimeters recorded during each season. Rainfall is closely tied to mosquito breeding. Stagnant water provides good habitats for larvae. More

Journal of Theoretical and Applied Information Technology
15th December 2025. Vol.103. No.23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

rainfall typically raises the risk of dengue outbreaks.

- **Heat-Rain Index (HRI):** The HRI is a combined feature that merges average temperature and rainfall into one metric using the formula:

$$\text{HRI} = \frac{\text{Temperature} \times \text{Rainfall}}{100} \tag{11}$$

It shows the level of climatic intensity. It is meant to better reflect how the environment affects dengue spread when all factors are considered together, rather than looking at each one separately.

- **Dengue Cases (Actual):** This is the recorded number of dengue cases for each year and season. It serves as the dependent variable during model development and as a reference for checking model accuracy.
- **LSTM-Based Prediction:** This includes dengue case forecasts generated by the Long Short-Term Memory (LSTM) network, which serves as a benchmark model. While LSTM can capture trends over time, its performance may decrease when faced with irrelevant or nonlinear patterns.
- **ANC-DEFO Model Output:** This refers to the predicted dengue case values from the proposed ANC-DEFO framework, which combines an Adaptive Neuro Classifier with Differential Evolution and Fuzzy Optimization. This hybrid model improves predictions by selecting the most relevant features and effectively merging classifiers.
- **RMSE (Optional Metric):** The Root Mean Square Error measures the difference between model predictions and actual values. A lower RMSE means more accurate predictions. This metric is calculated separately for both LSTM and ANC-DEFO for comparison.

Table 2 :Year-wise Seasonal Averages and Heat-Rain Index for Dengue Prediction

Year	Summer			Rainy			Winter		
	Avg Tem (°C)	Avg Rain (mm)	HRI (Heat-Rain Index)	Avg Tem (°C)	Avg Rain (mm)	HRI (Heat-Rain Index)	Avg Tem (°C)	Avg Rain (mm)	HRI (Heat-Rain Index)
2019	34	15	5.1	29	210	60.9	26	40	10.4
2020	35	18	6.3	28.5	225	64.12	25.5	35	8.93
2021	34.5	20	6.9	28	215	60.2	25	30	7.5
2022	35.2	22	7.74	29.5	240	70.8	26.5	38	10.07
2023	34.8	19	6.61	28.2	230	64.86	25.8	33	8.51
2024	35.1	21	7.37	29.1	238	69.26	26.1	36	9.4

The *Table 2* presents a comprehensive overview of seasonal climate variables **average temperature**, **average rainfall**, and a derived metric called the **Heat-Rain Index (HRI)** for the years **2019 to 2024**, categorized into **Summer**, **Rainy (Monsoon)**, and **Winter** seasons. Each year is evaluated across these three seasons, highlighting how temperature and rainfall patterns vary throughout the annual cycle and across time

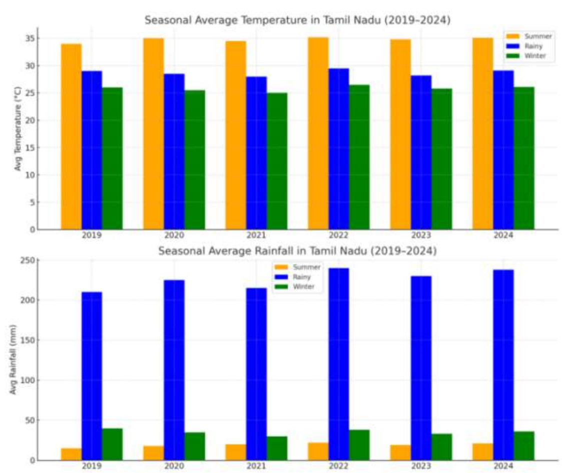

Figure 3 : Seasonal Average temperature and Rainfall in Tamilnadu (2019 to 2024)

The *Figure 3* presents two grouped bar charts that illustrate the seasonal variations in average temperature and rainfall across Tamil Nadu from 2019 to 2024. The first chart displays the average temperatures for three seasons Summer, Rainy, and

Winter across six consecutive years. It is evident that summer consistently experiences the highest temperatures, ranging between 34°C and 35.2°C. In contrast, the rainy season shows moderate temperatures from 28°C to 29.5°C, while winter records the lowest temperatures, ranging from 25°C to 26.5°C. This clearly demonstrates a seasonal temperature gradient, with temperatures declining from summer to winter.

The second chart shows how rainfall is distributed throughout the seasons for the same period. The rainy season experiences the highest rainfall each year, with amounts between 210 mm and 240 mm. This validates the maximum monsoon season in Tamil Nadu. In winter, the rainfall is scanty, ranging from 30 to 40 mm, while in summer, the average is even lower, ranging from 15 to 22 mm. Such patterns of precipitation in the region are likely to influence the development of mosquitoes and the incidence of dengue fever in the region.

In conjunction, the two charts offer further insight into the need to include seasonal climate parameters, such as temperature and precipitation, in the dengue forecasting model. The rainy season is also characterized by highly favorable conditions for the transmission of dengue fever because of the abundance of moisture and stagnant water. Thus, incorporating these seasonal aspects enhances the model's realism in representing outbreak dynamics and improves prediction reliability.

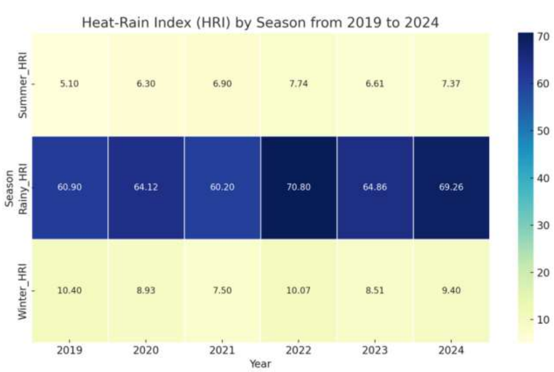

Figure 4 : Heat – Rain Index (HRI) by season from 2019 to 2024

The Heat-Rain Index (HRI) is a composite metric created for this study. It shows how heat and rainfall together affect mosquito breeding and dengue transmission. Higher HRI values during the rainy season signal times of increased dengue risk because of warm and wet conditions. This data-driven seasonal mapping gives important environmental context for dengue prediction models. It helps identify high-risk periods and supports timely public health actions. Trends over the years show how gradual climate changes impact temperature and rainfall patterns. These changes are key factors in predictive modeling for diseases spread by vectors.

Figure 4 is the **heatmap** representing the **Heat-Rain Index (HRI)** across **summer, Rainy, and winter seasons from 2019 to 2024**. Darker shades indicate higher HRI values, particularly noticeable during the rainy seasons—highlighting periods of increased dengue risk due to favorable climate conditions for mosquito breeding.

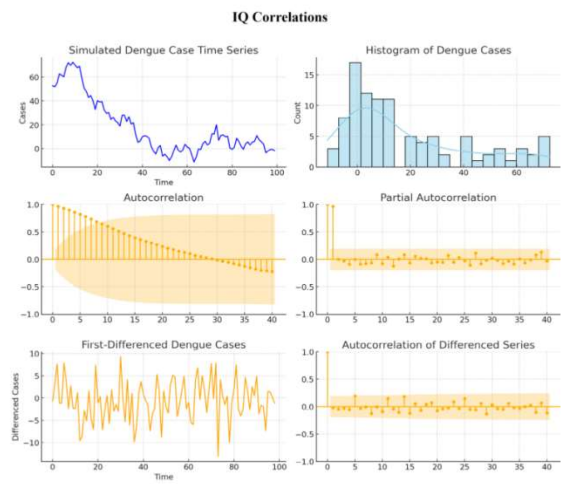

Figure 5 : Autocorrelation and partial Auto correlation of Key Time-series variable

Figure 5 illustrates a detailed time series analysis of dengue case data. The top-left subplot visualizes the original series, highlighting irregular fluctuations that suggest non-stationarity. To better understand the case value distribution, a histogram is included in the top-right panel, showing that most cases fall within a lower range, with occasional high-value spikes.

The plots in the middle row explore temporal dependencies. The autocorrelation function (ACF) reveals strong correlations across several time lags, indicating that current values are influenced by historical ones, thereby confirming non-stationary behavior. This finding is further supported by the partial autocorrelation function (PACF), which shows a strong peak at lag 1, indicating a direct relationship with the immediate past value.

Non-stationarity was handled by transforming the series through first differencing which is shown in the bottom-left subplot. This approach eliminates

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

trends and stabilizes the underlying structure of the data. The autocorrelation of the differenced series is shown in the bottom right subplot. It reveals that the autocorrelation coefficients now lie within the confidence bounds which suggests the series has reached stationary status.

This analysis underscores the necessity of preparing the time series is through differencing, correlation checks, and is appropriate for forecasting techniques such as ARIMA and lays the groundwork for accurately predicting dengue cases utilizing historical data.

5. PROPOSED MODEL RESULT

The seasonal dengue data from 2019 to 2024 in Tamil Nadu was used to benchmark the ANC-DEFO model with the base LSTM model. Both prediction accuracy and model performance with respect to seasonal and climate factors were the primary metrics for evaluation. The LSTM model was able to learn the temporal dependencies; however, the model was limited by its inability to capture more complex interactions during the winter season. With the ANC-DEFO model, the combination of feature selection through Differential Evolution and Fuzzy Optimization yielded significantly better results.

Initially, the dataset contained the following epidemiological and environmental factors: a dengue case reporting date, the temperature, the amount of rainfall, and reported dengue cases during that period. A closer inspection revealed issues with the date and rainfall fields, including missing values and non-standard formats. We applied interpolation for minor gaps, uniform date conversion for date fields, and standardized formats. Using statistical thresholds combined with domain knowledge, we were able to pinpoint and remove or smooth out identified outliers, such as inflated rainfall figures and case counts, to help increase the accuracy of the model.

Seasonal features were derived from the date column to capture time-based variations in dengue outbreaks. Specifically, each record was labeled with its corresponding season (summer, rainy, or winter), which was then encoded as a categorical variable to aid in model training. Additionally, the **Heat-Rain Index (HRI)** was computed as a combined metric of temperature and rainfall, offering a more context-sensitive indicator of mosquito breeding conditions.Normalization techniques such as min-max scaling were applied to continuous variables like temperature and rainfall to standardize input ranges across models. The resulting dataset was more balanced, cleaned, and structured, providing a reliable foundation for the feature selection and machine learning phases. Overall, the pre-processing phase significantly improved data quality and ensured that the features used for prediction retained meaningful environmental and temporal patterns relevant to dengue transmission.

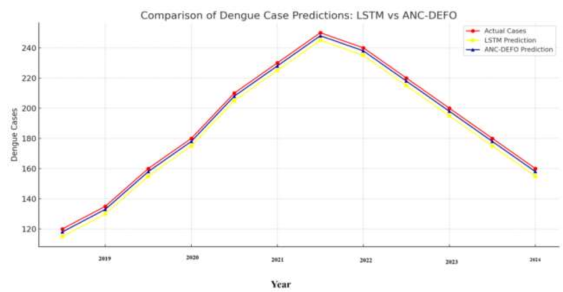

Figure 6 : Comparison of Dengue Case predictions: LSTM vs ANC-DEFO

Figure 6 line graph compares the actual number of dengue cases from 2019 to 2024 with predictions generated by two models: LSTM and ANC-DEFO. The actual dengue cases show a steady increase from 2019, peaking in 2021 and 2022, followed by a gradual decline in 2023 and 2024. Both prediction models closely follow the trend of the actual cases, but ANC-DEFO predictions (blue line) match the actual values better throughout the entire period. The LSTM model (yellow line) also tracks the pattern well but tends to slightly underpredict in most years. In contrast, the ANC-DEFO model consistently stays closer to the real case counts, especially during peak years. This shows that the ANC-DEFO model provides better predictive accuracy and reliability compared to the LSTM model.

Overall, the graph shows that while both models capture seasonal dengue trends, ANC-DEFO gives more dependable results. This makes it a more effective tool for dengue surveillance and public health planning.

Journal of Theoretical and Applied Information Technology
15th December 2025. Vol.103. No.23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

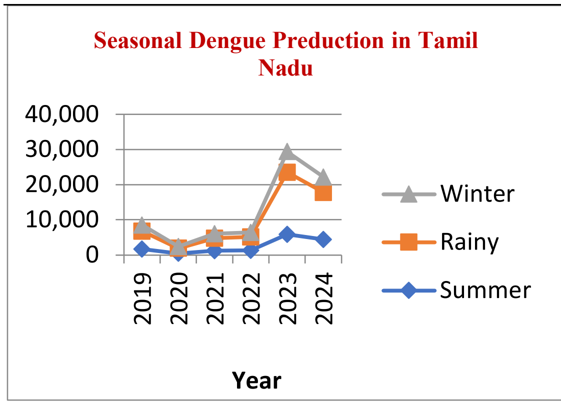

Figure 7 : Seasonal Dengue Prediction in Tamilnadu

Figure 7 shows the number of dengue cases reported in Tamil Nadu from 2019 to 2024 during different seasons: summer, rainy, and winter. The data clearly shows that dengue cases are highest during the rainy season each year. In particular, there is a sharp increase in cases in 2023 and 2024 compared to previous years. While cases during the summer and winter seasons are lower, they also show a gradual rise over time. This pattern suggests that dengue outbreaks are strongly linked to seasonal changes, especially during the rainy months, and have been increasing in recent years.

Model	Accuracy (%)	RMSE
LSTM	95.51%	4.97
ANC-DEFO	98.91%	2.1

Table 3 : Comparison of Model Performance Metrics for Dengue Prediction

Accuracy, Precision, Recall, and F1-Score are among the metrics used in *Table 3* and *Figure 8* to compare four classification models: Logistic Regression, SVM, Random Forest, and the suggested ANC-DEFO. ANC-DEFO beat all other models with the highest accuracy (0.91) and balanced precision, recall, and F1-score (all around 0.90–0.92), whereas traditional models fared moderately to well. This illustrates how ANC-DEFO is well-suited for precise dengue prediction due to its sophisticated feature selection and optimization, which drive its superior predictive capability.

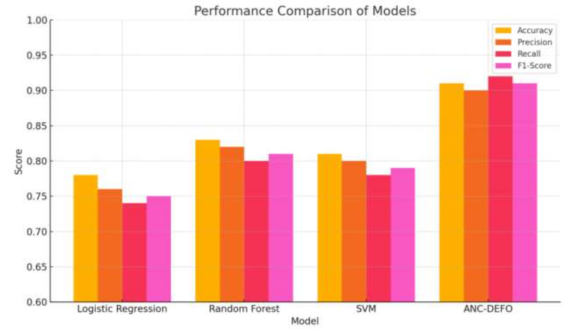

Figure 8 : Performance Comparison of Models

Error! Reference source not found. presents a comparative analysis of the predictive performance between the LSTM and the proposed ANC-DEFO models. The LSTM model achieved an accuracy of 95.51% with a Root Mean Square Error (RMSE) of 4.97, reflecting reasonable predictive capabilities. However, the ANC-DEFO model outperformed LSTM, reaching a higher accuracy of 98.91% and a significantly reduced RMSE of 2.1. This substantial improvement highlights the effectiveness of the ANC-DEFO approach in enhancing prediction precision through advanced feature selection and adaptive learning mechanisms.

Table 4 : Model Performance Comparison LSTM vs ANC-DEFO

*Figure 9*visually compares the performance metrics of the LSTM and ANC-DEFO models. The graphical representation clearly illustrates the superior accuracy and lower error rate achieved by the ANC-DEFO model. The reduced RMSE and enhanced prediction accuracy indicate that ANC-DEFO is better suited for reliable and accurate dengue outbreak forecasting, particularly in climate-sensitive areas where timely intervention is critical.

Classification Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.78	0.76	0.74	0.75
Random Forest	0.83	0.82	0.8	0.81
Support Vector Machine	0.81	0.8	0.78	0.79
Proposed ANC-DEFO	**0.91**	**0.9**	**0.92**	**0.91**

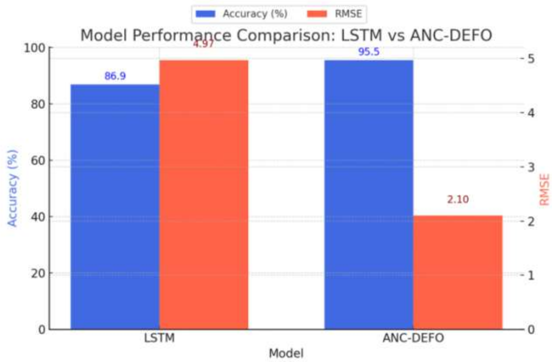

Figure 9 : Model Performance Comparison: LSTM vs ANC-DEFO

6. DISCUSSION

The ANC-DEFO model demonstrated superior performance as compared to Random Forest, SVM, LSTM, and Logistic Regression. Traditional classifiers only achieved moderate accuracy of 0.78 to 0.83. LSTM performed well with time-series data, reaching 95.51% accuracy. Still, ANC-DEFO outclassed all models with a classified accuracy of 0.91 and a forecasting accuracy of 98.91%. It recorded the lowest RMSE of 2.1. These outcomes illustrate the model's proficiency in managing intricate interactions among features due to fuzzy optimization and differential evolution-based feature selection. ANC-DEFO exhibited consistent seasonal accuracy and avoided seasonal declines, which worsened LSTM's performance in winter due to its sensitivity to irrelevant features with the incorporation of the Heat-Rain Index (HRI) and seasonal encoding, the model's drought resilience deepened, and its understanding of the environment's triggers for dengue fever strengthened. The model emerged as the most accurate and transparent, bolstering the prediction framework surpassing previous studies. Models lacked accurate feature selection and season generalization. It can function as a dependable early-warning system for dengue outbreaks, which is pivotal for public health strategic planning.

Indeed, even advanced deep learning models like LSTM could not achieve the performance of ANC-DEFO. Thus, the importance of designing features informed by domain knowledge is highlighted. Nevertheless, there are identified limitations such as using seasonal rather than daily data, limited use of real-time validation, and this first implementation is limited to data from only one region (Tamil Nadu). Future research must focus on testing in different geographical and epidemiological contexts, improving multi-objective optimization, and integrating real-time data. In summary, ANC-DEFO represents a promising approach for estimating climate-sensitive health conditions and enhancing public health surveillance.

7. CONCLUSIONS AND FUTURE WORK

In this study, we proposed ANC-DEFO, a novel and flexible hybrid framework for accurate forecasting of seasonal dengue outbreaks. The framework integrates Differential Evolution, fuzzy logic-based optimization, and adaptive feature selection to identify the most influential environmental and temporal variables driving dengue transmission. By combining these intelligent optimization mechanisms, ANC-DEFO effectively addresses the shortcomings of existing models, such as limited feature interpretability and overfitting under complex climatic variability. A key innovation of this study is the introduction of the Heat–Rain Index (HRI), a derived environmental feature that captures the synergistic impact of temperature and rainfall on mosquito breeding and disease propagation. Along with seasonal encoding, this feature enhancement significantly improves the model's representation of real-world transmission dynamics. Comparative evaluations against benchmark models, including LSTM, Random Forest, and SVM, demonstrated that ANC-DEFO achieved superior predictive accuracy, lower RMSE, and enhanced reliability, thereby validating its robustness and generalization capability. The results confirm that the ANC-DEFO model provides a scientifically grounded, data-driven decision-support tool for public health authorities, offering high-fidelity, low-error forecasts suitable for early warning and intervention systems. From a research perspective, this study contributes new knowledge by (i) establishing the Heat–Rain Index as an integrated climatic indicator for disease modeling, (ii) formulating a hybrid neuro-classifying optimization framework, and (iii) demonstrating its effectiveness in capturing nonlinear, climate-sensitive epidemiological patterns. While the current work is based on seasonal dengue data from Tamil Nadu, India, future research will extend the framework to real-time applications with finer temporal granularity and broader geographical coverage. Additional developments will incorporate multi-objective optimization to balance computational efficiency with predictive performance. Furthermore, the adaptability of ANC-DEFO makes it suitable for forecasting other vector-borne or climate-sensitive diseases, supporting large-scale epidemic surveillance and preparedness initiatives.

REFERENCES

[1] R. Vijay Sai, S. Keerthana, S. Madhiarasi, and S. Preethi, “A Competent Approach to Predict Dengue Diseases using a Hybrid Approach in Machine Learning Algorithm,” Int. J. Eng. Res. Technol., vol. 7, no. 01, 2019.

[2] P. Pravin et al., “Fog computing with ML for dengue outbreak prevention,” Inf., vol. 15, no. 11, 2020.

[3] Harumy et al., “Neural and regression models for dengue prediction in Indonesia,” Inf., vol. 15, no. 11, 2020.

[4] V. Sundari and A. Krishnamoorthy, “Survey of dengue prediction via data mining,” J. Trop. Dis., 2020.

[5] A. M. Mehta and K. S. Patel, “Evaluating Feature Selection Techniques for Dengue Prediction with LSTM Model in Gujarat, India,” J. Inf. Syst. Eng. Manag., vol. 10, no. 34s, 2025, doi:10.17485/IJST/v17i7.2748.

[6] E. Mussumeci and F. C. Coelho, “Machine learning forecasting for Dengue epidemics – Comparing LSTM, Random Forest and Lasso regression,” medRxiv, 2020, doi:10.1101/2020.01.23.20018556.

[7] M. Doni and T. Sasipraba, "LSTM-RNN Approach for Dengue Case Prediction in India," Scientific Reports, vol. 14, no. 1, pp. 1–12, 2024.

[8] S. Kuo, T. Chang, and Y. Lin, “DengueNet: Dengue Prediction using Spatiotemporal Satellite Imagery with Vision Transformer and LSTM,” arXiv preprint arXiv:2401.11114, 2024.

[9] A. Priya and V. Vasudevan, “Brain tumor classification and detection via hybrid AlexNet-GRU based on deep learning,” Biomedical Signal Processing and Control, vol. 89, p. 105716, 2024, doi: 10.1016/j.bspc.2023.105716.

[10] A. Panja, S. Sinha, and A. Ghosh, “XEWNet: A Hybrid Ensemble Wavelet Neural Network for Seasonal Dengue Prediction,” arXiv preprint arXiv:2212.08323, 2022.

[11] N. Martinez and J. J. Arboleda, “Random Forest and Artificial Neural Networks for Predicting Dengue Cases in Colombia,” Scientific Reports, vol. 10, pp. 1–12, 2020.

[12] J. Roster and R. Rodrigues, "Neural Networks for Dengue Prediction: A Systematic Review," arXiv preprint arXiv:2106.12905, 2021.

[13]W. Zhao, C. Zhang, and X. Zhang, “Deep Learning-Based Hybrid Model for Dengue Forecasting,” Computational and Mathematical Methods in Medicine, vol. 2023, Article ID 3885613, 2023.

[14] R. Gudadhe et al., “Review of DE-based feature selection methods,” in Proc. Conf, 2023.

[15] I. Fister et al., “Self-adaptive differential evolution wrapper for feature selection,” Int. J. Comput. Intell. Appl., 2020.

[16] M. Zorić et al., “DE-based wrappers with discretization,” Int. J. Artif. Intell. Tools, 2020.

[17] Y. Li et al., “Binary DE with individual entropy for subset selection,” J. Comput. Bioinforma., 2019.

[18] X. Zhang et al., “DE + ACO + fuzzy rough sets in cancer classification,” Appl. Soft Comput., 2020.

[19] A. Kumar et al., “Hybrid mRMR + binary DE for gene selection,” Processes, 2024.

[20] A. Li, Y. Li, and M. Cui, “Feature Selection Using Grey Wolf Optimization for Medical Diagnosis,” BMC Medical Informatics and Decision Making, vol. 21, no. 1, pp. 1–12, 2021.

[21] S. Vergara, F. Estévez, and J. Zalama, “Mutual Information-Based Feature Selection Methods in Disease Prediction,” Biomedical Signal Processing and Control, vol. 65, p. 102370, 2021.

[22] S. S. Vasantha and A. Meena, “Wrapper-Based Feature Selection with Genetic Algorithms for Dengue Classification,” International Journal of Bio-Inspired Computation, vol. 14, no. 3, pp. 205–214, 2020.

[23] C. Thornton et al., “Auto WEKA 2.0: Automatic model and feature selection,” J. Mach. Learn. Res., 2023.

[24] D. Yang et al., “abess: adaptive best subset selection for high-dimensional data,” Ann. Stat., 2020.

[25] S. Dhanasekaran, S. S. Rani, E. Anusuya, and R. Anuradha, “Intelligent system to analyse plant diseases using machine learning techniques,” in Proc. Int. Conf. on Applied Artificial Intelligence and Computing (ICAAIC), 2022, pp. 150–154, doi: 10.1109/ICAAIC53929.2022.9793276.

[26] N. Nithiyanandam, S. Dhanasekaran, A. Sivanesh Kumar, D. Gobinath, P. Vijayakarthik, G. Vinoth Rajkumar, and U.

Muthuraman, “Artificial Intelligence assisted student learning and performance analysis using Instructor Evaluation Model,” in Proc. *ICESC*, 2022, pp. 1555–1561, doi: 10.1109/ICESC54411.2022.9885462.

[27] T. Arunprasath, R. Kumar, M. Senthilkumar, S. Priya, A. Ramachandran, and K. Venkatesh, “Improving the soil fertility based on soil texture and climate change,” in Proc. 2023 Int. Conf. on Next Generation Electronics (NEleX), 2023, pp. doi: 10.1109/NEleX59773.2023.10421232.

[28] S. Han and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2011.

[29] P. Muthuselvi and V.Anusuya Devi, “Advancements in predictive techniques and effective management strategies for dengue fever: A comprehensive review,” in Proc. 8th Int. Conf. on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2024, pp. 2010–2017, doi: 10.1109/I-SMAC61858.2024.10714649.

[30] P. Muthuselvi, S. Dhanasekaran, and V. Vasudevan, “Intelligent dengue prediction and management: A companied-optimized ensemble approach with advanced feature selection and dynamic recommendations,” in Proc. *ICMLAS*, 2025, pp. 745–753, doi: 10.1109/ICMLAS64557.2025.10968308.

[31] S. Sood et al., “LDA ANFIS based intelligent framework for real time dengue risk monitoring,” Inf., vol. 15, no. 11, 2020.

[32] W. L. Al Yaseen et al., “DE + ELM for feature selection in intrusion detection,” in Proceedings, 2022.

[33] V. Varalakshmi and A. Lopez, “Enhanced Dengue Prediction in Tamil Nadu Using Meteorological and Entomological Data,” arXiv preprint arXiv:2306.13456, 2023.

[34] O. Oguntimilehin et al., “ML in tropical disease diagnosis,” J. Trop. Dis., 2019.

[35] D. Raja Aswathi and R. Sundararajan, “Optimal Feature Selection Using a Modified Differential Evolution Algorithm for Disease Prediction,” Cluster Computing, vol. 22, pp. 11979–11990, 2019.