

# st20212770CIS6035Writ1.docx

*by* Mahasen Ruhunuhewa Savindu

---

**Submission date:** 29-Apr-2024 08:48AM (UTC+0100)

**Submission ID:** 231256668

**File name:** 93787\_Mahasen\_Ruhunuhewa\_Savindu\_st20212770CIS6035Writ1\_2154672\_1135350804.docx  
(902.81K)

**Word count:** 5277

**Character count:** 31640

# Chapter 1

## Introduction

Dengue fever poses a persistent threat to public health in Sri Lanka, exacting a heavy toll on individuals and the healthcare system alike. With its tropical climate fostering the proliferation of the Aedes mosquito—the primary vector for dengue transmission—the country contends with recurrent outbreaks of the disease, precipitating widespread illness, economic strain, and societal disruption (Malavige et al., Changing Epidemiology of Dengue in Sri Lanka—challenges for the future 2021).

Conventional methods of diagnosing dengue hinge largely on clinical assessment, a process fraught with challenges due to the disease's variable and overlapping symptoms. This diagnostic complexity frequently results in delays in treatment initiation, heightening the risk of severe complications and mortality. Moreover, the sheer volume of dengue cases overwhelms healthcare facilities, exacerbating resource constraints and impeding effective disease management (Raafat et al., A review of Dengue diagnostics and implications for surveillance and Control 2019).

Recognizing the imperative for more efficient and precise diagnostic tools, there has been a burgeoning interest in harnessing machine learning technologies for dengue detection. Machine learning, a facet of artificial intelligence, holds promise in scrutinizing extensive and diverse datasets to unveil hidden patterns and insights. By leveraging this computational prowess, machine learning algorithms can facilitate early detection of dengue, stratification of risk, and allocation of resources, thereby ameliorating clinical outcomes and optimizing resource utilization (Saturi, Development of prediction and forecasting model for dengue disease using machine learning algorithms 2020).

This project endeavors to craft a bespoke machine learning-driven dengue detection system tailored explicitly to the sri lankan milieu. Through the integration of disparate datasets encompassing clinical, epidemiological, and environmental variables, our objective is to fashion predictive models adept at accurately discerning dengue cases .

## **Research Question**

"How machine learning techniques can effectively predict dengue fever cases among patients in Sri Lanka ?"

## **Problem statement**

- Based on symptoms, machine learning existing applications or systems are fully created for health care professionals, doctors, and other professionals. At that moment there is no specific machine learning application or system is available for patient to detect the dengue by their own smart device.
- Lot of urban areas don't have laboratory facilities for earlier detecting of dengue. Based on that urban area people need to go long distance for meet the doctor or health care professional to check the dengue or not. So at that time dengue fever can be highly increase for them. Based on this reason web app is needed for detecting the dengue fever. Because when using the web app, can be easily access and detect the dengue at earlier stage

## **Research Objective**

Explore Patient Needs and Challenges in Dengue Diagnosis

- Investigate the requirements and difficulties patients face regarding dengue diagnosis, particularly in urban settings where laboratory access is limited. Explore the obstacles patients encounter in obtaining prompt and accurate dengue diagnosis and their potential impact on public health.

## Chapter2

### Literature Review

#### Introduction

In now a days dengue is a common health issue in the world. Because dengue is spread though the word faster. Especially in srilanka. For an example in srilanka in may 2023 61,361 total dengue patients are reported (Sri Lanka: Dengue Outbreak - May 2023). Earlier days lot of laboratory testing methods were doing for detecting the dengue. For example <sup>19</sup> Reverse transcriptase-polymerase chain reaction (RT-PCR), viral isolation and identification, nuclear acid sequence-based amplification (NASBA), IgM and IgG seroconversion, PCR testing are used. But these tests have several disadvantages. for example <sup>35</sup> accuracy of results, cost, complexity, and takes lot of time to get the result etc (Wong et al., <sup>35</sup> Diagnosis of severe dengue: Challenges, needs and opportunities 2020). Based on these disadvantages different systems, and machine learning techniques already created for dengue detection. But now a days most of these created machine learning systems or applications, and normal application, and systems are created for assisting to doctors and healthcare professional to identify the patient has dengue or not. Which means there is no specifically created the any particular machine learning system or application for patient to dengue identify. In this literature review mainly focused on what are the already existing application, systems by using machine learning techniques and their problems. And also consider the other normal application, or systems. And their problems.

#### Problems

##### Dengue prediction and diagnosis using machine learning techniques

According to this <sup>1</sup> “A dengue disease prediction and diagnosis model using sentiment analysis and machine learning algorithms” research paper, for the predicting the dengue <sup>27</sup> KNN classifier, decision tree, random forest, Gaussian naive Bayes, and support vector classifier (SVC)(Gupta et al., <sup>1</sup> DDPM: A dengue disease prediction and diagnosis model using sentiment analysis and machine learning algorithms 2023).

### **The main limitations of this application**

Unbalance data set - when it comes to the dengue dataset. Where that contains positive cases of dengue may be rare compare to the negative cases.

Hyper parameter tuning problem - when using this models for increasing the model accuracy need to take lot of time to find out the best parameters.

Scalability - when using this KNN classifier, random, SVC are high expensive when we are dealing with larger amount of dataset (Jamwal & Bhatia, <sup>30</sup>Prevalence of vector borne diseases in Jammu division, Jammu and Kashmir, India 2021).

### **Classification of dengue application using machine learning techniques**

According to the <sup>45</sup>“Classification of dengue using machine learning techniques” research paper, Simple Cart, C-4.5, Multi-layer <sup>1</sup>perception algorithms used for classifying the dengue is infected or not (Sajana et al., Classification of dengue using Machine Learning Techniques 2018).

### **The main limitations of this models**

- Overfitting - These simple cart, C-4.5, willing to overfitting, when trees becomes to the more deep, and complex. But this leads to generate the inaccurate predictions.
- Sensitive to small changes - these simple cart, C-4.5 algorithms highly respond to the training dataset. That leading to the different tree structure and potentially different predictive performance.
- Imbalance dataset - when it comes to the dengue dataset. Where that contains positive cases of dengue may be rare compare to the negative cases.
- Privacy and security problems - extracting the human sensitive information raise some privacy problems.

- Model accuracy - According to the <sup>33</sup> “Dengue outbreaks prediction in bangladesh perspective using distinct multi-layer perceptron NN and decision tree” research paper. Predicting accuracy of Multi-layer perceptron <sup>7</sup> 68.5%. which means it not sufficient for using the predicting purpose (Khan et al., Dengue outbreaks prediction in Bangladesh perspective using distinct multilayer Perceptron NN and decision tree 2022).

#### <sup>44</sup> Presumptive diagnosis System for dengue fever by using machine learning

Presumptive diagnosis system used the <sup>5</sup> Decision Tree, Random Forest, Naive Bayes algorithms used for predicting the dengue fever. What ever this Random Forest , Naive Bayes, Decision Tree algorithms use for early detection of dengue in this system (Khan & Raza, <sup>5</sup> Development and evaluation of a predictive diagnostic system for dengue fever using Machine Learning Techniques 2023).

#### The main Limitation of this system

- Less number of data points are available
- Accuracy of this models are very low
- Sensitivity - presumptive diagnosis system is lack with sensitivity for early stage of detecting (Khan & Raza, <sup>5</sup> Development and evaluation of a predictive diagnostic system for dengue fever using Machine Learning Techniques <sup>2</sup> 2023).

#### Detection of dengue disease by using fused machine learning

<sup>1</sup> In this case use the PFDM model. In this PFDM model use two main ML-based procedures. For an example SVM and ANN. And this PFDM has two basic components. The training layer and the testing layer. In the training layer,

It has five steps <sup>2</sup> (collecting, preprocessing, classification, efficiency, and machine-level combination). If the model fails to meet learning criteria, it is reassigned until satisfactory results are achieved. <sup>2</sup> The outputs of ANN and SVM are then fused, and

<sup>2</sup> the trained model is stored in a cloud system. In the testing layer, <sup>2</sup> data is obtained from a database, and preprocessed trained models are loaded from the cloud.

The fused model is used to predict whether a Dengue diagnosis is progressive or destructive <sup>1</sup> with the predicted outcome compared to the actual result to measure accuracy (Al Nasar et al., Detection of dengue disease empowered with fused machine learning 2022).

### The main limitation

- One important factor in classification is class imbalance, and classification efficiency cannot be entirely determined by accuracy rates alone (Al Nasar et al., Detection of dengue disease empowered with fused machine learning 2022).
- Feature engineering - SVM (Support vector machine) is heavily depend on feature engineering to identify the relevant patterns (Hoyos et al., Dengue models based on machine learning techniques: A systematic literature review 2021).
- Complexity and overfitting - when dealing with the noisy and small datasets. Overfitting can generate unreliable predictions (Hoyos et al., Dengue models based on machine learning techniques: A systematic literature review 2021).
- Computational resources - when using ANN it required the computational resources including high performance GPU, and memory (Chakraborty et al., <sup>22</sup> Forecasting dengue epidemics using a hybrid methodology 2019).

### <sup>22</sup> Decision support system for dengue detection

The use of decision support systems (DSS) is essential for the early diagnosis of dengue. A viable method for creating DSS for dengue care is case-based reasoning (CBR), which uses clinical guidelines and historical cases <sup>18</sup> to forecast a patient's present state based on their vital signs and symptoms (binti Mohd Zainee & Chellappan, A preliminary dengue fever prediction model based on vital signs and blood profile 2020).

### **Limitation of decision support system**

- Imbalance dataset - which means dengue infected positive cases significantly are lower than negative cases.
- Data is limited - dengue dataset size is small .
- Complexity of system - Dengue has several symptoms. That symptoms can be overlapped with similar illness (Lopez et al., An intelligent decision support system to prevent and control of Dengue 2018).

### **Solution**

According to the above mentioned machine learning systems, applications and normal systems like decision support system there are significant limitations are available. Base on that machine learning system, applications limitations are complexity of model and system, low accuracy of models, overfitting, unbalance dataset etc. And also above mention systems are mainly created for doctor and healthcare professional. Then some urban area people need to go long distance for meet the doctor or laboratories for detecting the dengue is infected or not. According to that urban area people cannot be identify the dengue in earlier stage. These are the major problems existing in machine learning dengue systems and application. When it comes to the normal systems like decisions support system there is some issue like complexity of the system, data limitation, inbalance dataset available.

Based on above mentioned machine learning systems, applications and normal systems like decision support systems problems or limitations, dengue detection system is going to be implemented for patient by using decision tree, random forest, logistic regression algorithm while including the “grid\_search cv” for finding the best parameters, “ensemble technique” for generate the combined model, proper dataset and other models parameter like random state.



## Chapter 3

### Methodology

#### Technologies

For this dengue detection project by using machine learning used flask (python frame work) for back end, HTML, CSS, Bootstrap , and JavaScript for front end. And also testing for each software component of this application use the pytest( python framework).

#### Data Collection

For the dengue detection system based on symptoms by using machine learning technique, data sets are received by websites for example keggale website (Find open datasets and Machine Learning Projects 2024)

#### Splitting the dataset

This dengue dataset is going to be divided into the two part by using “random splitting function” Training set and testing set. Training set size is 80%. Test set size is 20%. This training set is used for model training and testing is used for evaluate model performance.

#### Data Cleaning

Checking the null values in created training and test dataset.

### <sup>43</sup> Split the dataset into the training and validation set

After spited <sup>7</sup> the prepared data set into the training and testing dataset, this training set split into the training and validation test again .

### Model selection and explanation of machine learning models used

For classifying this dengue, binary <sup>17</sup> logistic regression, decision tree algorithm, random forest algorithms are the models going to be used.

Binary logistic regression - According to this <sup>3</sup> “Machine learning models for early dengue severity prediction” research paper, Binary logistic regression is chosen as a model for dengue prediction due to its simplicity and ease of interpretation.

Unlike more complex models, logistic regression offers a clear understanding of <sup>40</sup> the relationship between independent variables and the binary outcome of dengue presence. Its straightforward nature makes it particularly valuable in medical research, where comprehensibility is crucial. Logistic regression is also adept at handling large datasets efficiently, making it suitable for the extensive data often encountered in dengue prediction studies. Furthermore, it accommodates both continuous and categorical predictor variables, allowing for the incorporation of diverse data types commonly found in such studies. By estimating the probability of dengue presence based on various <sup>47</sup> symptoms and risk factors, logistic regression serves as a valuable tool in the array of machine learning techniques for dengue prediction (Caicedo-Torres et al., <sup>3</sup> Machine learning models for early dengue severity prediction 2020)).

Decision tree model - According to this <sup>1</sup> “Machine Learning for Dengue Outbreak Prediction: A performance evaluation of different prominent classifiers” research paper, The decision tree model is a key component of our dengue detection system for several compelling reasons. Firstly, decision trees are renowned for their <sup>21</sup> simplicity and ease of use, making them accessible to a wide range of users, including healthcare professionals and public health officials. This simplicity

translates into ease of interpretation, allowing stakeholders to grasp the underlying decision-making process intuitively. Additionally, decision trees have demonstrated the potential to achieve high levels of accuracy in dengue prediction tasks. By efficiently partitioning the feature space into homogeneous regions, decision trees can capture complex relationships between environmental factors, demographic variables, and dengue incidence.

This ability to model intricate interactions empowers our system to generate more accurate predictions, enabling proactive interventions and resource allocation. Moreover, decision trees excel in producing interpretable rules that align with domain knowledge, facilitating actionable insights for disease prevention and control strategies.

Overall, the decision tree model represents a valuable asset in our quest to combat dengue fever effectively and mitigate its impact on public health. (Iqbal & Islam, Machine Learning for Dengue Outbreak Prediction: A performance evaluation of different prominent classifiers 2019)

Random forest model -

According to this “prediction of dengue using machine learning algorithm” research paper, random forest is a cornerstone of dengue detection system due to its multifaceted advantages. Firstly, random forest excels in enhancing model accuracy by leveraging the wisdom of crowds. By aggregating predictions from multiple decision trees trained on different subsets of the data, random forest mitigates the risk of overfitting and variance, resulting in robust and reliable predictions. This ensemble approach also enhances the model's generalization capability, enabling it to perform well on unseen data. Additionally, random forest is adept at handling large and complex datasets commonly encountered in dengue prediction tasks. Its inherent Parallelization and scalability make it well-suited for processing vast amounts of environmental, epidemiological, and clinical data, thereby facilitating comprehensive analysis and inference.

Furthermore, random forest offers built-in mechanisms for feature importance assessment, enabling us to identify the most informative variables driving dengue incidence prediction. This interpretability fosters a deeper understanding of the disease dynamics and informs targeted intervention strategies. Overall, the versatility,

robustness, and scalability of random forest make it an indispensable tool in our efforts to combat dengue fever effectively and safeguard public health. (Sarwar & Al Mamun,<sup>36</sup> Prediction of dengue using machine learning algorithms: Case study Dhaka 2022)

Ensemble technique to generate combine model including the random forest, decision tree classifier, and binary logistic regression -

According to this “<sup>3</sup> Machine learning algorithms for dengue risk assessment: A case study for São Luís do Maranhão ” research paper,<sup>17</sup> binary logistic regression, decision tree regression, random forest models are combined and created the model. It is known as combined model. This combined model advantages are easy of use, wide range of users, and this models is good for reducing the overfitting issues (Rocha & Giesbrecht,<sup>3</sup> Machine learning algorithms for dengue risk assessment: A case study for São Luís do Maranhão 2022).

When using above mentioned random forest model, binary logistic regression model, and decision tree model overfitting issue raised. But when using combined model overfitting issue is fixed. Based on that combined model is final model for making the dengue predictions.

## Techniques

In this methodology, utilize advanced techniques such as hyperparameter tuning and grid search CV to optimize the performance of our models. These techniques involve identifying the most effective parameters for binary logistic regression, random forest, and decision tree algorithms, thereby fine-tuning the models for dengue prediction.

## Parameters

Binary Logistic regression model best parameters  
C, penalty, l2, solver

<sup>37</sup>  
Decision tree classifier model best parameters  
max\_depth, and min\_samples\_split

<sup>11</sup>  
Random forest classifier model best parameters  
Max\_depth, max\_features, min\_samples\_leaf, min\_samples\_split, n\_estimators

By systematically exploring the above mentioned parameters, aim to enhance model accuracy and mitigate the risk of overfitting.

Additionally, employ ensemble methods to further improve model robustness and generalization. Ensemble techniques involve combining predictions from multiple models, such as random forest, decision tree, and binary logistic regression that choose for dengue prediction.

By leveraging the strengths of each individual model and mitigating their weaknesses, ensemble methods enable us to achieve superior predictive performance while reducing the likelihood of overfitting.

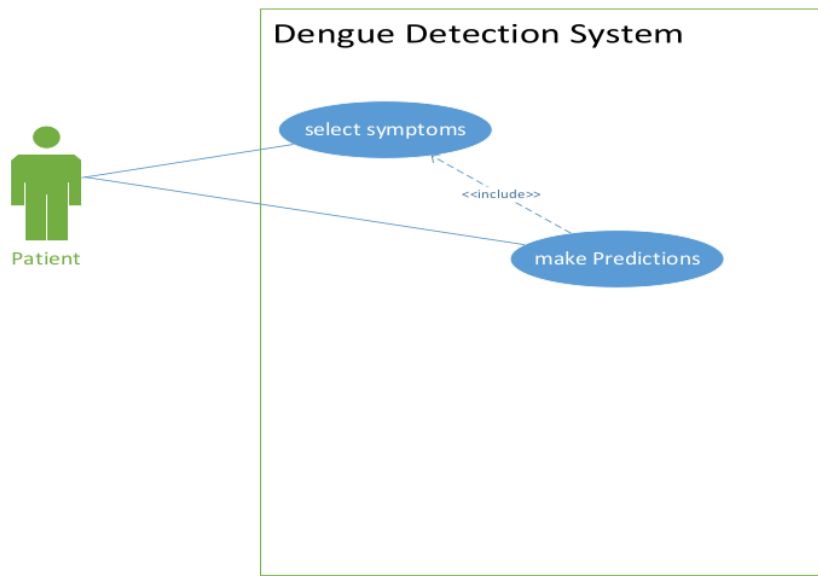
Overall, our approach integrates these advanced techniques within our methodology to optimize model performance, enhance accuracy, and ensure robust dengue prediction capabilities.

### **Model deployments**

After implementing ensemble techniques, the ensemble voting classifier (combined model) emerged as the final model for dengue prediction. To make this model accessible for real-world use, created a pickle file to store its parameters and import it into our application. With flask framework, seamlessly integrated this model into the back end, and allowing it to interact with the front end interface. This integration ensures that our dengue detection system is user-friendly and readily available for deployment, empowering users to make informed decisions about dengue risk and management.

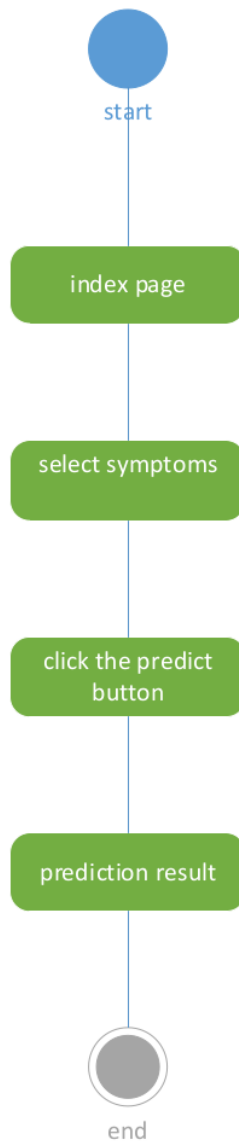
## Designing

### Use case diagram



According to the above use case diagram patient is the actor select symptoms, and make prediction are the activities. In between select symptoms and make predictions activities relationship is include. Which means to make the predictions for patient patient first of all need to select the symptoms.

### Activity Diagram

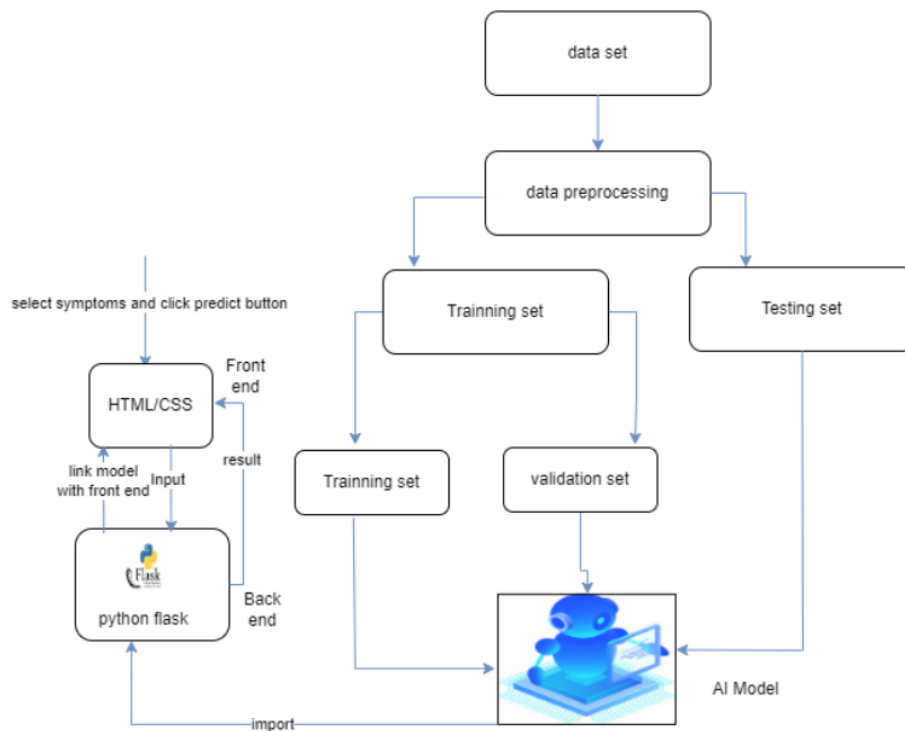


According to the above sequence diagram it describes the flow of system. Which means first of patient redirect to the index page and select the symptoms and get the prediction result.



## Architectural Diagram

### Architectural Diagram For Dengue prediction system



According to the above architectural diagram after training the and testing final AI model, that AI model is linked with UI by using flask. Which means flask is used for back end. Then if patient select the symptoms according to that symptoms prediction result display.

## Chapter 4

### Testing

#### Model results and final model evaluation

According to this dengue prediction project after doing the grid search cv for hyper parameter tuning, models metrices (accuracy, F1 score, precision, recall, F1 score) which means model evaluations are shown in below images.

```
Best Parameters: {'C': 0.1, 'penalty': 'l2', 'solver': 'lbfgs'}
Logistic Regression Classifier Metrics:
Accuracy: 1.0000
Precision: 1.0000
Recall: 1.0000
F1 Score: 1.0000
```

```
Classification Report:
      precision    recall  f1-score   support

     0       1.00      1.00      1.00       773
     1       1.00      1.00      1.00        15

 accuracy          1.00          1.00          1.00          788
 macro avg          1.00          1.00          1.00          788
weighted avg          1.00          1.00          1.00          788
```

```
Confusion Matrix:
[[773  0]
 [ 0 15]]
Logistic regression Validation Accuracy: 1.0
Logistic regression Mean Squared Error (MSE): 0.0000
Logistic regression Root Mean Squared Error (RMSE): 0.0000
```

---

```
Best Parameters: {'max_depth': 10, 'min_samples_split': 2}
DecisionTree Classifier Metrics:
Accuracy: 1.0000
Precision: 1.0000
Recall: 1.0000
F1 Score: 1.0000
```

```
Classification Report:
      precision    recall  f1-score   support

     0       1.00      1.00      1.00      773
     1       1.00      1.00      1.00       15

   accuracy          1.00      1.00      1.00      788
  macro avg          1.00      1.00      1.00      788
 weighted avg          1.00      1.00      1.00      788
```

```
Confusion Matrix:
[[773  0]
 [  0 15]]
Decision Tree Validation Accuracy: 1.0
Decision Tree Mean Squared Error (MSE): 0.0000
Decision Tree Root Mean Squared Error (RMSE): 0.0000
```

---

```
Random Forest Classifier Metrics:
Accuracy: 1.0000
Precision: 1.0000
Recall: 1.0000
F1 Score: 1.0000
```

```
Classification Report:
      precision    recall  f1-score   support

     0       1.00      1.00      1.00      773
     1       1.00      1.00      1.00       15

   accuracy          1.00      1.00      1.00      788
  macro avg          1.00      1.00      1.00      788
 weighted avg          1.00      1.00      1.00      788
```

```
Confusion Matrix:
[[773  0]
 [  0 15]]

Random Forest Mean Squared Error (MSE): 0.0000
Random Forest Root Mean Squared Error (RMSE): 0.0000
```

After doing the grid search cv and hyperparameter tuning model overfitting issue doesn't fix. Because above random forest, decision tree, binary logistic regression accuracies are still 1.0 (100%). So then, to avoid that overfitting issue use the ensemble technique to created combined model. Below images is the example for combined model metrices (accuracy, precision, recall, F1 score). which means model evaluation after using the ensemble technique.

```
Ensemble (Voting) Classifier Metrics:  
Accuracy: 0.9048  
Precision: 0.9810  
Recall: 0.9048  
F1 Score: 0.9360
```

```
Classification Report:  
              precision    recall  f1-score   support  
  
     0           1.00        0.91        0.95        773  
     1           0.15        0.87        0.26         15  
  
   accuracy          0.90        0.90        0.90        788  
  macro avg          0.57        0.89        0.60        788  
 weighted avg          0.98        0.90        0.94        788
```

```
Confusion Matrix:  
[[700  73]  
 [  2  13]]
```

```
Ensemble voting classifier Mean Squared Error (MSE): 0.0952  
Ensemble voting classifier Root Mean Squared Error (RMSE): 0.3085
```

When we consider the final model evaluation <sup>15</sup> random forest, binary <sup>15</sup> logistic regression, <sup>15</sup> decision tree models accuracy is 1.0 (100%), and <sup>15</sup> precision, recall, F1 score are 1.0 (100%). Mean squared error and root means squared error of this three models are 0.0000. which means overfitting. combine model by using ensemble technique <sup>34</sup> metrics are <sup>34</sup> accuracy is 0.9048 (90%), <sup>34</sup> precision is 0.9810 (98%), <sup>34</sup> recall is 0.9048 (90%), and <sup>34</sup> F1 score is 0.9360 (94%). Mean squared error of this combine model is 0.0952. And root means squared error of this combined model is 0.3085. So then final <sup>11</sup> and best model is combined model by combining the binary <sup>11</sup> logistic regression, <sup>11</sup> decision tree classifier, random forest classifier.

## Unit Testing

<sup>6</sup> Unit testing is a software testing method where individual units or components of a software application are tested in isolation to ensure they perform as expected. In Python, unit testing is commonly used to verify the correctness of functions, classes, or modules (Lukasczyk et al., <sup>31</sup> An empirical study of Automated Unit Test Generation for Python 2023). This Unit testing has three main types

### 1. White box testing

<sup>8</sup> In white-box testing, the tester has access to the internal structure, design, and implementation details of the software being tested. Test cases are designed based on an understanding of the code logic, paths, and data flows. The objective is <sup>6</sup> to ensure that the code behaves as expected by testing its internal workings (<sup>4</sup> Ehmer & Khan, A comparative study of white box, black box and grey box testing techniques 2019)

### <sup>16</sup> 2. Black box testing

In black-box testing, the tester treats the software as a "black box" and does not have access to its internal structure or implementation details. Test cases are designed based on the software's specifications, requirements, and external behavior. The focus is on testing the functionality, inputs, and outputs of the software without considering its internal workings (<sup>8</sup> Ehmer & Khan, A comparative study of white box, black box and grey box testing techniques 2019)

### <sup>13</sup> 3. Grey box testing

Grey-box testing is a combination of white-box and black-box testing. The tester has partial access to the internal structure and design of the software while also testing its external behavior. This approach allows the tester to <sup>42</sup> design test cases based on

both the internal logic and the external requirements of the software. Grey-box testing can provide a balance between the thoroughness of white-box testing and the independence of black-box testing (Ehmer & Khan, A comparative study of white box, black box and grey box testing techniques 2019).

#### **Benefits of Unit Testing:**

- **Early Detection of Bugs:** Unit tests allow you to catch bugs and issues in the early stages of development, making them easier and cheaper to fix.
- **Regression Testing:** Unit tests serve as a safety net, ensuring that changes made to the code base don't introduce new bugs or regressions.
- **Improved Code Quality:** Writing tests encourages writing modular, decoupled, and reusable code, which leads to better overall code quality (Mårtensson, Unit testing 2019).

In this unit testing there are several testings are available. One of the best testing is pytest that is python framework. It belongs to the white box testing. Because White-box testing is like peeking inside a clock to check if all the gears are turning as they should. So then when using the pytest tester knows how the software works on the inside, checking if every part of the code functions correctly (A research paper on white box testing 2022).

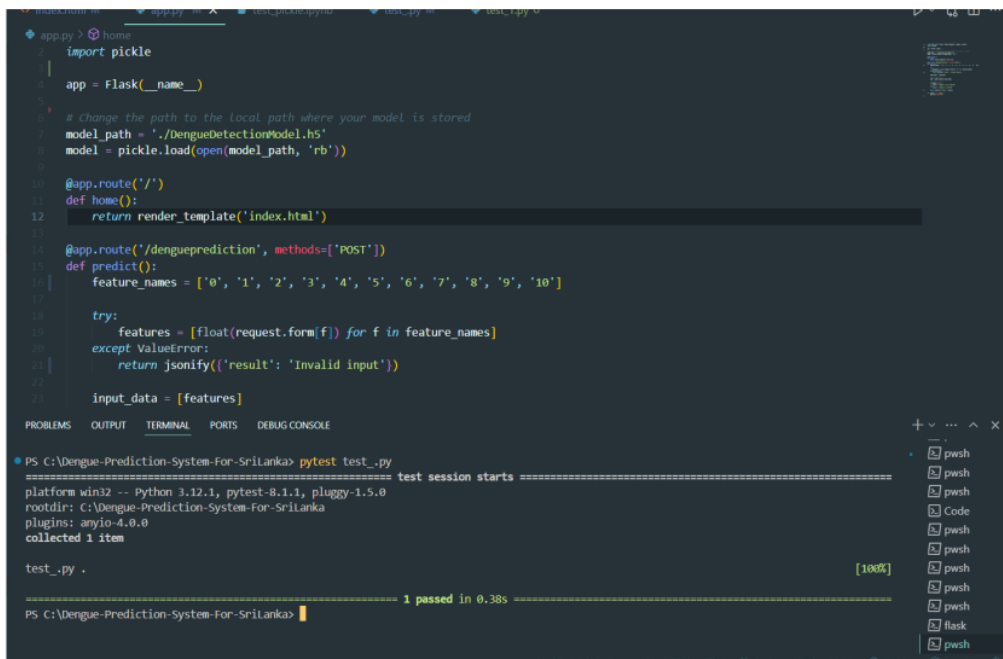
#### **Key feature of pytest**

- **Simple Syntax:** pytest offers a simple and intuitive syntax for writing test functions, making it easy to get started with writing tests.
- **Fixture Support:** pytest provides a powerful fixture mechanism for setting up and tearing down resources needed by your tests.

- Parameter Testing: pytest allows you to easily parameterize your tests, enabling you to run the same test with multiple input values.
- Plugins: pytest is highly extensible through plugins, allowing you to customize and extend its functionality to suit your needs (Pajankar, Pytest 2021).

According to the above key features, pytest is suitable for this project.

#### Test case 1 - Dengue is infected



The screenshot shows a VS Code editor with a Flask application file named `app.py` and a terminal window displaying the output of a pytest test session.

**app.py code:**

```
import pickle

app = Flask(__name__)

# Change the path to the Local path where your model is stored
model_path = './DengueDetectionModel.h5'
model = pickle.load(open(model_path, 'rb'))

@app.route('/')
def home():
    return render_template("index.html")

@app.route('/dengueprediction', methods=['POST'])
def predict():
    feature_names = ['0', '1', '2', '3', '4', '5', '6', '7', '8', '9', '10']

    try:
        features = [float(request.form[f]) for f in feature_names]
    except ValueError:
        return jsonify({'result': 'Invalid input'})

    input_data = [features]
```

**Terminal output:**

```
PS C:\Dengue-Prediction-System-For-Srilanka> pytest test_.py
platform win32 -- Python 3.12.1, pytest-8.1.1, pluggy-1.5.0
rootdir: C:\Dengue-Prediction-System-For-Srilanka
plugins: anyio-4.0.0
collected 1 item

test_.py . [100%]

===== 1 passed in 0.38s =====
PS C:\Dengue-Prediction-System-For-Srilanka>
```

Test case 1 is passed

## Test case 2 -Dengue is not infected

```
import pytest
import requests

# Define the URL of your Flask application
BASE_URL = 'http://127.0.0.1:5000'

# Sample input data for testing
SAMPLE_INPUT_DATA = {
    '0': '1', '1': '0', '2': '0', '3': '0', '4': '0', '5': '0', '6': '0', '7': '0', '8': '0', '9': '0', '10': '0'
}

def test_predict_route(capsys):
    # Send a POST request to the predict route with sample input data
    response = requests.post(f'{BASE_URL}/dengueprediction', data=SAMPLE_INPUT_DATA)

    # Check if the response is successful (status code 200)
    assert response.status_code == 200

    # Extract the result from the response text
    result = response.text.strip() # Remove leading/trailing whitespace

    # Print the result to the terminal

C:\Dengue-Prediction-System-For-Srilanka> pytest test_1.py
===== test session starts =====
platform win32 -- Python 3.12.1, pytest-8.1.1, pluggy-1.5.0
rootdir: C:\Dengue-Prediction-System-For-Srilanka
plugins: anyio-4.0.0
collected 1 item

test_1.py . [100%]

===== 1 passed in 0.49s =====
C:\Dengue-Prediction-System-For-Srilanka>
```

## Test case 2 is passed

## Test case 3 - dengue is not infected

```
test_2.py:
1 import pytest
2 import requests
3 from bs4 import BeautifulSoup
4
5 # Define the URL of your Flask application
6 BASE_URL = 'http://127.0.0.1:5000'
7
8 # Sample input data for testing
9 SAMPLE_INPUT_DATA = {
10     '0': '1', '1': '0', '2': '0'
11 }
12
13 def test_predict_route():
14     # Send a POST request to the predict route with sample input data
15     response = requests.post(f'{BASE_URL}/dengueprediction', data=SAMPLE_INPUT_DATA)
16     response = requests.post(f'{BASE_URL}/dengueprediction', data=SAMPLE_INPUT_DATA)
17     print(response.content)
18
19     # Check if the response is successful (status code 200)
20     assert response.status_code == 200
21
22     # Parse the HTML response
23
24     # Check if the response is successful (status code 200)
25     assert response.status_code == 200
26     E assert 400 == 200
27     + where 400 = <Response [400]>.status_code
28
29 test_2.py:20: AssertionError
30
31 ----- Captured stdout call -----
32 b'<doctype html>\nhtml lang=en>\n<title>400 Bad Request</title>\n<h1>\n<p>The browser (or proxy) sent a request that this ser-
33 ver could not understand.</p>\n'
34
35 ===== short test summary info =====
36 FAILED test_2.py::test_predict_route - assert 400 == 200
37
38 ===== 1 failed in 0.47s =====
39 PS C:\Dengue-Prediction-System-For-Srilanka>
```



Test case 3 is failed

According to the above mentioned test case 3, other test cases which have not completed input fields values like test case 3, that test cases are fails like test case 3. The reasons of the test cases are fail is assertion error.

### Sample UI Images

**DENGUE PREDICTION**

Please select all symptoms

\*Skin Rash: Select value

\*Chills: Select value

\*High Fever: Select value

\*Headache: Select value

\*Nausea: Select value

\*Loss Of Appetite: Select value

\*Pain Behind of Eyes: Select value

\*Back Pain: Select value

\*Malaise: Select value

\*Muscle Pain: Select value

\*Red Spots: Select value

Predict

## DENGUE PREDICTION

Please select all symptoms

*Skin Rash:	*Chills:
No	No
*High Fever:	*Headache:
No	No
*Nausea:	*Loss Of Appetite:
No	No
*Pain Behind of Eyes:	*Back Pain:
No	No
*Malaise:	*Muscle Pain:
Yes	Yes
*Red Spots:	
Yes	

Predict

Prediction Result
✕

Dengue Not Infected

## DENGUE PREDICTION

Please select all symptoms

*Skin Rash:	*Chills:
Yes	No
*High Fever:	*Headache:
Yes	No
*Nausea:	*Loss Of Appetite:
Yes	No
*Pain Behind of Eyes:	*Back Pain:
Yes	No
*Malaise:	*Muscle Pain:
Yes	No
*Red Spots:	
Yes	

Predict

Prediction Result
✕

Dengue Infected

## Chapter 5

### Conclusion and Recommendation

#### <sup>32</sup> Conclusion

In conclusion, this research explores the application of machine learning algorithms for dengue detection in Sri Lanka, a country where the disease poses a significant public health challenge. Through the utilization of various machine learning techniques such as classification algorithms, feature selection, and ensemble methods, aimed to develop accurate <sup>38</sup> predictive models for early detection of dengue cases.

Based on the findings demonstrate that machine learning models, particularly those based on binary logistic regression, decision tree classifier, and ensemble method such as random forest exhibit promising performance in dengue prediction.

These models leverage diverse sets of features including symptoms to accurately classify individuals at risk of contracting dengue fever.

Furthermore, feature selection techniques such as doing the correlational coefficient have been instrumental in identifying the most relevant predictors for dengue detection, thereby enhancing the efficiency and Interpretability of the models. And also, it is essential to acknowledge several limitations and challenges encountered during the course of this study. These include the availability and quality of data, the dynamic nature of environmental factors influencing dengue transmission, and the need for real-time data integration to enhance the predictive accuracy of the models.

Despite these challenges, the results of this research hold significant implications for public health authorities and policymakers in Sri Lanka. By leveraging machine learning algorithms for dengue detection, it is possible to improve the early warning systems, allocate resources more efficiently, and implement targeted interventions to mitigate the burden of dengue fever in the country.

Moving forward, future research directions may involve the integration of additional data sources such as satellite imagery, climate forecasts, and social media data to enhance the predictive capabilities of the models. Additionally, the development of user-friendly and scalable predictive tools tailored to the needs of healthcare practitioners and decision-makers could facilitate the adoption of machine learning-based approaches in routine dengue surveillance and control efforts.

In conclusion, while there are challenges to be addressed, the application of machine learning in dengue detection offers tremendous potential to advance our understanding of disease dynamics and ultimately improve public health outcomes in Sri Lanka.

### **Recommendation**

Currently, the dengue prediction system focuses on identifying whether an individual is infected with dengue fever at an early stage. However, future iterations of the system aim to expand its capabilities to categorize the specific serotype of the dengue virus, namely dengue virus serotype 1 (DENV-1), dengue virus serotype 2 (DENV-2), dengue virus serotype 3 (DENV-3), and dengue virus serotype 4 (DENV-4). This expansion is based on research findings reported in Nature news in 2022(Nature new, 2022). And also, in srilanka some areas, people don't have good knowledge about the English. So then this dengue detection system cannot be used without having good English knowledge. As result as in future iteration this system is going to implemented based on the sinhala and tamil language. Because By incorporating these features into the system, it can provide more detailed and precise information about the specific type of dengue virus affecting an individual, enabling more targeted and effective treatment strategies.

## References

- Malavige, G.N. *et al.* (2021) 'Changing Epidemiology of Dengue in Sri Lanka—challenges for the future', *PLOS Neglected Tropical Diseases*, 15(8). doi:10.1371/journal.pntd.0009624.
- Raafat, N., Blacksell, S.D. and Maude, R.J. (2019) 'A review of Dengue diagnostics and implications for surveillance and Control', *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 113(11), pp. 653–660. doi:10.1093/trstmh/trz068.
- Saturi, S. (2020a) 'Development of prediction and forecasting model for dengue disease using machine learning algorithms', *2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)* [Preprint]. doi:10.1109/discover50404.2020.9278079.
- Sri Lanka: Dengue Outbreak - May 2023* (no date) *ReliefWeb*. Available at: <https://reliefweb.int/disaster/ep-2023-000084-lka> (Accessed: 28 February 2024).
- Pruszczyk, P. (2018) 'Diagnosis: Clinical prediction rules and laboratory tests', *ESC CardioMed*, pp. 2758–2761. doi:10.1093/med/9780198784906.003.0658.
- Gupta, G. *et al.* (2023) 'DDPM: A dengue disease prediction and diagnosis model using sentiment analysis and machine learning algorithms', *Diagnostics*, 13(6), p. 1093. doi:10.3390/diagnostics13061093.
- Wong, P.-F., Wong, L.-P. and AbuBakar, S. (2020) 'Diagnosis of severe dengue: Challenges, needs and opportunities', *Journal of Infection and Public Health*, 13(2), pp. 193–198. doi:10.1016/j.jiph.2019.07.012.
- Jamwal, N. and Bhatia, S. (2021) 'Prevalence of vector borne diseases in Jammu division, Jammu and Kashmir, India', *Journal of Vector Borne Diseases*, 58(1), p. 28. doi:10.4103/0972-9062.321746.

- Khan, Md.A. *et al.* (2022) 'Dengue outbreaks prediction in Bangladesh perspective using distinct multilayer Perceptron NN and decision tree', *Health Information Science and Systems*, 10(1). doi:10.1007/s13755-022-00202-x.
- Sajana, T. *et al.* (2018) 'Classification of dengue using Machine Learning Techniques', *International Journal of Engineering & Technology*, 7(2.32), p. 212. doi:10.14419/ijet.v7i2.32.15570.
- Khan, J.R. and Raza, S.K. (2023) *Development and evaluation of a predictive diagnostic system for dengue fever using Machine Learning Techniques* [Preprint]. doi:10.21203/rs.3.rs-2473833/v2.
- Al Nasar, M.R. *et al.* (2022) 'Detection of dengue disease empowered with fused machine learning', *2022 International Conference on Cyber Resilience (ICCR)* [Preprint]. doi:10.1109/iccr56254.2022.9996009.
- Hoyos, W., Aguilar, J. and Toro, M. (2021) 'Dengue models based on machine learning techniques: A systematic literature review', *Artificial Intelligence in Medicine*, 119, p. 102157. doi:10.1016/j.artmed.2021.102157.
- Chakraborty, T., Chattopadhyay, S. and Ghosh, I. (2019) 'Forecasting dengue epidemics using a hybrid methodology', *Physica A: Statistical Mechanics and its Applications*, 527, p. 121266. doi:10.1016/j.physa.2019.121266.
- binti Mohd Zainee, N. and Chellappan, K. (2020) 'A preliminary dengue fever prediction model based on vital signs and blood profile', *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)* [Preprint]. doi:10.1109/iecbes.2016.7843530.
- Lopez, D. *et al.* (2018) 'An intelligent decision support system to prevent and control of Dengue', *Journal of Ambient Intelligence and Humanized Computing* [Preprint]. doi:10.1007/s12652-018-0829-9.
- Jayasundara, S.D.P., Perera, S.S.N. and Rathnayaka, N.S. (2017) 'Developing a decision support testing algorithm to detect severity level of Dengue', *European Scientific Journal*, 13(9). doi:10.19044/esj.2017.v13n9p137.

Rocha, F.P. and Giesbrecht, M. (2022) ‘Machine learning algorithms for dengue risk assessment: A case study for São Luís do Maranhão’, *Computational and Applied Mathematics*, 41(8). doi:10.1007/s40314-022-02101-z.

Katta, M. *et al.* (2022) ‘An efficient learning model selection for dengue detection’, *Smart Innovation, Systems and Technologies*, pp. 439–453. doi:10.1007/978-981-16-9873-6\_40.

*Find open datasets and Machine Learning Projects* (no date a) *Kaggle*. Available at: <https://www.kaggle.com/datasets/> (Accessed: 29 April 2024).

Rana, S.K. *et al.* (2022) ‘Dengue fever prediction using Machine Learning Analytics’, *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)* [Preprint]. doi:10.1109/com-it-con54601.2022.9850923.

(2022) *Nature news*. Available at: <https://www.nature.com/scitable/topicpage/dengue-viruses-22400925/> (Accessed: 20 April 2024).

Lukasczyk, S., Kroiß, F. and Fraser, G. (2023) ‘An empirical study of Automated Unit Test Generation for Python’, *Empirical Software Engineering*, 28(2). doi:10.1007/s10664-022-10248-w.

Mårtensson, B. (2019) ‘Unit testing’, *Testing in Xcode* [Preprint]. doi:10.1007/978-1-4842-5192-8\_2.

‘A research paper on white box testing’ (2022) *Data Analytics and Artificial Intelligence*, 2(6). doi:10.46632/daai/2/6/8.

Pajankar, A. (2021) ‘Pytest’, *Python Unit Test Automation*, pp. 123–141. doi:10.1007/978-1-4842-7854-3\_5.

Ehmer, M. and Khan, F. (2019) ‘A comparative study of white box, black box and grey box testing techniques’, *International Journal of Advanced Computer Science and Applications*, 3(6). doi:10.14569/ijacsa.2012.030603.

- Hoyos, W., Aguilar, J. and Toro, M. (2021) 'Dengue models based on machine learning techniques: A systematic literature review', *Artificial Intelligence in Medicine*, 119, p. 102157. doi:10.1016/j.artmed.2021.102157.
- Saturi, S. (2020) 'Development of prediction and forecasting model for dengue disease using machine learning algorithms', *2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)* [Preprint]. doi:10.1109/discover50404.2020.9278079.
- Sarwar, Md.T. and Al Mamun, Md. (2022) 'Prediction of dengue using machine learning algorithms: Case study Dhaka', *2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)* [Preprint]. doi:10.1109/icecte57896.2022.10114535.
- Iqbal, N. and Islam, M. (2019) 'Machine Learning for Dengue Outbreak Prediction: A performance evaluation of different prominent classifiers', *Informatica*, 43(3). doi:10.31449/inf.v43i3.1548.
- Caicedo-Torres, W., Paternina, Á. and Pinzón, H. (2020) 'Machine learning models for early dengue severity prediction', *Lecture Notes in Computer Science*, pp. 247–258. doi:10.1007/978-3-319-47955-2\_21.



ORIGINALITY REPORT

---

20%

SIMILARITY INDEX

17%

INTERNET SOURCES

13%

PUBLICATIONS

12%

STUDENT PAPERS

---

PRIMARY SOURCES

---

1	<a href="https://dspace.daffodilvarsity.edu.bd:8080">dspace.daffodilvarsity.edu.bd:8080</a> Internet Source	2%
2	<a href="http://www.open-access.bcu.ac.uk">www.open-access.bcu.ac.uk</a> Internet Source	2%
3	<a href="http://ijream.org">ijream.org</a> Internet Source	1%
4	<a href="http://etheses.whiterose.ac.uk">etheses.whiterose.ac.uk</a> Internet Source	1%
5	<a href="https://assets.researchsquare.com">assets.researchsquare.com</a> Internet Source	1%
6	<a href="https://dev.to">dev.to</a> Internet Source	1%
7	<a href="https://link.springer.com">link.springer.com</a> Internet Source	1%
8	Submitted to Manipal University Student Paper	1%
9	Submitted to Monash College Pty Ltd Student Paper	1%

---

10	Submitted to Universidade Estadual de Campinas Student Paper	1 %
11	www.frontiersin.org Internet Source	1 %
12	Submitted to CSU, San Jose State University Student Paper	<1 %
13	www.browserstack.com Internet Source	<1 %
14	Submitted to Huddersfield New College Student Paper	<1 %
15	"Data Analytics and Management", Springer Science and Business Media LLC, 2021 Publication	<1 %
16	www-csfy.cs.ualberta.ca Internet Source	<1 %
17	dokumen.pub Internet Source	<1 %
18	onlinelibrary.wiley.com Internet Source	<1 %
19	www.researchgate.net Internet Source	<1 %
20	Submitted to Campbellsville University Student Paper	<1 %

21	<a href="http://www.gminsights.com">www.gminsights.com</a> Internet Source	<1 %
22	Aboobucker Ilmudeen. "Design and development of IoT-based decision support system for dengue analysis and prediction: case study on Sri Lankan context", Elsevier BV, 2021 Publication	<1 %
23	Submitted to Bocconi University Student Paper	<1 %
24	<a href="http://cybersecurity.springeropen.com">cybersecurity.springeropen.com</a> Internet Source	<1 %
25	<a href="http://nlistsp.inflibnet.ac.in">nlistsp.inflibnet.ac.in</a> Internet Source	<1 %
26	<a href="http://publications.aston.ac.uk">publications.aston.ac.uk</a> Internet Source	<1 %
27	<a href="http://www.mdpi.com">www.mdpi.com</a> Internet Source	<1 %
28	Submitted to Oklahoma State University Student Paper	<1 %
29	Submitted to University of Wales Institute, Cardiff Student Paper	<1 %
30	<a href="http://www.ijmrhs.com">www.ijmrhs.com</a> Internet Source	<1 %

31	Submitted to IUBH - Internationale Hochschule Bad Honnef-Bonn Student Paper	<1 %
32	pu.edu.pk Internet Source	<1 %
33	pubmed.ncbi.nlm.nih.gov Internet Source	<1 %
34	www.scilit.net Internet Source	<1 %
35	actascientific.com Internet Source	<1 %
36	ieeexplore.ieee.org Internet Source	<1 %
37	Submitted to Asian Institute of Management Student Paper	<1 %
38	Archana Thirugnanam, Faritha Banu Jahir Hussain. "Exploring Machine Learning Algorithms for the Prediction of Dengue: A Comprehensive Review", Revue d'Intelligence Artificielle, 2023 Publication	<1 %
39	Submitted to Karunya University Student Paper	<1 %
40	earsiv.cankaya.edu.tr:8080 Internet Source	<1 %

41

[repository.eafit.edu.co](https://repository.eafit.edu.co)

Internet Source

&lt;1 %

42

[vb.mruni.eu](https://vb.mruni.eu)

Internet Source

&lt;1 %

43

"Data Engineering and Intelligent Computing", Springer Science and Business Media LLC, 2022

Publication

&lt;1 %

44

Jibran Rasheed Khan, Syed Kawish Raza. "Development and Evaluation of a Predictive Diagnostic System for Dengue Fever using Machine Learning Techniques", Research Square Platform LLC, 2023

Publication

&lt;1 %

45

Salim G. Shaikh, B. Suresh Kumar, Geetika Narang, N.N. Pachpor. "Several Classification and Recommendations Methods Used in Dengue Fever Prediction System", 2023 International Conference on Integration of Computational Intelligent System (ICICIS), 2023

Publication

&lt;1 %

46

Fernanda Paula Rocha, Mateus Giesbrecht. "Machine learning algorithms for dengue risk assessment: a case study for São Luís do Maranhão", Computational and Applied Mathematics, 2022

Publication

&lt;1 %

---

47

Supreet Kaur, Sandeep Sharma, Ateeq Ur Rehman, Elsayed Tag Eldin, Nivin A. Ghamry, Muhammad Shafiq, Salil Bharany. "Predicting Infection Positivity, Risk Estimation, and Disease Prognosis in Dengue Infected Patients by ML Expert System", Sustainability, 2022

Publication

---

<1 %

---

Exclude quotes Off

Exclude matches Off

Exclude bibliography On