MODELING AND FORECASTING THE DENGUE INCIDENCE IN SRI LANKA

DILINI THEJANI KARUNASENA

M.Sc. 2023

MODELING AND FORECASTING DENGUE INCEDENCE IN SRI LANKA

A PROJECT REPORT PRESENTED BY

DILINI THEJANI KARUNASENA

to the board of Study in Industrial Mathematics of the

POSTGRADUATE INSTITUTE OF SCIENCE

in partial fulfillment of the requirement

for the award of the degree of

MASTER OF INDUSTRIAL MATHEMATICS

of the

UNIVERSITY OF PERADENIYA

SRI LANKA

2023

DECLARATION

I do hereby declare that the work reported in this Independent Study project report was exclusively carried out by me under the supervision of Dr. Lakshika S. Nawarathna. It describes the results of my own independent research except where due reference has been made in the text. No part of this project report/thesis has been submitted earlier or concurrently for the same or any other degree.

Date: 14.09.2023 ....

Signature of the Candidate

Certified by:

Supervisor (Name): Dr. (Mrs.) L.S Nawarathna Date: 14/09/2023

(Signature): ....

MODELING AND FORECASTING DENGUE INCIDENCE IN SRI LANKA

D.T. Karunasena

Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka

Department of Mathematics, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka

Dengue fever remains a significant public health concern in Sri Lanka, with its incidence varying across provinces and years. This study delves into a comprehensive analysis of Dengue incidence data from 2010 to 2022, considering all provinces separately. Initially, we examined the normality of the data using histograms and the Shapiro-Wilk test, revealing that the data did not follow a normal distribution. Subsequently, we investigated the correlations between Dengue incidence and climate variables (precipitation, temperature, and humidity) for each year using heatmaps and Spearman's correlation test. Our findings unveiled positive correlations between Dengue incidence and both precipitation and humidity, while temperature exhibited a negative correlation. These insights contribute to a better understanding of the complex relationship between climate and Dengue transmission. To gain a spatial perspective, we conducted hotspot analyses for each province annually from 2010 to 2022. This allowed us to categorize provinces into "High-Intensity Regions," "Moderate-Intensity Regions," and "Low-Intensity Regions," providing valuable information for targeted intervention strategies. Exploring the temporal dimension, we analyzed monthly Dengue incidence data. Surprisingly, no distinct monthly patterns emerged, emphasizing the unpredictable nature of Dengue outbreaks. To enhance preparedness and control efforts, we developed a forecasting model using Random Forest Regressor. Validation for 2021 and 2022 demonstrated the model's accuracy in predicting Dengue incidence. Encouraged by these results, we utilized the model to forecast Dengue incidence for 2023 and 2024. In conclusion, this study combines spatial and temporal analyses, climate correlations, and predictive modeling to provide a comprehensive view of Dengue transmission in Sri Lanka. The findings emphasize the need for dynamic intervention strategies that consider both climate factors and regional variations. As Dengue continues to pose a public health threat, the insights gained from this study are invaluable for policymakers and healthcare professionals working towards effective Dengue prevention and control in Sri Lanka.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Dr. Lakshika S. Nawarathna, Senior Lecturer in the Department of Statistics and Computer Science, Faculty of Science, University of Peradeniya. Dr. Nawarathna's unwavering support, guidance, and expertise were instrumental in shaping the course of this research. Her profound knowledge and commitment to academic excellence have been a constant source of inspiration throughout this study.

I would also like to acknowledge the resources and facilities provided by the University of Peradeniya, which played a crucial role in the execution of this research.

Furthermore, I am grateful to my friends and family for their unwavering support and encouragement throughout this academic journey. Their belief in my capabilities and their constant encouragement have been a source of motivation.

Finally, I extend my thanks to all the participants and stakeholders whose contributions and cooperation made this research possible.

This work would not have been achievable without the collective effort and support of all those mentioned above. Thank you for being an integral part of this endeavor.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CMC	Colombo Municipal Council
CSV	Comma Separated Values
DENV-1	Dengue Virus 1
DENV-2	Dengue Virus 2
DENV-3	Dengue Virus 3
DENV-4	Dengue Virus 4
DF	Dengue Fever
DHF	Dengue Hemorrhagic Fever
GIS	Geographical Information System
Hum.	Relative Humidity at 2 meters
MAE	Mean Absolute Error
Prec.	Average Precipitation (mm/Day)
RH	Relative Humidity
R2	R-Squared Score
Tem.	Temperature at 2 meters
WHO	World Health Organization

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

Dengue is a mosquito-borne viral disease that has rapidly spread to all regions of WHO in recent years, and it is found in tropical and sub-tropical climates worldwide, mostly in urban and semi-urban areas. (World health organization 2022). In Sri Lanka Dengue is an increasing heath concern, along with other mosquito-borne diseases such as zika and chikungunya. Sri Lanka has history of over 40 years of Dengue Fever (DF) and Dengue Hemorrhagic Fever (DHF). 51 cases of Dengue hemorrhagic fever and 15 deaths were found in the period 1965-1968. Therefore from 1969 up to 1988, multiple similar outbreaks with endemic Dengue Fever were reported in urban areas. (National Dengue Control Unit 2022). Total of 31162 suspected Dengue cases for the year 2020 were reported and 20207 suspected cases were reported from all over the island from January 2022 to up to now. In the year 2017 there was an outbreak situation and the highest number of Dengue cases were reported in 29th week of 2017. Approximately 36.9% of Dengue cases were reported from the western province. (Epidemiology unit 2022).

Out of 140 species of mosquitoes in Sri Lanka, *Aedes aegypti* and *Aedes albopictus* mosquitoes transmit the Dengue virus to human. These two types of mosquitoes can be differentiated by using the marking of their bodies. Dengue vectors breeding sites are found both indoors and outdoors such as discarded receptacles, water storage containers, automobile tires, natural breeding sites etc. (National Dengue Control Unit 2022). Dengue virus appears as four different but associated serotypes of virus of viruses namely, DENV-1, DENV-2, DENV-3, and DENV-4. (Dengue Virus Net). High fever, severe headache, pain behind eyes, skin rashes, muscle and joint pain, vomiting are the main symptoms of Dengue Fever. In addition, widespread bleeding, low blood pressure and body organ such as Liver and kidney failure can occur in Dengue hemorrhagic fever and Dengue shock syndrome. (Piyatilake et al., 2020). Both Dengue and Hemorrhagic Fever incidences are commonly found among younger population, less than 19 years old. (Sumith Pathirana et al., 2009).

There are lots of factors that impact on Dengue fever. Some of them are meteorological factors such as precipitation, humidity, air pressure, wind speed, temperature, and population mobility, large family environment, poverty, informal garbage disposal etc. Based on climate data from 2010 to 2018 Chandrasena et al. (2019) suggested that rainfall was the only significant factor that are affected the likelihood of having higher Dengue incidences. Furthermore, the influence of rainfall on dengue increase is expected to be visible after some lag months. (Chandrasena et al., 2019). The national Dengue control unit says that Dengue has a seasonal transmission in Sri Lanka, with two peaks occurs with the monsoon rains in June-July and October-December respectively and the major cases occur during June-July the summer monsoon. (National Dengue Control Unit 2022).

Precipitation and air pressure, wind speed and bioclimate zone are identified as the main influential factors for the Dengue outbreak in 2017. Growing humidity, air pressure, and wind speed increase the Dengue illness. (Faruk et al., 2022).

Using epidemiological data from December 2015 to November 2017 of Gampaha district, Gayan et al. (2018) have analyzed that there was a positive correlation between Dengue incidences and climate variable, monthly rainfall, number of rainy days, minimum and maximum RH, and minimum temperature, and negative correlation between Dengue incidences and maximum temperature and wind speed. (Gayan et al., 2018). Analyzing the forecast of Dengue incidences in urban Colombo, Erandi et al. (2021) have conducted that if the weekly rainfall value for Colombo Municipal center area is in between 14mm to 454mm then Dengue incidences have correlation with 8-14 weeks lag. And average temperature of Sri Lanka varies from 17^{o}C to 35^{o}C which is ideal for Dengue transmission. (Erandi et al., 2021).

Human mobility factors are often overlooked when considering preventive aspects. Preventive measures targeting human movements may enhance effectiveness of Dengue control programs. (Kumanan et al., 2019). It is found that high population movements in the area plays a dominant role to transmit the disease to other areas. (Piyatilake et al., 2020). Dharmawardana et al. (2018) have shown that the information derived from human mobility model also can be used to predict Dengue incidences. (Dharmawardana et al., 2018).

After decays of researches the first Dengue vaccine was licensed in 2015. Recent analysis has shown that the vaccine performance is dependent on serostatus. In those who have had a previous Dengue infection, i.e., who are seropositive, the efficacy is high and vaccine is safe. (Wilder Smith., 2020).

But for countries with limited resources like Sri Lanka, it is necessary to identify the dynamics of the Dengue spread thoroughly to determine more efficient control strategies. (Erandi et al., 2021). Therefore, the aim of this study is to develop a mathematical model to predict the general population of Dengue incidences in next two years in all the provinces of the country using previous data from 2010 to 2022. In this study, our primary objectives are to comprehensively examine the multifaceted relationship between Dengue fever outcomes and key climate variables, specifically precipitation, humidity, and temperature. We aim to rationalize and elucidate the intricate interplay between these environmental factors and Dengue incidence across various provinces. Additionally, we conduct a hotspot analysis to identify geographic regions of heightened Dengue activity, shedding light on areas requiring targeted intervention. Moreover, we delve into a temporal analysis, scrutinizing the monthly variations in Dengue incidence from 2010 to 2022 to discern long-term trends and seasonal patterns. As a forward-looking aspect of this research, we endeavor to develop predictive models to forecast Dengue incidence for the years 2023 and 2024. Through these multifaceted investigations, this study aims to contribute valuable insights into Dengue epidemiology and climate-related factors, with implications for public health planning and preparedness.

CHAPTER 2

LITARETURE REVIEW

Dengue is a viral infection transmitted to humans through the bite of mosquitoes. (World health organization). In Sri Lanka Dengue is an increasing health concern along with other mosquito-borne diseases such as Zika and Chikungunya. (World mosquito program). Mainly there are meteorological and population factors that impact on Dengue Fever. If we can identify the dynamics on the Dengue spread in Sri Lanka, and develop a mathematical model to predict the Dengue cases, that would be so important for reducing the number of Dengue cases in Sri Lanka. There are many previous studies done on the impact of some factors on Dengue cases and also prediction of Dengue cases in Sri Lanka.

When we consider the environmental factors, Faruk et al. (2022) have done a study on the impact of environmental factors on the spread of Dengue Fever in Sri Lanka. The aim of that study was to investigate the effect of environmental, seasonal and special variations on the spread of Dengue Fever in Sri Lanka. The author has used secondary data of monthly Dengue infection and the monthly average of environmental parameters 26 Sri Lankan regions from January 2015 to December 2019. In this study the country has divided into 3 bioclimatic zones as dry zone, intermediate zone, and wet zone. All 5-year data were merge into single data file for the aggregate regression model. Descriptive statistics that author has used for the chosen environmental variables were mean, standard deviation, maximum, minimum, skewness and kurtosis. For testing the linear of the variables, analysis of variance (ANOVA) tables was computed for each of the explanatory variables considering the dependence variable. This study has showed that precipitation and air pressure had a significant effect on spread of Dengue Fever in each year, and the wet zone had a significant impact on the vast Aedes infection. And also, this study determined that the Northen monsoon season greatly influenced the spread of Dengue in Sri Lanka. (Faruk et al., 2022).

Sirisena et al. (2017) have done a study to show the effect of climatic factors and population density on the distribution of Dengue in Sri Lanka. They have focused on the use of geographical information systems (GIS) to map and evaluate the spatial and temporal distribution of Dengue in Sri Lanka from 2019 to 2014. Mainly they have divided the country into 3 Argo-climatic zones like, wet, dry and intermediate zones. After collecting the data of Dengue cases from Epidemiology unit of ministry of health, the population data from the census and statistic report, department of census and statistics and the data of annual rainfall, temperature and humidity from the department of meteorology, the have analyzed them using the SPSS, version 20(2011) and R studio, (2012) software. And the spatial distribution of Dengue incidence and climate factors were mapped using the Arc GIS, version 10.2(2012) software. As the results of their study, they have showed that the Dengue incidences were high in areas where the population density is high and that was clearly seen in the three major districts, Colombo, Kandy and Jaffna. The correlation between Dengue incidence and climate factors including temperature, rainfall and humidity has evaluated using spearman's correlation. And also, the have showed that Dengue incidence had a positive correlation with rainfall and no positive correlation between Dengue incidence and temperature or humidity. (Sirisena et al., 2017).

Edussuriya. (2020) has developed a mathematical model for predicting the number of Dengue cases in tropics using machine learning technique using Island wide Dengue epidemiology data, weather data and population density data. She has considered rainfall, humidity, wind speed and temperature as weather data. But Kalutara, Matale, Matara, Kilinochchi, Mulative and Kegalle districts were not available for her analysis. (Edussuriya., 2020). Piyatilake et al. (2020) have done a study to develop a clustering technique to identify the Dengue hot spots in Sri Lanka. In their study, they have identified eight different predictive factors. They are average value of daily temperature, rainfall, rainy days per week, humidity, collection of garbage, population density, percentage of urbanization and the number of population movements. By considering the data they have further subdivided into three risk levels, namely high, moderate and law. Finally, they have identified the population movements in the area as a significant factor to increase the Dengue risk. (Piyatilake et al., 2020).

Weisun. (2017) has done a study on spatial-temporal distribution of Dengue and climate characteristics for two clusters in Sri Lanka from 2012 to 2016. Using the hot spot analysis and the spatial-temporal clustering method, he has investigated the spatial-temporal distribution of Dengue in Sri Lanka to identify spatial-temporal clusters and elucidate the association of climatic factors with Dengue incidence. He has considered that targeting hot spots during outbreaks instead of all the regions could save resources and time for public health authorities.

Association of Dengue incidences with local climatic factors were evaluated by the spearman's correlation test for all 24 districts. The results of his study showed that there was certain correlation between Dengue incidences and local climates and the correlations were different for different districts. (Weisun., 2017)

There are studies which are done targeting the risk areas such as Colombo and Gampaha. Attanayake et al. (2020) have done their study applying exponential smoothing technique in order to model and forecast Dengue cases in Colombo, Sri Lanka. They have used the data consist of monthly reported Dengue cases in Colombo district from January to May 2019. They have built the model using the data from January 2019 to February 2019 and have validated the model using the rest of data. R software was mainly used for their data analysis. By considering the forecasted values of their model, they have concluded that monthly Dengue cases to be computed in the upcoming months would increase slowly in Colombo district. The forecasted values generated by their model for March to August 2019 were 580, 536, 786, 1342, 1892, and 1192. But according to the data reported in Epidemiology unit of Ministry of health, the actual values for those months were 815, 579, 866, 1126, 1857, and 1903. (Attanayake., 2020)

Gnanapragasam. (2016) has done a study to fit a statistical model to forecast the Dengue cases in Colombo Municipal Council (CMC) areas of Colombo district, using the data from March to June 2016 for the purpose of model validation and the data from January 2010 to February 2016 for model development. As the resalts, over 1500 Dengue cases were expected from upcoming 6 months in the year 2016 and it was 31.5% increase from the Dengue cases reported in last 6 months in 2015. (Gnanapragasam., 2016)

Chandrakantha. (2019) has done his study focused on the City of Colombo to develop a risk prediction model for Dengue transmission based on climate data using logistic regression methodology. He has used monthly data from 2010 to 2018 for model building. The climate factors he has used were monthly average temperature, cumulative rainfall per month and monthly relative humidity. The result of his study has suggested that rainfall was the only significant factor that the affected the likelihood of having higher Dengue increase. (Chandrakantha., 2019)

Since the district of Gampaha is also a risk area of Dengue incidences, Withanage et al. (2018) have done a study to develop and validate a simple accurate forecasting model for the district of Gampaha. They have developed a Three time-series regression models using monthly rainfall, rainy days, temperature, humidity, wind speed and retrospective Dengue incidences over the period of January 2012 to November 2015. As the results, during the analysis of correlation between Dengue incidences and climate variables, monthly rainfall, number of rainy days, minimum and maximum RH and minimum temperature have showed positive correlations while maximum temperature and wind speed have showed negative correlations. (Withanage., 2018)

CHAPTER 3

METHODOLOGY

3.1 Study Area

In this study, the analysis has done using the previous data of Dengue incidence and the 3 climate variables which are Precipitation, Temperature and Humidity considering 13 years from 2010 to 2022. The relations between Dengue incidence and those climate variables were analyzed yearly. The environmental parameters here measured in different units such as average Precipitation in mm/Day, Temperature at 2 meters, Relative Humidity at 2 meters.

Data from January 2010 to December 2020 were used for the model fitting and data from January 2021 to December 2022 were used for the model validation.

3.2 Data Collection

The data of Dengue incidence were collected from the website of Epidemiology Unit, Ministry of Health Care and Nutrition, Sri Lanka. The data of the 3 climate variables were collected from the website, Data Access Viewer-NASA POWER. It provides solar and meteorological datasets from NASA research for support of renewable energy, building energy efficiency and agricultural needs.

3.3 Data Analysis

The data were statistically analyzed in python language using the Google.Colab environment.

3.3.1 Time Series Plots

Time Series plots were plotted for the data from 2010 to 2022 of all the variables using python language.

```python
# Import the libraries
import pandas as pd
import matplotlib.pyplot as plt

# Read the data from the CSV file
df1 = pd.read_csv("/content/Annual Dengue Incidence.csv")

# Extract the year and Dengue Incidence data
x = df1['Year']
y = df1['Dengue Incidence']

# Create the bar plot
plt.figure(figsize=(15, 6))  # Adjust the figure size as needed
plt.bar(x, y, width=0.5, color='b', align='center')
plt.title("Time Series Plot of Dengue Incidence", fontsize=15)
plt.xlabel("Year", fontsize=14)
plt.ylabel("Dengue Incidence", fontsize=14)
plt.grid(axis='y', linestyle='--', alpha=0.7)

# Set x-ticks to display all years
plt.xticks(x, rotation=45, ha='right')

# Add data labels above the bars
for i, v in enumerate(y):
    plt.text(x[i], v + 50, str(v), ha='center', va='bottom')

# Show or save the plot
plt.tight_layout()
plt.show()  # Display the plot in the Python environment
```

Fig. 3.1 Python codes for Time Series Plot of Dengue data from 2010 to 2022.

A time series plot is a graph that shows how a variable change over time. Typically, time is plotted on the x-axis and the variable being measured is plotted on the y-axis.

3.3.2 Normality Test

A normality test is a statistical method used to determine whether a set of data follows a normal distribution, also known as a Gaussian distribution or bell curve. In a normal distribution, the majority of the data is clustered around the mean, with fewer and fewer data points occurring at greater distances from the mean. Normality tests are used to assess the assumption of normality in statistical analysis, which is often required for certain methods like parametric tests. If the data is normally distributed, then parametric tests can be used. However, if the data is not normally distributed, non-parametric tests may be more appropriate. There are several methods for testing normality, including graphical methods such as histograms and Q-Q plots, as well as statistical tests like the Shapiro-Wilk test, the Kolmogorov-Smirnov test, and the Anderson-Darling test. In this study Histograms and the Shapiro-Wilk test were used to test the Normality of the data.

3.3.2.1 Histograms

Time Series plots were plotted for the data from 2010 to 2022 of all the variables using python language.

```
import numpy as np  # Import the NumPy library for numerical operations
import pandas as pd  # Import the Pandas library for data manipulation
import matplotlib.pyplot as plt  # Import the Matplotlib library for creating plots

# Read a CSV file named "pro-2010.csv" into a Pandas DataFrame and store it in the variable df10
df10 = pd.read_csv("/content/pro-2010.csv")

# Extract the 'Total' column from the DataFrame and store it in the variable Total
Total = df10.Total

# Create a histogram using the data in the 'Total' variable with blue bars
plt.hist(Total, color='b')

# Add a label to the x-axis of the histogram
plt.xlabel("Dengue Incidence", fontsize=20)

# Add a label to the y-axis of the histogram
plt.ylabel("Frequency", fontsize=16)

# Add a title to the histogram
plt.title("2010", fontsize=16)

# Display the histogram plot
plt.show()
```

Fig. 3.2 Python codes for Histogram of Dengue data in 2010.

A variable that is normally distributed has a histogram (or "density function") that is bell-shaped, with only one peak, and is symmetric around the mean. The terms kurtosis ("peakedness" or "heaviness of tails") and skewness (asymmetry around the mean) are often used to describe departures from normality. In this study Histograms were plotted considering the dependent variable. The histograms were not showed bell shaped curves.

3.3.2.2 Shapiro-Wilk Test

```python
import pandas as pd  # Import the Pandas library for data manipulation
from scipy.stats import shapiro  # Import the Shapiro-Wilk test from SciPy

# Read a CSV file named "pro-2010.csv" into a Pandas DataFrame and store it in the variable df1
df1 = pd.read_csv("/content/pro-2010.csv")

# Extract the 'Total' column from the DataFrame and store it in the variable Total
Total = df1.Total

# Apply the Shapiro-Wilk test to the data in the 'Total' variable
shapiro(Total)
```

Fig. 3.3 Python codes for Shapiro-Wilk Test for the Dengue data of 2010.

Shapiro Wilk Test is a statistical test used to assess the normality of a dataset. It measures the departure of a sample from a normal distribution and provides a P-value to indicate the level of significance. If P-value is less than a certain threshold (0.05), the null hypothesis that the data follows a normal distribution is rejected, and the data is considered to be non-normal Shapiro wilk test was applied for the data of Dengue incidence of 12 years separately to check the normality of the datasets.

3.3.3 Correlation Analysis

Correlation analysis is a statistical technique used to examine the relationship between two or more variables. It is commonly used in research to determine the extent to which two variables are related or associated with each other. Correlation analysis calculates a correlation coefficient, which is a statistical measure that indicates the strength and direction of the relationship between two variables. The correlation coefficient can range from -1 to +1, where a value of -1 indicates a perfectly negative correlation, 0 indicates no correlation, and +1 indicates a perfectly positive correlation.

3.3.3.1 Spearman's Correlation Test

Spearman's correlation test is a statistical method used to measure the strength and direction of the relationship between two nonlinear variables. To assess the bivariate relationship between the spread of Dengue infection and the environmental variables, Spearman's Correlation test was applied. First the test was applied for the 12 years separately.

3.3.3.2 Heatmaps

```
import pandas as pd  # Import the Pandas library for data manipulation
import matplotlib.pyplot as plt  # Import the Matplotlib library for creating plots
import seaborn as sns  # Import the Seaborn library for data visualization

# Read data from CSV files into Pandas DataFrames
df10 = pd.read_csv("/content/pro-2010.csv")
df10p = pd.read_csv("/content/2010-Precipitation.csv")
df10t = pd.read_csv("/content/2010-Temperature.csv")
df10h = pd.read_csv("/content/2010-Humidity.csv")

# Rename columns in each DataFrame for clarity
df2010 = df10.rename(columns={"Total": "Dengue"})
df2010p = df10p.rename(columns={"Annual": "Prec."})
df2010t = df10t.rename(columns={"Annual": "Tem."})
df2010h = df10h.rename(columns={"Annual": "Hum."})

# Concatenate the selected columns from different DataFrames into a single DataFrame
Df10 = pd.concat([df2010['Dengue'], df2010p['Prec.'], df2010t['Tem.'], df2010h['Hum.']], axis=1)

# Create a correlation heatmap using Spearman correlation method
sns.heatmap(Df10.corr(method="spearman"), annot=True, square=True, cmap="coolwarm")

# Add a title to the heatmap
plt.title('2010')
```

Fig. 3.4 Python codes for Heatmap for the data of Dengue and climate variables in 2010.

A heatmap is a graphical representation of data where individual values are represented as colors. The values are usually represented in a two-dimensional matrix, where the rows and columns correspond to variables and the color of each cell represents the value of a particular observation. To get a clear visualization of the correlation between Dengue incidence and the climate variables, Heatmaps were developed using python language.

3.3.4 Model Fitting

Model fitting refers to the process of training a model on a given dataset to learn patterns and relationships in the data. During this process, the model uses an algorithm to adjust its parameters to minimize the difference between the predicted outputs and the actual outputs in the training data. In this study, years from 2010 to 2020 are used to fit the model.

3.3.5 Model Validation

Model validation, on the other hand, refers to the process of evaluating how well the trained model performs on new, unseen data. This is done to ensure that the model is able to generalize to new data and make accurate predictions. In this study, years 2021 and 2022 are used for the Model Validation.

3.3.6 Hotspot Analysis

In this study, a comprehensive hotspot analysis of Dengue incidence was conducted across the nine provinces of Sri Lanka, utilizing data spanning the years from 2010 to 2022. With the utilization of the Python programming language, an in-depth analysis was performed for each individual year, applying spatial and temporal techniques to identify and map the hotspots of Dengue activity. By considering the unique characteristics of each province, the evolving dynamics of Dengue within Sri Lanka are illuminated through this hotspot analysis.

```python
#Import the pandas library:
import pandas as pd
```

```python
#Load the CSV dataset into a pandas DataFrame:
data10 = pd.read_csv('/content/pro-2010.csv')
data10
```

	Province	January	February	March	April	May	June	July	August	September	October	November	December	Total
0	Western	1313	1353	719	543	776	1375	2636	1513	608	424	197	252	11709
1	Central	377	298	166	141	163	224	483	317	135	106	47	64	2521
2	Southern	121	251	212	189	140	294	537	401	234	98	33	26	2536
3	Northen	1435	1166	295	233	111	325	369	375	139	124	125	152	4849
4	Easten	550	610	372	263	79	103	178	92	44	30	25	59	2405
5	North Western	405	331	185	104	151	257	437	296	127	78	46	71	2488
6	North Central	300	338	165	98	81	99	134	97	81	41	31	47	1512
7	Uva	121	107	104	114	122	211	686	538	185	93	68	64	2413
8	Sabaragamuwa	183	277	310	319	327	549	653	542	310	143	75	67	3755

```python
#Sort the Dataframe:
sorted_data10 = data10.sort_values(by='Total', ascending=False)
```

```python
#print the sorted DataFrame:
print(sorted_data10)
```

Fig. 3.5 Python codes for Hotspot Analysis in 2010.

3.3.7 Analysis of risky months

Incorporating all months of the year, an analysis of the months with higher risk of Dengue incidence in Sri Lanka was conducted for the years from 2010 to 2022. Utilizing Dengue data and the Python programming language, this study delved into identifying the months that exhibited elevated susceptibility to Dengue outbreaks. Through a meticulous examination for each individual year within the specified timeframe, the risky months were discerned.

+ Code + Text

```
#Import the pandas library:
import pandas as pd
```

```
#Load the CSV dataset into a pandas DataFrame:
df2010 = pd.read_csv('/content/2010-den.Tr.csv')
df2010
```

	Month	Western	Central	Southern	Northen	Easten	North Western	North Central	Uva	Sabaragamuwa	Total
0	January	1313	377	121	1435	550	405	300	121	183	4805
1	February	1353	298	251	1166	610	331	338	107	277	4731
2	March	719	166	212	295	372	185	165	104	310	2528
3	April	543	141	189	233	263	104	98	114	319	2004
4	May	776	163	140	111	79	151	81	122	327	1950
5	June	1375	224	294	325	103	257	99	211	549	3437
6	July	2636	483	537	369	178	437	134	686	653	6113
7	August	1513	317	401	375	92	296	97	538	542	4171
8	September	608	135	234	139	44	127	81	185	310	1863
9	October	424	106	98	124	30	78	41	93	143	1137
10	November	197	47	33	125	25	46	31	68	75	647
11	December	252	64	26	152	59	71	47	64	67	802

```
#Sort the Dataframe:
sorted_df2010 = df2010.sort_values(by='Total', ascending=False)
#print the sorted DataFrame:
print(sorted_df2010)
```


Fig. 3.6 Python codes for Analysis of months.

3.3.8 Forecasting the Dengue Incidence.

In this study, a forecasting model for Dengue incidence in Sri Lanka was developed utilizing Dengue data spanning the years 2010 to 2022. Employing the Random Forest Regressor method, annual forecasts were generated for each province, contributing to a comprehensive assessment of Dengue trends across the nation. Through the meticulous application of this predictive technique, the study presents insights into the potential trajectory of Dengue cases for each year and province. The integration of predictive analytics provides a valuable tool for anticipating and addressing the challenges posed by Dengue outbreaks, thereby facilitating more effective disease management and mitigation efforts.

3.3.8.1 Random Forest Regressor Analysis.

Random Forest Regressor is a machine learning algorithm that belongs to the ensemble learning family. Ensemble learning involves combining multiple models to improve overall predictive performance and robustness. The Random Forest Regressor is specifically designed for regression tasks, where the goal is to predict a continuous numeric value rather than a categorical label.

3.3.8.2 Forecasting Model building.

Data Loading and Preparation: -

The dataset was loaded using the Pandas library. The data, contained in the CSV file, was read to extract the 'Year' and 'Total' columns. The 'Year' column, representing the years of data, was reshaped into a format suitable for analysis, ensuring compatibility with the subsequent machine learning operations. The 'Total' column, which holds the Dengue incidence data, was extracted as the target variable 'y' for prediction.

+ Code + Text

```
#Import the necessary libraries:
import pandas as pd
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
```

```
#Load the CSV dataset into a pandas DataFrame:
dataset = pd.read_csv('/content/Western.csv')
```

Fig. 3.7 Python codes to import necessary libraries and load the dataset.

Data Splitting: -

The dataset was split into training and testing sets using the ‘train_test_split’ function from the ‘sklearn.model_selection’ module. An 80-20 split ratio was used, with 80% of the data allocated to training and 20% to testing. This division ensures that the machine learning model can be trained on a subset of the data and evaluated on unseen data to assess its predictive performance.

```
[ ] #Split the dataset into input features (Year) and target variable (Total):
    X = dataset['Year'].values.reshape(-1, 1)
    y = dataset['Total'].values
```

Fig. 3.8 Python codes to Split the data.

Random Forest Regressor Initialization and Training: -

An instance of the RandomForestRegressor class was created from the sklearn.ensemble module. The hyperparameter 'n_estimators' was set to 100, indicating the number of decision trees to be used in the ensemble. The 'random_state' parameter was set to 42 to ensure reproducibility of results. The created regressor model was then trained using the training data ('X_train' and 'y_train') using the fit method. The model learned from historical data to make predictions based on the given features.

```
#Create and train the Random Forest Regressor model:
rf_regressor = RandomForestRegressor(n_estimators=100, random_state=42)
rf_regressor.fit(X_train, y_train)
```

```
▾ RandomForestRegressor
RandomForestRegressor(random_state=42)
```

Fig. 3.9 Python codes to create and train the Random Forest Regressor model.

Prediction: -

To predict the Dengue incidence for the year 2022, a single data point ('year_2022') representing the year 2022 was created. This data point was formatted as a list containing the value 2022. Using the trained random forest regressor model, the "predict" method was applied to the 'year_2022' data point, yielding an estimated value for the Dengue incidence in 2022. The predicted value was then converted to an integer for clarity.

```
[ ]  #Perform the prediction for 2022:
     year_2022 = [[2022]]
     predicted_2022 = rf_regressor.predict(year_2022)
     predicted_2022_integer = int(predicted_2022)
```

Fig. 3.10 Python codes to Perform the prediction.

Printing the Prediction: -

The final predicted Dengue incidence value for the year 2022 was printed to the console, providing an estimate of the number of Dengue cases based on the trained random forest regressor model. This step concludes the predictive analysis and provides insight into the expected Dengue incidence for the specified year.

```
[ ]  #Print the result:
     print("Predicted number of dengue incidence in 2022:", predicted_2022_integer)

     Predicted number of dengue incidence in 2022: 20860
```

Fig. 3.11 Python codes to print the result.

3.3.8.3 Testing the Accuracy: -

Mean Absolute Error (MAE) and the R-squared (R2) score were tested for the model.

```
[14] # Import the mean_absolute_error and r2_score functions from scikit-learn metrics module
     from sklearn.metrics import mean_absolute_error, r2_score

     # Use the trained regression model (rf_regressor) to make predictions on the test set (X_test)
     y_pred = rf_regressor.predict(X_test)

     # Calculate the Mean Absolute Error (MAE) between the true target values (y_test) and predicted values (y_pred)
     mae = mean_absolute_error(y_test, y_pred)

     # Calculate the R-squared (R2) score to assess the goodness of fit of the model
     r2 = r2_score(y_test, y_pred)

     # Print the Mean Absolute Error (MAE) to evaluate the average absolute difference between true and predicted values
     print("Mean Absolute Error (MAE):", mae)

     # Print the R-squared (R2) score to assess the proportion of variance explained by the model
     print("R-squared (R2) Score:", r2)
```

Fig. 3.12 Python codes for testing the accuracy of the model.

‘mean_absolute_error(y_test, y_pred)’ calculates the Mean Absolute Error (MAE), which measures the average absolute difference between the true target values (y_test) and the predicted values (y_pred). It's a measure of how close the model's predictions are to the actual values, with lower MAE values indicating better performance.

‘r2_score(y_test, y_pred)’ calculates the R-squared (R2) score, which assesses the goodness of fit of the regression model. It quantifies the proportion of the variance in the target variable that is explained by the model. A higher R2 score (closer to 1) indicates a better fit, while a lower score (closer to 0) suggests that the model does not explain much of the variance in the target variable.

3.3.8.4 Testing the model for 2021 and 2022: -

The model was trained using historical Dengue incidence data spanning previous years. Subsequently, the model's predictive capabilities were evaluated by applying it to the years 2021 and 2022. The predicted values for Dengue incidence in these years were compared against the actual recorded values. This comparative analysis allowed for an assessment of the model's accuracy and its ability to capture the nuances of Dengue transmission patterns.

+ Code   + Text

```python
# Import the pandas library for data manipulation and the pyplot module from matplotlib for plotting
import pandas as pd
from matplotlib import pyplot as plt
```

```python
#Load the CSV dataset into a pandas DataFrame:
df2021 = pd.read_csv("/content/Test_2021.csv")
df2021
```

	Province	Predicted	Real
0	Western	16114	12683
1	Central	5365	2351
2	Southern	3680	2987
3	Northen	5088	3835
4	Easten	6641	5453
5	North Western	3416	2335
6	North Central	1064	637
7	Uva	1205	921
8	Sabaragamuwa	4362	3261

```python
# Create a new figure for the plot with a specific size (6 units wide and 4 units tall).
plt.figure(figsize=(6,4))
# Rotate the x-axis labels by 40 degrees for better readability.
plt.xticks(rotation=40)
# Set the title of the plot.
plt.title("Number of predicted and real Dengue incidence in 2021")
# Set the label for the x-axis.
plt.xlabel("Province")
# Set the label for the y-axis.
plt.ylabel("Number of Dengue incidences")
# Add gridlines to the plot with a line width of 0.5.
plt.grid(linewidth= 0.5)
# Plot a line chart with square markers ('s-') for the 'Predicted' data.
plt.plot(df2021.Province, df2021.Predicted,'s-')
# Plot a line chart with circle markers ('o-') for the 'Real' data.
plt.plot(df2021.Province, df2021.Real,'o-')
# Add a legend to the plot, labeling the two plotted lines as 'Predicted' and 'Real'.
plt.legend(["Predicted","Real"])
# Display the plot.
plt.show()
```

Fig. 3.13 Python codes to plot and compare the predicted values and the real values in 2021.

3.3.8.5 Forecasting the Dengue incidence in 2023 and 2024:

Having established the model's credibility, the number of Dengue incidences for the years 2023 and 2024 was forecasted using the validated Random Forest Regressor model. By leveraging the model's predictive capabilities, projections were made for these future years, offering insights into the anticipated trends of Dengue transmission.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Time Series Plots

Time series plots were used in order to get an idea of the spread of all the variables.

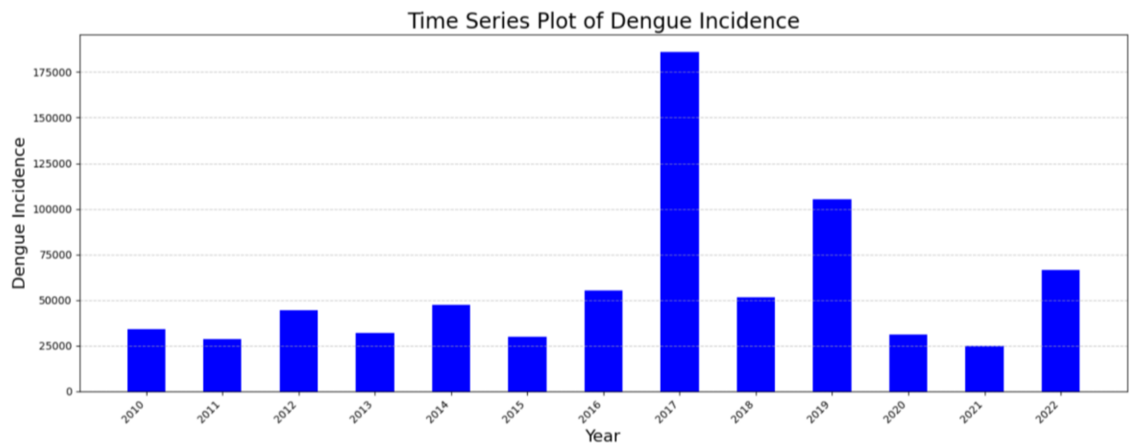

Fig. 4.1Time Series Plots of Dengue incidence from 2010 to 2022.

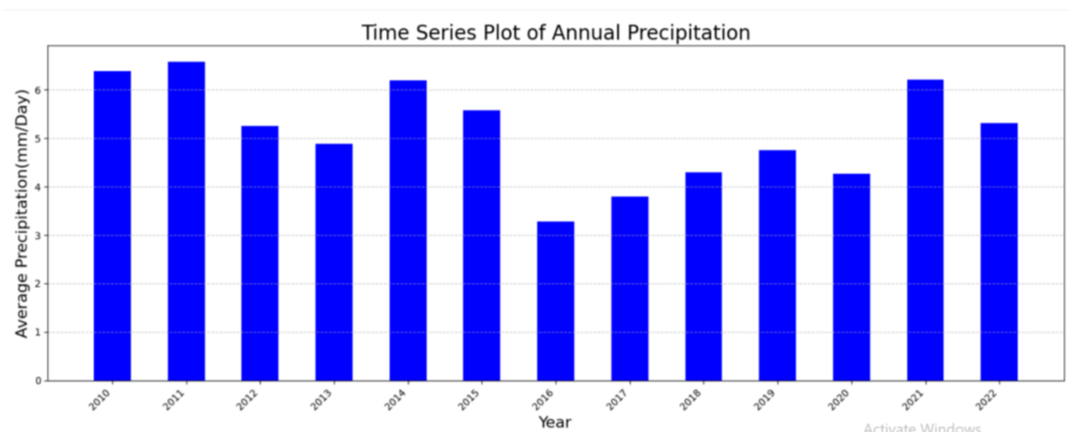

Fig. 4.2 Time Series Plots of Annual Precipitation from 2010 to 2022.

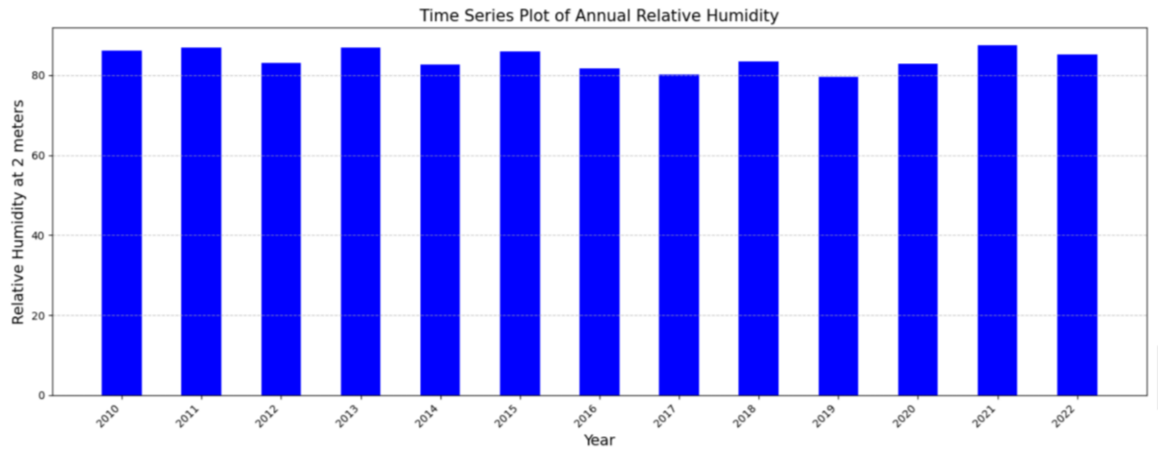

Fig. 4.3 Time Series Plots of Annual Relative Humidity from 2010 to 2022.

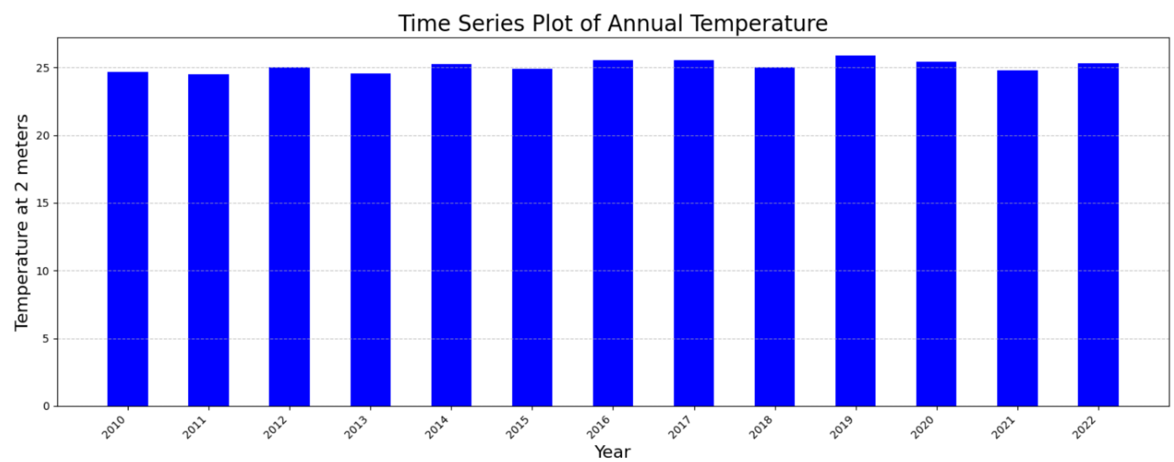

Fig. 4.4 Time Series Plots of Annual Temperature from 2010 to 2022.

4.2 Normality Test

4.2.1 Histograms

For testing the normality of the dataset, Histograms were plotted considering the data of Dengue incidence. But it was clearly seen that those Histograms did not resemble a bell-shaped curve.

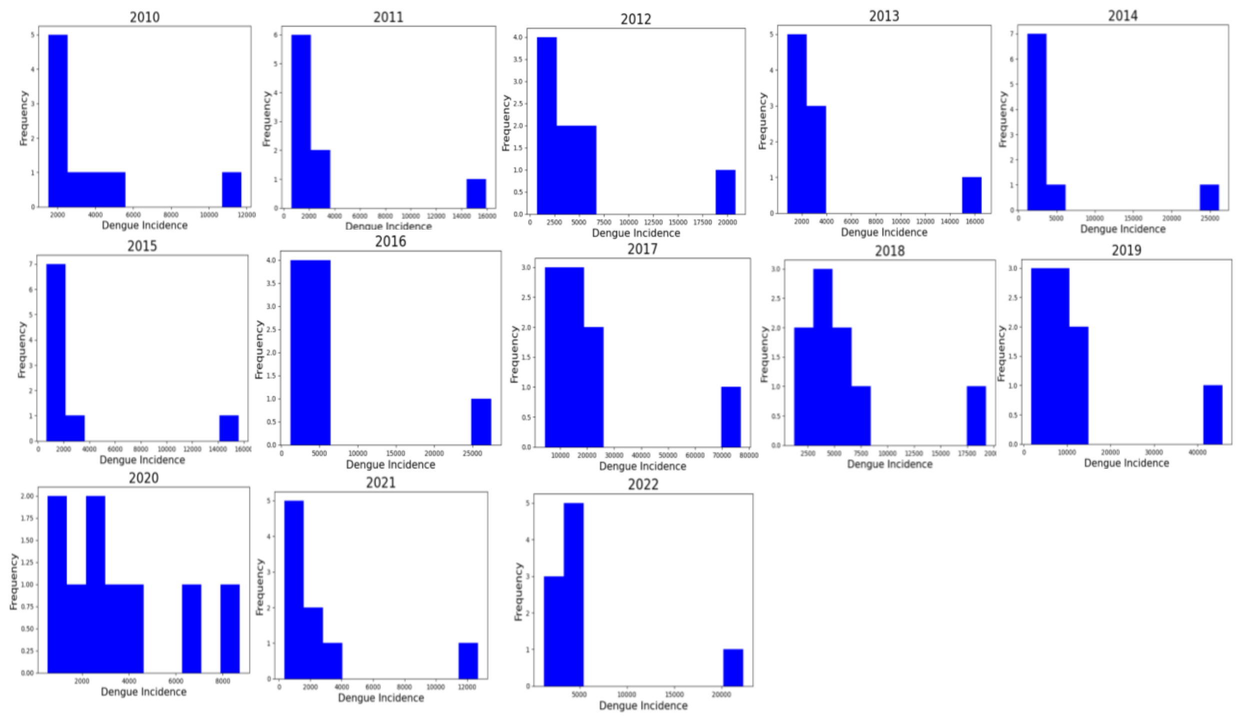

Fig. 4.5 Histograms for the data of Dengue incidence from 2010 to 2021.

4.2.2 Shapiro-Wilk Test

Hypothesis:

Ho: Data are from the normal population

Ha: Data are not from the normal population.

If the p-value < 0.05, H0 is rejected at 5% significance level, then the series is not normally distributed.

Table 4.1 Shapiro-Wilk Normality Test for the data of Dengue incidence from 2010 to 2022.

Year (Dengue Incidence)	P-value	Interpretation
2010	3.007e-04	Not normal
2011	9.758e-06	Not normal
2012	9.826e-04	Not normal
2013	4.154e-05	Not normal
2014	1.045e-0.5	Not normal
2015	1.562e-05	Not normal

2016	6.160e-05	Not normal
2017	2.822e-04	Not normal
2018	3.560e-03	Not normal
2019	6.214e-04	Not normal
2020	4.063e-02	Not normal
2021	9.282e-0.5	Not normal
2022	1.432e-04	Not normal

Since the variables considered in this study did not follow normality assumptions, non-parametric statistical methods were used in order to do the analysis.

4.3 Correlation Analysis

Since the data was distributed Not-normally, Spearman's Correlation test was done for all the datasets of Dengue incidence and the 3 climate variables from the year 2010 to 2022 to identify the Correlations between Dengue and climate factors.

Prec. – Average Precipitation (mm/Day)

Tem. – Temperature at 2 meters

Hum. – Relative Humidity at 2 meters

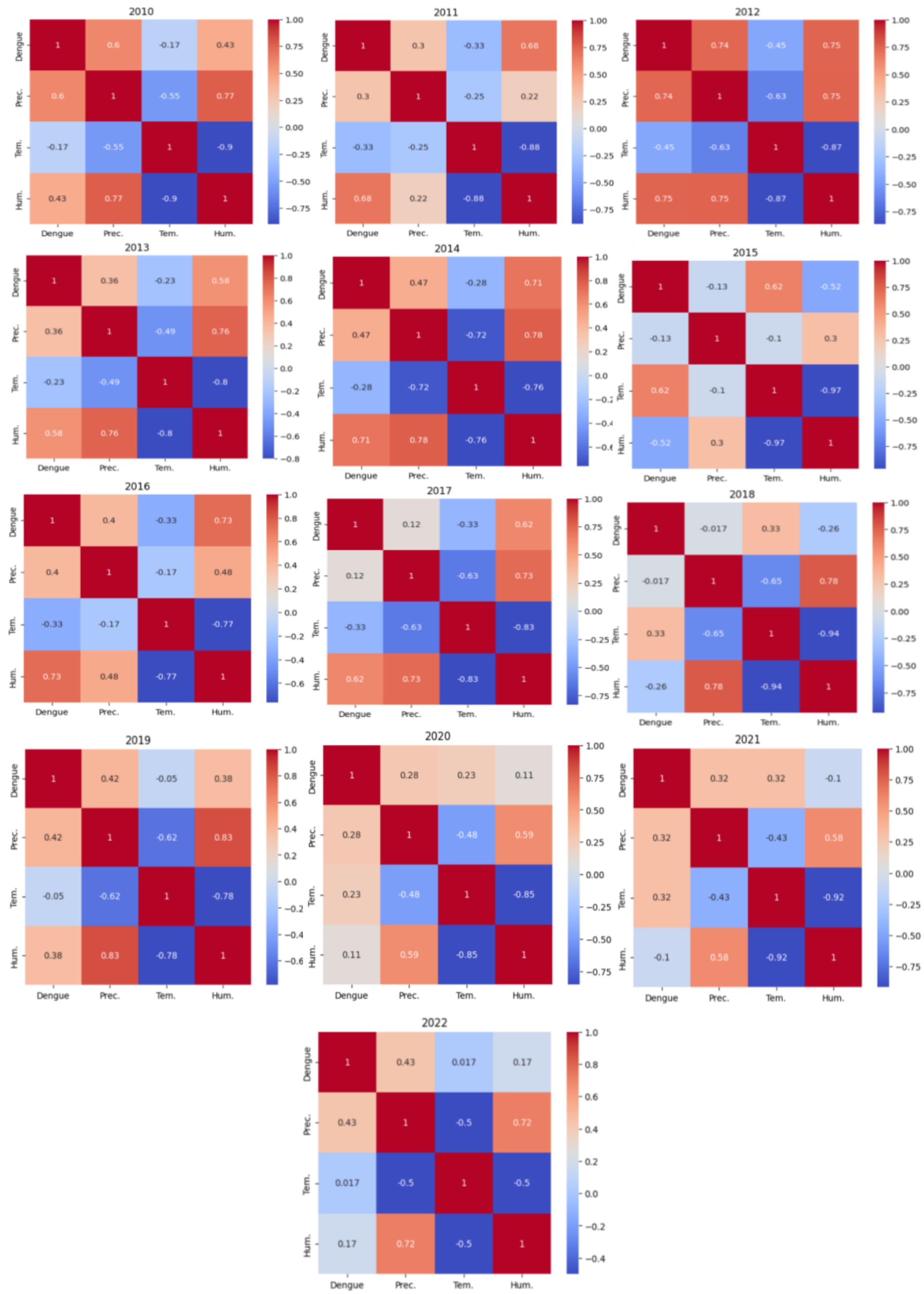

Fig. 4.6 Heatmaps for the data of Dengue incidence and the climate variables from 2010 to 2022.

Year	P-value and the Correlation between Dengue and Precipitation	P-value and the Correlation between Dengue and Temperature	P-value and the Correlation between Dengue and Humidity
2010	0.6 (Positive)	-0.17 (Negative)	0.43 (Positive)
2011	0.3 (Positive)	-0.33 (Negative)	0.68 (Positive)
2012	0.44 (Positive)	-0.23 (Negative)	0.67 (Positive)
2013	0.36 (Positive)	-0.23 (Negative)	0.58 (Positive)
2014	0.47 (Positive)	-0.28 (Negative)	0.71 (Positive)
2015	-0.13 (Negative)	0.62 (Positive)	-0.52 (Negative)
2016	0.4 (Positive)	-0.33 (Negative)	0.73 (Positive)
2017	0.22 (Positive)	-0.27 (Negative)	0.7 (Positive)
2018	-0.17 (Negative)	0.33 (Positive)	-0.26 (Negative)
2019	0.42 (Positive)	-0.05 (Negative)	0.38 (Positive)
2020	0.28 (Positive)	0.23 (Positive)	0.11 (Positive)
2021	0.32 (Positive)	0.31 (Positive)	-0.1 (Negative)
2022	0.43 (Positive)	0.017 (Positive)	0.17 (Positive)

Table 4.2 P- values and the correlations between Dengue and climate variables from 2010 to 2021.

4.4 Hotspot Analysis

Dengue hotspots were analyzed considering all the provinces of Sri Lanka from 2010 to 2022.

2010	2011	2012	2013	2014	2015
Western	Western	Western	Western	Western	Western
Northen	Central	Sabaragamuwa	North West	Sabaragamuwa	Northen
Sabaragamuwa	Sabaragamuwa	North West	Sabaragamuwa	North West	Easten
Southern	Southern	Southern	Central	Central	North West
Central	Easten	Central	Southern	Northen	Central
North West	North West	Northen	Northen	Southern	Southern
Uva	Uva	Easten	North Central	Easten	Sabaragamuwa
Easten	Northen	North Central	Easten	Uva	Uva
North Central	North Central	Uva	Uva	North Central	North Central

2016	2017	2018	2019	2020	2021	2022
Western	Western	Western	Western	Western	Western	Western
Central	Sabaragamuwa	Northen	Southern	Easten	Easten	North West
Southern	North West	Easten	Northen	Central	North West	Easten
Sabaragamuwa	Central	Central	Central	Northen	Southern	Sabaragamuwa
Northen	Southern	North West	Sabaragamuwa	Sabaragamuwa	Central	Central
North West	Northen	Sabaragamuwa	Easten	Southern	Sabaragamuwa	Southern
Uva	Easten	Southern	North West	North West	Northen	Northen
Easten	Uva	Uva	Uva	North Central	Uva	Uva
North Central	North Central	North Central	North Central	Uva	North Central	North Central

High-Intensity Regions
Moderate-Intensity Regions
Low-Intensity Regions

Fig. 4.7 Hotspot Analysis of Dengue incidence from 2010 to 2022.

Armed with this knowledge, the government and health authorities can take necessary actions, such as intensified vector control, public awareness campaigns, and focused healthcare services, aimed at reducing the number of Dengue incidences within these identified hotspots. Ultimately, harnessing the power of geospatial analysis offers a promising approach to enhance disease surveillance and facilitate informed decision-making, paving the way for a more effective and proactive approach to Dengue prevention and control.

4.5 Analysis of risky months

The months were listed out according the number of Dengue Incidence in Sri Lanka from 2010 to 2022.

2010	2011	2012	2013	2014	2015
July	July	June	January	June	January
January	December	August	August	July	February
February	June	July	February	November	December
August	September	November	March	December	November
June	November	January	July	October	July
March	October	December	December	March	August
April	August	October	May	August	March
May	May	February	November	January	May
September	April	September	June	September	August
October	March	March	April	February	June
December	February	May	September	April	April
November	January	April	October	May	September

2016	2017	2018	2019	2020	2021	2022
July	July	January	November	January	December	January
December	June	July	December	February	November	June
January	August	December	October	June	July	February
August	May	June	August	June	October	July
June	March	November	July	March	April	December
February	April	February	September	August	March	October
September	January	August	June	May	January	May
April	December	May	January	December	February	March
March	September	March	May	September	June	August
May	November	April	March	October	August	November
November	February	September	February	November	May	April
October	October	October	April	April	September	September

Fig. 4.8 Analysis of months according to the number of Dengue incidence from 2010 to 2022.

The absence of a clear temporal pattern suggests that factors influencing Dengue transmission in Sri Lanka might exhibit complexities that elude straightforward monthly correlations. This finding underscores the intricate interplay of various environmental, climatic, and socio-economic variables contributing to the unpredictable nature of Dengue outbreaks in the region. Further investigation and integration of additional contextual factors are warranted to unravel the nuanced dynamics underlying the temporal distribution of Dengue cases in Sri Lanka.

4.6 Accuracy of the forecasting model: -

Mean Absolute Error (MAE) and the R-squared (R2) score were tested for the model.

Mean Absolute Error (MAE): 6966.143333333333

The MAE of approximately 6966.143 means that, on average, this model's predictions for the number of dengue incidences are off by approximately 6966.143 cases when compared to the actual observed values.

R-squared (R2) Score: 0.7621698529434118

An R2 score of 0.7622 means that approximately 76.22% of the variance in the number of dengue incidences is explained by this model. In other words, this model captures a significant portion of the variability in the data.

To enhance the accuracy of this dengue incidence prediction model, the following steps can be considered. Firstly, relevant features, such as climate data, population factors, or public health interventions, should be incorporated into the dataset to provide the model with more information. Secondly, the model's hyperparameters, such as the number of trees and their depth, can be fine-tuned through hyperparameter optimization techniques. Thirdly, ensemble methods that combine various regression algorithms can be explored to capture diverse patterns in the data. Fourthly, data quality should be ensured by addressing outliers, handling missing values, and rectifying inconsistencies. Fifthly, cross-validation can be applied to assess generalization and identify instances of overfitting. Lastly, domain expertise should be sought, and additional data collection efforts can be considered to further refine the model's accuracy and utility.

4.7 Testing the forecasting model: -

The model was tested for the years 2021 and 2022 comparing the number of predicted Dengue Incidence and the number of real Dengue Incidence considering every province separately.

Table 4.3 The predicted and real values of Dengue incidence in 2021

Province	Predicted value	Real value
Western	16114	12683
Central	5365	2351
Southern	3680	2987
Northen	5088	3835
Easten	6641	5453

North West	3416	2335
North Central	1064	637
Uva	1205	921
Sabaragamuwa	4362	3261

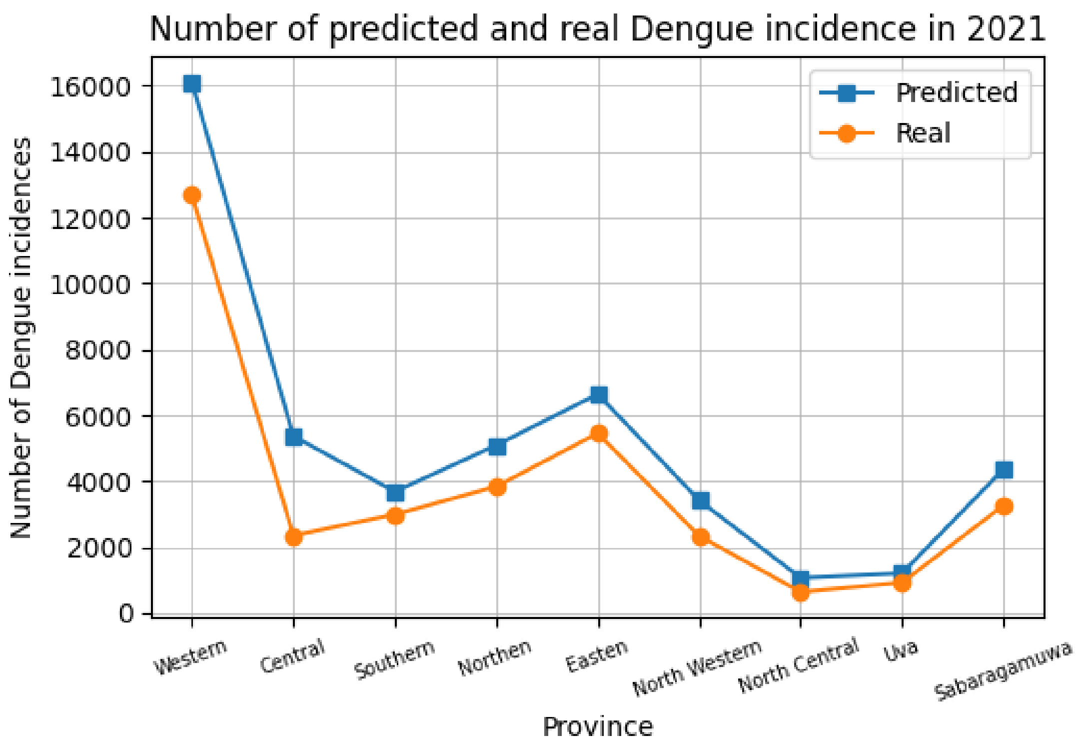

Fig. 4.9 Comparison of Predicted and Real values in 2021.

Table 4.4 The predicted and real values of Dengue Incidence in 2022

Province	**Predicted value**	**Real value**
Western	20860	22256
Central	3949	4243
Southern	3448	3489
Northen	3683	2732
Easten	4913	5325
North West	4535	5360
North Central	956	1223
Uva	1664	1439
Sabaragamuwa	3863	4344

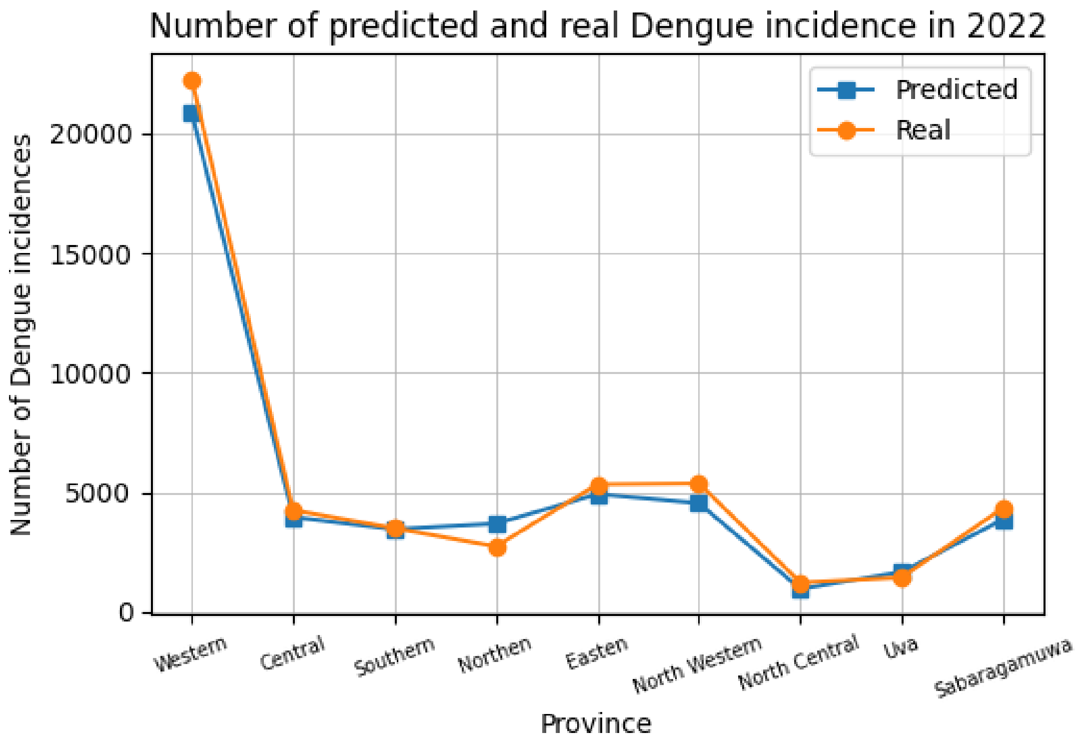

Fig. 4.10 Comparison of Predicted and Real values in 2022.

4.8 Forecasting the number of Dengue Incidence in 2023 and 2024

The model's accuracy was confirmed through an evaluation of its performance in predicting Dengue incidence for the years 2021 and 2022, demonstrating close alignment with the actual reported values. Encouraged by the favorable comparison between predicted and actual data, the model was further extended to forecast Dengue incidence for the years 2023 and 2024.

Table 4.5 The predicted number of Dengue incidence in 2023 and 2024.

Province	**2023**	**2024**
Western	18423	17761
Central	4374	4319
Southern	3033	3058
Northen	3316	3378
Easten	5706	5741
North West	4222	4044
North Central	1080	1053

Uva	1205	1157
Sabaragamuwa	4006	3897

By leveraging this forecasting approach, the research contributes a valuable tool for health authorities and policymakers in Sri Lanka to anticipate and proactively address potential fluctuations in Dengue incidence. The benefits of such forecasting are manifold: it enables early resource allocation, targeted public health interventions, and timely deployment of vector control measures, thereby reducing the impact of Dengue outbreaks on public health. This forward-looking approach empowers decision-makers with actionable insights to curb the spread of the disease, potentially minimizing its socio-economic burden and fostering a more resilient healthcare system in Sri Lanka.

CHAPTER 5

CONCLUSION

In conclusion, of this study, titled "Modeling and Forecasting the Dengue Incidence in Sri Lanka," entailed a comprehensive examination of the interplay between Dengue incidence and climate variables across all provinces from 2010 to 2022. It was determined during the initial assessment that the data did not follow a normal distribution, necessitating the utilization of specialized analysis techniques. Subsequently, an in-depth exploration of the relationship between Dengue incidence and climate variables, comprising precipitation, temperature, and humidity, was carried out via heatmap visualization and Spearman's correlation test. Notably, positive correlations were ascertained between Dengue incidence and both precipitation and humidity, whereas temperature exhibited a negative correlation.

Furthermore, spatial analysis encompassed hotspot analysis for each province, revealing distinct patterns in Dengue intensity, resulting in the categorization of provinces into "High-Intensity Regions," "Moderate-Intensity Regions," and "Low-Intensity Regions." A temporal analysis of monthly Dengue incidence was executed, albeit without the emergence of discernible patterns, thereby highlighting the intricate nature of Dengue's seasonality in Sri Lanka. The culmination of this research was the development of a robust forecasting model employing the Random Forest Regressor. The model's accuracy was validated against actual data for 2021 and 2022, revealing satisfactory predictive capabilities. Leveraging this validated model, forecasts for Dengue incidence in 2023 and 2024 were generated.

Collectively, this research has contributed valuable insights into the intricate relationship between climate variables and Dengue incidence, alongside the spatial and temporal patterns of Dengue occurrence in Sri Lanka. This study contributes to our understanding of the relationship between climate factors and Dengue incidence, the spatial distribution of Dengue risk, and the potential for accurate short-term forecasting. These insights have practical implications for public health authorities in Sri Lanka. By leveraging these insights, we aim to contribute to more effective Dengue prevention and control measures in the years to come.

LIST OF REFERENCES

Attanayake, A., Perera, S., & Liyanage, U. (2020, July 17). Exponential smoothing on forecasting dengue cases in Colombo, Sri Lanka. *Journal of Science*, *11*(1) (2602–9030). https://doi.org/10.4038/jsc.v11i1.24

Chandrakantha, L. (2019, May 10). Risk Prediction Model for Dengue Transmission Based on Climate Data: Logistic Regression Approach. *Stats 2019*, *2*(2) (272–283). https://doi.org/10.3390/stats2020021

Erandi, K., Perera, S., & Mahasinghe, A. (2021, January 7). Analysis and forecast of dengue incidence in urban Colombo, Sri Lanka. *Theoretical Biology and Medical Modelling*, Article 3 (2021). https://doi.org/10.1186/s12976-020-00134-7

Faruk, M., Jannat, S., & Rahman, M. (2022, January 14). Impact of environmental factors on the spread of dengue fever in Sri Lanka. *International Journal of Environmental Science and Technology (2022*. https://doi.org/10.1007/s13762-021-03905-y

Gnanapragasam, S. (2018). Modeling and Forecasting Dengue Cases in the area of Colombo Municipal Council in Sri Lanka. *International Journal of Scientific and Research Publications*, *6*(2250–3153). https://www.ijsrp.org/research-paper-0816.php?rp=P565692

NASA POWER | Data Access Viewer. (n.d.). Retrieved September 22, 2022, from https://power.larc.nasa.gov/data-access-viewer/

National Dengue Control Unit. (2022b). National Dengue Control Unit. http://www.dengue.health.gov.lk/

Pathirana, S., Kawabata, M., & Goonatilake, R. (2009). STUDY OF POTENTIAL RISK OF DENGUE DISEASE OUTBREAK IN SRI LANKA USING GIS AND STATISTICAL MODELLING. *Rural and Tropical Public Health*, *8*. https://jrtph.jcu.edu.au/vol/JRTPH_Vol8_p8-17_Pathirana.pdf

Sirisena, P., Noordeen, F., Kurukulasuriya, H., ALAR Romesh, T., & Fernando, L. (2017, January 9). Effect of Climatic Factors and Population Density on the Distribution of Dengue in Sri Lanka: A GIS Based Evaluation for Prediction of Outbreaks. *PLOS ONE*, *12(1)*. https://doi.org/10.1371/journal

Sun, W., Xue, L., & Xie, X. (2017, October 10). Spatial-temporal distribution of dengue and climate characteristics for two clusters in Sri Lanka from 2012 to 2016. *Scientific Reports*, Article 12884 (2017). https://doi.org/10.1038/s41598-017-13163-z

Withanage, G., Viswakula, S., & Nilmini Silva Gunawardena, Y. (2018, April 24). A forecasting model for dengue incidence in the District of Gampaha, Sri Lanka. *Parasites Vectors*, *11*(262 (2018)). https://doi.org/10.1186/s13071-018-2828-2

World Health Organization (WHO). (2022b). World Health Organization. https://www.who.int/

APPENDICES

Appendix 1: Dataset of all the variables of Western province

Year	**Number of Dengue Incidence**	**Average Precipitation (mm/Day)**	**Average Temperature at 2 meters**	**Relative Humidity at 2 meters**
2010	11709	7.81	26.16	84.81
2011	12913	6.22	26.01	85.5
2012	15814	6.67	26.26	83.31
2013	16540	6.07	26.09	85.31
2014	22153	6.52	26.41	83.5
2015	15582	6.69	26.42	84.56
2016	23442	4.49	26.65	82.5
2017	63882	5.04	26.53	82.5
2018	40270	5.6	26.27	83.62
2019	45686	6.52	26.85	82.06
2020	20234	5.36	26.8	82.69
2021	12465	7.61	26.4	84.94
2022	15653	7.34	26.31	85.5

Appendix 2: Dataset of all the variables of Central province.

Year	Number of Dengue Incidence	Average Precipitation (mm/Day)	Average Temperature at 2 meters	Relative Humidity at 2 meters
2010	2521	6.38	24.68	86.25
2011	2252	6.58	24.5	86.94
2012	3455	5.25	24.99	83
2013	2530	4.88	24.55	87
2014	3299	6.2	25.23	82.75
2015	1906	5.58	24.91	85.94
2016	5632	3.28	25.51	80.25
2017	18471	3.79	25.55	80.25
2018	4845	4.3	25.01	83.38
2019	9823	4.75	25.9	79.62
2020	4206	4.26	25.4	82.81
2021	1351	6.21	24.76	87.5
2022	4243	6.32	24.56	81.34

Appendix 3: Dataset of all the variables of Southern province.

Year	Number of Dengue Incidence	Average Precipitation (mm/Day)	Average Temperature at 2 meters	Relative Humidity at 2 meters
2010	2536	6.88	26.9	83.25
2011	2090	6.5	26.69	84.12
2012	3952	6.4	26.8	82.75
2013	1771	5.47	26.75	84.06
2014	2637	5.92	26.97	82.75
2015	1887	7.01	27.03	83.56
2016	5370	4.7	27.23	81.88
2017	16174	5.74	27.12	81.88
2018	3096	5.42	26.79	83.38
2019	13481	6.64	27.38	81.56
2020	2584	5.23	27.3	82.56
2021	1587	6.9	27.06	83.56
2022	3489	6.67	26.86	83.93

Appendix 4: Dataset of all the variables of Northen province.

Year	Number of Dengue Incidence	Average Precipitation (mm/Day)	Average Temperature at 2 meters	Relative Humidity at 2 meters
2010	4849	4.49	27.38	78.94
2011	754	5.27	27.03	80.38
2012	1721	4.19	27.65	76.12
2013	1665	3.69	27.05	80.31
2014	3081	5.02	27.58	77.44
2015	3090	5.28	27.05	82.5
2016	4216	2.85	27.85	76
2017	11751	2.89	27.78	76
2018	7111	3.14	27.48	77.25
2019	11597	3.78	28.18	75
2020	3751	3.95	27.74	77.56
2021	1139	5.14	27.27	81.19
2022	2732	4.56	28.33	82.2

Appendix 5: Dataset of all the variables of Easten province.

Year	Number of Dengue Incidence	Average Precipitation (mm/Day)	Average Temperature at 2 meters	Relative Humidity at 2 meters
2010	2405	5.03	27.73	78.69
2011	2066	7.54	27.47	79.62
2012	1040	5.22	27.94	76.06
2013	1005	4.39	27.53	79.12
2014	1789	5.39	27.83	77.44
2015	2128	6.13	27.62	80.5
2016	1375	3.68	28.21	76.44
2017	11548	3.58	28.02	76.44
2018	6307	3.67	27.59	78.44
2019	6205	4.65	28.37	75.5
2020	6331	4.11	28	77.75
2021	3453	5.31	27.73	79.94
2022	5325	5.87	28.01	76.89

Appendix 6: Dataset of all the variables of North West province.

Year	Number of Dengue Incidence	Average Precipitation (mm/Day)	Average Temperature at 2 meters	Relative Humidity at 2 meters
2010	2488	4.59	26.97	81.19
2011	1611	3.98	26.66	82.5
2012	5337	4.05	27.15	79.12
2013	3696	3.31	26.74	82.19
2014	3380	4.9	27.23	79.69
2015	1992	4.73	26.92	82.94
2016	3602	2.66	27.43	77.88
2017	19110	2.73	27.45	77.88
2018	4474	3.34	27.13	79.5
2019	5475	3.98	27.85	76.94
2020	1469	3.5	27.5	79.25
2021	2335	5.09	26.9	83.25
2022	5360	4.68	27.89	82.11

Appendix 7: Dataset of all the variables of North Central province.

Year	Number of Dengue Incidence	Average Precipitation (mm/Day)	Average Temperature at 2 meters	Relative Humidity at 2 meters
2010	1512	4.46	26.68	80.19
2011	603	5.54	26.26	82.31
2012	782	4.52	26.99	77.19
2013	1077	3.69	26.35	81.75
2014	1190	5.3	26.99	78.19
2015	651	5.42	26.27	84.38
2016	1210	2.96	27.22	76.56
2017	4325	3.06	27.19	76.56
2018	1209	3.44	26.84	78.56
2019	1725	4.02	27.66	75.31
2020	684	3.88	27.05	78.94
2021	337	4.96	26.5	82.94
2022	1223	4.21	26.72	80.92

Appendix 8: Dataset of all the variables of Uva province.

Year	Number of Dengue Incidence	Average Precipitation (mm/Day)	Average Temperature at 2 meters	Relative Humidity at 2 meters
2010	2413	6.32	24.39	84.25
2011	944	7.24	24.26	84.88
2012	717	5.25	24.55	82.12
2013	813	4.86	24.4	84
2014	1426	5.79	25.14	79
2015	789	6.94	24.23	87.19
2016	1660	3.4	25.19	80.62
2017	6986	5.12	24.96	80.62
2018	1467	4.65	24.21	85.06
2019	2255	5.48	25.37	79.81
2020	525	4.64	24.87	82.56
2021	921	7.1	24.29	87.06
2022	1439	6.39	25.56	84.44

Appendix 9: Dataset of all the variables of Sabaragamuwa province.

Year	Number of Dengue Incidence	Average Precipitation (mm/Day)	Average Temperature at 2 meters	Relative Humidity at 2 meters
2010	3755	6.96	22.33	89.25
2011	2240	6.61	22.24	89.75
2012	2240	5.67	22.44	87.69
2013	2966	5.55	22.3	89.75
2014	4547	6.45	22.59	87.81
2015	1752	6.12	22.59	89.19
2016	4643	3.57	22.91	85.31
2017	20854	4.54	22.93	85.31
2018	3780	4.7	22.49	87.69
2019	6802	5.37	23.17	85.38
2020	2879	4.63	22.92	87.12
2021	1261	6.6	22.55	89.88
2022	4344	6.56	23.01	88.68