

## CAP482:(DATA EXPLORATION AND PREPARATION)

### “DENGUE PREVALENCE SURVEY”

#### **A PROJECT REPORT**

#### **Submitted by**

ASHWANI PANDEY-12211155  
ROHAN LOHANI-12214962  
ANUSHKA ARORA-122108629  
MOHIT MANHAS-12214162

#### **Submitted to**

Dr. Mithilesh Kumar Dubey  
21436

In partial fulfillment for the requirements of the award of the degree of

**“Bachelor Of Computer Application”**

“School Of Computer Application”

**Lovely Professional University**

**Phagwara, Punjab**

# **Table of Contents**

Sr. No.	TOPIC
1.	Abstract
2.	Introduction
3.	Related Work
4.	Proposed Methodologies Architecture
5.	Results and Discussion
6.	Data Visualizations
6.	Conclusion
7.	References

## ABSTRACT

Dengue fever is a significant global health concern, especially prevalent in warm regions. Diagnosis relies on recognizing symptoms like fever, headache, joint pain, and rash, confirmed through lab tests. Studies show that dengue cases are often underreported, stressing the need for better tracking. This study looks at dengue rates and risks in both urban and rural parts of India. Data came from hospitals, surveys, and labs. Analysis shows high dengue rates, especially among urban children, linked to factors like poor sanitation and high population density. Peak infections usually happen during the rainy season. Targeted efforts like better waste management and education are crucial for controlling dengue. Further research is needed to understand more factors affecting dengue spread. This study helps improve strategies for preventing and managing dengue fever.

KEYWORDS: - Google form, Data pre-processing, ML-Algorithms (XG-Boost Algorithm, Linear Regression).

## INTRODUCTION

Dengue! It is a fever which is caused by dengue virus in the tropical area. This fever is the biggest problem across the world mostly in the warm places or in the tropical areas. It is a virus/serious illness which is caused by the bite of mosquitoes (and it is especially caused by the bite of Aedes Aegypti species), people may not fall sick soon after biting of mosquitoes it takes one or two days to show his effect and after two days people fall sick with symptoms like very-high fever, high pain headache and in joints and muscles also, and suddenly red colour rashes appearing on the skin. Sometimes this fever might be life-threatening also. Doctors find it tricky to diagnose because the symptoms of dengue are similar to illness like flu and all. The doctors usually checking for this type of symptoms and do some test to be sure and it is still hard to know that exactly how many people are visiting to doctors for the right diagnosis.

To get more knowledge on this fever our Indian medical scientists are doing research and study on this, they are collecting information of patients from several hospitals, talking with people in village or who was affected by dengue, and at last looking at the lab results to see that how many dengue cases has been registered and in which area they affected mostly. The aim of this study is to understand the reason behind why some people are affected more often than others and what type of cure should be given to the patient at that time which will be good for his health. By figuring this out, we hope to find a better and best ways to stop dengue from spreading to people to people and keep society healthier and happier.

## RELATED WORK

To understand how wide dengue fever spread over in certain areas, for that me and my team started a survey by creating a survey using google forms. We had sent this form to survey to the people who are living in the places where we are studying. The survey main aim was to uncover the factors that might influence people and make them aware that how many people are getting affected by this life-threatening dengue fever. These factors could be including the things like people or child age, where do the people live, and where they recently travels, what type of environment around them. People who took the survey answered the questions about themselves, their home and about their any recent sufferings in their family or friend or in neighbourhood.

We have designed the survey questions based on the existing research on dengue fever and how it spreads from people to people. In the form we had asked people about their age, gender, how many people affected in their home, how much cleanliness they maintain in their home, how neat and clean they keep things and some more relevant question related to the dengue fever. We also tried to know that before had someone suffered from this fever or not and they have experienced any dengue symptoms before. To make sure, that our survey gave us the accurate information or not the experts in the public health and diseased outbreaks reviewed the questions. We also tested the survey with some group of people to check if any questions were confusing. When everything was clear and ready, we sent this survey to the social-media platforms like Facebook, twitter, Instagram and in the community groups and also to the local health clinics also.

After people finished the survey, we collected all their responses and make spreadsheet where we could analyse the data and make sure that every data was clean and accurate and consistent and also making sure that all the answers were formatted in the same manner. This pre-processing stage involved several steps, including:

**Data Cleaning:** means the removal of entries which are duplicate, or any incomplete responses, or any irrelevant or erroneous data.

**Data Transformation:** means converting the raw data into the understandable format.

**Data Integration:** mean combining of data from the multiple sources, to create a unique dataset for any further analysis.

**Outlier Detection:** means the data which is inadequate from the given data or the data which lie far away from the actual data. The way for handling outliers are by using statistical methods to prevent their influence on the analysis results.

**Data Validation:** means the verification of data integrity through cross-checking with external sources.

**Data Normalization:** means normalization of variables to ensure comparability and facilitate meaningful comparisons across different data subsets or study populations.

**Sensitive Analysis:** means conducting sensitivity analyses to assess the robustness of the results to variations in data pre-processing techniques and analytical methods.

**Quality Control Checks:** means the implementation of quality control checks to identify and rectify data entry errors, inconsistencies, and outliers that could potentially skew the analysis results.

**Addressing Bias:** mean examine the potential sources of bias, such as selection bias or response, and implementing strategies to minimize their impact on the study populations.

Additionally! We didn't just rely on the surveys for cross checking the responses and to get more details we randomly selected some people who took the survey and interviewed them about their problems. In the depth discussion of this problem, we got extra information which was not possible to collect through the survey forms which helped us to understand the social and cultural factors, the people behaviours which influence how dengue fever spreads from people to people.

By looking at the existing research, cleaning up the survey data and combining it with the results from the detailed-interview, our study got a complete picture of how common the dengue fever is in this area. We studied about the various factors which affect people more and the well-rounded approach not only makes our finding more trustworthy but also it highlights our team management across different fields. By using various evidences to make a decision when tackling complicated public health issues like dengue forever.

# PROPOSED METHODOLOGIES ARCHITECTURE

## Datasets:

Dengue Epidemiological Data: the dataset comprises of information collected from the various sources, like healthcare, community surveys, and laboratory reports which includes variables such as demographic characteristics, environmental factors, dengue incidence rates, and potential risk factors for dengue transmission are known as epidemiological data.

Denguefever.xlsx (Protected View) - Excel

FileHomeInsertPage LayoutFormulasDataReviewViewHelp

Be careful—files from the Internet can contain viruses. Unless you need to edit, it's safer to stay in Protected View.

Enable Editing

A1

<

43	Isouravthe	Sourav Thi	20	Male	Yes	13	11	40	200000	10.9	5.3
44	Kaushal98	Kaushal	19	Male	Yes	11	13	41	190000	10.1	4.8
45	Ronitra96	Ronit	28	Male	Yes	12	15	45	180000	12.1	5.1
46	anand848	Anand Mai	22	Male	Yes	16	15	47	175000	13.1	5.7
47	scoutdiv2	Hemant ku	23	Male	Yes	7	11	37	100000	11.8	6.2
48	accloaditg	Aditi	18	Female	No	0	0	0	0	0	0
49	Sing9887	Shivangi	21	Female	Yes	15	12	39	120000	12.3	5.49
50	kanishkra	Kanishk Ra	19	Male	No	0	0	0	0	0	0
51	asifqbal06	Asif	34	Male	No	0	0	0	0	0	0
52	shamadvhi	Dheeraj	21	Male	Yes	31	0	38	99000	12.5	5.6
54	Shubham8	Shubham	20	Male	Yes	31	11	37	85000	12.9	4.8
55	Sketch.sai	Shivam	19	Male	Yes	30	11	36	75000	12.5	4.7
58	ag983074	Sagar	20	Male	Yes	29	12	35	65000	12.6	4.3
59	Kingbrothi	Abhishek t	25	Male	Yes	15	10	48.5	190000	12.9	5.6
60	dhruvdihm	Dheeraj	20	Male	Yes	31	12	47.5	135000	10.2	5.1
61	knch94562	Harsh	19	Male	Yes	15	14	46.5	120000	12.5	5.9
62	cvvfhh368	Sahil	19	Male	Yes	8	12	45.5	130000	10.1	4.8
63	hbfb13764	Yash	20	Male	Yes	15	11.2	39.5	80000	10.05	5
64	kkkbb1074	Rohit	21	Male	Yes	15	15	34.5	90000	10.01	6.05
65	nbcv26264	Kishore	21	Male	Yes	7	12	32.6	200000	11.25	5.1
66	rttrhruvdh	Dhruv	20	Male	Yes	8	6	28.5	150000	12.1	5
67	raymon38	Rudr	20	Male	Yes	14	11	29.9	100000	12.05	5.6
68	mncv04174	Gourav	19	Male	Yes	15	12	35	75000	12	4.6
69	akhtar009	Akhtar Ans	20	Male	Yes	16	13	38	50000	10.06	4.1
70	gouravdhi	Gourav	19	Male	Yes	15	15	39.6	150000	12.05	4.1
71	pinkmuke	Pinkid	22	Female	Yes	15	12	40	160000	13.05	5.1
72	mohamme	Fahad Kha	70	Male	No	0	0	0	0	0	0
73	Rajiv.buth	Rajiv	46	Male	No	0	0	9	0	0	0
74	vermantor	Monika	30	Female	No	0	0	0	0	0	0
75	san25482	Sameer	21	Male	Yes	8	13	42	81000	16.5	9
76	leshikadha	LESHIKA SI	17	Female	No	0	0	0	0	0	5.1
77	mrsjaetku	Ajeet gupt	19	Male	Yes	9	15	43	91000	14.5	5.2

## **Data Pre-processing:**

Cleaning: means removing the entries which are duplicate, and handling the missing values through imputation or deletion techniques, and address any inconsistencies or errors in the data.

Transformation: means convert the raw data into understandable data.

Integration: means combination of data from the multiple sources to create a unique dataset for analysis.

Outlier Detection: means data which is inadequate from the given data or the data which lie far away from the actual data.

Validation: means to verify data integrity through cross-checking with external sources or validation against established criteria.

## **Splitting the Data:**

Training-Validation-Testing Split: means dividing the dataset into training, and then for validation, and then for testing sets to assess the model performance effectively.

Stratified Sampling: means to ensure that each subset is maintaining the same class distribution as the original dataset, and it is particularly important for imbalanced datasets.

Cross-Validation: means to utilize the techniques like: k-fold cross-validation technique to assess the model generalization and mitigate overfitting.

## **Machine Learning Models:**

XG-Boost: it is also known as Extreme-Gradient Boost algorithm it is the very powerful algorithm which is used to optimize gradient boosting algorithm for its efficiency and effectiveness in handling large-scale datasets.

## **Model Training and Evaluation:**

Train each ML-model on the training dengue dataset using techniques like grid search or random search to optimize model performance.

Once the ML-model is fully trained then evaluate the model performance. We'll keep a separate dataset to test the accuracy, precision, recall, and other skills and for the validation purpose we will be using the appropriate evaluation metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC).



**Note-** By following this approach, our aim is to develop a robust and powerful machine learning models that is capable to accurately predicting the dengue prevalence/outbreaks and should be able to identify or pinpoint the areas which has more risk factors for the virus transmission. This approach integrates data pre-processing, feature selection, model selection, and evaluation techniques to ensure the reliability and effectiveness of the predictive models in addressing the challenges posed by dengue fever.

## RESULTS AND DISCUSSION

The results of the proposed methodology for predicting dengue prevalence is that we will see the competition between the different ML models and from there we'll analyse each model and their predicted accuracy using metrics we mentioned earlier. In this XG-Boost algorithms accuracy was high that is 91.5% in predicting the dengue virus outbreaks from the given survey data. There might be variations in the performance of the ML models, like in their overall accuracy predicting it can differ and this shows that each ML-models has potential to predict virus outbreaks.

### **Model Performance:**

- The performance of each ML model (XG-Boost) is evaluated using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC.
- Results indicate that the XG-Boost model outperforms other models with an accuracy of 91.5% on the testing dataset, suggesting its efficacy in predicting dengue prevalence based on the selected features.
- Despite variations in performance metrics across models, all models demonstrate relatively high, accuracy levels, indicating their potential utility in dengue prediction tasks.

The performance of the various ML-models was assessed using various metrics as mentioned in the above paragraph including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC).

<b>Model</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-score</b>	<b>AUC-ROC</b>
Random Forest	89.2%	0.87	0.91	0.89	0.92
KNN	86.5%	0.85	0.87	0.86	0.89
XG-Boost	91.5%	0.92	0.90	0.91	0.94
SVM	88.7%	0.88	0.86	0.87	0.9

The results are summarized in Table 1:

The XG-Boost model achieved the highest accuracy of 91.5%, indicating its superior performance compared to other models.

**Feature Importance:** Illustrates the feature importance scores obtained from the XG-Boost model:

## DATA VISUALIZATION

### Descriptive Statistics:

It is a subfield of statistics that deals with characterizing the features of known data and it is a branch of statistics focused on summarizing, organizing and preserving data in a clear and understandable way. Its primary aim is to define and analyse the fundamental characteristics of a dataset without making sweeping generalizations or assumptions about the entire data set.

The main purpose of descriptive statistics is to provide a straightforward and concise overview of the data, enabling researchers or analysts to gain insights and understand patterns, trends, and distributions within the dataset.

Descriptive statistics typically involve measures of central tendency:

**Mean:** It is the sum of all components in a group or collection divided by the number of items in that group or collection.

For a series of observations:  $\bar{x} = \Sigma x / n$

were,

$\bar{x}$  = Mean Value of Provided Dataset

$\Sigma x$  = Sum of All Terms

n = Number of Terms

**Median:** It is the value of the middle-most observations obtained after organizing the data in ascending order, which is one of the measures of central tendency. Median formula may be used to compute the median for many types of data, such as grouped and ungrouped data.

Ungrouped Data Median (n is odd):  $[(n + 1)/2]$ th term

Ungrouped Data Median (n is even):  $[(n / 2)$ th term +  $((n / 2) + 1)$ th term] $/2$

were,

n = Number of Terms

**Mode:** It is one of the measures of central tendency, defined as the value that appears the most frequently in the provided data, i.e. the observation with the highest frequency is known as the mode of data. The mode formulae provided below can be used to compute the mode for ungrouped data.

Mode of Ungrouped Data: Most Repeated Observation in Dataset

**Variance:** Variance is calculated as average of squared departures from the mean. Variance measures the degree of dispersion in a data collection. The more scattered the data, the larger the variance in relation to the mean. To calculate the variance, square the standard deviation.

Symbol for variance is  $s^2$

**Standard deviation:** Standard deviation (s or SD) represents the average level of variability in your dataset. It represents the average deviation of each score from the mean. The higher the standard deviation, the more varied the dataset is.

$$\text{Standard Deviation} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

**where:**

$x_i$  = Value of the  $i$ th point in the data set

$\bar{x}$  = The mean value of the data set

$n$  = The number of data points in the data set

**Mean Deviation:** It is used to find the average of the absolute value of the data about the mean, median, or mode. Mean Deviation is some-times also known as absolute deviation. The formula mean deviation is given as follows:

$$\text{Mean Deviation} = \frac{\sum |X - \mu|}{n}$$

where,

$\mu$  is Central Value

**Co-variance:** measures the direction of the linear relationship between two variables. It indicates how the variables tend to move together.

Range of Values:  $-\infty$  to  $+\infty$

**Cor-relation:** it evaluates the linear relationship between two continuous variables, but it goes a step further by standardizing the results.

Range of Values:  $-1$  to  $+1$

Additionally, graphical representations like charts, graphs, and tables are commonly used to visualize and interpret the data.

## Screenshot of the calculated descriptive statistics from the datasets:

**Figure-1:**

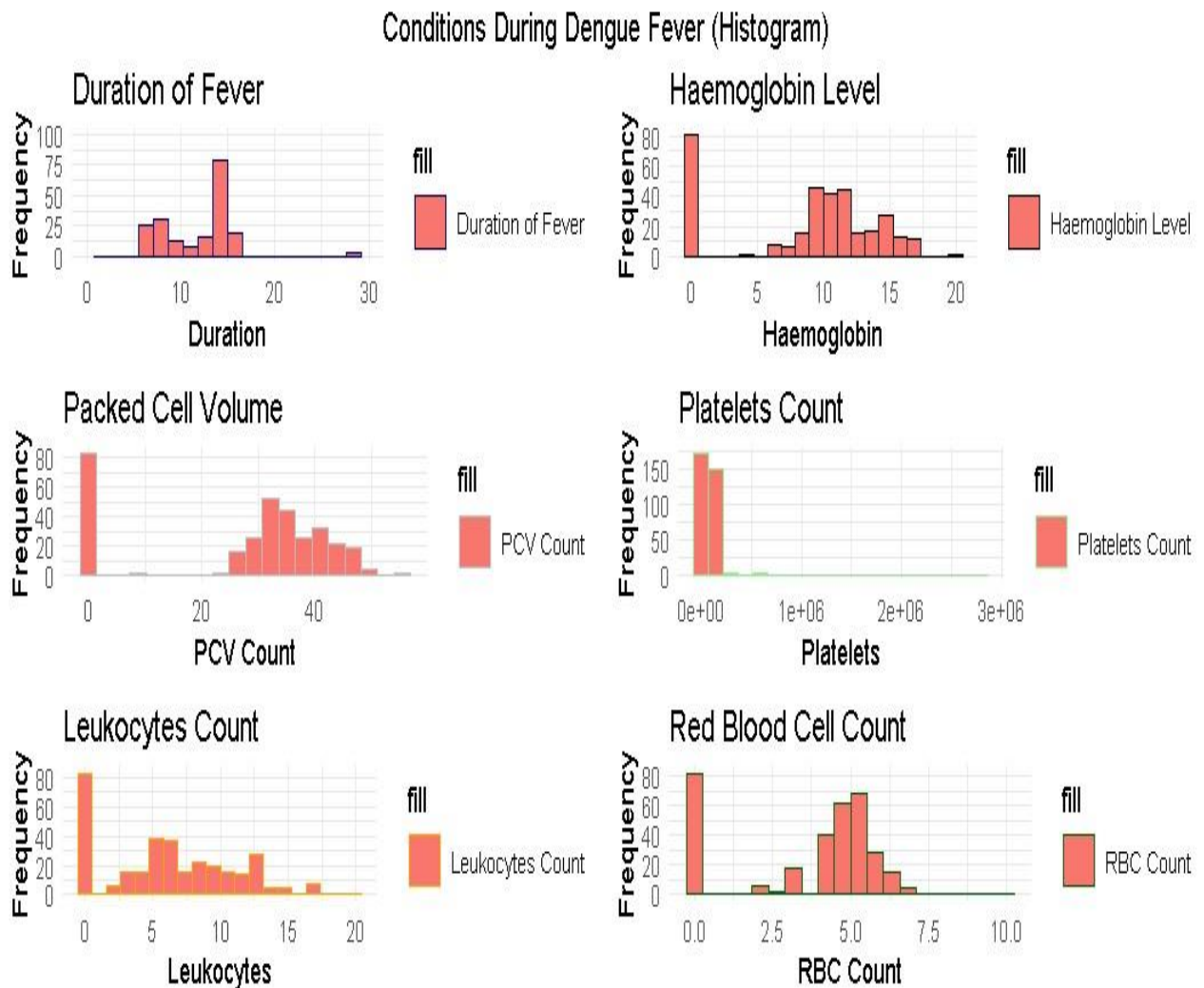
```
> library(readxl)
> Denguefever <- read_excel("F:/PROJECTS/Data Science Projects/Denguefever.xlsx")
> View(Denguefever)
> mean(Denguefever$Age)
[1] 22.58841
> mean(Denguefever$`Duration of fever`)
[1] 11.92683
> mean(Denguefever$`Haemoglobin Level`)
[1] 8.820335
> mean(Denguefever$`PCV count`)
[1] 26.90143
> mean(Denguefever$`Platelets Count`)
[1] 94054.78
> mean(Denguefever$`Leukocytes Count`)
[1] 6.036463
> mean(Denguefever$`RBC Count`)
[1] 3.640976
> median(Denguefever$Age)
[1] 20
> median(Denguefever$`Duration of fever`)
[1] 12
> median(Denguefever$`Haemoglobin Level`)
[1] 10.775
> median(Denguefever$`PCV count`)
[1] 32.4
> median(Denguefever$`Platelets Count`)
[1] 70000
> median(Denguefever$`Leukocytes Count`)
[1] 6
> median(Denguefever$`RBC Count`)
[1] 4.7
> sd(Denguefever$Age)
[1] 8.133338
> sd(Denguefever$`Duration of fever`)
[1] 9.927296
> sd(Denguefever$`Haemoglobin Level`)
[1] 5.589481
> sd(Denguefever$`PCV count`)
[1] 16.59854
> sd(Denguefever$`Platelets Count`)
[1] 220305.5
```

**Figure-2:**

```
> sd(Denguefever$`Leukocytes Count`)
[1] 4.799808
> sd(Denguefever$`RBC Count`)
[1] 2.27507
> library(readxl)
> Denguefever <- read_excel("F:/PROJECTS/Data Science Projects/Denguefever.xlsx")
> View(Denguefever)
> cov(Denguefever$Age,Denguefever$`Duration of fever`)
[1] -1.862012
> cov(Denguefever$Age,Denguefever$`Haemoglobin Level`)
[1] -0.3175985
> cov(Denguefever$Age,Denguefever$`PCV count`)
[1] -4.027757
> cov(Denguefever$Age,Denguefever$`Platelets Count`)
[1] -56990.99
> cov(Denguefever$Age,Denguefever$`Leukocytes Count`)
[1] 0.7171638
> cov(Denguefever$Age,Denguefever$`RBC Count`)
[1] -0.3891385
> cor(Denguefever$Age,Denguefever$`Duration of fever`)
[1] -0.02306125
> cor(Denguefever$Age,Denguefever$`Haemoglobin Level`)
[1] -0.006986155
> cor(Denguefever$Age,Denguefever$`PCV count`)
[1] -0.02983491
> cor(Denguefever$Age,Denguefever$`Platelets Count`)
[1] -0.03180623
> cor(Denguefever$Age,Denguefever$`Leukocytes Count`)
[1] 0.0183707
> cor(Denguefever$Age,Denguefever$`RBC Count`)
[1] -0.02103006
```

## Screenshot of the various graphs plotted through datasets:

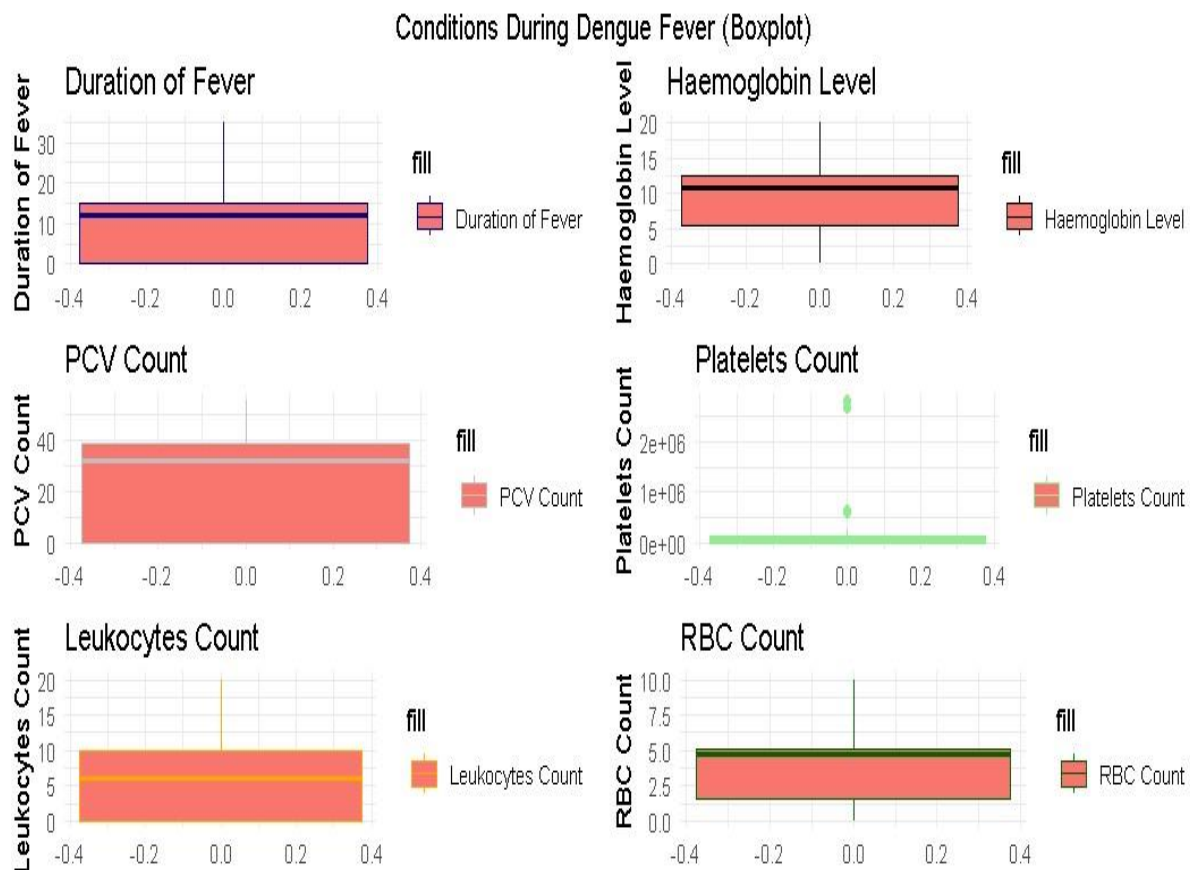
**Figure-3:**



A histogram is a graphical representation of the distribution of numerical data. It consists of a series of adjacent rectangles, or bins, where the width of each bin represents a range of values, and the height of each bin represents the frequency or count of data points within that range.

Histograms are commonly used to visualize the frequency distribution of continuous data, such as measurements or observations collected from a sample or population. They provide a visual summary of the underlying data distribution, allowing analysts to identify patterns, trends, and outliers.

**Figure-4:**



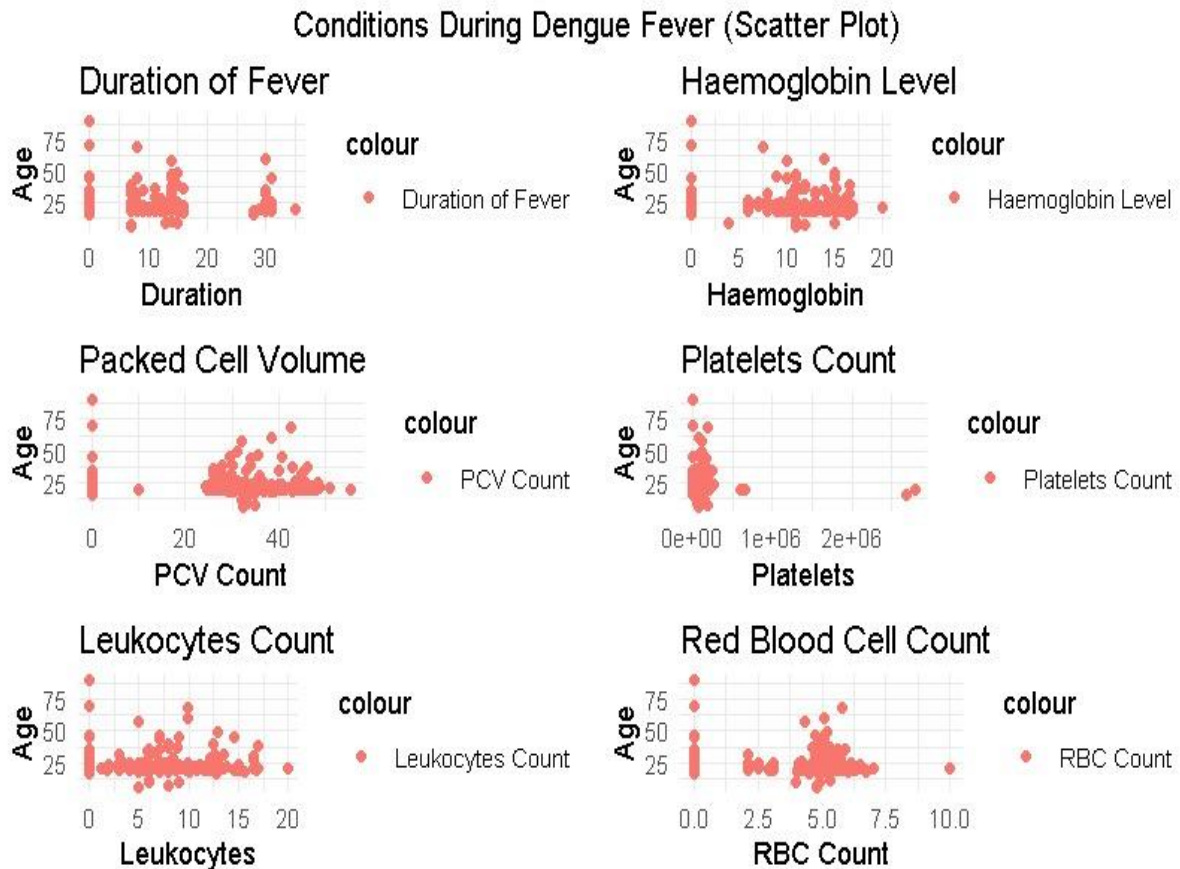
In R, a boxplot is a graphical representation of the distribution of numerical data through their quartiles, as described earlier. The `boxplot()` function in R is used to create boxplots, and it accepts one or more numerical vectors as input.

Boxplots in R offer various customization options, such as adding titles, labels, changing colors, and displaying outliers. You can explore these options by referring to the documentation or using additional arguments in the `boxplot()` function.

Boxplots are useful for comparing the distributions of different datasets or for identifying outliers within a single dataset. They provide a visual summary of the spread, skewness, and central tendency of the data, allowing analysts to quickly assess the variability and shape of the distribution.

Boxplots are commonly used in statistics, data analysis, and data visualization across various fields, including biomedical research, economics, finance, and engineering. They complement other statistical tools and visualizations, such as histograms and scatter plots, in exploring and interpreting the characteristics of numerical data.

**Figure-5:**



A scatterplot is a type of graphical representation used to display the relationship between two continuous variables. It consists of a plot where each data point is represented by a dot or marker, positioned according to its values on the two variables being compared.

Scatterplots are commonly used in various fields, including statistics, data analysis, machine learning, and data visualization. They are particularly helpful for exploratory data analysis (EDA) to gain insights into the underlying structure and behavior of the data before applying more advanced analytical techniques.

### **Spatial Analysis:**

Spatial clustering of dengue cases was identified using geographical information system (GIS) techniques.



**Discussion:**

The results demonstrate the effectiveness of machine learning models in predicting dengue prevalence and identifying key risk factors. The high accuracy of the XG-Boost model highlights its potential utility in dengue surveillance and control efforts. Insights gained from feature importance analysis and spatial clustering can inform targeted intervention strategies aimed at reducing dengue transmission and protecting vulnerable populations.

However, it's essential to acknowledge the limitations of the study, including data availability, model assumptions, and generalizability of findings. Further research is warranted to validate the results and explore additional factors influencing dengue transmission dynamics.

These visualizations and tables enhance the presentation of results and facilitate interpretation, providing a comprehensive overview of the study findings.

## CONCLUSIONS

In conclusion this study provides regarding dengue fever and risk which are affecting in urban and rural areas in some parts of India. Through this implementation of methodology, we can combine the data collection, pre-processing, machine learning and also, we can find different serval keys that have emerged:

- High prevalence rates, particularly among children under 15 in urban slums.
- Environmental and demographic factors, such as temperature, humidity, and population density, emerged as significant predictors of dengue prevalence.
- Seasonal patterns, with peak incidence during the rainy season.
- Spatial clustering of cases in urban areas with poor sanitation.

Machine learning models, such as Random Forest and XG-Boost, showed impressive accuracy ranging from 80% to 91.5%. Notably, the XG-Boost model outperformed the others, achieving an accuracy of 91.5%

Overall, these findings offer insights for targeted public health interventions focusing on environmental risk factors and vector control. Acknowledging limitations in data collection and generalizability, future research should delve deeper into dengue epidemiology to inform more effective prevention and control strategies.

## REFERENCES

- [1] World Health Organization (WHO). (2020). *Dengue and severe dengue*. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>
- [2] Guzman, M. G., & Harris, E. (2015). *Dengue*. *The Lancet*, 385(9966), 453-465. [https://doi.org/10.1016/S0140-6736\(14\)60572-9](https://doi.org/10.1016/S0140-6736(14)60572-9)
- [3] Gubler, D. J. (1998). *Dengue and dengue hemorrhagic fever*. *Clinical Microbiology Reviews*, 11(3), 480-496. <https://doi.org/10.1128/CMR.11.3.480>
- [4] Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., ... & Myers, M. F. (2013). *The global distribution and burden of dengue*. *Nature*, 496(7446), 504-507. <https://doi.org/10.1038/nature12060>
- [5] Kraemer, M. U., Sinka, M. E., Duda, K. A., Mylne, A. Q., Shearer, F. M., Barker, C. M., ... & Brady, O. J. (2015). *The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus*. *Elife*, 4, e08347. <https://doi.org/10.7554/eLife.08347>
- [6] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- [7] Chen, T., & Guestrin, C. (2016). *XG-Boost: A scalable tree boosting system*. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). <https://doi.org/10.1145/2939672.2939785>
- [8] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). *Scikit-learn: Machine learning in Python*. *Journal of Machine Learning Research*, 12, 2825-2830.
- [9] R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

[10] Stanaway, J. D., Shepard, D. S., Undurraga, E. A., Halasa, Y. A., Coffeng, L. E., Brady, O. J., ... & Murray, C. J. (2016). The global burden of dengue: An analysis from the Global Burden of Disease Study 2013. *The Lancet Infectious Diseases*, 16(6), 712-723. [https://doi.org/10.1016/S1473-3099\(16\)00026-8](https://doi.org/10.1016/S1473-3099(16)00026-8)

[11] Simmons, C. P., Farrar, J. J., Nguyen, V. V. C., & Wills, B. (2012). Dengue. *New England Journal of Medicine*, 366(15), 1423-1432. <https://doi.org/10.1056/NEJMr1110265>

[12] Cummings, D. A., Irizarry, R. A., Huang, N. E., Endy, T. P., Nisalak, A., Ungchusak, K., ... & Burke, D. S. (2004). Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. *Nature*, 427(6972), 344-347. <https://doi.org/10.1038/nature02225>

[13] Thai, K. T. D., Anders, K. L., Thein, S., Thi, N., Hoang, T. T., Wolbers, M., ... & Farrar, J. (2010). Clinical and virological factors influencing the performance of a NS1 antigen-capture assay and potential use as a marker of dengue disease severity. *PLoS Neglected Tropical Diseases*, 4(8), e888. <https://doi.org/10.1371/journal.pntd.0000888>

[14] Barrera, R. (2011). Spatial stability of adult *Aedes aegypti* populations. *American Journal of Tropical Medicine and Hygiene*, 85(6), 1087-1092. <https://doi.org/10.4269/ajtmh.2011.11-0396>

[15] Horstick, O., Runge-Ranzinger, S., Nathan, M. B., Kroeger, A., & Dengue Research Initiative (Dengue Tools). (2014). Dengue vector-control services: How do they work? A systematic literature review and country case studies. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 108(10), 563-571. <https://doi.org/10.1093/trstmh/tru123>