SC-209

Vulnerability Index mapping of DA-IICT for Dengue

GROUP - 2

GROUP MEMBERS:

Vatsal Gujarati : 201801162
Swapnil Dhola : 201801163
Prayag Bhatt : 201801164
Divya Vazirani : 201801165
Rushi Patel : 201801166
Dhanvi Shah : 201801167
Param Modi : 201801168
Astha Shah : 201801169
Urvish Pandya : 201801170
Ved Joshi: 201801171
Jainit Patel : 201801172
Arshit Mangukiya : 201801173
Shreya Joshi : 201801174
Yagnik Sakhiya : 201801175
Meet Shah : 201801434
Sarthak Patel : 201801435 (Group Leader)
Vishal Patel : 201801436

TABLE OF CONTENT:

ABSTRACT

This report is about the case study of the Dengue outbreak in DA-IICT campus last year to which over a hundred people fell victim. We have developed tools that will help decide precautions to be taken in advance to avoid future outbreaks. Dengue spreads because of the mosquitoes' breeding at the water logging sites aided with favorable weather conditions like temperature and humidity. To predict and prevent such an outbreak, we have developed different prediction models. The first of which is the differential model. It has some limitations to overcome which, we have designed a predictive model based on Machine Learning using Artificial Neural Networks. As we can make predictions, proper care and preventive measures can be practiced to avoid such spread in the future. We have plotted the topographical map and analyzed Drainage Basins of campus so that we can find out the most prominent breeding and spread sites. We developed a Vulnerability Map from the results of our analysis. Along with this, we have made a Dengue App for smartphones which is a handy tool. It uses the current location of the user and helps the user to take safety measures accordingly. This report gives a thorough idea of last year's outbreak and an in-depth look into the various predictive models which we have designed to help control and curb future outbreaks of Dengue.

WHAT IS DENGUE?

Dengue is a mosquito-borne viral disease caused by the dengue virus. It is widely spread in tropical and subtropical regions. Dengue virus has four serotypes (DEN-1, DEN-2, DEN-3, and DEN-4) which belong to the genus Flavivirus. The risk of severe diseases from secondary infection increases if someone previously exposed to DEN-1 contracts with DEN-2 or DEN-3 or someone previously exposed to DEN-3 acquires DEN-2. In Asian regions, mainly DEN-2 and DEN-3 are associated with the severe disease along with secondary dengue infections.

How Does It Spread?

Dengue is transmitted by the Aedes aegypti mosquito which bites during daylight hours. Infected Aedes mosquitoes are the main vector that transmits the dengue virus. It can't spread directly from one person to another person. Once the human is infected, he or she becomes the main carrier and multiplier of the virus. They act as a source of virus for uninfected mosquitoes. The virus circulates in the blood of an infected person for 2-7 days. During this time if a mosquito bites the infected person, the mosquito becomes infected with the dengue virus. This transmission from an individual to a vector can take place within 4-5 days (maximum 12 days). When this infected mosquito bites uninfected human, the virus enters into the human body through its saliva, it circulates in human blood and reproduces itself. The infected mosquito can transmit the virus for the rest of its life.

Aedes aegypti mainly prefers to lay its eggs in artificial containers to live close to humans as it is highly dependent on human blood for food. Uncovered tanks, drums filled with stagnant water, discarded plastic buckets and containers, old tires, mud pots, etc. are breeding places for dengue mosquitoes.

Dengue virus can be transmitted via infected blood and organ donation and also sometimes includes sexual transmission, but these are very unusual.

Symptoms:

A person having dengue virus has flu-like symptoms in general. Symptoms of dengue fever are:

- Severe headache and fever
- Pain behind eyes
- Nausea / vomiting
- Muscle and joint pains
- Rashes

Severe dengue is a deadly complication due to severe organ impairment, severe bleeding, or plasma leakage (bloodstream leaks into body cavities through the wall of vessels due to capillary force). As a result, less blood circulates into vessels and blood pressure becomes very low. It leads to a reduced number of platelets, which are necessary for blood clotting.

Current Challenges:

There have been major advances in dengue diagnosis. Still there exists some problems for the timely development of new solutions.

- Increased global travel expands the area of dengue virus and provides many opportunities for the successful selection of viral variants of high epidemic potential.

- Poor awareness of dengue symptoms (generally people ignore symptoms at an early stage), misdiagnosis (happens due to having the same symptoms of influenza, malaria, etc.) and poor medical facilities in rural areas.

- PCR (Polymerase Chain Reaction) test can be used to determine which of the 4 serotypes is causing infection. But it requires specific laboratory equipment and facilities as well as extensive evaluation of the different protocols under field conditions.

- Since the dengue virus is caused by four serologically related viruses, the first major problem in developing the dengue vaccine is to develop not just one immunogen but four immunogens that will induce a protective immune response against all four viruses simultaneously.

- The development of the dengue vaccine has also been challenged because of the critical issues like lack of animal models for the disease and the absence of suitable markers of protective immunity.

- Another challenge in developing a vaccine is in determining the reasonable cost as those who need the vaccines must not be hindered from accessing them due to the high price.

- Dengue is highly dependent on the infected people as well as the density of mosquitoes. So it is necessary to identify susceptible populations and areas but because of the poor surveillance system, it becomes hard to identify. Due to this, the reporting system for dengue cases becomes poor.

- Following aspects require attention for better diagnosis:
    I. Tests for early clinical diagnosis of individuals.
    II. Serological tests that can differentiate dengue from other Flavi-virus infections and even more specifically determine the infecting dengue serotype.
    III. Tools that can suggest a prognosis, allowing for better management of clinical follow-up.

Future:

Dengue virus has become a serious issue in India as it has caused many endemics. There is limited success in reducing the transmission of dengue. There is a prediction that this disease is going to increase and people are at risk of getting infected. Therefore, there is an urgent need to develop an alternative solution for this disease to reduce infections. WHO, through its Tropical Diseases Research Programme and the Initiative for Vaccine Research, is working toward the goals.

Scientists are working to develop antiviral drugs (alternative to vaccinations) that could be used to treat the attacks of dengue fever and prevent severe complications. The most effective way to control dengue disease in the future will include the use of a safe and effective vaccine. Several vaccine candidates are under development, including live attenuated virus vaccines, live chimeric virus vaccines, inactivated virus vaccines, etc. The live chimeric virus vaccine is undergoing a phase III clinical trial.

DENGUE OUTBREAK AT DA-IICT

The first case of dengue on DAIICT campus was allegedly detected on the 5th of August, 2019. A lot of cases followed in the upcoming 3 months. The major cause of this outbreak was water accumulation at various places on the campus. Due to rainfall, water got accumulated at various depressions formed in the land, at the outlets from the washroom pipes, at and around the construction sites, etc. Due to water-logging at multiple places, mosquitoes with aid of suitable temperature and weather conditions, started breeding and gradually infecting people.128 cases were reported over the entire season with many instances of multiple reports on a single day. Preventive measures were deployed by the institute to slow down and eventually curb the spread. Hence, to better help prevent the same outbreak from happening again, we have designed predictive models for this year’s outbreak, so that precautionary measures can be taken well in advance in time order to avoid similar situations

Statistics of 2019’s Outbreak At DA-IICT:

- Below are the graphs which depict the number of cases registered during different intervals.

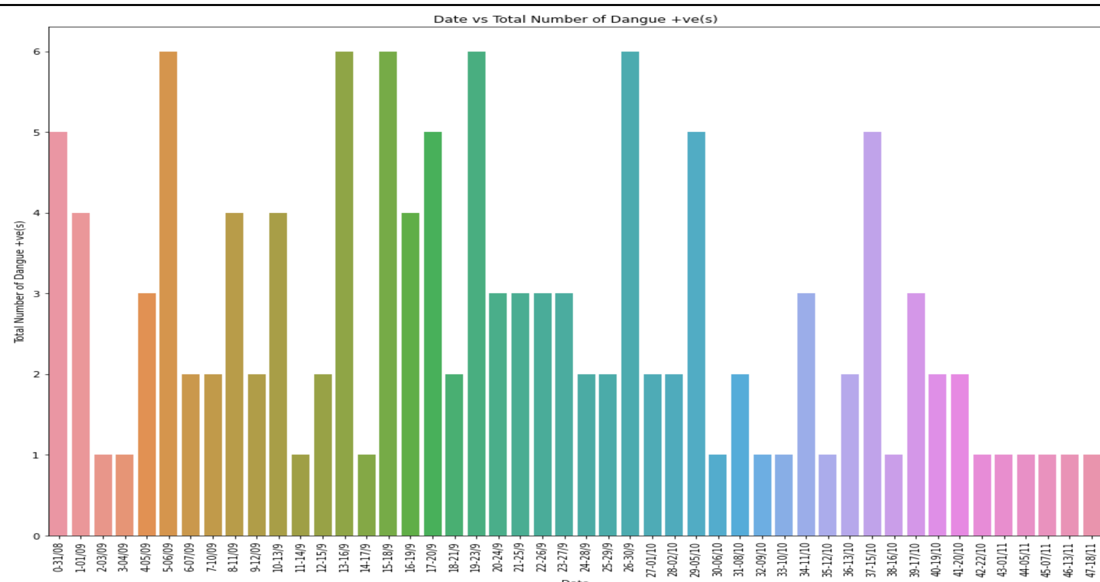


Figure 1 Number of Cases Registered Every Day During the Outbreak Period.

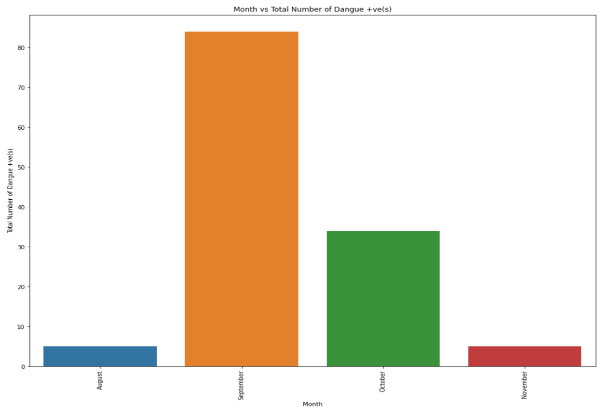


Figure 2 Total Cases Per Month

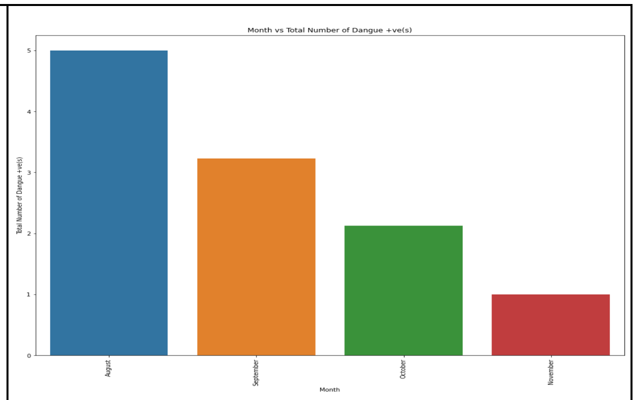


Figure 3 Average Cases Per Day For Each Month

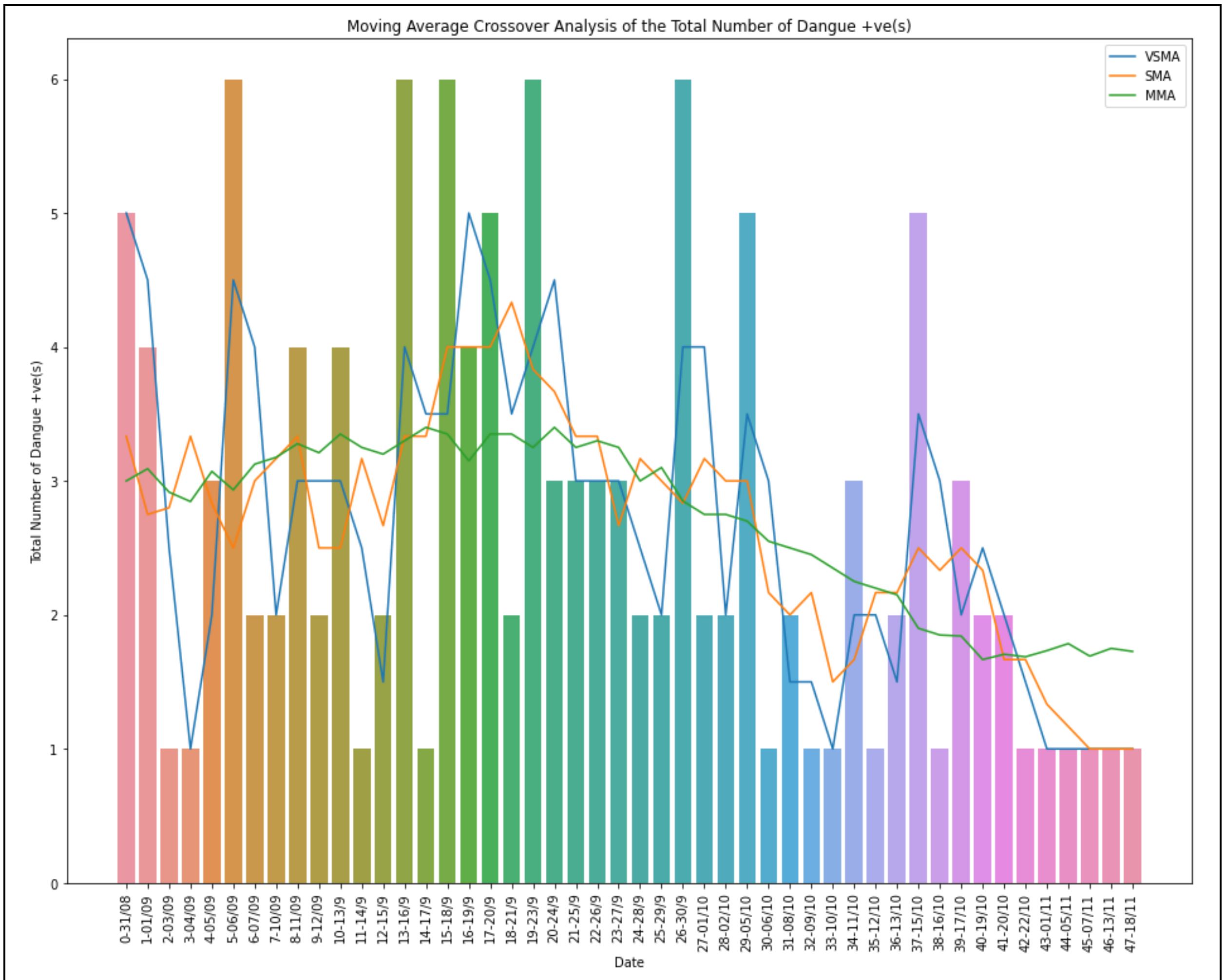


Figure 4 Moving Average Analysis of the season’s reportings.

DANGER INDEX

The favorability for the breeding of dengue mosquitoes is directly a function of the current weather conditions. We can analyze the current state of weather using various parameters to calculate the favorability index for their breeding, which in turn, will be the danger index for hosts i.e. us. The more supportive the weather conditions are for the mosquitoes' breeding, the increased danger there is for us. Hence, we have established a mathematical correlation of mosquitoes' breeding with the weather conditions which will provide us, in advance, with the degree of caution and the preventive measures that we need to practice.

The Parameters:

The breeding of dengue mosquitoes is extremely sensitive to changes in temperature and humidity. This fact was established by an extensive study that was conducted in Thailand using data of 1.2 million dengue reports over a period of around 20 years. The study also showed that the breeding of mosquitoes and hence the frequency of reporting peaked at 29℃ mean temperature and 62% mean humidity. The study also showed that the breeding rate and survival rate of mosquitoes were more dependent on temperature and humidity while almost being independent of rainfall. The empirical disease and weather data used for the study spanned over 72 provinces, 228 months, and was collected using 84 principles and almost 1000 secondary weather stations throughout the entire nation.

Support Vectors For Analysis And Calculations of Danger Index:

Since the survival and breeding rate of mosquitoes is extensively dependent upon temperature and humidity, we used these two weather parameters to mathematically derive a danger index.

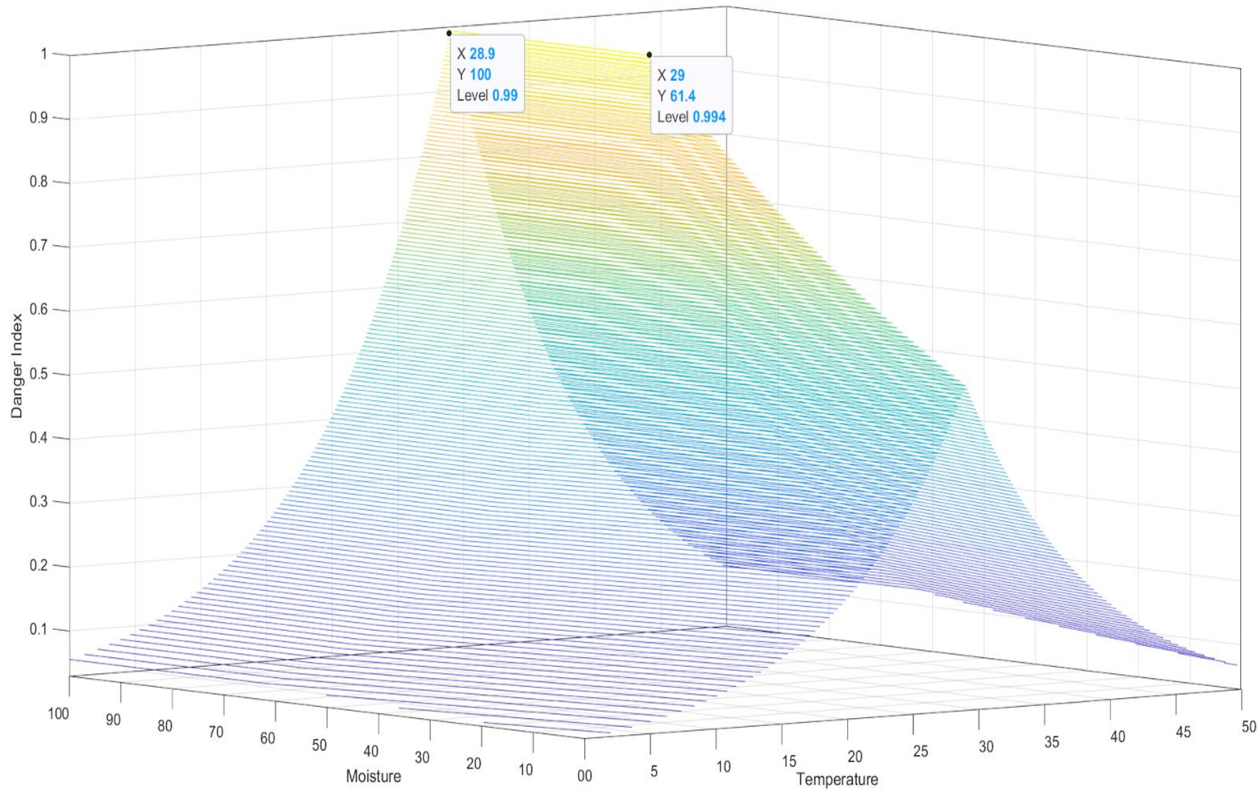
The statistical and laboratory studies obtained the most favorable values of temperature and humidity for their sustainment are:

Temperature = 29℃ Humidity > 62%

Hence, we use these two values of temperature and humidity as our support vectors.

Mathematics In Calculating Danger Index:

For hot or cold temperatures over or below 29℃, the probability of an egg's survival falls exponentially. Similarly, if the humidity drops below 62%, then also the probability falls exponentially. For humidity above 62%, the entire range up to 100% is found to be favorable in their breeding. This behavior can be modeled by the use of RBF (Radial Basis Function). These functions peak at a certain value, in our case the support vectors, and decay exponentially according to our scale on either side.



Our danger index tends to a value of 1 for temperature ~ 29°C and moisture > 62%. This is evident from the 3D plot above with temperature, humidity, and danger index on X, Y, and Z axes respectively.

Now, the danger factor due to temperature and humidity individually are-

Temperature Factor: $T_f = e^{(-|current\_temperature - 29| * 0.1)}$

Humidity Factor: $H_f = 1$ $\quad$ $if\ humidity > 62\%$

$H_f = e^{(-|current\_humidity - 62| * 0.01)}$ $\quad$ $if\ humidity < 62\%$

Now, the final danger factor will simply be a multiplication of both individual factors and will lie between 0 and 1.

$$Danger\ Index = T_f * H_f \qquad (2a.)$$

Also, the eggs can survive up to a period of 6-7 days. So we also need to consider the weather conditions of the previous days in our analysis. To achieve this, we take a moving average at the last data point, i.e. the current day. We take into consideration the temperature and humidity of the past 5 days and the current day's weather.

Let's say today is Day-0, then the danger index for today will be:

$$Danger\ Index\ =\ (D_0 + D_{-1} + D_{-2} + D_{-3} + D_{-4} + D_{-5}) / 6 \qquad (2b.)$$

Hence, if it had been raining for the past four days which would've led to a large number of eggs being laid, and it's been sunny today, then the equation will also take into consideration the eggs that might've been laid in the past few days.

Short Term Model:

This model takes into account the temperature and humidity of the current day only. It takes into input the temperature and humidity, calculates the danger index according to the formula (2a.), and provides us with the danger index correspondingly.

Medium Term Model:

This model takes into account the span of the past few days along with the current day to calculate the danger index. This model calculates the danger index according to the formula (2b.). It will calculate the danger index according to the short term model for each day, and our final danger index will be the average of all of them.

Results:

Our danger index model shows how strong the inflow of dengue mosquitoes and their breeding is expected to be on a given day according to the weather conditions of that day. And as discussed above, the weather conditions of the past few days also contribute to the current day vulnerability, so we also take into account the weather data of those past few days.

We also have designed a mobile application which, according to the weather data of your location, shows you the vulnerability index. The same has also been explained in an upcoming section.

For the year of 2019, we gathered the weather data of DAIICT over the entire year using Google Earth Engine and applied the medium-term model on it. The results are as follows:

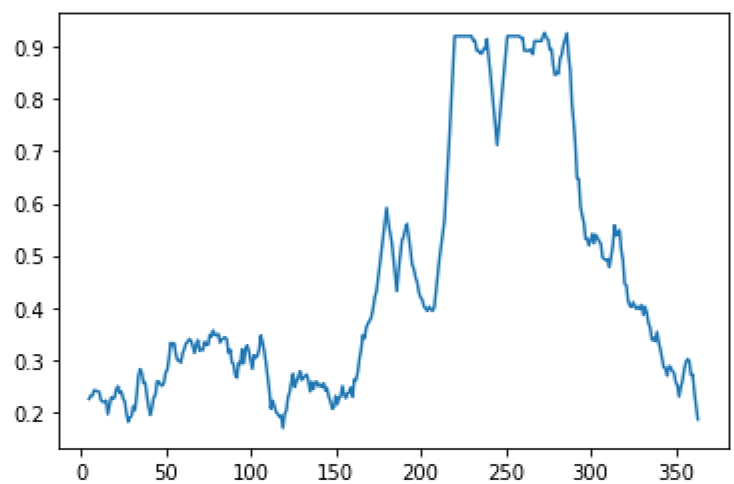


Figure 2.0

The X-axis represents the day of the year and the Y-axis shows the vulnerability index on a scale of 0 to 1.

We can assume the degree of infection for a given day from the graph. If the index is low on a given day, then there is less probability that a person might get infected with dengue. And if there is a high peak then there is a high probability of getting infected.

As we can observe from the graph that during winter and summer seasons, the index lies below the value 0.5. This is obvious because the breeding of mosquitoes wouldn’t be feasible during those seasons. Since the temperatures are either over 35℃ or less than 20℃, which are highly unfavorable for mosquitoes' breeding. While in monsoon it has high peaks, as there are suitable weather conditions for mosquitoes breeding.

Comparison of Our Results With Last Year's Outbreak:

In our graph, the danger index is very high(>0.9) for the months; August, September, and October. In comparison with the last year's outbreak period and the cases reported during that period, we can verify that our alarm system provides accurate details regarding the imminent danger and necessary precautions. And hence, our model provides predictions that are very consistent with the statistical data.

Following is the graph of the actual data of number of cases:

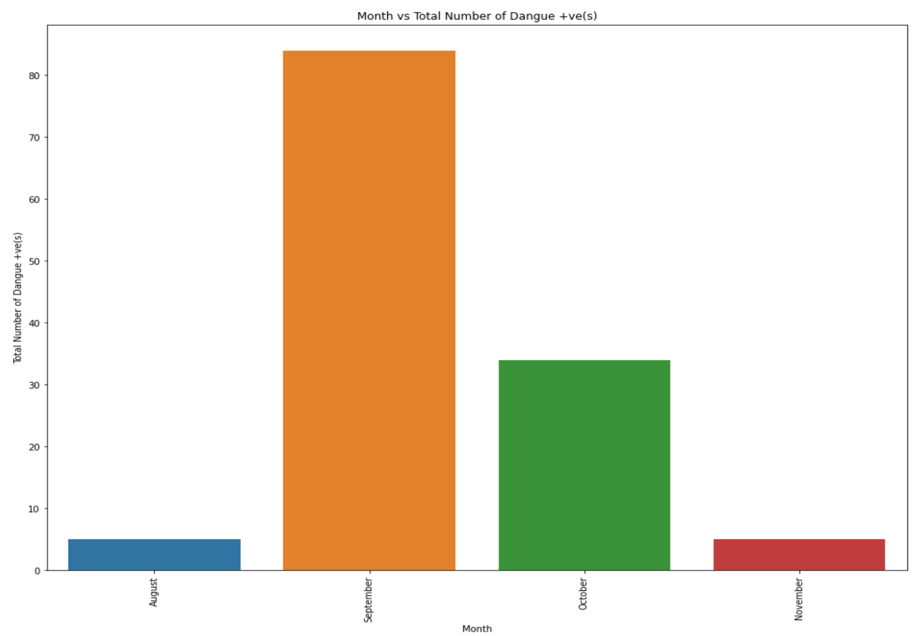


Figure 2.1

As we can observe from the graphs(fig. 2.0 & 2.1) that our alarm model is quite consistent with the actual data. As if the vulnerability index is higher then there is a high probability of a person getting infected with dengue which results in more cases on that day.

DIFFERENTIAL MODEL

Dengue has been a widely spread disease for decades. There have been major outbreaks of dengue over the course of time in different areas. Generally, the disease spreads in the lower land areas. Taking into consideration the large outbreak last year, we have made a differential model to help better control the situation.

In this model, we have established various differential equations on the parameters causing the outbreak of dengue and affecting its spread. We have calculated different rates with the help of the above-mentioned parameters and equations. And with this calculation by having the rate of spread of dengue we can estimate the future conditions roughly and we can take actions according to that and cease the spread of this disease.

How Does It Work?

Since dengue has no vaccines, it has to be taken care of as soon as possible. We have developed a S.I.R.(Susceptible, Infected, Recovered) model since there are two stages in this disease. Here in our model, we help to calculate the rate of patients getting infected and the susceptible students on our campus. We have also taken the infection rate and death rate of the mosquito under consideration. With the help of these rates, we can calculate the currently infected and susceptible hosts and currently infected vectors.

The main transmitter of dengue virus is Aedes aegypti mosquito. This virus is passed to humans through the bites of an infective female Aedes mosquito. The Aedes mosquitoes generally get this virus by biting an infected human. Thus a mosquito also gets infected with the dengue virus, such an irony! Some of the infected mosquitoes even die of this virus.

So here in our model, we have taken into account the death rate of infected mosquitoes. We acquired it from lab data, the death rate of the vector (mosquito) = 1/12.

The total number of students in our college is approximately equal to 1600, which is the total number of susceptible hosts.

Through the dynamics of dengue disease obtained from laboratory studies, we have established four different differential equations connecting these parameters. We have acquired the equations of the rate of susceptible hosts, which decreases day by day as the number of infected hosts increases. We have also acquired the rate of infected hosts, the rate of infected vectors, and the recovery rate of patients.

Mathematics Involved:

- Euler's Method: All the equations that we have used in this model are first-order differentials equations. And hence, we have used Euler's method to solve them.

$$\frac{dy}{dx} = f(x)$$

$$\frac{\Delta y}{\Delta x} \approx f(x)$$

$$\frac{y_2 - y_1}{x_2 - x_1} = f(x)$$

$$y_2 = y_1 + (x_2 - x_1) f(x) \qquad \ldots(1)$$

Differential Equations Used in The Model:

$$\frac{dS_v}{dt} = A - B_v S_v I_h - D_v S_v$$

$$\frac{dS_h}{dt} = -B_h S_h I_v$$

$$\frac{dI_v}{dt} = B_v S_v I_h - D_v I_v$$

$$\frac{dI_h}{dt} = B_h S_h I_v - r I_h$$

$$\frac{dR}{dt} = +r I_h$$

Where the parameters that are used in the equations along with their initial values used for our model based on DA-IICT campus are as follows:

- A = Constant Recruitment Rate of Vector Population = 8
- Sv0 = Initial No. of Susceptible Vector = 10
- Sh0 = Initial No. of Susceptible Host = 1600
- Iv0 = Initial No. of Infected Vector = 1
- Ih0 = Initial No. of Infected Host = 0
- R0 = Initial No. of Recovered Patients = 0

These are the different rates that we have used for our calculation.

- Dv = Death Rate of Vector = 1.0/12
- Dh = Death Rate of Host = 0
- Bv = Transmission Rate from Host to Vector = 0.000045
- Bh = Transmission Rate from Vector to Host = 0.00097
- r = Recovery Rate of Patients = 0.03

```
double A = 8;                 // Constant Recruitment Rate of Vector Population
double totalCases = 0;        // Intially Total Number of Cases to be 0

// Initial Values
double Sv0 = 10;               // Initial No. of Susceptible Vector
double Sh0 = 1600;            // Initial No. of Susceptible Host
double Iv0 = 1;               // Initial No. of Infected Vector
double Ih0 = 0;               // Initial No. of Infected Host
double R0 = 0;                // Initial No. of Recovered Patients

// Rates
double Dv = 1.0/12;           // Death Rate of Vector
double Dh = 0;                // Death Rate of Host
double Bv = 0.000045;         // Transmission Rate Host -> Vector
double Bh = 0.00097;          // Transmission Rate Vector -> Host
double r = 0.03;              // Recovery Rate of Patients
```

Based on these initial values, rates, and differential equations we implemented this model in C++. We use Euler's method as described above to calculate the solution of these differential equations.

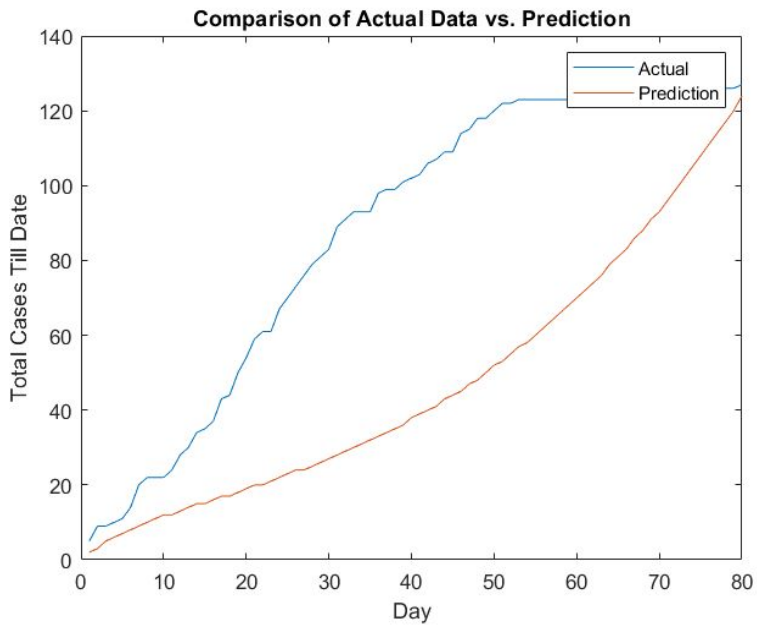
To use Euler's method we have used a simple looping structure and get the output (number of positive cases) based upon the day. (as mentioned in the code $dt = 1$).Here we have implemented this differential model for the specific period of 80 days of monsoon when the Dengue outbreak occurred

```
double next_Sv = prev_Sv + (A - Bv*prev_Sv*prev_Ih - Dv*prev_Sv) * dt;
Sv.push_back(next_Sv);

double next_Sh = prev_Sh - Bh * prev_Sh * prev_Iv * dt;
Sh.push_back(next_Sh);

double next_Iv = prev_Iv + (Bv * prev_Sv * prev_Ih - Dv * prev_Iv) * dt;
Iv.push_back(next_Iv);

double next_Ih = prev_Ih + (Bh * prev_Sh * prev_Iv * dt - r * prev_Ih) * dt;
Ih.push_back(next_Ih);

double next_R = prev_R + r * prev_Ih * dt;
R.push_back(next_R);
```

Output:

In the following graph, we have compared the real number of dengue cases to our predicted ones. We have plotted the graph with "no. of days (length of dengue season)" on the x-axis and "no. of Dengue cases till date" on the y-axis.

With the help of the above-explained parameters, constants and more importantly the differential equations, our model predicts the number of infected students, susceptible students, infected mosquitoes, and recovered students on each day from day one to the end of the season.



(optimized output)

Comparison With Statistical Data and Limitations:

We obtained various differential equations that establish a relation between all the involved parameters. But to solve these differential equations, especially the integration part we require some values of constants. Certain data regarding the vectors requires extensive studies in controlled environments, so in order to solve these equations, we had to take values of some of the parameters directly from detailed research that have been done.

The differential model lacks accuracy in predicting the number of cases of an individual day. It works better at giving us an idea of the total cases over the entire season.

PREDICTION USING MACHINE LEARNING

Infection by dengue is a public health issue in most of the countries worldwide. Most of the developing countries have made efforts to prevent this infection. In recent years, many countries are looking for different techniques to avoid this infection. The prediction models of ML are also techniques to prevent infection. However, some of these prediction models included certain laboratory test results that are not available in remote or poor areas. But there are also some prediction models that predict in a short time.

We have developed the prediction model that predicts the total number of cases of a day using temperature and humidity. We have developed this model using Deep learning and specifically Artificial Neural Networks (ANN). This model is completely based on the previous year's data. Using the date from the dataset of the previous year, we found temperature (in Celsius) and humidity(in %) of that day. So the input of the prediction model is

- Highest temperature during the day
- Lowest temperature during that day
- Average humidity

How Does This Model Work?

The model has been trained from the dataset of the previous year. For this, the data was divided into two parts.
1) 80% data for training the model,
2) 20% data for testing the model.

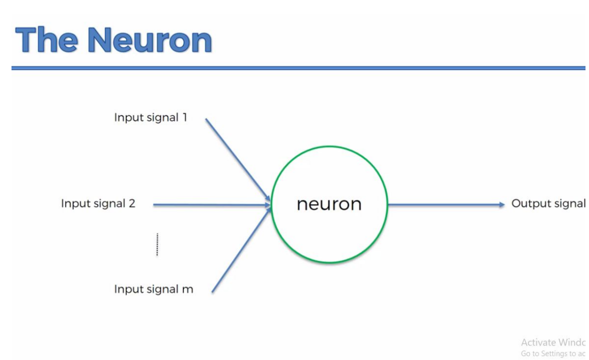
Then the model will compare the answer of the testing part with the training part. This is how the train and test data are used. Using some deep learning techniques like Keras-Dense layer, ‘elu’-activation function, optimizer, the model will give the total number of cases of a day. This is how this prediction model is related to temperature and humidity.

Why These Inputs?

The effect of temperature and humidity on dengue virus’s propagation, as discussed during danger index, is a substantially contributing factor to dengue fever. Results show that the infected mosquitos kept under the conditions of the rainy season and under the simulated conditions had a significantly higher viral titer (concentration of virus) when compared with the other two seasons. So it is thought that the temperature and relative humidity of the rainy season favors dengue virus propagation in the mosquito and is one of the contributing factors to the occurrence of dengue outbreaks.

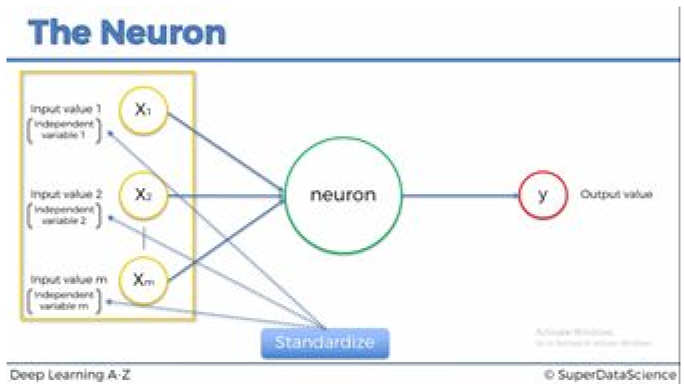
Mathematics Behind Artificial Neural Networks (ANN):

So as introduced above the prediction model uses ARTIFICIAL NEURAL NETWORKS (ANN) to predict the results.
ANN is the algorithm that mimics a human brain. So we can relate its components and process with a human mind.
Now what kind of process is behind the prediction?
The central and primary unit of a NN(Neural Network) is called a node, which is like a neuron in a human mind and NN isn't possible without neurons.



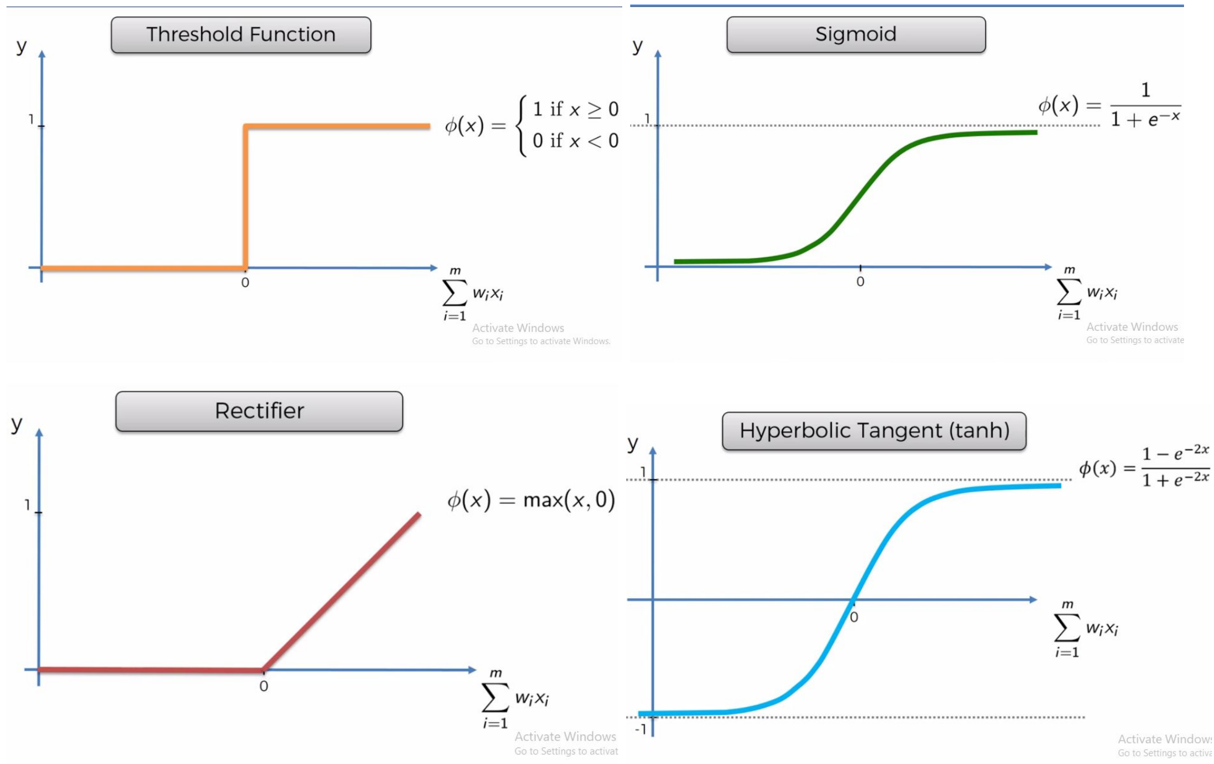
This is the simplest representation of the simplest NN.

So, these nodes combine into a layer and NNs have several such layers. The most outer layer is called INPUT LAYER, the layer which consists of prediction or the output is called the OUTPUT LAYER and layers except the input layer and output layer are called HIDDEN LAYERS.
Now, we have to feed this NN with our data so it can predict the results and NN only takes nodes for the count. We have to feed the data through nodes into NN and the image below describes that.



As per the requirement of the output, there can be multiple outputs instead of single.
Now, the blue lines which are connecting two nodes are called weights and the weights of NN are randomly assigned initially and they take values between 0 and 1.
Now, the next thing is how we calculate the values of each node except the input nodes? That's where the activation functions come into play and there are so many types of activation functions. Now what these functions do is they take a combination of previous nodes' values and weights and as per their characteristics, they measure the output.

Some of the **activation functions** are shown below:

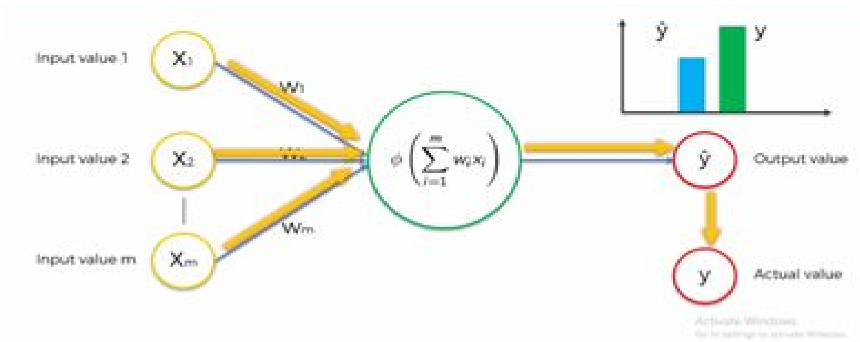


Here as per figure Y is the output or the value of the next layer's node and what sigma(WiXi) means is that each node is dependent on each of the previous layers (More or less).

How Does The Neural Network Work?

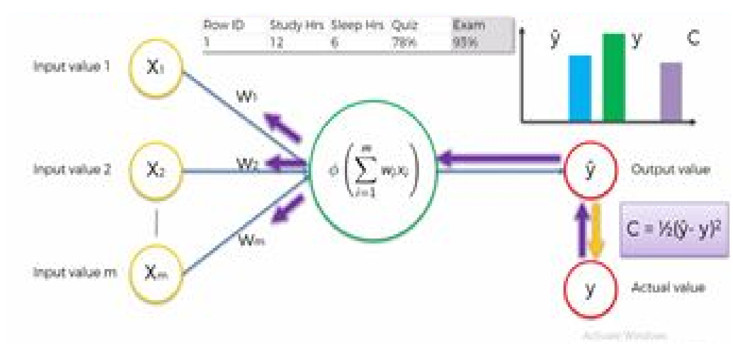
So as these both figures describe the working of NN, in the right figure, we can see that all of the layer's nodes' values are not dependent on every node of previous layers' nodes. Here shown are the most impactful nodes connected to the next layer's nodes and also shown which activation function is used.

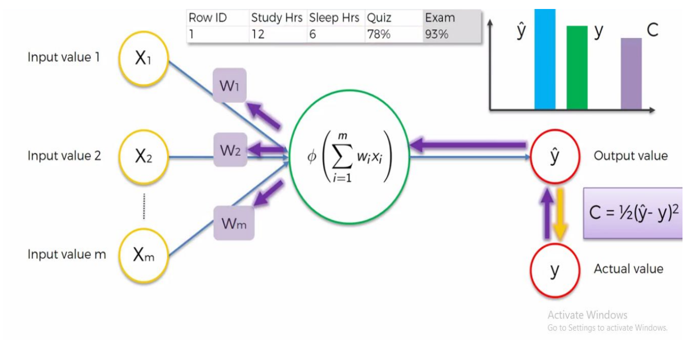
As we can think the NN has to predict the output which it doesn't know, so it can just predict the value which is closer to or maybe accurate as of the actual output. So what it does is that it runs the above process and then predicts a result(for training examples) and it compares it with the output(of training examples) as shown in the figure.

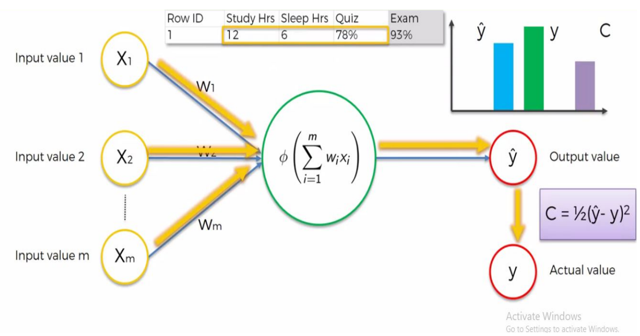


The process of getting output by weights and activation function is called FORWARD PROPAGATION.

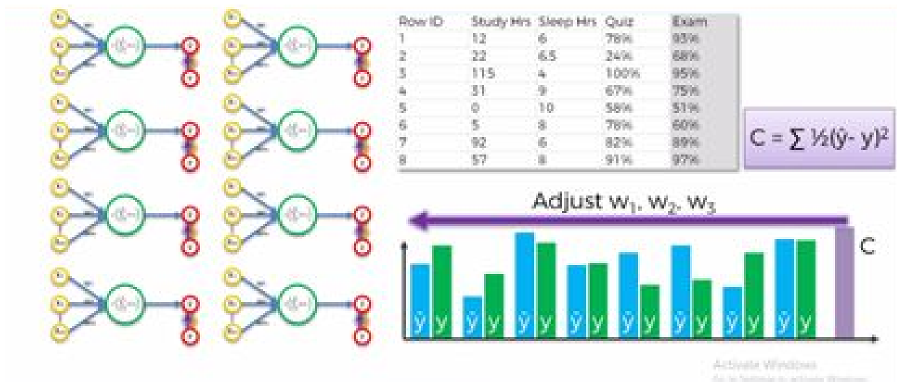
After comparing it uses a function called gradient descent which is basically the squared error of predicted and actual output value.
And the goal of NN is obviously to try to minimize the value of gradient descent or to get predictions closer to actual output.
So, after this process of calculating the value of gradient descent (until it becomes minimum), the NN itself changes its assigned weights to get more accurate predictions.







In short, this has to be done for all input examples, as shown below:



This process of adjusting weights and repeating it until getting the minimum value of gradient descent is called BACK PROPAGATION.

The combined process of FORWARD and BACKWARD PROPAGATION is called an EPOCH and depending on the dataset choosing an ideal number of epochs leads to more accurate results/prediction.

Here, mentioned Gradient Descent (GD) is of two types:
- Batch GD: It takes Batch (i.e. batch of 10 input examples) over all of the dataset for epochs.
- Stochastic GD: It takes individual input training examples for an epoch.

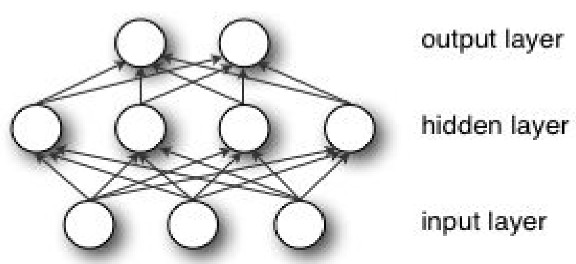
Application Of Machine Learning Algorithms:

We first imported our CSV file in the form of a data frame in python using the pandas library. Then we divide these data sets into two parts: one as input which contains parameters like low temperature, high temperature, average temperature, and humidity. And another is output which contains a number of cases per day.

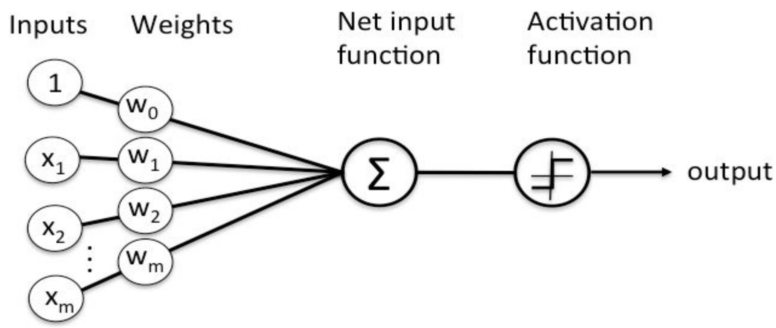
We have to divide this data into two parts: training and testing data. For machines to learn the patterns we have to provide them with the data that has the right answers. This training data is used for the training of the model. The testing data is to evaluate the performance of the model. We used train_test_split to divide 80% of our data as training and 20% data as testing.

Now if all the features are scaled down to the same range like 0 to 1, then it would prevent one feature from dominating over other features. So from sklearn, we imported the StandardScaler function. In that, we used the Fit_Transfrom method to transform the data. Now our data has been scaled.

Then we did the deep learning part. Deep learning finds the correlations. It tries to map the input to output. This deep learning model consists of layers that are made of nodes. It has an input layer, hidden layers, and an output layer.



A node assigns significance to input by combining input with a set of coefficients, or weights and the algorithm then tries to analyze which input is helpful to classify data without error. These input-weight products are summed and then passed to the activation function to determine whether and to what extent that signal should progress further through the network to affect the ultimate outcome.



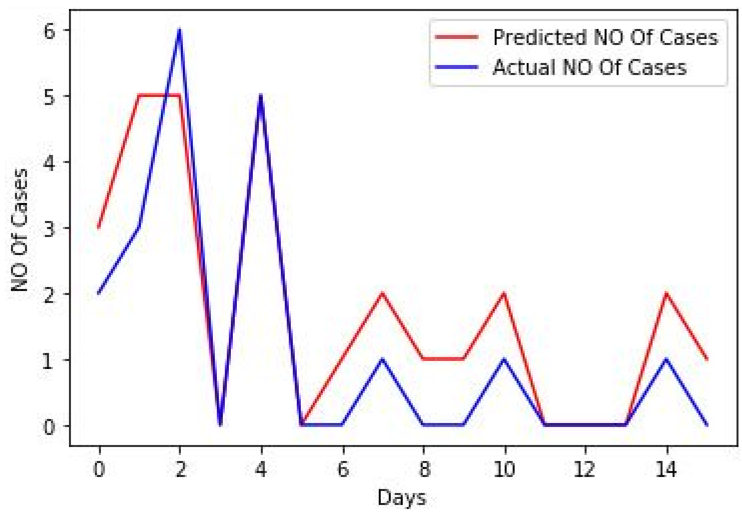
To build our model further we used Keras sequential model. In the input layer, we used 6 nodes as our input parameters. Then 3 hidden layers and one output layer, which consists of one node, i.e. number of dengue cases.

Now we have divided our training data sets into batches. And to get accurate results we tried to pass these data sets multiple times. We came to the conclusion that 1000 epochs are good for our model because if we increase the number then it will overfit our data so we might get ambiguous results for new data.

Now the error is calculated as the difference between our prediction and actual output and the function which does this is known as the loss function. In our model, we used mean_squared_error function which gives the average of the square of prediction minus actual outcome.
We have used Adam optimizer to update network weights.

Output And Comparison:

After that, we have just passed test data to the model to get the results and then plotted the graph of days vs the number of cases for both: actual outcome and predicted outcome.



From the graph, we can see that the red line shows the predicted number of cases and the blue line shows the actual number of cases. And the predicted output is very close to the actual outcome. Maximum delta is only 1 case per day.

To check our model's accuracy, we used accuracy metrics which counts the frequency at which y_pred matches y_true and returns the accuracy. It turns out that our model's accuracy is 70-75%. So we can say that the model can predict the 2019's outbreak quite accurately.

Now with this model, we can also predict future outbreak for DAIICT, for which we will only have to provide temperature and humidity data to the model.

Future Improvements And Challenges:

As the model mainly focuses on the temperature and humidity and the model is trained on the last year's data, but During the last year, many construction activities were going on in the campus due to which there were many small pits where the water could have been accumulated and would result in breeding of the mosquitos. While this year this won't be a problem so the number of cases might be reduced than last year if we see on a microscopic level. This would also be a challenging factor for the model.

One of the other challenging factors for the model is human behavior. Like there are some people who do not clean their rooms, and the washrooms are also not clean, and that would lead to breeding spots for dengue, and these cases cannot be predicted by the model.

The model can have a better prediction, and the accuracy can also be increased if the data provided to the model is large enough so by which we can train the model more efficiently and have some better results.

TOPOGRAPHICAL MAP OF DA-IICT

Here on the map, different colors are used to represent the altitude of land.

.

- Meaning of colors
  Green: ~76 meter
  Orange: ~72 meter
  Red: ~71 meter

According to research Mosquito breeding requires stagnant water. Any place with still water or an ecosystem can be enough of a breeding ground, and mosquitoes can especially thrive after flooding because of all the still water. During monsoon season places which have lower altitude are more likely to have water loggings and these places are more suitable for mosquito breeding. To get an idea which places in DA-IICT are more suitable for mosquito breedings, we have plotted a topographical map of DA-IICT.

How Did We Do The Topographical Mapping?

First, we took a 2D array of 16,000+ coordinates confined to our campus and exported it using Google Earth in KML format.

Then we put that data into a GPS visualizer, which took that KML file and added elevation data corresponding to each coordinate.

Then we plotted that elevation data using ArcGIS's ArcMap. We then exported the colored topographical map from there.

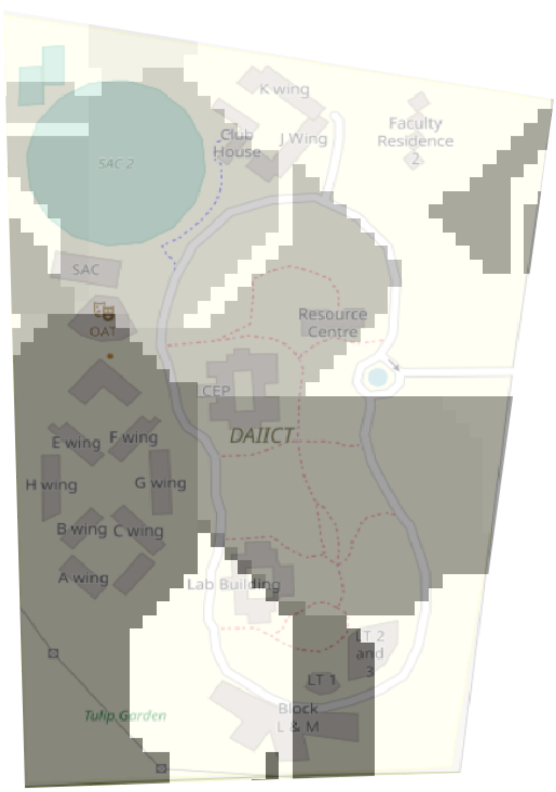
We then took that colored topographical map, took a JPEG image of our campus from Google Earth, and overlaid the map on top of it using Photoshop. And Photoshop because we had to scale and align the topographical map with the satellite image.

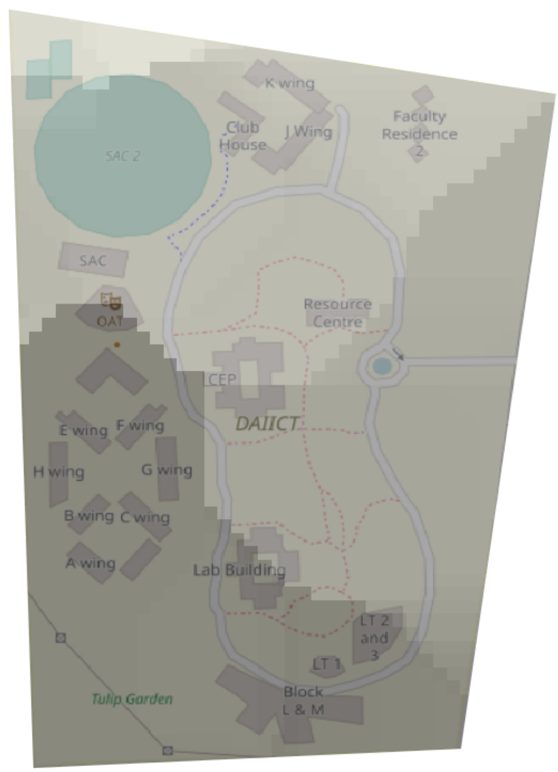
Observations:

In DA-IICT the normal ground level is around 76 meters. So, the places which have a lower height than 76 meters, those places have high chances of waterlogging. From the map, we can clearly see that the boy’s hostel, girls hostel, lecture theaters, and lab building are nearer to the red and orange zone. As red and orange zones have a lower height than ground level there are higher chances of waterlogging. And these places under red and orange zones are more likely to have stagnant water during the monsoon season which is most favorable for mosquito breeding.

It is important to take preventive measures at places that have a higher possibility of waterlogging. As we can observe a boys hostel and girls hostel are near to the red and orange zones, it may be one of the main reasons for last year's dengue outbreak, and in addition, it is a good move to take preventive measures near residence areas to prevent dengue outbreak like last year.

Drainage Basins:





Drainage basin 4	Drainage basin 0.03

We have made a vulnerability index map of DA-IICT, in which we have used the Malaysian model of vulnerability mapping method for water-associated disease Dengue for reference. In their model, they have categorized components into two parts. The first one is exposure, and the other one is susceptibility. Now in the exposure component, they have used the climate and human environment. Furthermore, in the susceptibility component, they have used individual communities. Now we have tried to implement the same model. In climate, component covered factors are minimum and maximum temperature and precipitation. They have set threshold values and relationships between values of factor and its contribution in structuring the index by the research of many years on the mapping of water-associated disease index and dengue virus.

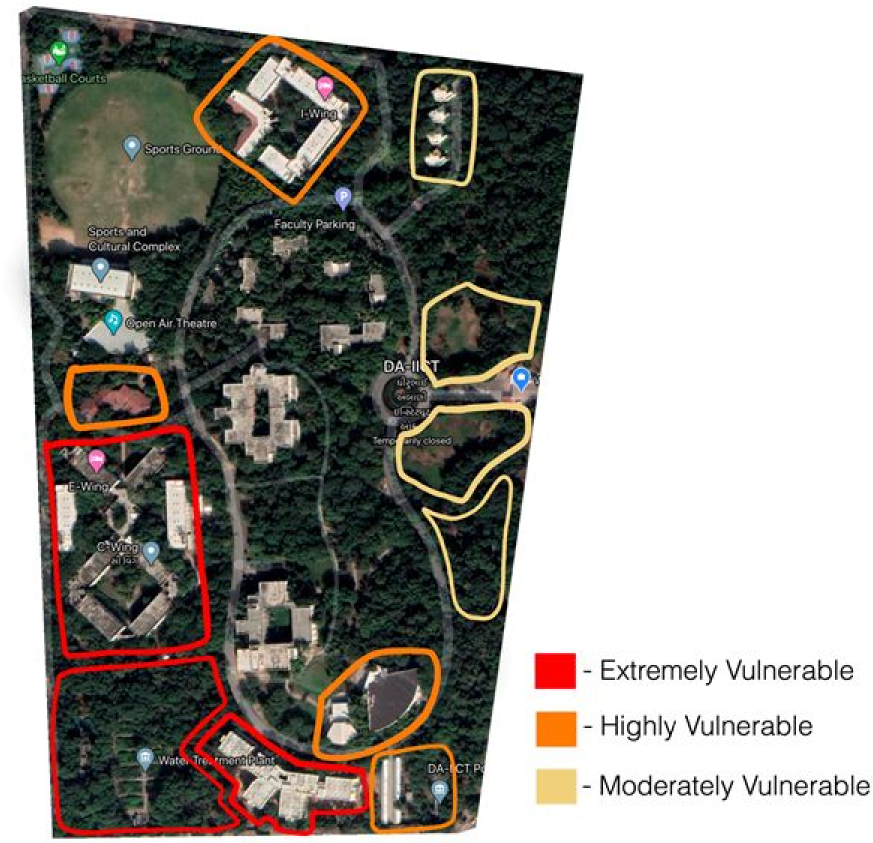
For precipitation, the same way the index is linearly increasing from 0 to 300 mm and zero otherwise. The Malaysian government has considered population density and type of land in the construction of the index. Now for the susceptibility part, there are various factors such as availability of health care, female education, age criteria, and housing arrangements for people. For the index of any individual factor in susceptibility, they have come up with an equation for every factor. Let us suppose the value of any factor is "x" at any place, and "x_{min}" is a minimum value among all places, and the same is "x_{max}" is the maximum value among all places. Then index for that factor is given as

$$Susceptibility(x) = (x - x_{min}) / (x_{max} - x_{min})$$

After getting all data, in construction for the index, they have considered the weighted average of the overall exposure index and susceptibility index. The ratio is 3 for exposure and 1 for susceptibility.

Now, if we want to implement this model on the DA campus, some changes have to be made. In this model, the susceptibility component has various factors like population under 15 years, Water and sanitation, Health Care Access, and Female education level. So, for DA campus, all individuals are above 15 years, so that percentage will be zero and not counted in the index. Now, water, sanitation, and health care access are provided, so the population is not getting that will be zero, and the same way all females are educated. So, any factor of susceptibility component will not contribute to the construction of the final index. Now that leaves us with an exposure component, in the exposure component, population density, and climate. Given the threshold by the Malaysian government's experiment, if we calculate DA-IICT's population density, it turns out to be more than one, so it gives 1 for index construction. Now in the climate part, there are two underlying factors: temperature and precipitation. Now for small places like DA-IICT, we can assume that temperature is almost constant. So it will reflect the same effect in every place DA. It leaves us with precipitation to be an essential factor for mapping at different places in DA-IICT. There are breeding of dengue mosquitoes in clean water. Now rainwater gets accumulated at someplace, which happens to be essential for mapping. Now we have taken DEM of DA-IICT with 10m resolution horizontally. With the help of this DEM and water catchment algorithm, we mapped that where all water gets accumulated. Furthermore, it also supports our elevation map of DA-IICT.

Vulnerability Index Mapping Of DA-IICT:



Dengue Index App:

We have developed a mobile application that can be installed on any modern Android smartphone. This app provides the user with the Danger Index according to his current location and the current weather conditions.

Inherently, the danger index also quantifies the type and degree of caution and precautions an individual needs to exercise.

The app utilizes the device's GPS and cellular capabilities to identify the user's current location. It then utilizes the user's location to obtain the weather data for that location.

The location is passed to a Weather API. The app requests and collects the weather data of the current day and the past five days corresponding to the user's location.

The app then applies the short-term model(2a.) on the temperature and humidity values of each day. It then applies the medium-term model(2b.) over the instantaneous values of the danger index of each day to give a cumulative danger index for the current day.

The whole process including opening the app, grabbing onto the user's location, obtaining the weather data, calculating the index, and displaying it to the user takes less than around six seconds.

We analyzed last year's data through our model against the statistical data of reported cases. We then used the analysis to classify the index into 4 classes. These 4 classes express varying degrees of danger and the corresponding necessary precautions.

Danger Index:	< 0.35	Green Zone
	< 0.55	Yellow Zone
	< 0.72	Orange Zone
	< 1.00	Red Zone

The app also provides the user with advice and precautions to be safe from the threat of Dengue Mosquitoes. The precautions and preventive measures are shown corresponding to the class that the Danger Index lies in.

Use Cases For The Application:

1. **By Authorities:**

   The most useful use case would be by local authorities. Local authorities can use the Danger Index of their locality to issue ward wide flags and notices and keep their population aware of the imminent danger. They can use it to issue precautions for the masses which might include advised clothing type, use of indoor mosquito repellent, keeping windows and doors closed, going outdoors unless necessary, use of mosquito repellent cream and ointment when going out of the house, frequent checking and cleaning of containers that might have stagnant water, fumigating and keeping our surroundings clean.

2. **By Institutions:**

   Institutions can use the index to deploy preventive measures like regular fumigation, sprinkling gammaxene in the damp areas, where water usually gets accumulated, and in corners and surroundings of buildings, leveling the potholes, and restricting access to certain highly prone areas.

3. **By Individuals:**

   Individuals can use the app to assess the type of clothing they may wear, the necessity of repellent creams, and danger in going outdoors and to high-risk places.

Precautions:

“Prevention is better than cure”.

The particular vaccine for dengue has not been invented yet. So it is better to take precautions to stay away from this disease.

Precautions to be taken:

- Protect yourself against mosquito bites.
- Avoid visiting areas prone to mosquitoes.
- Strictly wear full-body clothing.
- Keep windows and doors closed at all times.
- Use mosquito repellent in rooms.
- Apply mosquito repellent cream when going outdoors.
- Frequently check and empty objects and places that might have stagnant water.
- Keep your surroundings clean.

Steps the Institute should take to protect the students from disease:

❖ The places where water accumulation is taking place are the spots for mosquitoes breeding. Those places should be emptied regularly.

❖ Cleanliness, a very important thing. Each and every spot of campus must be cleaned frequently.

❖ Fumigation should be done on a regular basis in residential buildings.

The Requirement Of Technology To Battle This And How Our Efforts Will Help In Controlling This Disease:

Dengue is a mosquito-borne viral infection that affects millions of people worldwide every year. Due to the rapid rate of dengue spread, particularly in tropical and subtropical areas, professionals are implementing new strategies to diagnose better, treat, and contain this disease. There was a high mortality rate due to dengue in the past. The main reason was the lack of medical treatment and technological data for preventions. Nowadays, Information technology is growing at a rapid rate. Despite such a significant medical service, dengue has a high infection rate. Using ML algorithms, professionals are predicting the outcome of dengue in regions. Models are being developed based on meteorological data. Humidity and temperature are chief factors affecting the breeding state of mosquitoes; using them, many predictive models are being prepared using Machine learning. Apart from that, stagnant water is also a prime factor. Using humidity and temperature data from the previous year's outbreak of dengue, we have developed a predictive model, which predicts the number of people infected on that day. We also developed an application that predicts the vulnerability index of your surrounding environment. According to it, you can act on precautions. Now using the predictive model, we can predict the day, so steps to prevent spreading can be taken. Other such predictive models are being developed at the macro scale. The results can also be different as something unexpected phenomena have taken place. In our case, construction was carried out, which can be a factor for such an outbreak. We can use these technologies for a better and healthy future for humankind.

ACKNOWLEDGEMENT

We are very thankful to Teaching Assistants who supported us for this project and gave their guidance to complete our thesis work effectively on time. We would also like to express our special thanks to our senior Srikumar bhaiya for their able advice and support in completing our project. We are equally grateful to our course instructor Prof. Ranendu Ghosh, who guided us in different matters and gave solutions regarding the topic. He has been a source of inspiration for us, and his experience and knowledge have helped us in learning and giving this project the shape it has assumed. We would also like to thank him for giving this project in a group which raised our skills to coordinate with other members of the group which will be an important quality that will help us further in life. This report has played an important role in giving us a wide-angle view of a topic which before was thought as very narrow and something less important.

REFERENCES:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3648565/#!po=0.595238
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6678723/#!po=0.943396

https://www.ijidonline.com/article/S1201-9712(03)00039-0/fulltext#secd33634667e667

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3991153/#:~:text=Vector%20control%20has%20achieved%20only,a%20safe%20and%20effective%20vaccine.

https://en.wikipedia.org/wiki/Dengue_fever

https://www.researchgate.net/profile/Bhavna_Gupta6/publication/235776956_Fight_against_dengue_in_India_Progresses_and_challenges/links/00b7d5189bcbdbf2d9000000/Fsssight-against-dengue-in-India-Progresses-and-challenges.pdf?origin=publication_detail

https://www.hindawi.com/journals/cmmm/2018/9861572/

https://www.who.int/denguecontrol/human/en/

https://pathmind.com/wiki/neural-network