International Journal of Scientific Research and Engineering Development-– Volume 2 Issue 4, July – Aug 2019

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Factors to Predict Dengue Fever using Data Mining Techniques: A Review

B. Sundari [1],M. Krishnamoorthy[2]

[1] (Research Scholar, Department of CSA, SCSVMV University, Enathur, Kanchipuram.)

[2](Assistant Professor, Department of CSA, SCSVMV University, Enathur, Kanchipuram.)

----------**********----------

Abstract:

Dengue fever is a virus infection which is caused by female Aedes mosquitoes. It is mostly found in urban and suburban areas, as they like to breed in man- made containers like tires, flowerpots and household water. The dengue viruses occur in 4 serotypes (DENV1 to DENV 4). Its ranges from mild febrile disease to severe hemorrhagic fever. This research will analyze various dengue factors and reviewed about 40 research papers in the aim of identifying the data mining techniques and models used to predict the dengue.

Keywords —Data mining, prediction, Dengue

----------**********----------

I. INTRODUCTION

Dengue fever is a painful, debilitating mosquito-borne disease caused by any one the four dengue virus. It is transmitted by the bite of an Aedes mosquito infected with dengue virus. It can be spread directly from one person to another person. As estimated 390 million dengue infection occur worldwide each year. Most cases occur in tropical and sub-tropical areas of the world.Dengue fever is classified into two categories namely type 1 and 2. The first type is a classical dengue which is referred as dengue fever and the other type is known as dengue hemorrhagic fever. The disease might develop further into dengue hemorrhagic fever (DHF) which results in decrease in the number of blood platelets.

Symptoms of Dengue: Symptoms which usually begin four to six days after infection and last for up to 10 days, may include sudden high fever, severe headaches, Pain behind the eyes, severe joint pain, Muscle pain, Fatigue, Nausea, Vomiting, Skin rash which appears two to five days after the onset of fever.

II. REVIEW OF LITERATURE

1. [P.Manivannan, Dr.P.Isakki] Provided k-means clustering algorithm for prediction of dengue fever. K-means clustering: K-means is a clustering algorithm used to classify or group the objects based on attributes and that are divided into k-number of group. In this paper, k-means clustering algorithm used to partition the data sets into k-clusters. This work is done on predicting the dengue based on categorization of age group using k-means clustering. This technique is one of the best technique for predicting the dengue patients with serotypes.

2. [Panuwat mekha, khukrit osathanunul et al]
Genome sequences: In molecular biology and genetics, the term pertains to the complete set of genetic material in an organism. The genome of eukaryotes is contained in a single, haploid set of

chromosomes. Dengue genes are extracted from the virus pathogen resource. They are divided into four distinct types. In this work we compared the different methods for classification and finally concluded that using codon usage is the best one for dividing the viruses. It can be support of gene expression and molecular classification. It achieves 96.22% of prediction accuracy by using a neural network.

3. [Abdul mahatir najar, Mohammad isa irawan et al]

Dengue hemorrhagic fever: Dengue hemorrhagic fever (DHF) is a disease that is transmitted through mosquito's bites, especially Aedes aegypti or Aedes albopictus. It is commonly founded in tropical and sub-tropical areas. In this work focused on prediction related to the risk level of DHF outbreaks. Because each region needs to get treatment according to its risk levels. The result shows that extreme learning machine method can predict the risk level of dengue hemorrhagic fever by using 50 hidden neurons.

4. [priyo s sasongko, helmie a wibawa et al]

In this paper focused on to find the best back propagation algorithm to find the detection of dengue and to determine the possibility of DHF. Problems that can be arises from early detection of dengue disease is the data used is YES or No. To solve this problem by using optimization on multi-level perceptron. The result of this work is Levenberg Marquart algorithm shows the best performance in detection of dengue disease early.

5. [Buchado omkar, Dalsania preet et al]

In this paper they are stratifying dengue into dengue fever (DF), Dengue hemorrhagic fever and healthy patients. Here we are using spider monkey optimization and also used probabilistic neural network for classification. They were used greedy feed forward selection algorithm for selection of the specific gene. The primary task is to find whether the person is suffering from dengue or not. This proposed model gives 90.91% accuracy.

6. [Norhayati binti Mohd Zainee et al] In this research the proposed work is based on vital signs containing of blood pressure (BP), heart rate and body temperature are the factors of physical test for dengue fever. The reference being used in the implementation was taken from the Hospital Canselor Tuanku Muhriz (HCTM) in Malaysia. It shows that the trained linear discriminant model using median value gives the highest TPR with the lowest FNR but provide a low accuracy during the verification. It provides low accuracy during the verification.

7. [Nirbhay Mathur et al] This study was used to visualize the dengue incidences on weekly basis in Selangor, Malaysia. This research work is done on by using the geographical information system based on k- means clustering and expectation maximization algorithm. This study finally found that some areas of Selangor are on the high risk of dengue incidences.

8. [Marimuthu, T., et al] exposed new bio-computational model for mining the dengue gene sequences. They proposed a bio-computational model called sequence miner to interpret the relationship among the dengue viruses. The accuracy of the proposal model is 96.74%. The relationship between dengue serotypes are predicted via the proposed tool. It helps to the biotechnologies and drug designers for discovering an effective vaccine for dengue.

9. [M Mufli Muzakki et al] In this paper they proposed prediction of DHF in Bandung Regency using K-Means Clustering as preprocessing method and Support Vector Machine (SVM) algorithm. The data being used in the research from Meteorological, Climatological, and Geophysical Agency in Bandung Regency. Weather data can be used for predicting the DHF disease because there is a relation between weather attribute and DHF disease. K Means Clustering label method shows accuracy more than 86%. The implementation of this research work can be useful for Health Department of Bandung Regency and to increase awareness about DHF disease.

10. **[Iwan Inrawan Wiratmadja, Siti Yaumi Salamah et al]** In this research work focused on predicting the hospital length of stay from the time of admission, which is used for hospital management. With the use of decision tree an accuracy of 71.57% is achieved. The prototype pf

dengue patient length of stay prediction system was developed using the resulting decision tree classification rules.

III.Influencing Factors to Predict Dengue Fever

Temperature

Temperature is an important climatic factor of affecting biological processes pf mosquito including their interactions with viruses. Increasing temperature increases the valuable habitat for the dengue fever vector, the Aedes aegypti mosquito while concurrently increasing both the longevity of the virus and the mosquito. If temperature increases by approximately 30-degree Celsius mean incidence rates during epidemics can double. The mosquito feeding rate also increases; dengue fever viruses in adult Aedes aegypti mosquitoes requires shorter incubation period to salivary glands. temperature increase the Aedes Ageypti displays shorter period of development in all stages of their life cycle, which leads to increased population growth. Temperature is also positively associated with pre adult mosquito maturation, oviporation rate, and virus incubation rate in mosquitoes.

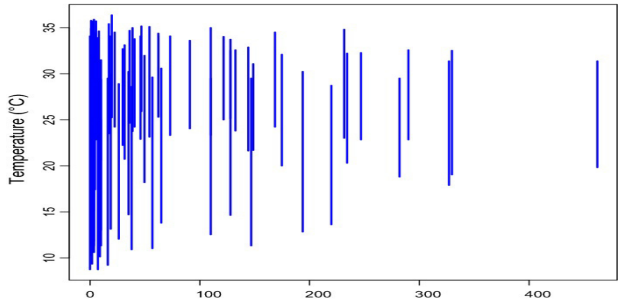

Figure 1: Dengue Fever Cases and Temperature

Rainfall

Some studies reported that rainfall can lead to increases in dengue fever transmission. They suggested that rainfall creates abundant outdoor breeding sources for Aedes aegypti, and the water storage container also can serve as breeding habitats. Rainfall leads to an increase in breeding sites of the mosquito's vectors, which would contribute to the increase in dengue fever occurrence. When more consecutive wet days occurred in a period dengue fever incidence increased.

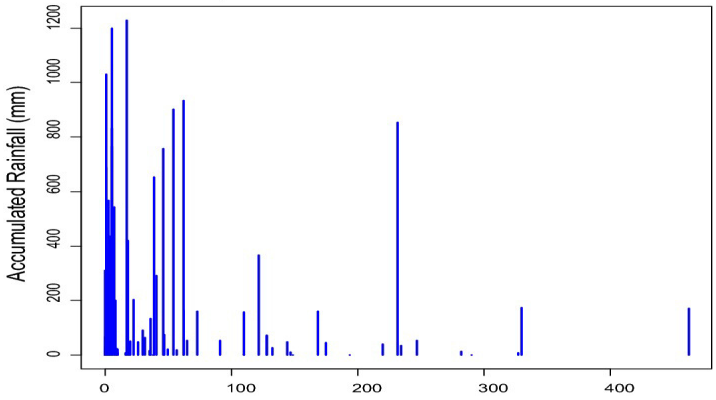

Figure 2: Dengue Fever Cases and Accumulated Rainfall

Sunshine

Sunshine is also closely related to other ecological factors such as temperature and humidity and thereby might affect the dengue fever incidence. Correlation studies carried out on monthly dengue fever cases have found the risk of dengue to be inversely associated within duration of sunshine.

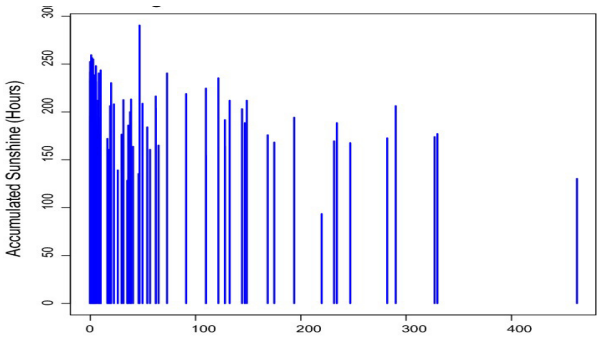

Figure 3: Dengue Fever Cases and Accumulated Sunshine

El Nino Oscillation

El Niño begins with a rise in surface seawater temperature over the tropical eastern Pacific Ocean that later extends as afar as the western pacific and has a warming effect on surface air. The el Niño southern oscillation(enso) an ocean atmosphere phenomenon of the Pacific Ocean with a semi periodic multi-annual cycle has been hypothesized to be a driving force behind the dengue epidemics in regions at risk through its profound influence on the local climate. Warmer temperature reduces the larval size of Aedes aegypti, the vector of dengue, ultimately affecting the adult size. Coupled with increased rainfall, which might happen during the el Niño oscillation, leading to increased breeding sites for mosquitoes and to enhanced survival of adults, it is not surprising that there is an increase in dengue fever during these periods. el Niño is one of the important driving forces of dengue epidemics across the geographically diverse regions.

Mosquito density

Aedes albopictus is native to tropical and subtropical regions, they are successfully adapting themselves to cooler regions. They can also tolerate snow and temperature regions they hibernate over winter. Aedes aegypti is the primary vector of the virus that causes dengue, a disease that remains a serious public health problem in tropical and subtropical countries. The control of this disease is directed towards the reduction of mosquito density in the urban environment. The endophagic behavior of ae. aegypti adults, which can be found throughout the residential environment, is directly implicated in the successful transmission of the virus.

Human mobility

The introduction of dengue into new populations is mediated by travel of infected individuals to areas that can support transmission. The effects of human mobility on the spatial temporal dynamics of degree is explored. Human mobility strongly affects the spread of infection. Because of human mobility, multiple foci appear throughout the evolution of the outbreaks. The coalescence of these foci with the main one generates a fast spreading of the outbreaks. This effect increases the propagation of the outbreak before arrival of the winter season.

High humidity

High humidity results from high rainfall combined with the high temperatures. High humidity is associated with increased feeding activity. Survival and development of eggs in an aegypti.Moreover, the daily minimum temperature and an increase in the rainfall from the previous months were associated with increase in the larval abundance.

IV. Classification of Journals

The total of 40 journals has been selected and analyzed. The objective of this analysis is to get the

clear perspective on the following:

i) To investigate the level of research basis on the concept of data mining techniques for dengue prediction.

ii) To identify the factors associated with dengue fever.

iii) To study about the influencing factor of dengue.

4.1 Journals by country of authors

Irrespective spread of the dengue prediction is the most important task for the research. Data mining is one of the best techniques for prediction. In this perspective, to know the importance of data mining in dengue prediction, the 40 journals are classified based on the authors and year of publication as shown in table 1

Table 1: No of journals based on authors

Year of publication	Indian authors	Others	Total
2012	1	1	2
2013	1	2	3
2014	1	-	1
2015	1	4	5
2016	5	6	11
2017	6	4	10
2018	4	4	8
Total	19	21	40

4.2 Journals based on technique

To predict the dengue, different prediction algorithms were used. In this, Decision Tree, Neural Networks, Support Vector Machine, machine learning and Combination of above are the some of the well-known techniques. To know about the usage of above data mining techniques in the area of dengue prediction, the 40 journals were classified based on the techniques are shown in table 2

Table 2: No of journals based on technique

Technique	Journals
k-means	4
Support vector machine	2
Decision tree	6
Artificial neural network	3
Machine learning	7
Ensemble prediction	2
Naïve Bayes	1
Hybrid	4
Others	11
Total	40

4.3 Journals based on technique and year of publications

To know the recent techniques of data mining in dengue prediction and to know about the effective techniques are shown in table 3.

Table 3: Journals based on techniques and year of publications

Technique	2012	2013	2014	2015	2016	2017	2018	Total
K -means	-	-	-	-	-	3	1	4
Support vector machine	1	-	-	-	-	-	1	2
Decision tree	1	1	1	1	-	-	1	5
Artificial neural network	-	1	-	-	-	1	1	3
Machine learning	-	-	-	-	3	3	1	7
Ensemble prediction	-	-	-	1	-	1	-	2
Naïve Bayes	-	-	-	1	-	-	-	1
Hybrid models (classification+clustering)	-	-	-	1	2	-	2	5
Others	-	1	-	2	5	2	1	11
Total	2	3	1	6	10	10	8	40

V. Conclusion

Dengue is a fast-emerging pandemic –prone viral disease in many part of the world. This disease is spread by several species of female mosquitoes of the Aedes type. There is no vaccine to prevent human infection by this virus. Personal protection and environmental management of mosquitoes are important in preventing illness. In this paper, the various factors like temperature, sunshine, rainfall, humidity, etc. that causing dengue fever were studied. The result of our analysis shows that the risk of dengue fever is positively associated with high temperature and inversely associated with the period of rainfall, sunshine. This result is consistent with findings of previous studies on dengue factors.

VI. Limitations of Research

As reveal that, not all the research journals are considered for the review. But the research journals which are reliable in all the perspective are considered for the analysis. So, the review has some limitations.

i) As first 40 journals form the last one decade has been considered for the research purpose.

ii) As second, journals are searched based on dengue prediction using data mining techniques. So, the search based on “dengue prediction” or “dengue factors”.

VII.References

[1] P. Manivannan, Dr.P.Isakki, Dengue Fever Prediction Using K-Means Clustering Algorithm, IEEE INTERNATIONAL CONFERENCE ON INTELLIGENT TECHNIQUES IN CONTROL,OPTIMIZATION AND SIGNAL PROCESSING, 978-1-5090-4778-9/17/©2017

[2] Panuwat Mekha, Khukrit Osathanunkul et al,Gene Classification of Dengue Virus Type Based on Codon Usage ,978-1-5090-4420-7/16/ ©2016 IEEE

[3] Abdul mahatir najar,Mohammad isa irawan et al, Extreme Learning Machine Method for Dengue Hemorrhagic Fever Outbreak Risk Level Prediction,IEEE International Conference On Smart Computing And Electronic Enterprise (ICSCEE2018) ©2018

[4]priyo s sasongko, helmie a wibawa et al, Performance Comparison of Artificial Neural Network Models for Dengue Fever Disease Detection , 1st International Conference on Informatics and Computational Sciences (ICICoS) 2017

[5]Buchado omkar, Dalsania preet et al,International Research Journal of Engineeringand Technology (IRJET) Volume: 04 Issue: 10 | Oct - 2017

[6] Norhayati binti Mohd Zainee , Kalaivani Chellappan , A Preliminary Dengue Fever Prediction Model Based On Vital Signs And Blood Profile , IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES) 2016

[7] Nirbhay Mathur, Vijanth S. Asirvadam, Visualization of Dengue Incidences using Expectation Maximization (EM)Algorithm , IEEE 978-1-5090-0845-2/16/© 2016

[8] T. Marimuthu,V . Balamurugan, A NOVEL BIO-COMPUTATIONAL MODEL FOR MINING THE DENGUE GENE SEQUENCES , International Journal of Computer Engineering & Technology (IJCET) Volume 6, Issue 10, Oct 2015, pp. 17-33, Article ID: IJCET_06_10_003

[9] M.Mufli Muzakki , Fhira Nhita , The Spreading Prediction of Dengue Hemorrhagic Fever (DHF) In Bandung Regency Using K-Means Clustering and Support Vector Machine Algorithm 6th International Conference on Information and Communication Technology,2018

[10] Iwan Inrawan Wiratmadja, Siti Yaumi Salamah & Rajesri Govindaraju, Journal of engineering and Technology Science., Vol. 50, No. 1, 2018, 110-126

[11] www.webmd.com/a-z guides/denguefever/ref1

[12] www.ncbi.nlm.govw

