DENGUE PREDICTIVE MODEL 2020 FOR RAWALPINDI DISTRICT PAKISTAN

Dr. Rizwana Shahid, Dr. Shazia Zeb, Prof. Dr. Muhammad Umar (Vice Chancellor RMU), Dr. Asim Noor, Dr. Syed Usama Khalid, Prof. Dr. Muhammad Khurram and Dr. Muhammad Mujeeb Khan

Rawalpindi Medical University, New Teaching Block, Holy Family Hospital, Satellite Town, Rawalpindi.

Article Info

*Corresponding Author:

Dr.Rizwana Shahid

Email:
drriz_shahid@yahoo.com

Abstract

Objectives: To draw dengue predictive model 2020 for Rawalpindi district in response to dengue epidemic 2019 for timely arrest and mitigation of dengue cases.

Methods: Predictive model was drawn by using machine learning technique. Numerous tools like Pandas, Numpy, Matplotlib, Sklearn, Pylab were use for data wrangling. Residential data of 12,192 dengue cases admitted in 3 teaching hospitals (Holy Family Hospital, Benazir Bhutto Hospital and District Head Quarters Hospital) affiliated with Rawalpindi Medical University was employed for this purpose. Weather data for the Rawalpindi was extracted from Texas A&M university weather data archive. Ordinary Linear Sequential (OLS) Regression was used to estimate the relationship of dengue cases reported during 2019 with weekly average temperature and cases to be reported during 2020 in Rawalpindi district. R^2 was computed as an indicator of model fit. Significance of association between reported cases and air temperature was statistically confirmed by application of t-test.

Results: Fitness of predicted model was determined in terms of R^2 and P-value by application of t-test of correlation. R^2 of our predicted model is 0.79, presenting positive relationship between active dengue cases and average weekly air temperature. Moreover, statistically significant association of reported cases with air temperature (P < 0.086) was determined with 97.5% Confidence Interval (CI).

Conclusion: Dengue predictive model 2020 drawn for Rawalpindi district would really be helpful to reduce dengue cases by application of appropriate preventive measures in high risk zones by concerned authorities.

Keywords: dengue predictive model, Ordinary Linear Sequential regression, t-test, machine learning technique.

Introduction:

Dengue is a disastrous health problem found to be prevalent worldwide. About 128 countries are found to be suffering from this disease. Pakistan is facing a drastic dengue outbreak menace since 2005. Propagation and survival of 4 serotypes of Aedes mosquito in Pakistan is mainly attributed to climatic conditions, urbanization, communication gap and poor surveillance, thus paving the way towards emergence of peak incidence of this disease[1].

Approximately 50% of global population is residing in dengue prone regions and 100 million dengue cases are reported annually[2]. High mortality and morbidity are mainly attributed to dengue hemorrhagic fever and dengue shock syndrome that are considered to be the most severe form of this ailment[3]. Even no internationally authorized vaccine is available for its prevention[4].

Globally dengue virus transmission is found to be endemic in about 128 countries of tropical and subtropical region that constitute approximately 3.9 billion of the world population[5].Dengue is a multifaceted disorder with varied symptoms ranging from mild to severe. Patients suffering with this complex disease are diagnosed as having dengue fever, dengue hemorrhagic fever, dengue shock syndrome and extended dengue syndrome[1]. Dengue fever results in both primary and secondary infections and commonly prevails among adults and older children[6]. Dengue hemorrhagic fever is markedly seen during secondary dengue infection. However, infants may primarily be infected via maternally attained dengue immunoglobulins[7]. High mortality is attributed to dengue shock syndrome mostly due to severe shock and multi-organ failure[8].

Pakistan is hotspot for numerous vector borne diseases in addition to dengue fever and this disease is attributed to circulation of all four serotypes of dengue virus[9]. WHO regional office is seriously concerned with outbreak of dengue in Pakistan during 2019 and is providing technical support in this regard to manage this alarming situation[10]. Despite the support provision by WHO for vector control, case

management and community awareness, about 47,120 confirmed cases of dengue fever including 75 deaths were reported from all 3 provinces of Pakistan. A total of 9,676 confirmed dengue fever cases including 16 deaths were testified from Punjab province. Generally the risk of dengue in Pakistan was appraised to be elevated[11]. It seemed that all control measures failed to arrest the spread of dengue in Pakistan.

The present study is therefore intended to draw a dengue predictive model 2020 for Rawalpindi District by analyzing the data of dengue cases registered in Teaching hospitals (Holy Family Hospital, Benazir Bhutto Hospital and District Head Quarters Hospital) affiliated with Rawalpindi Medical University, Rawalpindi during dengue epidemic 2019. Dengue predictive model designed in this regard will definitely facilitate our policy makers, strategic planners and respective health work force to take appropriate measures for timely management of dengue cases in 2020 by getting intimation of expected cases from our dengue predictive model.

Methodology:

The data-set comprised of hospitalized dengue cases for the year of 2019 in the three teaching hospitals of the Rawalpindi Medical University i.e; Holy Family Hospital, Benazir Bhutto Hospital and District Head Quarters Hospital. Data was gathered after securing informed consent from respective Hospital administration. Residential data of 12192 cases diagnosed with dengue fever, dengue hemorrhagic fever and dengue shock syndrome admitted in all three teaching hospitals during September – November 2019

---

was included in this research. This research was intended to mitigate the chances of dengue epidemic in future by focusing on number of active dengue cases reported during 2019, the residence of patients and geographic or climatic conditions (relative humidity, rainfall, temperature) prevailing during dengue outbreak 2019 in Rawalpindi district. Weather data for the Rawalpindi was extracted from Texas A&M university weather data archive (https://globalweather.tamu.edu/).

Results:

Data Analysis

For data wrangling and later analysis python stack was used. Following tools have been frequently used for this project, Pandas, Numpy, Matplotlib, Sklearn, Pylab, Ipyleaflet, Seaborn, Geopanda, Geopy, jupyter notebook, etc.

*Fig 1: Reporting of Dengue cases on weekly basis in Rawalpindi District 2019*
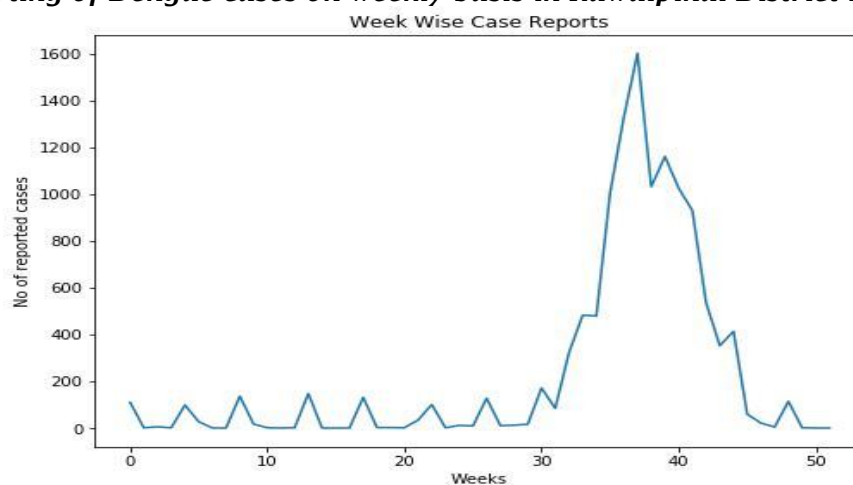
The above plotted graph clearly validate that dengue spread started increasing in the month of July, it peaked in October then started declining by mid of November and reached back to its minimum level.

Area Wise Dengue Concentration

Map clearly shows the areas affected by the dengue. More preventive focus should be placed on these areas in terms of spray, fumigation and awareness.
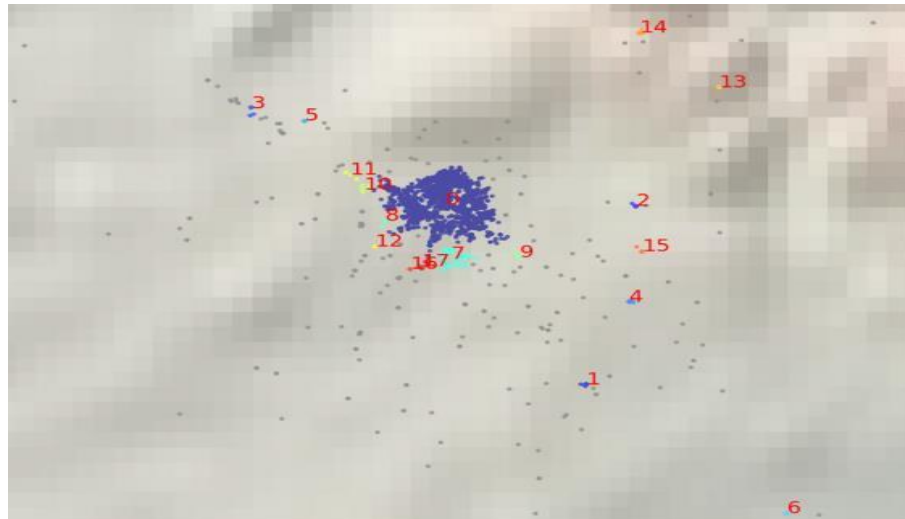
Fig 2: Map of Rawalpindi District depicting cluster of dengue cases in 2019
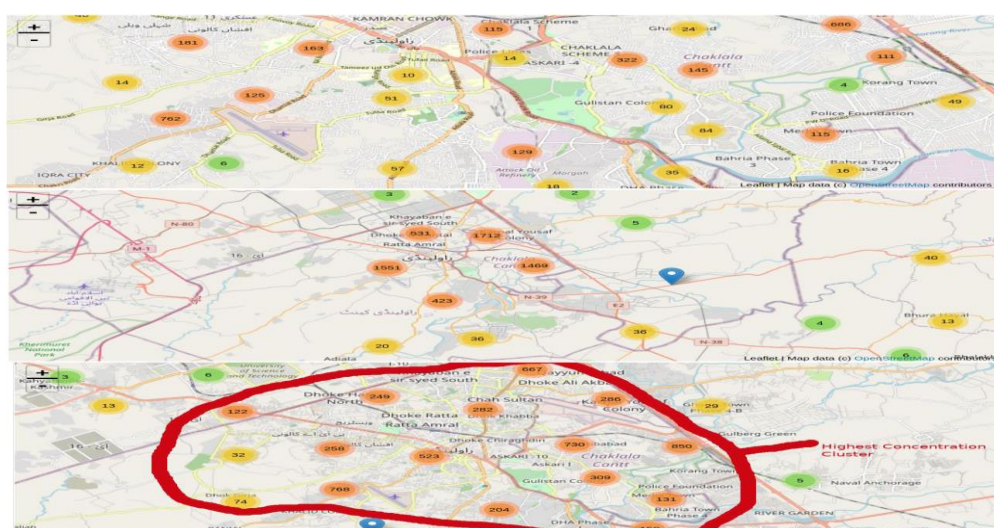
Disease Cluster:

Higher concentration of disease is visible through this map. The figure below is obtained through DBSCAN state of the art clustering algorithm. It is also showing the same cluster as visible in above map.

Fig 3: Cluster of Dengue cases

Cluster 0 is in the center of the city and is highlighted through red circle in the top map. Table below shows the number of reported cases in each cluster.

Table 1: no. Of dengue cases reported in each cluster during dengue epidemic 2019

S.No.	Zones	Areas of Rawalpindi District (5km radius)	No. of Dengue cases recorded in 2019
1.	Red Zone	Khayaban-e-Sir Syed / Dhoke Mangtal / Ratta Amral	531
2.		Colonel Yousaf colony	1712
3.		Rawalpindi city	1551
4.		Chaklala Cantt	1469
5.		Rawalpindi Cantt	423
6.		Dhamyal Road / Tulsa Road	762
7.	Others	Adyala Road	60
8.		Police Foundation	115
9.		Bahria Town	140
10.		Afshan Colony	181
11.		Morgah	128
Total			**7072**

- 58% form Rawalpindi district
- 34 % form Islamabad Capital Territory (ICT)
- 8% Azad Jammu Kashmir (AJK), Abbottabad, Attock.

Predictive Model

We took inspiration from literature and experimented with following independent variables:

- Average weekly temperature
- Average weekly precipitation
- Average weekly wind speed
- Active Cases (last two weeks total reported cases)

We applied Ordinary Linear Sequential Regression analysis. Our experiment showed two variables: *'Active Cases '*and *'Average weekly temperature'* as significant variables. Other two were found to be insignificant.

Following equation represents our predictive model:

Reported Cases = α *+* β1 *(average weekly temperature) +* β1 *(number of active cases) ---->EQ1*

Table 2: OLS Regression Results

```
==============================================================================
```

Dep. Variable:	y	R-squared:	0.794
Model:	OLS	Adj. R-squared:	0.785
Method:	Least Squares	F-statistic:	94.23
Date:	Mon, 20 Apr 2020	Prob (F-statistic):	1.62e-17
Time:	15:11:22	Log-Likelihood:	-343.74
No. Observations:	52		
Df Residuals:	49		
		Df Model:	2
		Covariance Type:	Non-robust

```
==============================================================================
```

	coef	std errt	P>|t|	[0.025	0.975]
const	-113.0093	82.852	-1.364	0.179	-279.506	53.487
x1	5.4860	3.132	1.752	0.086	-0.807	11.779
x2	0.4442	0.034	13.252	0.000	0.377	0.512

R^2 (coefficient of determination) of our predicted model is 0.79, thus illustrating the amount of variance in active dengue cases due to variance in average weekly air temperature. Moreover, P-value computed with 97.5% CI on application of t-test of correlation was computed to be significant showing statistically significant association of reported cases (x2) with air temperature (x1).

Goodness of the Model Fit:

The figure in this section depicts the goodness of the fit model. As apparent blue dots are real values, whereas red squares are predicted values. Blue and red dots are moving together that means our model fits well.

Fig 4: Linear Regression of reported dengue cases with average weekly temperature & cases to be reported during 2020
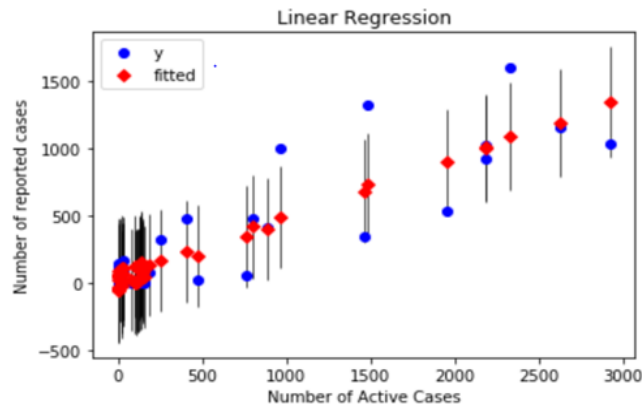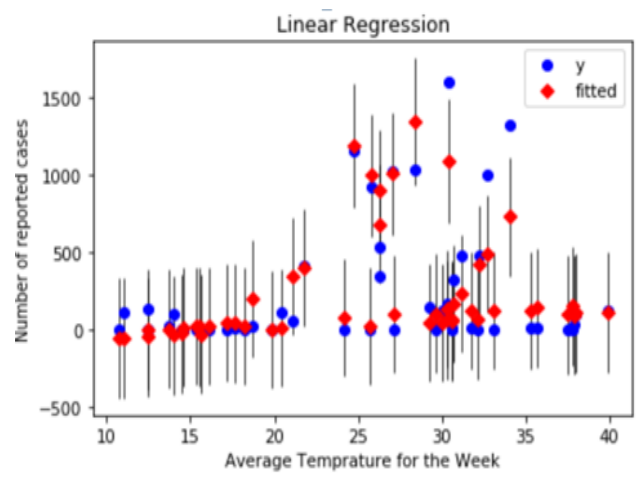
Fig 5: Dengue cases to be reported in 2020 in correlation with active cases of 2019

In the following figure solid line represents the predicted values whereas blue dots represent real values. Again these plots in Figure 5 show model is good fit

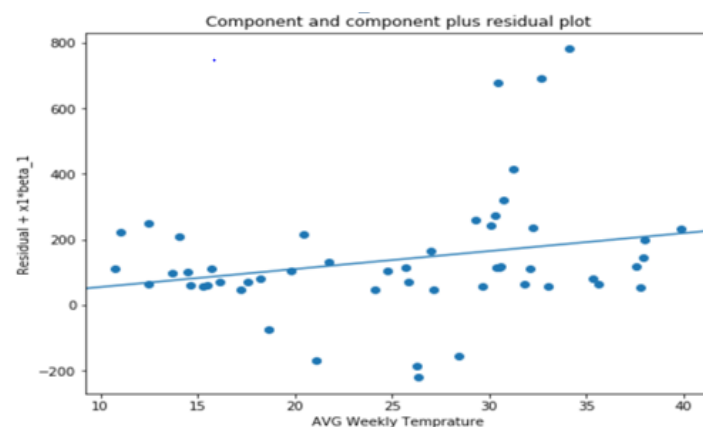*Fig 6: Estimation of active cases in compliance with average weekly temperature*

Figure 6 depicts that average weekly temperature (blue solid line) is reasonably estimating cases reported (blue dots), so it is good estimator of reported cases.

Disease Spread Scenarios for 2020

Our predicted model relied on the two variables (number of active cases and average weekly temperature). Since temperature is an exogenous variable out of anyone control, all efforts should be made to reduce the number of active cases. If double the number

of active cases is reported during July, then worst scenario would be to have around 14025 cases. However, better management and timely action to arrest the expansion of dengue epidemic will reduce the cases to about 1/3 (instead of 12,192 dengue cases reported during 2019, cases can best be mitigated to around three thousand) as reflected below in Table 3.

Table 3: Dengue cases predicted during 2020 in Rawalpindi District

	Best Cases	Average Cases	Worst Cases
Expected Number of Dengue Cases in 2020	3072	7806	14025

Discussion:

Despite the self-limiting capacity of dengue virus infections, they have emerged as a public health challenge in tropical and sub-tropical regions of the world. Their diverse clinical manifestations ranging from asymptomatic to alarming complications drag our attention towards timely and accurate diagnosis. This facet is imperative for apt control in peak season and reduction in case fatality rate[12].

In current study dengue predictive model is developed for Rawalpindi district by using machine learning technique. Rawalpindi district is located in northernmost part of Punjab with Rawalpindi city constituting capital of this district[13]. This district occupies total area of 5286 km^2 with a population of approximately 5.4 million[14]. Apart from active dengue cases reported during 2019, various meteorological attributes (temperature, precipitation, wind speed) were taken into account in Ordinary Linear Sequential (OLS) regression analysis. However, active cases of 2019 registered in Rawalpindi district and air temperature were found to be significant. On review of average air temperature from July to November 2019, minimum temperature range was 25-35.8^{0}C while maximum temperature was determined to be 7.5-24.8^{0}C[15]. Our model used cases of Dengue Fever (DF), Dengue Hemorrhagic Fever (DHF) and Dengue Shock Syndrome (DSS) to foresee upcoming cases during 2020. Similarly a dengue predictive model was designed through multiple regressions considering the Dengue Hemorrhagic Fever (DHF) cases in Makassar (capital of Indonesia) in addition to climatic attributes of that locality. Contrary to reporting of dengue cases in Pakistan from May onwards, increased dengue cases in capital of Indonesia are accounted during first few months of the year[16]. Likewise our research, occurrence of dengue cases in Makassar did not depict significant correlation with rainfall despite facilitation of mosquito breeding in early stages during rainy season[17,18].

Similar to the present study, A Chinese research to foresee dengue cases in future by Guo P et al revealed same methodology (use of active cases and climatic conditions as predictor variables) but unlike our research, Chinese scientists have employed multiple regression models. One of those models Support Vector Regression (SVR) model conferred trivial prediction errors[19] and therefore labeled as one of the modern and potent machine learning algorithms in forecasting time series[20]. A group of international researchers evaluated the accuracy of three regression models in forecasting dengue incidence and finally decided that SVR model devised for estimation of dengue cases in Singapore was more authentic in comparison with other regression models. Pearson correlation, r^2 and area under the Receiving Operating Characteristic (ROC) curve were

computed to determine the authenticity of this model[21]. Comparative analytical studies should be carried out to explore the efficiency of various regression models drawn by different countries in context of their climatic attributes and prediction of dengue incidence.

The current research is aimed to predict the occurrence of dengue cases in Rawalpindi district of Pakistan 2020 because implementation of widespread dengue prevention and control strategies as specified in WHO Report (2012-2020)[22] could not prevent us from facing dreadful consequences of dengue epidemic 2019. We secured residential data of dengue cases reported in Rawalpindi district during 2019 along with meteorological statistics. A similar research was planned in Bangkok, capital of Thailand by taking into account climatic conditions apart from socio-economic data of dengue cases[23]. Likewise our study, data pertinent to cases of dengue fever, dengue hemorrhagic fever and dengue shock syndrome was taken into consideration, machine learning methodology was employed and fitness of forecasted model in district was depicted in terms of adjusted R-squared value. Moreover, researchers of Thailand also displayed discriminating ability of final model in terms of sensitivity, specificity, positive and negative predictive value. Their research confirmed the impact of population mobility patterns on the escalation of dengue cases[23].

Our study did not reveal any relationship of humidity and rainfall with occurrence of dengue cases. Conversely, a research by Muurlink et al on multiple dengue outbreaks in Bangladesh concluded that number of rainy days in previous two months and average weekly air temperature in preceding one month apart from mean humidity level about six months prior to epidemic month are the best predictors of dengue outbreak[24]. Another international research was carried out by Park S et al among Thai children who presented with febrile illness during 2018 by using structural equation models. This study was aimed to revise clinical management guidelines specified for dengue suspects before development of critical illness. Predictors of dengue cases varying from dengue fever to dengue shock syndrome were age, tourniquet test, aspartate aminotransferase, white blood cell and platelet counts[25]. These attributes can also be utilized by our researchers for revision of dengue management guidelines and arrest the progression of suspects to critical phase of illness. This would really be constructive in reducing workload particularly on our infectious diseases department during dengue season.

Despite the endemicity of dengue in most of the tropical and subtropical regions of the world, denuge incidence both within and between years reflect substantial disparity[26]. Reporting of more dengue cases in Mexico was primarily attributed to increased population due to urbanization[27,28] resulting in inappropriate water management[29] and increased foreign trade and tourism[30]. These attributes could also be the probable contributing factors for dengue outbreaks in Rawalpindi district. Apart from climatic conditions, environmental sanitation and trading capacities can also be taken into account as independent variables to forecast the occurrence of dengue cases rationally.

Conclusion & Recommendations:

Occurrence of dengue cases during dengue epidemic 2019 in Rawalpindi district of Pakistan was in correlation with air temperature of the region. Peak of dengue cases during 2020 will most likely be in the same localities that reported escalated number of cases earlier. This dengue predictive model would really be very helpful to precisely track the dengue dynamics in Rawalpindi district. Moreover, high accuracy and robustness of this dengue predictive model could also be validated by using data of

dengue cases of previous 5-10 years and also by retrival of data from other provinces of Pakistan.

Peak in dengue cased during 2020 can drastically be suppressed by strict control on build up of dengue cases. This could best be achieved by implementation of integrated vector control measures, social mobilization and inter-sectoral coordination. Moreover, capacity building of our entomologists, vector control personnel, environmental and social scientists is the need of time to intelligently tackle the situation. These measures coupled with epidemiological surveillance would help the concerned authorities to much extent to arrest the future occurrence of dengue outbreak.

References:

[1.] Ahmed S, Aziz MS, Aftab A, Ullah Z, Ahmad MI, Mustan A. Epidemiology of dengue in Pakistan, present prevalence and guidelines for future control. International Journal of Mosquito Research 2017; 4(6): 25-32.

[2.] Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature, 2013; 496:504-507.

[3.] Khawsak p. Phantana s. Chansiri K. Determination of dengue virus serotypes in Thailand using PCR based method. Southeast*Asian J Trop Med Public Health* 2003: 34: 781-785.

[4.] Yousaf MZ, Siddique A, Ashfaq UA, Ali M. Scenario of dengue infection & its control in Pakistan: An up-date and way forward. Asian Pacific Journal of Tropical Medicine 2018; 11(1): 15-23.

[5.] Murray NEA, Quam MB, Wilder-Smith A. Epidemiology of Dengue: Past, present and future prospects. Clin Epidemiol 2013; 5(1): 299-309.

[6.] Narayanan M, Aravind MA, Thilothammal N, Prema R, Sargunam CS, Ramamurty N. Dengue fever epidemic in Chennai-a study of clinical profile and outcome. Indian Pediatr. 2002;39:1027–33.

[7.] Halstead SB, Lan NT, Myint TT, Shwe TN, Nisalak A, Kalyanarooj S, et al. Dengue hemorrhagic fever in infants: Research opportunities ignored. Emerg Infect Dis. 2002;8:1474–9.

[8.] Shivpuri A, Shivpuri A. Dengue-An overview. Dent Med Probl. 2011;48:153–6.

[9.] Paul RE, Patel AY, Mirza S, Fisher H, Luby SP. Expansion of epidemic dengue viral infection to Pakistan. Int J Infect Dis 1988; 2: 197-201.

[10.] 10. WHO support to Pakistan on dengue fever. Available at: http://www.emro.who.int/pak/pakistan-infocus/world-healtday.html.

[11.] Outbreak update – Dengue in Pakistan 19 November 2019. Available at: https://reliefweb.int/report/pakistan/outbreak-update-dengue-pakistan-19-november-2019.

[12.] Hasan S, Jamdar SF, Alalowi M, Al Beaiji SMA. Dengue Virus: A global human threat: Review of Literature. J Int Soc Prev Community Dent 2016; 6(1): 1-6.

[13.] District wise Census results Census 2017. www.pbscensus.gov.pk. Archived from the original (PDF) on 29 August 2017.

[14.] District Profile: Northern Punjab - Rawalpindi Archived 16 May 2008 at the Wayback Machine.

[15.] Monthly weather forecaset and climate Rawalpindi, Pakistan. Available at:

[16.] Halide H, Ridd P. A predictive model for Dengue Hemorrhagic Fever epidemics. Int. J. Environ. Health Res 2008; 18(4): 253-265.

[17.] Alto BW, Juliano SA. Precipitation and temperature effects on populations of Aedes albopictus: Implications for range expansion. J Med Entomol 2001; 38: 646-656.

[18.] Costanzo KS, Kesavaraju B, Steven A, Juliano SA. Condition-specific competition in container mosquitoes: The role of noncompeting life history stages. Ecology 2005; 86: 3289-3295.

[19.] Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, et al. Developing a dengue forecast model using machine learning: A case study in China. PLoS Negl Trop Dis 2017; 11(10): e0005973. https://doi.org/10.1371/journal.pntd.0005973.

[20.] Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing. 2004;14(3):199–222.

[21.] Althouse BM, Ng YY, Cummings DA. Prediction of dengue incidence using search query surveillance. PLoS neglected tropical diseases. 2011 Aug;5(8):e1258. PubMed Central PMCID: PMCPmc3149016. pmid:21829744.

[22.] Global strategy for dengue prevention and control, 2012-2020 (WHO Report). Available at: https://www.who.int/denguecontrol/9789241504034/en/.

[23.] Jain R, Sontisirikit S, Iamsirithaworn S, Prendinger H. Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data. BMC Infectious Diseases 2019; 19(1): 272-287. DOI: 10.1186/s12879-019-3874-x.

[24.] Muurlink OT, Stephenson P, Islam MZ, Taylor-Robinson AW. Long-term predictors of dengue outbreaks in Bangladesh: A data mining approach. Infect Dis Model. 2018; 3: 322–330. doi: 10.1016/j.idm.2018.11.004

[25.] Park S, Srikiathachorn A, Kalayanarooj S, Macareo L, Green S, Friedman JF, et al. Use of structural equation models to predict dengue illness phenotype. PLoS Negl Trop Dis 2018; 12(10): e0006799. https://doi.org/10.1371/journal.pntd.0006799.

[26.] Dantes HG, Farfan-Ale JA, Sarti E. Epidemiological trends of dengue diseases in Mexico (2000-2011): A systematic literature search and analysis. PLoS Negl Trop Dis 2014; 8(11): e3158. https://doi.org/10.1371/journal.pntd.0003158.

[27.] San Martín JL, Brathwaite O, Zambrano B, Solórzano JO, Bouckenooghe A, et al. The epidemiology of dengue in the americas over the last three decades: a worrisome reality. Am J Trop Med Hyg 2010; 82: 128–135

[28.] Diaz-Quijano FA, Waldman EA. Factors associated with dengue mortality in Latin America and the Caribbean, 1995–2009: an ecological study. Am J Trop Med Hyg 2012; 86: 328–334.

[29.] Tauil PL. Urbanization and dengue ecology. Cad Saude Publica 2001; 17: 99–102.

[30.] Cuddehe M. Mexico fights rise in dengue fever. Lancet 2009; 374: 602.