www.arpnjournals.com

ENHANCED QUALITY OF SERVICE IN VISUALIZING THE MALARIA DATA USING CLOUD COMPUTING

A. Vijayalakshmi and B. Rajesh Kanna
School of Computing Science and Engineeing, VIT University, Chennai, India

ABSTRACT

Malaria disease is a major issue in public health problem and its spread in various tropical countries. Malaria forecasting was conducted in many tropical countries and typically uses data on environmental factors, human factors. There are variations in time on malaria causing incidence respect to geographical regions. The goal of the proposed work is to develop a cloud computing based application to insight the root cause of malaria. It provides the enhanced quality of service like data filter, statistical data model, data view, data relation, which helps to forecast the environmental risk of malaria. This forecast is inevitable because malaria prediction data are independent to each other, static in nature, hierarchical and complex. The proposed malaria forecasting application has been programmed using 'R' script and integrated with built-in IBM Bluemix container. Finally, it has been deployed as platform as a service under IBM cloud. The developed application assists the malaria predictors to do seamless view of dense graphical informatics simultaneously.

Keywords: malaria, forecasting, visualization, cloud computing, quality of service, predictors.

1. INTRODUCTION

Malaria disease is caused by plasmodium parasite which infects the human red blood cells. The malaria parasites can be classified by types. They are P. falciparum, P. vivax, P. ovale, and P. malariae [1]. It transmitted to humans by female *Anopheles* mosquitoes [1]. The malaria victims are children below 5 years. As per the WHO report 3.4 billion people are at the risk of malaria, 207 million malaria infected cases recorded in the year 2013, and the estimated death is 6, 27, 000 [2] Malaria caused by P. falciparum and P. vivax is a serious public health problem and its impact is variable according to the geographical region [3]. The highest transmission rate is found in tropical and sub tropical countries [3]. The other two species causing human malaria P. malariae, P. ovale are much less common. Approximately malaria infection in Asia due to the P. vivax is about 70-90% [1]. To control the spread of the malaria disease diagnosing is the primary step. There are number of diagnostic methods currently available; they are clinical diagnosis, laboratory diagnosis and molecular diagnosis [4]. These tests will help us to identify the four different malaria species. The Routine methods used for diagnosis malaria parasites are rapid diagnostic test kits (RDT) and by means of light microscopy [1]. Malaria diagnosis involves identifying malaria parasites or antigens/products in patient blood [4]. Different species diagnosis at different stages like ring stage, trophozoite, schizont and gametocyte [1].

Malaria survey has been conducted in many countries to know the infectivity and mortality ratio of the people [5]. It is a second most serious disease after tuberculosis and endemic in areas of Africa, South Asia, America, Europe and Middle East [6]. So researchers and practitioners have in need to determine of spatial and temporal variability of malaria incidence to improve the forecasting system [5].The primary goal of data visualization is to capture the dynamic qualities in forecast the environmental risk However, public health professionals and health scientists are lacking in technical skills, computational skills and time to acquire, process and analyse the data [7]. Several approaches have been proposed to forecast the malaria data. Graphical tools help them to visually understand the data. These tools use computer graphics techniques to transform columns of data into images. These images enable analysts to assimilate the enormous amount of data online, and it accelerates the identification of hidden patterns contained within the malaria data set.

For online forecasting, Cloud Computing provides flexible and scalable services without having the computing resources on their system [11]. It moves data and application away from desktop and PC to a large data center [11]. The malaria visualization application can be delivered as services over the cloud to large number of users.

2. BACKGROUND

Forecasting of malaria on the geographical spaces depends on the following environmental factors like susceptible humans, infected humans and non-infected humans [8] which are depicted in Figure-1. Here, the more concern will be on the suspected and infected human. The suspected human may or may not infect by malaria parasite. On the other hand Genetic Factors present from birth can protect against certain types of malaria [10]. Inherited blood disorders (IBDs) also the common public health problem which is particularly related to malaria [9]. IBDs are passed down from the parents and affect the blood. Malaria parasites enter into the red blood cells and make some changes in the structure of our red blood cells. Some changes to red blood cells make more resistant to malaria infection. The few IBDs are Duffy blood group, Glucose-6-phosphate dehydrogenase (G6PD) deficiency, sickle cell [9].

Sickle-cell is called haemoglobin S and occurs in people who have inherited the haemoglobin S (HbS) gene

from both parents. The sickle cell trait protected against *P. falciparum* malaria [10]. Duffy blood types differ in the structure of the Duffy glycoprotein on the surface of the RBC. It is the resistant to the infection of *P. vivax* [10]. G6PD deficiency is a hereditary disease which can cause haemolytic anaemia, this can protect against p. *vivax* parasite [10].

The above discussed data related to malaria disease are currently available in the Malaria Atlas Project (MAP). Spatial medical data is central to the effective planning and controlling of malaria disease [9]. It assembled a unique spatial database based on medical intelligence, satellite-derived climate data. The map produce the largest ever archive of community-based estimates of parasite prevalence [9]. The MAP focus has been centered on predicting the endemicity of P. falciparum, the most deadly form of the malaria parasite, as well as P. vivax. It gives two separate sets of survey information for 100 countries. We can visualize the geographical wise spread of malaria infection and download the data at different perspective such as P. falciparum (Pf), Anopheles DVS, G6PD Deficiency, Sickle cell, Duffy Negativity, *P. vivax* (PV).
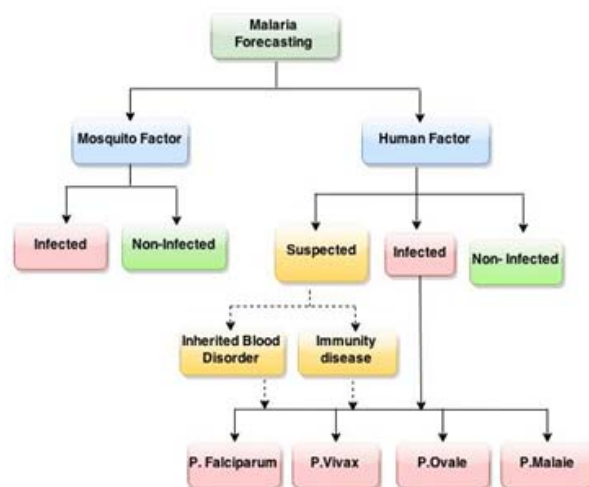

Figure-1. Ecology of Malaria forecasting.

3. MOTIVATION

Currently, MAP data sets consist of information about longitude, latitude, county name, the type of diagnostic method used, age range, month and year of diagnosing, number of population examined, number of population infected by the malaria parasite [9]. To forecast the malaria disease the available raw information are in adequate. If these data have been assembled and visualized by software it will be helpful to the geographers, statisticians, epidemiologists, biologists and public health specialists. The data are isolated, static and were scattered based on the location. Further, it is having a collection of various data type like numerical, descriptive, Boolean and location data. The functional characteristics

of those data involved in forecast are mostly independent each other.

Single visualization will not be the representative to all other parameters. The parameters which involve in the malaria occurrence are relatively varied due to the space and time. So, consolidating the result in single visual canvas is highly challenging. However, the purpose of forecasting is, the data can be viewed through multiple views for global prediction of malaria. The malaria survey data are spread over 100 counties, providing an online visual application need to share the resources to a large collection of users. So it is highly challenging to provide demanding infrastructure with respect to simultaneous access.

From the literature it is observed that the IBDs and malaria infection is inversely proportional. It means that the persons having IBD are less resistant to malaria. However, in MAP the Inherited blood disorder data sets such as Sickle cell, Duffy and G6PD are isolated. Further, integrating all these IDB data set together against the malaria victim with respect a geographical location is vital information for forecasting. It will provide the global study for intersecting disorder data set with respect to rate of the growth of malaria. This sophisticated feature is not available in existing system. In Figure-2, this operation was done as pilot study and explored the importance of this purpose. The efficiency of forecasting also depends on finding the inter relationship between different parasites and blood disorders. The existing system does not employ any mathematical models or statistical techniques to find the data distribution and relationship.
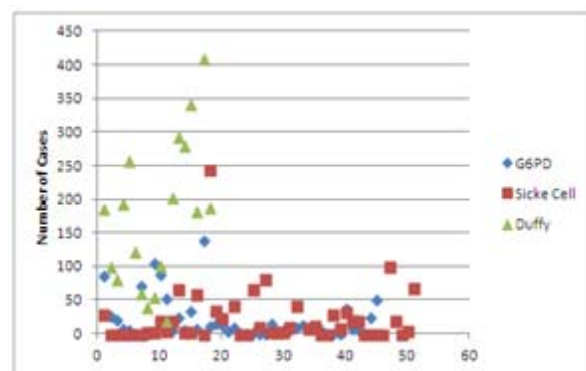

Figure-2. Comparison of infected number of cases though having Duffy, Sickle cell, G6PD from the region of India.

To prove the importance of this inter relation we have correlated those data for specific country. We observed that the number of infected human by *P. falciparum* is negatively correlated with P.vivax. The number of sickle cell infected human having malaria is negatively correlated with P. vivax. Table-1 shows the pair wise correlations between P. falciparum, P.vivax, Sicke cell, Duffy, G6PD.

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Table-1. Correlated values of P.falciparum, *P. vivax*, Sickle, Duffy, G6PD for India.

	P.falci parum	P.vivax	Sickle cell	Duffy	G6PD
P.falci parum	----	-0.040	-0.005	0.015	-0.045
P.vivax	----	----	-0.029	-0.048	-0.040
Sickle cell	----	----	----	0.065	0.033
Duffy	----	----	----	----	0.544
G6PD	----	----	----	----	----

Identifying the Malaria risk due to climate trigger and exploring total number of cases of malaria positives for a particular country with respect to year or month is lacking in the existing MAP. Figure-3 explores the comparison of predicted and observed number of cases infected by *P. falciparum* and *P. vivax*. More victims of malaria are observed to be children below 5 years. Considering the age as the decision variable and calculating the percentage of infection for a particular time period is vital analysis in Malaria forecast. It is constructive information to public health services for appropriate awareness and to take selective preventive measures. This service is not incorporated in existing on-line Malaria data portal. Figure-4 shows the graphical depiction of *P. falciparum* infection by the lower age people.
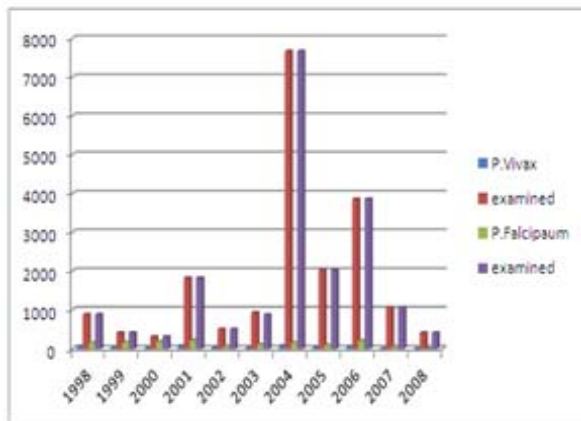

Figure-3. Comparison of predicted and observed cases infected by *P. falciparum* and *P. vivax* from India during 1998-2008.
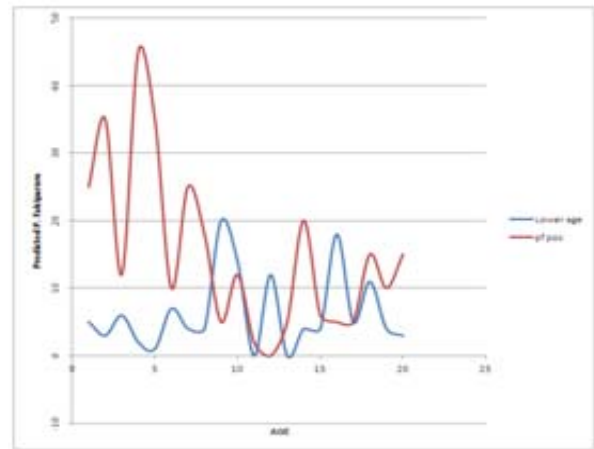

Figure-4. Incidence of falciparum malaria cases between the age group of 0 to 25 in India.

4. PROPOSED SYSTEM DESIGN

The aim of this work focuses on developing novel tool to forecast the malaria risk factors, and designing a customized tool to visualize the MAP data. It helps to esthetic access and easier to interpret the risks for the people, who involved in the malaria control.

The proposed system consists of five model viz. Data acquisition, View of data model, Development of visualization applications, Selection of application, Deploy application in cloud. We have used IBM DB2 database software for storing the structured malaria data. We employed java scripts for data selection; build the view of statistical data model and the selection of appropriate visual applications. Visualization procedures are carried out using R script. For building the cloud application we have utilized the IBM Bluemix's liberty for JAVA. It is a scalable java application, highly composable, ultra-light profile of IBM web sphere server application designed specific to cloud. Figure-5 depicts the flow of the process of our proposed system.

In data acquisition, the data used for malaria forecast have been collected from the survey records of MAP. MAP has given the provision to download all the survey data via Data explorer. Data explorer provides the open access of data in the form of .xls file to users. We used IBM DB2 software to import this file contents to data base. Thereafter, we have normalized the data and made a relationship among it.
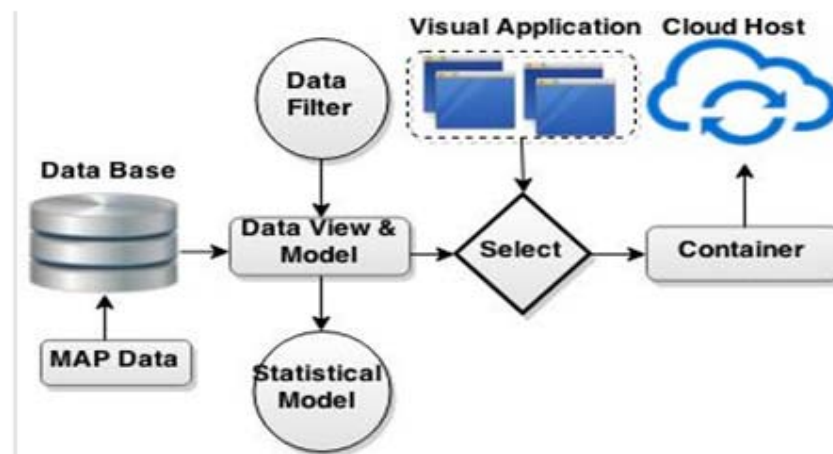
Figure-5. Process flow of the proposed system.

In view of data model, first the user will select the required data field as well the range of their wish. They can do all the basic mathematical operation to filter or remove the outliers to analyze the malaria epidemic risk. This includes multi level data selection, set filters, constrains and range values from database. To accomplish this task graphical user interface have developed and java scripts are used to perform the all sort of data filtering operations. Moreover, to find out the relationship among the selected data, we developed a set of programs to perform some common statistical calculations like mean, median, mode and few complex statistical distribution and probability functions. Figure-6 shows the overall design of data view model.

In developing visualization application, we used R programming tool, because it is an environment and a language to produce data graphics. It is also having provision to export the developed application as XML, can be easily in-corporate with web application. Here, we developed graph based visualization using R script, generally a graph consists of two axes called horizontal and vertical axis. However, in malaria forecast the relationship between one attribute with multiple attributes are also essential. Hence, we developed few applications it takes more than two attributes as input and provide its corresponding visuals. The developed visual applications exported as a XML package to deploy as web applications. In building and deploying the selected visualization application in cloud, first we need to fork sample application of Liberty for Java runtime from IBM Bluemix. Then, set up local Eclipse IDE to test and debug application locally. By using this local development environment, we modified the application code by incorporating the selected XML package for visualization. The updated source code can be deployed on to IBM Bluemix. The Bluemix manages the elasticity when the number of simultaneous usage of application, it also provides load balancing.
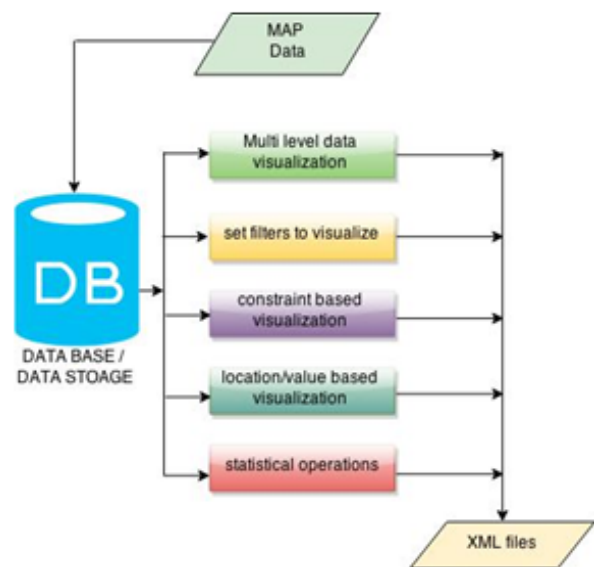

Figure-6. Data view and modeling.

5. CONCLUSION AND FUTURE WORKS

This work provides enhanced QOS to malaria predictors by addressing the customized malaria data manipulation and provides the cloud based elastic environment to produce customized data interpretation and visualization. It will support to control malaria at national and international scales. The people who involved in the malaria forecast can get benefit from this new tool. Our work completely depends on the MAP data. The data's are static and not a stream in nature. There is no automatic data crawling from the MAP website. So the periodic updating should be done manually. We can extend our work to automatically acquire the data from MAP and incorporate with our application

ARPN Journal of Engineering and Applied Sciences

REFERENCES

[1] May Z, Azreen Mohd Aziz S.S, Salamat R. 2013. Automated quantification and classification of malaria parasites in thin blood smears. Signal and Image Processing Applications (ICSIPA), 2013 IEEE International Conference on. vol., no., pp.369, 373, 8-10.

[2] World Health Organization. Factsheet on the World MalariaReport 2013. http://www.who.int/malaria/media/world malaria report 2013/en/.

[3] Dian Anggraini, Anto Satriyo Nugroho, Christian Pratama, Ismail Ekoprayitno Rozi, Aulia Arif Iskandar, Reggio Nurtanio Hartono. 2011. Automated Status Identification of Microscopic Images Obtained from Malaria Thin Blood Smears. IEEE International conference on Electrical Engineering and Informatics 17-19 July.

[4] Tangpukdee N, Duangdee C, Wilairatana P, Krudsood S. 2009. Malaria Diagnosis: A Brief Review. The Korean Journal of Parasitology; 47(2): 93-102. doi:10.334.

[5] Kate Zinszer, Aman D Verma, Katia Charland, Timothy F Brewer, John S Brownstein, Zhuoyu Sun, David L Buckeridge. 2012. A scoping review of malaria forecasting: past work and future directions. BMJ Open. 2(6): e001992.

[6] Kareem S, Kale I, Morling RCS. Automated P.falciparum Detection System for Post-treatment Malaria Diagnosis Using Modified Annular Ring Ratio Method. 14th International Conference on Modelling and Simulation.

[7] Wimberly, Michael C., et al. 2012. A computer system for forecasting malaria epidemic risk using remotely-sensed environmental data. Proceedings of the International Congress on Environmental Modelling and Software. Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty, Leipzig, Germany.

[8] Athithan S, Mini Ghosh. Mathematical modelling of Malaria with bed-net effect and treatment. Proceedings of National Conference on Pure and Applied Mathematics (NCPAM- 2013), pp. 153-159.

[9] Malaria atlas project (MAP). http://www.map.ox.ac.uk/

[10] CDC–Centers for Disease Control and Prevention. http://www.cdc.gov/malaria/about/biology/human_factors.html.

[11] Jadeja Y, Modi K. 2012. Cloud computing - concepts, architecture and challenges. 2012 International Conference on Computing, Electronics and Electrical Technologies [ICCEET]. vol., no., pp. 877, 880.