FORECASTING DENGUE PATIENT COUNTS IN A DISTRICT BASED ON NEIGHBORING DISTRICTS AND RAIN-FALL VIA MACHINE LEARNING

A. Nugaliyadde[1], G. T. U. V. Somaratne[2], R. Kumarasinghe[3], and C. J. Wikramasinghe[4]

[1]Department of Information Technology, Sri Lanka Institute of Information Technology, Email: anupiya.n@sliit.lk
[2] Department of Information Systems Engineering, Sri Lanka Institute of Information Technology, Email: usomaratne@gmail.com
[3] Department of Information Technology, Sri Lanka Institute of Information Technology, Email: roshk6@gmail.com
[4] Department of Information Technology, Sri Lanka Institute of Information Technology, Email: jagath.w@sliit.lk

ABSTRACT

Dengue is one of the irregularly spread diseases in Sri Lanka with a significant number of deaths per year; making it complicated to be prepared for it in advance. In order to minimize the risks of the situation, it is well identified that the disease must be predicted. Such prediction will prevent the government and the hospitals spending an additional expense of treating the patients and the required additional resources. Therefore, there is a high demand for a predicting module to forecast the number of patients for each district weekly. The research focuses on delivering a module that is capable of predicting the number of dengue patients in a particular district in the upcoming week via the rain patterns and the neighboring district's patient counts. The research focused on using Self-Organizing Maps (SOM) to forecast the patient counts. Data was gathered from over ten years of dengue patients, from each district weekly from the epidemiology unit Sri Lanka. The data was categorized using SOM for each district. The clustered patterns in each district were matched with the other neighboring districts. This showed that the disease spreads from district to district. This provides the ability to predict the patient count from district to district depending on the neighboring district's patient count. The research also revealed that the district rainfall affects the number of patients where it will also provide a considerable effect on the predictions for the patient count. This also generated a connection with the rain pattern of each district from district to the patient pattern.

***Keywords*:** Self organizing Maps, Clustering, Forecasting

1. INTRODUCTION

The first recognized dengue epidemics were occurred simultaneously in Asia, Africa, and North America in 1780s. By late 1990s close to 40 million cases of dengue fever and several hundred thousand cases of dengue hemorrhagic fever each year were reported. By now, about half of the world's population is at risk [1]. National Dengue Control Unit of Sri Lanka states that 5372 cases were reported for the month of January 2015 [2]. Dengue is becoming an intimidating disease since there is no specific treatment for it. Severe dengue is leading in the number of illness and death reported among children [3, 4]. Since there is no specific treatment, dengue prevention and control solely depends on effective vector control measures. Further number of research is being conducted worldwide to prevent and control dengue. According to World Health Organization, the disease is spread in tropical countries with local variations of rainfall, temperature and unplanned rapid urbanization [1]. Considering all these factors, the research focused on information related to rainfall and dengue count. Available resources pointed out that there may be a pattern of the rainfall and the dengue count since the rainfall is already accepted as a factor to increase the spread of dengue virus. Therefore, the researchers focused on methods to forecast dengue patient counts in a district based on neighboring districts rainfall. Once, a pattern is identified the knowledge generated from the research can be used to do the preparations for the upcoming dengue patients. The government, health sector, population of the country as well as other organizations can come up with solutions to minimize the spread of the disease.

2. METHODOLOGY

Recent research done in Sri Lanka has proven that dengue is widely spread. Nearly 90% of patients are children [5].Unpredicted occurrences of the disease create an unorganized treatment scheme for the patients. The scheme prevents the patients receiving the best treatment and will also establish a potential for higher fatality rate. Therefore, a methodology must be used in order to predict the number of dengue patients in a district. The prediction will also support in preventing the patient numbers rising from time to time.

Machine learning has been used in recent research on medicine. Machine learning has allowed to dynamically predict patient numbers as well as to detect cancer traits in human cells, which discovered more traits than medical science has identified [6]. Machine Learning has shown many patterns in medical science which has not been discovered.

The work was carried out on developing a neural network in order to predict the number of dengue patients. The research focus is on deriving a connection between each district and the migration of dengue, with the rainfall. The study used Self-Organizing Maps (SOM) to discover patterns in patient count in each district.

Data on dengue patient counts are gathered in each of the districts on a weekly basis from the epidemiology unit of Sri Lanka. Weekly data for eight years were gathered. The relevant data on the rainfall for the corresponding week was also collected to build the relationship between the two entities.

The common myth regarding the connection between the rainfall and dengue was discussed and studied, though a proper relationship was not created earlier between the two entities. A focus on generating a relationship between the two entities and creating a prediction based on the entities was considered.

Cory W. Morin, Andrew C. Comrie and Kacey Ernst have developed a framework [7] in which they discuss on the increased count of the mosquitoes in relation to the weather factors mainly focused on temperature and rainfall. Siti Morni Umor, Mazlin B.Mokhatar, Noraini Surip and Anizar Ahmad have built a risk map of dengue based on the environmental factors and GIS technologies [8] which manages to track the dengue distribution through the Asian countries. The above-mentioned research managed to identify that there is a connection between the weather and the spread of dengue. The research has proven that there is a direct effect on the rainfall and the spread of dengue. The research conducted by Khoa T.D. Thai and Katherine L. Anders on “The role of climate variability and change in the transmission dynamics and geographic distribution of dengue” [9] proven that with the dynamic change of the rainfall and other climate variables the distribution of dengue moves from a location to another location. A relationship was built on the research.

With the evidence found the researches focused on predicting the movement of dengue from district to district, with the use of machine learning. Self-Organizing Maps (SOM) was used in the research in order to cluster the dengue patient patterns of each district. With the use of

“Tanagra” the dengue patient counts were clustered for each district. Using the SOM a 7 x 7 metrics were created. Using SOM for each district a pattern was created. This pattern was created for five years which ranged from 2005 – 2010.

Using the patterns generated; a flow of dengue was discovered. 2011-2013 data was used as the testing data in predicting the dengue patient counts using the generated SOM pattern. The research predicted the number of dengue patients from the test data. The accuracy will be discussed in the results section. The movement of dengue was monitored and found that peeks of patient counts increased from district to district as the disease spread according to a pattern. The pattern of the disease transfer was identified discovered that it was similar to the rainfall movement of the monsoon rains.

The weekly based rainfall and the movement of the rain from one district to another. The rain pattern and the dengue patient counts resembled a correlation-ship. Therefore to predict the patient counts rainfall was used as a factor in which to identify which district the rainfall effects. This will be further discussed in the results section of the paper. The movement of the rapid increase of the disease from one district to another could be mapped with the direct pattern of the monsoon rain pattern. In the results it will discuss on the pattern and how the pattern has been related to the monsoon rain pattern.

With the rain patterns considered a model was created to predict the patient counts from a district to a district. The model designed is able to predict the number of patients with relevant to district population and the rainfall distribution.

3. RESULTS

The data was analyzed using the SOM. Each district's patient counts are created into a simple cluster. A cluster shown in each district for each year on a weekly basis show that dengue patient pattern has been transferred from one district to another closer district. The data clusters showed a pattern in which some closer districts did not directly get the dengue patient patterns that was expected.

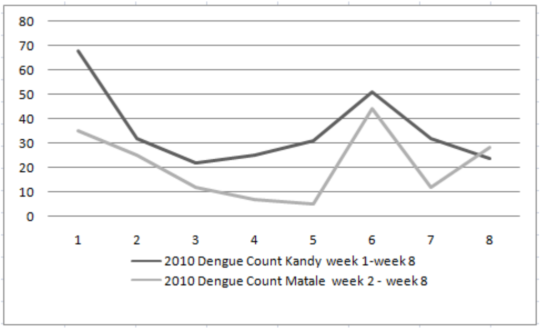

Figure 1: Sample data set which shows the transfer of dengue from district to district

The expected dengue spread pattern was similar to a spider web in which dengue spreads from the first cited pattern to every neighboring district. It was noticed that the dengue spread in a different pattern than expected. A pattern that was generated from one district did not appear in all the neighboring districts. The pattern repetition occurred on mostly one or two neighboring districts only. Similar to branching out in a tree the disease spread rather than web spread.

The Disease spread had no connection between the neighboring districts. Some instances indicated that the disease did not even appear in the neighboring districts.

The research was focused on the myth of the connection with the rain and the dengue spread. The monsoon rain pattern was considered. A connection between monsoon rain pattern and dengue was identified.

The movement of the monsoon rain and the dengue spread was similar, with this the proper connection could be drawn with the monsoon and the dengue spread. The pattern generated by

SOM was an indicator that the rain pattern and the dengue count pattern matched with the originator location of dengue. According to figure 1, The dengue transfer from Kandy to Matale has been shown for the first weeks of the year. The figure shows the dengue spread transfers within a week gap to the neighboring districts which with the rain distribution.
The Neighboring districts which were not affected by the monsoon did not report dengue outbreaks [10].

4. CONCLUSION

The research was focused on discovering a pattern in the spread of dengue from a district to another district. The focus was to discover a pattern with dengue outbreaks transferring and spreading throughout the country and to prevent the spread via the prediction. Working with the Epidemiology Unit of Sri Lanka, researchers gathered data from 2005 to 2014 on a weekly basis. In order to predict the patient uprisings Self-Organizing Maps (SOM), a machine learning algorithm was used. SOM was used to generate patterns in the gathered patient data for each district.

The researchers tried to match the each pattern generated for each district through SOM to identify a pattern of disease movement. Since an immediate pattern was not found the created patterns for each neighboring district was generated with a 2 weeks gap. The generated pattern using SOM was matched with the neighboring district. With this it was identified that the disease takes around 2 weeks to transfer from one district to another neighboring district.

It was also discovered that the disease did not spread through to every neighboring district and that the pattern of the disease spread was not as expected. The disease spread was mapped to the monsoon rain patterns. Outbreaks were mapped with the monsoon. If an outbreak occurred in one district that is effected by the monsoon rains, the outbreak transfers to neighboring districts that is with the monsoon rain. A direct transfer of dengue through monsoon rains was monitored. Which concludes that the effect of rain as well as the neighboring district's dengue status effects on the dengue spread.

5. REFERENCES

[1]"*Dengue and severe dengue*", http://www.who.int/media centre/factsheets /fs117/en/.

[2]"*National Dengue Control Unit*", http://www.dengue.health.gov.lk/.

[3]"*Alarming increase in number of Dengue patients*", http://www.dengue.lk/index.php/45-landscape-project-at-the-grayslake-campus.

[4]"*Factsheets*",http://www.epid.gov.lk/web/index.php?option=com_content&view=article&id=146&Itemid=459&ang=en.

[5] G. N. Malavige, P. K. Ranatunga, V. G. N. Velathanthiri, S. Fernando, D. H. Karunatilaka, J. Aaskov, and S L Seneviratne, "*Pattern of disease in Sri Lankan Dengue patients*", Arch Dis Child, vol. 91,pp. 396–400, 2006.

[6]"*Machine learning for cancer diagnosis & prognosis*",pages.cs.wisc.edu/~olvi/uwmp/cancer.html.

[7] C. W. Morin, A. C. Comrie, and K. Ernst", Climate and dengue transmission: *Evidence and implications*", in:Proceedings of Environmental Health Perspectives, vol. 121, pp 11 – 12, 2013.

[8] S. M. Umor, M. B. Mokhtar, N. Surip, and A. Ahmad, "*Generating a dengue risk map based on environmental factors using remote sensing and gis technologies*" in:Proceedings of Conference on the Human Microbiome.

[9] K. T. D. Thai, and K. L. Anders "*The role of climate variability and change in the*

transmission dynamics and geographic distribution of dengue”, in:Proceedings of Experimenimental Biology and Medicine, vol.1. pp 944 – 954, 2011

[10] H. Zhang, Z. Li, S. Lai, A. C. A. Clements, L. Wang, W. Yin, H. Zhou, H. Yu, W. Hu, and W. Yang. “*Evaluation of the performance of a dengue outbreak detection tool for China*”, vol. 9(8). pp. 1 – 5, 2014.