Forecasting and Data Visualization of Dengue spread in the Philippine Visayas Island group

Abstract— **Dengue is a rapidly spreading, mosquito-borne viral disease with an estimated incidence of 50 million cases occurring annually around the world [1]. In the Philippines, from January to August 2015 alone, 55,079 cases were reported [3]. Dengue continues to be a tremendous health issue due to several implementation [4], and environmental health factors [2]. The study proposes the use of an Artificial Neural Network (ANN) to accurately forecast future dengue cases with input parameters of *previous month's dengue cases*, and averaged *temperature* and *rainfall* with the highest correlation value of 0.80. An online platform was implemented to contain all records of dengue cases and weather data, and the neural network was integrated into the website. The data visualization of the predicted dengue cases is generated through the use of dot density and chloropleth maps that are then displayed on the online platform to allow for the general public, future researchers, government, and health officials to access, analyze, and interpret the predicted dengue cases as one sees fit.**

I. INTRODUCTION

Dengue is a rapidly spreading viral disease with an estimated incidence of 50 million cases occurring annually around the world. The mosquito vector, *Aedes aegypti*, is common in both rural and urban areas in the South East Asian Region [1]. In the Philippines, a total of 585,342 cases, with a fatality rate of 0.55%, were reported from 2008 to 2012 [2]. From January to August 2015 alone, 55,079 cases were reported, and the number of cases had increased by 9.12% from the previous year [3]. With insufficient government funding [4], inadequate public health infrastructure, and additional environmental and health risk factors, dengue continues to be a tremendous health issue that is challenging to tackle in the Philippine setting [2].

The study has two main objectives. It aims to be able to accurately forecast future dengue cases, and also to be able to develop a system that allows for the visualization of both recorded and predicted dengue data. Ultimately, the research aims to enable government and health officials to make intelligent and informed decisions in the deployment of manpower and resources in order to properly address the predicted number of dengue cases.

Previous researches have used various different models and factors to predict dengue cases. MSEIRS is a commonly used mathematical model that takes into account the number of vector and hosts in a disease system [5]. The model has been used by several studies that incorporate the use of the vector-host relationship with additional factors such as spatial distribution [8][9], vector offspring [8], and vector-host mobility [10]. Due to the flexibility of the MSEIRS model and its derivatives, it is also commonly used alongside a Cellular Automata model [10-14]. Regression models are also used in order to assess vulnerability and forecast dengue cases using various environmental and socioeconomic factors such as population density, land use, and weather [15-17]. Due to the difficulty in collecting mosquito vector data, a model that could forecast dengue cases without the use of a vector parameter was selected. The use of weather parameters was selected due to the correlation of weather factors such as temperature, rainfall, and humidity to the incidence of dengue cases [18][19].This research opts to use an Artificial Neural Network (ANN) that takes in dengue incidence and weather data as parameters to forecast dengue cases [19].

Artificial Neural Networks are commonly used in pattern recognition and classification problems. It is a single processing unit composed of several interconnected nodes. These nodes are divided into three main layers, input, hidden, and output, and are connected through weights that allow the neural network to transform the input parameters into a desired output. The neural network is trained to generate the desired output through learning. It does this by repeatedly processing several examples and altering the weights of the interconnected notes to produce the desired outcomes [19].

II. METHODOLOGY

The Visayas island group was chosen as the area of study due to its archipelagic nature. It comprises of Regions VI, VII, and VIII and consists of 337 cities and nine weather stations in total.

The data sets acquired consist of monthly records of weather and dengue data. The weather data includes the values *total rainfall (mm)*, *average temperature (°C)*, and *percentage relative humidity* for each station available in the vicinity. The dengue data includes total dengue cases per city. The data sets cover a five year period from January 2010 to December 2015. Any missing weather data over the study period was interpolated by averaging the available weather data. Data was acquired from the Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAGASA) and National Epidemiology Bureau (NEB) under the Department of Health (DOH) for the weather and dengue data, respectively.

In preparation for feeding the data into artificial neural network, the data was normalized using Gaussian Normalization and divided into two categories. Data spanning from 2010 to 2013 was used for training, while 2014 to 2015 data was used for testing and validation.

An Artificial Neural Network was used to predict the subsequent dengue cases. The network consists of three layers, the *input layer*, one *hidden layer* and an *output layer*. The input layer is fully connected to the hidden layer, and the hidden layer is connected to a single node output of the output layer. The output of the neural network is a normalized value for the predicted number of dengue cases. The input and hidden nodes of the neural network vary according to the number of input parameters per model.

Eight different models of neural networks were divided into two sets. The first set of models were trained using actual weather data, while the second set of models were trained using average monthly weather data over the five year period. Each set contained four different neural networks with varying input parameters. The standard model consists of input parameters of *last month's dengue cases*, and this month's *total rainfall* and *average temperature*. The second variation of the model consists of the standard model's input parameters, and includes a fourth input parameter of *percentage relative humidity*. This variation of the model is indicated by the capital letter H. The third variation of the model consists of the standard model's input parameters, and a *Province Index* as a fourth input parameter. This variation is indicated by the capital letter P. The fourth variation of the model contains the standard model's input parameters, and includes both additional parameters of *percentage relative humidity* and a *Province Index*. The model's variation is indicated by the capital letters HP. As indicated above, the input and hidden layers vary according to the number of input parameters for each model, with the standard model's layers containing three nodes each, the second and third containing four, and the fourth containing five nodes.

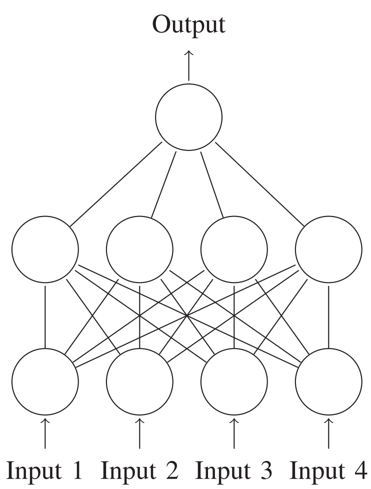

Fig. 1: A variation of the standard neural network with four input and hidden nodes.

III. RESULTS & DISCUSSION

All neural networks were trained for 600 epochs each. The resulting neural networks were then made to predict the subsequent dengue cases for the test years 2014 to 2015. The results were compared to a normalized value of the actual number of dengue cases. The positive relationship between the resulting set and the actual set was measured through *correlation value*.

Actual Set	Correl Value	Averaged Set	Correl Value
Standard	0.77	Standard	0.80
H	0.79	H	0.78
P	0.77	P	0.78
HP	0.61	HP	0.77

Fig. 2: Displays the corresponding correlation values for each neural network.

The highest correlating result was selected from each set. Among the set trained with actual weather data, the neural network with standard input parameters and an additional parameter of humidity scored the highest with a correlation value of 0.79. Whereas, among the neural networks trained with averaged weather data, the network with the standard input parameters scored the highest with a correlation value of 0.80. The correlation values of the remaining neural networks are as displayed in Figure 3.

The neural networks that contained a Province Index as an input parameter scored lower than the over variations among the set. The variations trained using the actual weather data that contained a Province Index scored a correlation value of 0.77 and 0.61, whereas the variations trained over the averaged weather data scored a value of 0.78 and 0.77, each containing the lowest from their set. This may be a result of the negligibility of provincial and geographical factors.

All, if not most, of the neural networks can closely predict the general incidence of dengue cases over the testing period of 2014 to 2015. Similarly, the networks that scored the highest correlation values from their respective sets can also predict the incidence of dengue cases. Although both networks can predict the fluctuations of the subsequent dengue cases, the networks have difficulty predicting exponential increases in dengue cases should there be a spike in the fluctuation. Spikes can be observed in both actual and predicted dengue cases. However, the magnitude of the increase for the predicted dengue cases does not match, but greatly surpasses, that of the actual dengue cases that occurs. The discrepancy in high volume spikes of occurrence in dengue cases is a problem that arose when the neural network was tested with Philippine dengue cases. In comparison to that of the original research situated in Singapore [19], where over the span of 10 years a rough total of 15,000 cases were recorded, the Philippines experiences more than 50,000 cases in less than a year. These discrepancies can be addressed through further training of the neural network or through changing the initial structure of the neural network model.

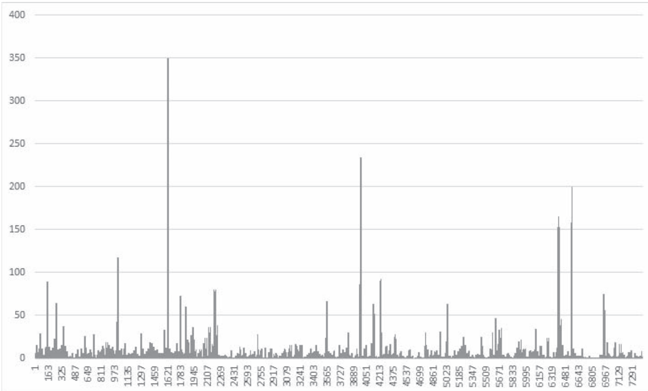

Fig. 3: Displays the actual dengue cases over the years 2014 to 2015.

Among the two highest scoring networks from the two sets, the averaged neural network with standard input parameters illustrated stronger correlation with a value of 0.80. Although the network trained with actual weather data and an additional humidity input parameter, follows closely with a correlation of 0.79. The neural network trained over average weather data more closely predicts the actual incidence of dengue cases. The set trained over the averaged weather data also scored better over its counterparts. This may be a result of either the negligibility in the change in weather over then span of a few years ,the cyclic and seasonal nature of weather, or due to overfitting of the network to the averaged weather data.

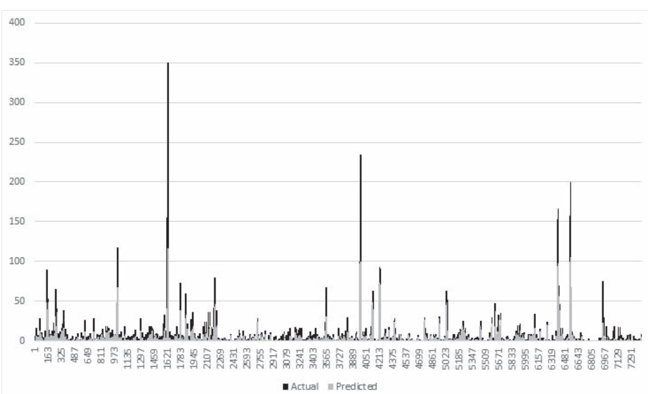

Fig. 4: Actual dengue cases overlayed by predicted cases of H actual neural network.

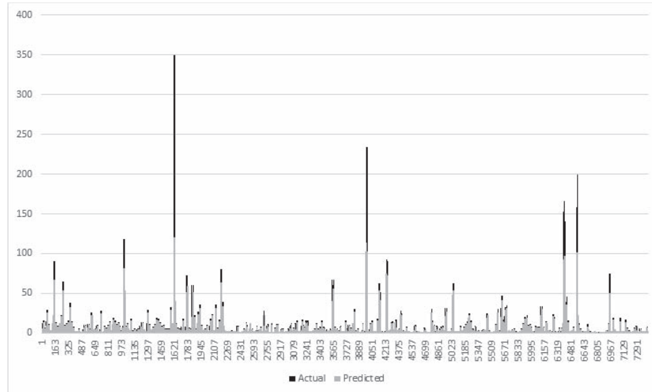

Fig. 5: Actual dengue cases overlayed by predicted cases of standard actual neural network.

IV. DESIGN

An online web system was selected as it serves as the most suitable platform to allow accessibility of information to the general public. Similar data visualization systems such as HealthMap [20], also use an online platform in order to make their data more accessible to the public.

All data and records are stored on the basis of a single geophysical unit, the city. Each city stores three types of data, geographic, weather, and dengue data. The geographical data of a city consists of a single coordinate denoting the center of the city, and an ordered set of coordinates that generates the shape and borders of the city. Both weather and dengue data consists of the input parameters of temperature, rainfall and humidity, and dengue incidence for a month. The recorded data is then used by the designed Artificial Neural Network integrated into the system to predict the next month's dengue cases. The process will run on the last day of each month in order to minimize the service downtime.

A visualization of the resulting predicted dengue cases is then generated in the form of dot density and chloropleth maps. These visualizations are publicly accessible without the need to register for an account. Any visitor of the online platform can select any previous month to view recorded dengue cases or select the upcoming month to view the predicted dengue cases and their corresponding visualization.

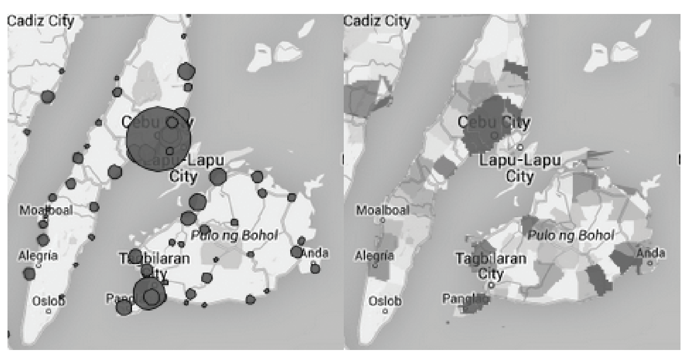

Fig. 6: Sample Visualization: Dot density map (left), Chloropleth map (right).

V. SCOPE & LIMITATIONS

All data collected for the research spans over the time period from January 2010 to December 2015. The geophysical scale of the research is limited to the Visayas island group spanning Regions VI, VII, and VIII. Each city of the specified area is used as one heterogenous geophysical unit. The factors incorporated as input parameters are limited to dengue cases, relative humidity, temperature, rainfall, and a province index.

VI. CONCLUSION & RECOMMENDATIONS

The use of an Artificial Neural Network with input parameters of *previous month's dengue cases*, and averaged *temperature* and *rainfall* proved to be the most effective in predicting dengue cases with a correlation value of 0.80. An online platform was implemented to contain all records of dengue cases and weather data, and the neural network was integrated into the website. The data visualization of the predicted dengue cases is generated through the use of dot density and chloropleth maps that are then displayed on the online platform. The platform is publicly accessible to all to allow for the general public, future researchers, government, and health officials to access, analyze, and interpret the predicted dengue cases as one sees fit.

Expanding the geophysical scope of the research to other regions such as Luzon or Mindanao is recommended. Including a larger time period of 10 to 15 years of both weather and dengue data is also recommended to improve the robustness of the model. Inclusion of other environmental or geographic factors such as terrain, cleanliness, and vector of disease is also recommended for future study. Finally, re-evaluation of a neural network model that can more accurately detect high volume spikes in dengue cases is a point of expansion for study.

REFERENCES

[1] World Health Organization (2009). Dengue, Guidelines for Diagnosis, Treatment, Prevention and Control, New Ed. 2009. Geneva, Switzerland: WHO press

[2] Edillo, F., Halasa, Y., Largo, F., Erasmo, J., Amoin, N., & Alera, M. et al. (Feb 2015). Economic Cost and Burden of Dengue in the Philippines. American Journal Of Tropical Medicine And Hygiene, 92(2), 360-366. doi: 10.4269/ajtmh.14-0139

[3] Geronimo, J.Y. (Sep 2015). Dengue on the rise: 55,000 reported cases so far in 2015. Rappler. Retrieved from http://www.rappler.com/nation/104736-dengue-cases-january-august-2015

[4] Geronimo, J.Y. (Mar 2015). PH spends less than P1B to fight neglected tropical diseases. Rappler. Retrieved from http://www.rappler.com/nation/86346-ph-2015-budget-neglected-tropical-diseases

[5] Hethcote, H.W. (2000). The Mathematics of Infectious Diseases. Society for Industrial and Applied Mathematics. Vol. 42, No. 4, 599-653.

[6] Pandey, A., Muyabi, A., & Medlock, J. (Dec 2013). Comparing vectorhost and SIR models for dengue transmission. PubMed - NCBI. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24427785

[7] Derouich, M., Bouyateb, A., & Twizell, E.H. (Feb 2003). A model of dengue fever. Biomed Eng Online. 2003;2:110. doi: 10.1186/1475-925X-2-4

[8] de Castro Medeiros L.C., Castilho C.A., Braga C., de Souza W.V., Regis L., Monteiro A.M, (Jan 2011). Modeling the dynamic transmission of dengue fever: investigating disease persistence. PLoS Negl Trop Dis. 5(1):e942. doi: 10.1371/journal.pntd.0000942

[9] Chowella, G., Diaz-Dueasc, P., Millera, J.C., Alcazar-Velazcoc, A., Hymana, J.M., & Fenimoreb, P.W., et. al. (Aug 2007). Estimation of the reproduction number of dengue fever from spatial epidemic data. Mathematical Biosciences, Volume 208, Issue 2, 571589. doi:10.1016/j.mbs.2006.11.011

[10] Enduri, M.K., & Jolad, S. (Apr 2015). Dynamics of Dengue with human and vector mobility. Retrieved from: http://arxiv.org/pdf/1409.0965.pdf

[11] Soemsap, T., Wongthanavasu, S., & Satimai, W. (Mar 2014). Forecasting number of dengue patients using cellular automata model. 2014 International Electrical Engineering Congress (iEECON), 1 - 4. doi: 10.1109/iEECON.2014.6925876

[12] Fu, S.C., & Milne, G. (n.d.). Epidemic Modelling Using Cellular Automata. Retrieved from: http://smr.csse.uwa.edu.au/pdf/EpidemicModellingUsingCA.pdf

[13] Holko, A., Medrek, M., Pastuszak, Z., & Phusavat, K. (Aug 2015). Epidemiological modeling with a population density map-based cellular automata simulation system. Expert Systems with Applications. doi: 10.1016/j.eswa.2015.08.018

[14] Merchant, F. (Feb 2007). Simulating the Spread of Malaria: A Cellular Automaton based Mathematical Model & a Prototype Software Implementation. Retrieved from: https://theses.lib.vt.edu/theses/available/etd-02222007-110205/unrestricted/Farid_Thesis.pdf

[15] Dickin, S.K., Schuster-Wallace, C.J., & Elliot, S.J. (2013). Developing a Vulnerability Mapping Methodology: Applying the Water-Associated Disease Index to Dengue in Malaysia. PLoS ONE 8(5): e63584. doi: 10.1371/journal.pone.0063584

[16] Chiu, C.H., Wen, T.H., Chien, L.C., & Yu, H.L. (2014). A Probabilistic Spatial Dengue Fever Risk Assessment by a Threshold-Based-Quantile Regression Method. PLoS ONE 9(10): e106334. doi: 10.1371/journal.pone.0106334

[17] Hii, Y.L., Zhu, H., Ng, N., Ng, L.C., & Rocklv, J. (Nov 2012). Forecast of Dengue Incidence Using Temperature and Rainfall. PLoS Negl Trop Dis 6(11): e1908. doi: 10.1371/journal.pntd.0001908

[18] Su, G.L. (Jun 2008). Correlation of Climatic Factors and Dengue Incidence in Metro Manila, Philippines. Ambio, A Journal of the Human Environment. doi: 10.1579/0044-7447(2008)37[292:COCFAD]2.0.CO;2

[19] Aburas, H.M., Cetiner, B.G., & Sari, M. (2010). Dengue confirmed-cases prediction: A neural network model. Expert Systems with Applications, 37.

[20] HealthMap (n.d): About. In: HealthMap. Retrieved from: http://www.healthmap.org/site/about