Using Clustering and Predictive Analysis of Infected Area on Dengue Outbreaks in Malaysia

Najihah Ibrahim, Nur Shazwani Md. Akhir, Fadratul Hafinaz Hassan
School of Computer Sciences, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia.
najihah.ibrahim@student.usm.my

***Abstract*—Machine learning and data mining have a great impact on the predictive analysis process. The features classification on machine learning can be used to adopt the clustering method to define further analysis on the targeted issues. Nowadays, the epidemic disease outbreaks have caused a great concern towards Malaysian community as the diseases can cause great fatality. One of the common killer epidemic diseases in Malaysia is dengue fever. Dengue fever is caused by dengue virus that spreads by Aedes mosquitoes. The outbreaks cause several cases of death and it varies throughout the states in Malaysia. The factors that cause this epidemic disease were determined and the data on the dengue outbreaks in Malaysia were gathered. To predict the infected area of dengue, data were mined and the machine learning method was implemented. In this study, the clustering method in machine learning for predictive analysis is proven to be an effective method in determining the most infected area of dengue outbreaks in Malaysia: Selangor and W.P. Kuala Lumpur/ Putrajaya. The selected areas were identified as the busiest place in Malaysia with a great number of population that had caused high physical contact and promoted the dengue outbreaks.**

***Index Terms*— Clustering; Dengue Outbreaks; Epidemic Disease; Machine Learning; Predictive Analysis.**

I. INTRODUCTION

Epidemic diseases are the contagious diseases that are possible to be spread into the entire nation if the contagion measurement had reached the outbreak level and managed to wipe out the entire population. There are some well-known epidemic outbreaks that happen in the entire world such as dengue, yellow fever, cholera, diphtheria, influenza, bird flu and many more [1-8]. These contagious diseases had caused a major world health issues and believed to be one of the major factors that had caused 43% of the global fatality [7].

Malaysia also had experienced some epidemic diseases outbreak such as dengue, hepatitis, chikungunya and many more. However, recently from 2013 to 2016, there were reports on the outbreak of diphtheria in Malaysia [9]. Diphtheria may cause the inflammation of the nerves, paralysis and bleeding problem to the host. Even though diphtheria involves the sole host, the infection may spread in many ways and multiple types of agents. The transmission usually happens via direct contact and contaminates the air. However, this disease also is believed to have been caused by the irregular of vaccination.

The infection of epidemic disease can spread vigorously by the active mobilization of the pathogen and the rapid production and stimulation of the pathogens [6]. This disease outbreak stimulated by several factors that can be identified as the natural factors and the man-made factors. These factors are almost impossible to be measured by classic statistic numeration. They may contribute towards the identification, detection, prediction and controlling of the disease via features classification. However, due to the integration of the features, a clustering method of the features has been introduced, and this method causes weight adjustment on the input.

This features' clustering method may result in less accurate and approximate detection and prediction on the epidemic disease. Hence, the study of finding the optimal result of the integration of clustered features was introduced via the back propagation method that is able to identify and correlate the factors that have impact on the epidemic disease dissemination [2].

This research paper discusses the effectiveness of predictive analysis in determining the epidemic disease infected area by enlightening the process of analysis using back propagation method of machine learning process that introduces the cross-validation technique. A case study on the implementation of the clustering method in machine learning is also introduced to predict the infected area of dengue outbreaks in Malaysia. The rest of this research paper is organized as follows: Section 2 consists of a review on the targeted machine learning process; the predictive analysis and the impact on the epidemic disease prediction analysis. The classifiers of the epidemic diseases dissemination are given in the Section 3. Section 4 introduces the case study of this research, and the analysis of this case study will be provided in Section 5. Conclusion is provided in Section 6.

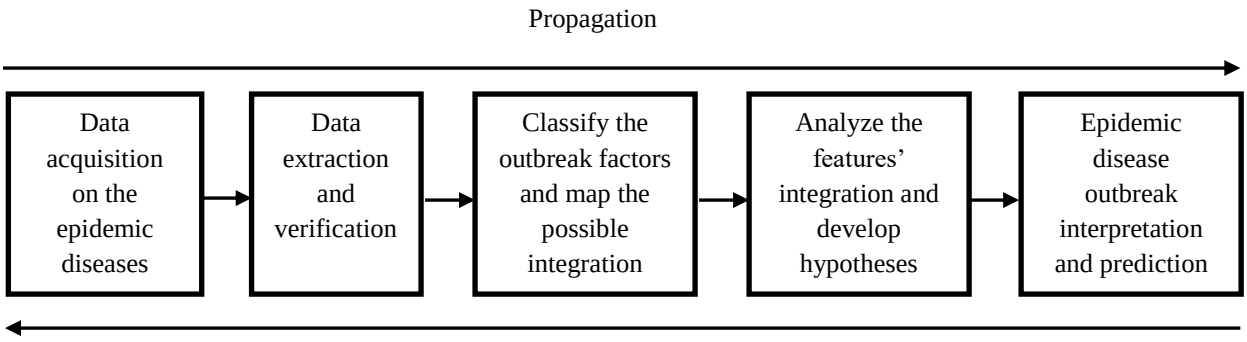

Figure 1: The data pipeline of machine learning process on epidemic diseases' data

II. MACHINE LEARNING: PREDICTIVE ANALYSIS

Classical analysis of the epidemic disease had taken place for years, before the advanced development of technology in computing intensive and data manipulation. In this context, computer scientists have developed advanced and more urban methods to describe the epidemic pattern, predict the future outbreaks or the impact size of a population over the infectious outbreaks and control the counter measure of the infectious diseases to reduce or overcome the outbreaks via data acquisition [2]. This urban method however, needs a machine to learn and train itself for a more dynamic input that leads to the need for cross validation techniques. Hence, this method is known as machine learning process.

Machine learning is a data based analysis process, in which the analysis can only be done by gathering data from data mining process and going through the data pipeline process for knowledge discovery. Machine learning can be used to describe, predict and control the mining data. This learning method promotes pattern recognition, data classification and features selection in order to produce accurate and reliable results.

This data learning activity can be done via supervised learning and unsupervised learning. The supervised learning is the directed learning activities, in which the desired result is known and makes a continuous comparison value to the corrected output in order to find errors. This learning method uses patterns recognition to predict correct output. The unsupervised learning is a heterogeneous analysis that clusters data functions and implement data segmentation for features selection. This unsupervised learning makes predictive analysis becomes more scalable and high dimensionality that leads towards inter-correlation of the selected functions' features [2]. However, the unsupervised learning needs high and complex computational process due to the self-organizing and the nearest-neighbour mapping of the data. Figure 1 shows the pipeline of the learning process using the epidemic diseases' data.

Based on Figure 1, this computational learning method is the most well known process of finding the classification of data, while inbound the numbers of data features in order to determine data correlation. The classification of the data is then analyzed and the hypothesis of the data is made for the epidemic diseases outbreak prediction. The classification can be made based on propagation method or back propagation method. The propagation method is the normal methods, in which the classification can be done from the root to the tips. However, the back propagation is the backward process, in which the classification can be done from the tips to the root.

Machine learning process was introduced to predict the epidemic disease dissemination by cross validating the impact of the epidemic diseases' training data using backpropagation. The backpropagation method then identified several main factors that caused the dissemination of epidemic disease. The factors then were clustered and analyzed , in which the correlation of the classifiers had been detected.

III. PREDICTIVE ANALYSIS: EPIDEMIC DISEASE FACTORS TO PREDICT THE INFECTED AREA

Health predictive analysis is the new outbreak of advanced technology that can prevent the contagious epidemic disease [2, 10]. Figure 2 shows the relation of the predictive analysis and its role to predict the factors of epidemic disease dissemination.

Based on Figure 2, the classification of epidemic diseases dissemination can be categorised as physical network, geographical location, clinical studies and social media [2, 5, 7, 8]. The infectious disease dissemination is caused by the harmful virus attack. The viruses have the capability to replicate inside the cells of the host and slowly invade the host's cell ecosystem. However, the viruses have no ability to mobilize themselves from one host to another host without any help from the third party [6]. This mobilization is known as physical network. The physical network is the chain of the virus interchanges between the hosts. There are two factors that cause the physical network that are: 1) population density and 2) hotspot.

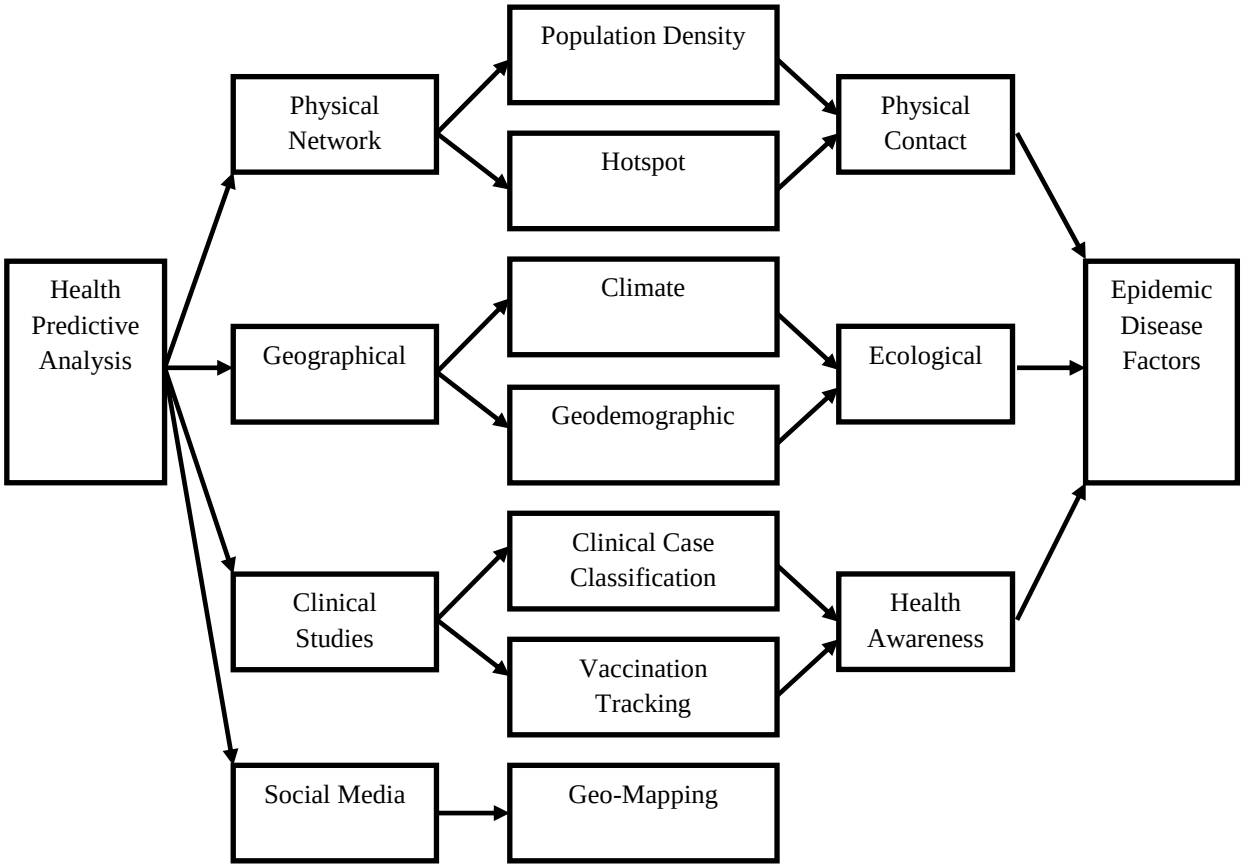

Figure 2: The correlation of health predictive analysis and epidemic disease dissemination factors

The population density is the percentage of human population over a specific area. The highest population density in a specific space leads to the higher potential occurrence of the epidemic disease breakouts. The migration, urbanization and productivity affect the health condition of the population. The high density of population is able to increase the potential of the infection's transmission.

The hotspot is the point of place, which has great human attraction that involves the high ingress and egress of human. The place of interest offers various kinds of activities such as tourist visit, amusement park, historical significant, business center, flight connecting point and etc. This point of attraction is the best factor for the rapid spreading of contagious epidemic disease because there are a lot of physical contacts that are possible to happen and become one of the best vulnerability factors in infectious disease dissemination.

The physical network that promotes the hosts' interchanges can be stimulated by the direct contact and the indirect contact of an object with the contagious agent. The direct contact of infection contributes to the network epidemiology via the vectors, such as human or animal that creates a social structure of contagion network that is peer-to-peer contacts such as the human-to-human contacts, human-to-animal contacts and etc. [6, 7]. This direct contact is the physical interchanges of the vector agents that continuously and diversely distribute the pathogen of the viruses via saliva, blood, body fluids and etc. [6, 7].

A virus, which mostly do not have a lifeform, is known as a molecule that consists of protein, such as DNA or RNA used to carry its genetic information. Hence, the molecule of viruses also can be spread via indirect contact through the natural distributor such as wind, air, water and etc. This indirect contact usually causes the viruses to penetrate the respiratory system or any first level filtration system of human body.

Natural distributor is one of the influenced factors in predicting the epidemic disease dissemination. This natural distributor has strong engagement with the geology factors [3, 4]. Based on Figure 2, the predictive analysis according to geographical location can be determined by two factors that are: 1) climate and 2) geodemographic.

Nowadays, the greenhouse effect had caused climate changes due to the deforestation and desertification. The rise in global temperature causes warm air that promotes the productivity of insect vectors and other air and water contamination factors [3, 4, 6]. Hence, it is highly anticipated that there will be a spread of some of the vector agents and natural-based vectored diseases, such as malaria and dengue [3, 4, 6].

It is believed that the occurrence of the greenhouse effect is due to man-made disasters. This particular factor is caused by geodemographic of one's population. Geodemographic is the clustering techniques in the classification of a community in a location based on their socio criteria. This classification can be done based on the sanitation level, family structure, neighborhood economic activities, education level and etc. [2, 4, 8]. The poor management of geodemographic features such as improper waste disposal, deforestation, and etc. cause the vulnerability of the ecology; hence, nurturing the epidemic disease outbreaks.

The epidemic disease outbreaks based on the physical network and geology can be overcome or reduced by the introduction of advanced health awareness. This advanced

health awareness is able to predict the epidemic disease dissemination based on the clustering method that is promoted by the clinical studies. The clinical studies on the epidemic disease can be done by conducting: 1) clinical case classification and 2) vaccination tracking.

The epidemic disease outbreaks are caused by heterogeneous factors that involve a vast area of study and tremendous data collection activities. One of the data collection activities that can be done is by using the Electronic Health Record (EHR). EHR is the patient's clinical records and it has already been implemented by almost all of the health organizations in the developed country for tracking, data storage and clustering purposes.

The clustering method of machine learning in clinical case classification can help the medical officer to describe, predict and control the epidemic disease outbreaks by collecting the EHRs and classifying data correlation based on the features selection [1, 2]. The clinical study is also able to promote the vaccination tracking used in finding the classification of the vaccine type and the most used vaccine in the infected area. Hence, the clinical studies play a significant role in health predictive analysis process that promotes machine learning method, which initially requires the training data for the backpropagation method for cross validation.

Nowadays, the predictive analysis of epidemic disease is enhanced by the usage of social media [2, 5, 7]. The social media has become the selected and important medium for information sharing in the twenty-first century. Due to the technology advancement, the social media is the metadata collection platform that is able to perform the geo-tagging, words' classification and personalization [2, 5]. The words' classification can be clustered with geo-tagging for a personalization of a user. This personalization can be analyzed and the infected area of the epidemic disease can be determined with fine granularity [5].

The new outbreaks of computer services for data segmentation based method are able to help data scientists in finding the classification of the epidemic diseases factors and will be able to predict the epidemic disease outbreaks faster than the classic clinical studies due to the live feeds of information from the users. This social media medium is able to predict the potential epidemic disease outbreak in a particular area.

The percentage of the infected people per area can be obtained by setting the area as a grid and setting for the targeted area as the cells of the grid. The percentage of infection disease outbreak in a day can be determined by Equation 1:

$$Infection\ \%\ in\ an\ Area\ = \frac{Negative\ Update\ Post\ in\ an\ i\ Area}{Total\ Numbers\ of\ Live\ Feed\ in\ i\ Area}\ x\ 100 = \frac{N_T}{L_T}\ x\ 100 \quad (1)$$

Where T is the number of day and i is the selected area to be analyzed.

The epidemic disease can be considered as contagious and heavily spread if the symptoms are remained constant or increase after 7 days of infection [5]. Based on the percentage of infection disease outbreak in a day, the percentage measurements using the geo-tagging and mapping can be monitored for at least the first 7 days. Hence, the prediction of the spreading of the contagious disease can be determined by:

$$Prediction\ of\ Contagious\ Disease\ = \frac{N_1}{L_1}\ \%,\quad \frac{N_2}{L_2}\ \%, \ldots\ldots\ldots, \frac{N_7}{L_7}\ \% \quad (2)$$

From the prediction of contagious disease, the contagious percentage of the disease can be predicted if the percentage value of the infection disease outbreak is constant or increases for over 7 days monitoring process. If there are some continuations on the infections, the selected area can be quarantined and the counter measure can be implemented based on this early prediction.

IV. Case Study: Dengue Outbreaks

Dengue is a break-bone fever caused by mosquito-borne viral infection. This virus is spread by Aedes mosquitoes. The Aedes mosquitoes have the same genus as the mosquitoes that spread the yellow fever viruses. The dengue virus generally takes charge on the human's immune system over three to fourteen days after the infection happens, which eventually results in dengue fever. The symptoms of dengue fever infection are headache, muscle and joint sore and pains, rashes on the skin and high fever. The recovery of dengue fever reaches over a week. However, dengue fever is life-threatening as the disease is able to evolve, causing hemorrhagic symptoms such as bleeding, irregular low in blood plasma and platelets. It can also cause low blood pressure, leading to fatality.

This epidemic disease management on dengue fever involves several features identified in three phases: pre-outbreaks, during outbreaks and post-outbreaks. The pre-outbreaks management on dengue fever involves the health awareness, the ecological environment and the physical contact. The dengue fever outbreaks can be controlled by administrating a vaccination for dengue virus to stimulate the development of immunity in an individual's immune system for resisting the infection by the virus' pathogen. This health awareness occasionally controls the dengue outbreaks and creates a more strong population in fighting the disease.

Aedes mosquitoes can be found living in the tropical and subtropical areas with high humidity together with warm temperature. The present wet and warm climate can help the Aedes mosquitoes to breed especially in stagnant water. Hence, a clean environment plays a great impact on preventing the Aedes mosquito from breeding. The home ground as well as any place that have stagnant water must be treated with larvicides or cleaned to avoid the breeding of Aedes' larvae. Hence, the geographical area plays a great impact in the dengue fever outbreaks. The climate and the geodemographic play the root-featured parameters in this disease outbreak. The climate is based on the nature of the designated area. However, the geodemographic of an area can be adjusted by educating the community to change their culture and the socio criteria to be cleaner and practicing the right sanitary level. The geodemographic issues can be overcome with the awareness of the community to enhance their quality of life.

During the pathogen transmission, the Aedes mosquitoes will transmit the dengue virus into any hosts that have suitable environment for the virus to breed. These mosquitoes are the vectors that transmit from one primates to other primates. However, dengue also can be spread by

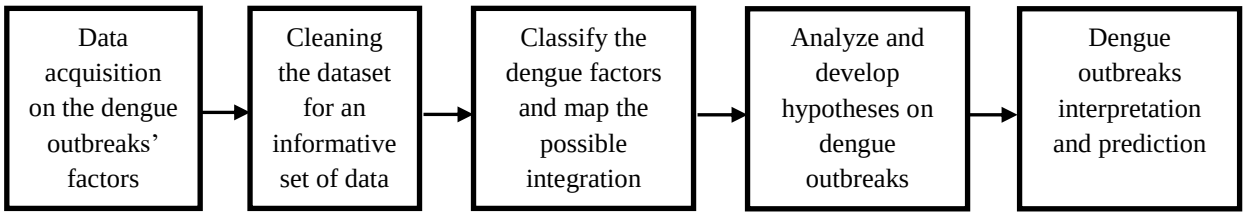

Figure 3: Methodology in using clustering method to predict the infected area of dengue outbreaks

blood transfusion, organ replacement and mother-infant vertical transmission. Hence, the physical contact of human-Aedes mosquitoes must be avoided and the process that involves organ transplant and blood transmission must go through a high screening process.

During the outbreak of dengue fever, there are some countermeasures that can be taken into consideration. The infected host should take the laboratory tests to diagnose the impact of dengue virus and to keep the EHR data. The infected host usually has a low blood plasma or blood platelet and requires treatment to increase the fluid's volume and antibodies. The patient is considered cure when the vital signs are stable and the body fluid reaches the right level.

The post-outbreaks management is the most important part of managing the outbreaks. The post management is able to help the right authorities to detect and predict the dengue outbreaks in the future. Hence, health informatics can be used to collect the information data from the pre- and during dengue outbreaks data and the data can be stored for EHR. This information acquisition and analysis can be classified as the data mining process and machine learning exploration on the pattern recognition and features classification for the clustering method. Recognizing the importance of data mining and machine learning, this research paper analyzed some features of the dengue fever outbreaks in Malaysia. Some analysis were made for classification of the infected area and the correlation and high contribution factors were highlighted, clustered and discussed for future predictive analysis on the dengue fever outbreaks.

V. ANALYSIS

Dengue outbreaks have taken a tremendous effect and become one of the most killer epidemic diseases in Malaysia. It is believed that the clustering method of the dataset from features classification is able to predict the factor parameters that cause the dengue virus to outbreaks; hence, some suggestions on the epidemic disease prevention management can be made. For this case study, a method to analyze a dengue fever was designed based on Figure 1 epidemic disease pipeline. Figure 3 shows the data pipeline design for this case study.

Figure 3 shows the data pipeline for features' clustering in order to find the integration of the features, while predicting and interpreting the data on the factors of dengue virus dissemination. Based on the pipeline in Figure 3, the first step of data mining process is the data acquisition. In this case study, the data were gathered by determining the parameters in a population related to human's health quality. The data were gathered from the data center provided by Malaysia government (http://www.data.gov.my/). One of recently completed data that can be determined in the data center was from the Ministry of Health Malaysia and the Ministry of Economic Planning and Development Malaysia. However, the available recorded dataset were from 2010 until 2015 only. Table 1 shows the number of death in Malaysia (2010 – 2015) and Table 2 shows the number of population in Malaysia (2010 – 2015).

Table 1
Number of deaths in Malaysia (2010 – 2015) due to the dengue fever outbreaks

	2010	2011	2012	2013	2014	2015
Johor	12	6	1	24	23	2
Kedah	2	0	4	1	6	0
Kelantan	11	0	0	2	15	1
Melaka	11	0	0	7	4	1
Negeri Sembilan	7	3	2	1	10	3
Pahang	4	1	0	1	3	2
Pulau Pinang	5	2	1	5	10	2
Perak	6	2	5	2	19	4
Perlis	0	0	1	1	0	0
Selangor	31	13	13	20	63	29
Terengganu	2	1	0	0	2	1
W.P. Kuala Lumpur/ Putrajaya	5	2	5	9	18	6
Sabah	2	2	2	5	3	2
Sarawak	11	0	1	4	3	1
W.P. Labuan	0	0	0	0	0	0

Table 2
Number of populations in Malaysia (2010 – 2015)

	2010	2011	2012	2013	2014	2015
Johor	3,362.90	3,401.80	3,439.60	3,477.20	3,515.30	3,553.60
Kedah	1,949.30	1,973.10	1,996.80	2,021.10	2,046.20	2,071.90
Kelantan	1,589.90	1,615.20	1,640.40	1,665.90	1,691.90	1,718.20
Melaka	823.6	833	842.5	852.4	862.5	872.9
Negeri Sembilan	1,029.50	1,042.90	1,056.30	1,070.10	1,084.10	1,098.40
Pahang	1,501.90	1,524.80	1,548.40	1,572.70	1,597.70	1,623.20
Pulau Pinang	1,575.90	1,593.60	1,611.10	1,628.40	1,645.70	1,663.00
Perak	2,379.00	2,397.60	2,416.70	2,436.40	2,456.70	2,477.70
Perlis	235.8	237.5	239.4	241.4	243.6	246
Selangor	5,502.10	5,577.40	5,650.80	5,725.30	5,800.10	5,874.10
Terengganu	1,055.40	1,074.00	1,092.90	1,112.50	1,132.70	1,153.50
W.P. Kuala Lumpur/ Putrajaya	1748.2	1770.9	1792.8	1814.5	1835.7	1856.3
Sabah	3,260.00	3,316.40	3,371.70	3,428.00	3,485.30	3,543.50
Sarawak	2,487.10	2,516.20	2,545.80	2,575.50	2,605.50	2,636.00
W.P. Labuan	88.2	89.8	91.6	93.3	95.1	96.8

Based on Table 1 and Table 2, the data were visualized as the heat map in Figure 4 and Figure 5. Figure 4 shows the heat map of the number of death due to dengue fever in Malaysia from 2010 to 2015 and Figure 5 shows the heat map of the number of population in Malaysia from 2010 to 2015.

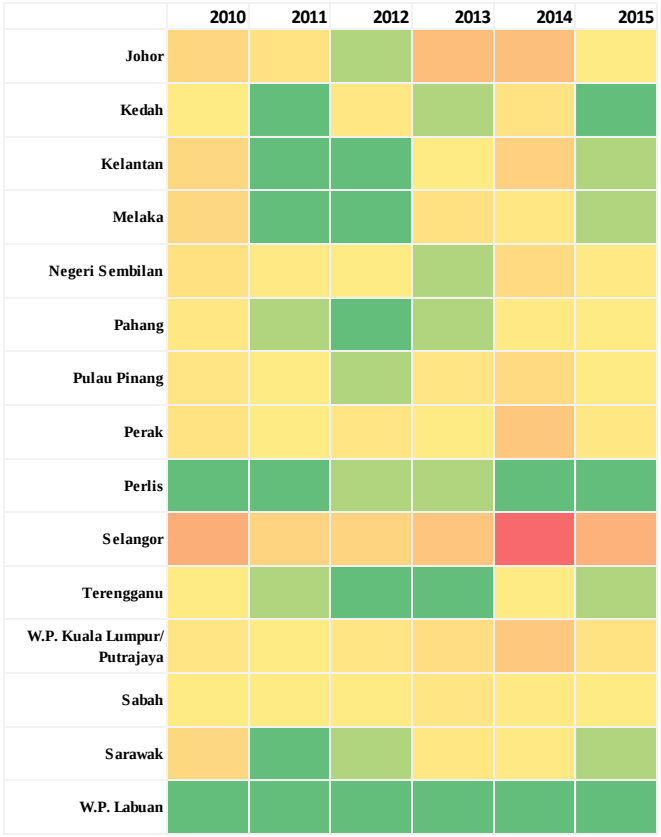

Figure 4: The heat map on number of deaths in Malaysia due to the outbreaks of dengue virus from 2010 until 2015

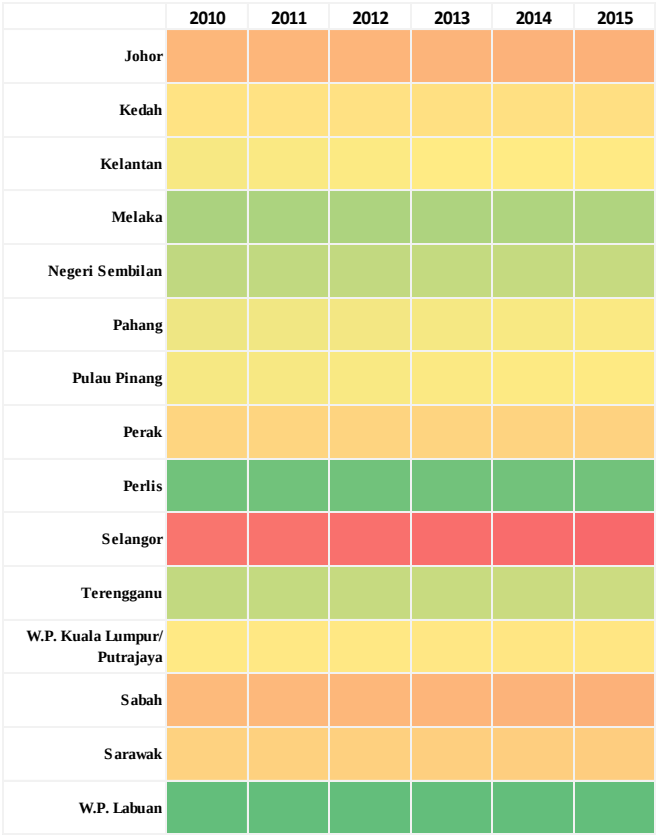

Figure 5: The heat map on number of population in Malaysia from 2010 until 2015

Based on Figure 4 and Figure 5, the heat maps show the results from red (maximum number) to light green (minimum number). There are 15 states (including federal territories) involved in this case study. Figure 4 shows that Malaysia was reaching towards high health quality country due to the lower number of death over the epidemic disease, especially dengue fever. The overall heat map shows that the dengue fever temperature was low (light green) in several states, which were formerly considered as the risky infection areas (yellow and orange) such as Kedah, Kelantan, Melaka, Terengganu and Sarawak. There were also states considered as the least infected areas; Perlis and W.P. Labuan. The higher risk states that had high percentage of dengue virus infected area were Johor, Perak, W.P. Kuala Lumpur/ Putrajaya, Sabah and the most "hottest" area for the dengue outbreaks was Selangor. Based on the correlation of the epidemic disease dissemination factors and the predictive analysis in Figure 2, at least one of the features suggested must be determined, and the Figure 5 shows the other heat map on the number of population in Malaysia from 2010 until 2015.

Based on Figure 5, the heat map shows the number of population in Malaysia and the population density of states in Malaysia can be determined from the heat map. Figure 5 shows that Johor, Perak, Selangor, Sabah and Sarawak had a great number of population while Kedah, Kelantan, Pahang, Pulau Pinang, W.P. Kuala Lumpur/ Putrajaya had average number of population and Melaka, Negeri Sembilan, Perlis, Terengganu and W.P. Labuan had minimum number of population.

Based on Figure 4 and Figure 5, the correlation of both data was verified and the features that contribute to the detection of the infected area were clustered. The clustering of features on both number of deaths data and number of

population can result in the percentage of the infection in an area from 2010 until 2015. The result can predict the spreading of the dengue fever virus based on Equation 1. Equation 3 shows the calculation of the percentage of infected area.

$$Infection\ \%\ in\ an\ Area\ = \frac{Number\ of\ Death\ per\ State}{Total\ Numbers\ of\ Population\ per\ State}\ x\ 100 \quad (3)$$

Figure 6 shows the heat map of the percentage of infection area of dengue virus.

Based on Figure 6, the heat map shows the infected states in Malaysia from 2010 until 2015. It shows that the dengue fever outbreaks can be predicted for future counter measure. In Figure 6, all of the states in Malaysia show the enhancement of health quality over dengue outbreaks. All of the states had improved the range in the heat map. However, several states can still improve and countermeasure the dengue virus dissemination factors. Further, the most infected areas, namely Selangor and W.P. Kuala Lumpur/ Putrajaya had shown some improvements in term of dengue outbreaks due to the decreasing number of deaths. As a summary from the clustering of the data from Figure 4 and Figure 5, the infected area can be related to the population density in Malaysia. Selangor and W.P. Kuala Lumpur/ Putrajaya have high population. W.P. Kuala Lumpur/ Putrajaya is the capital city of Malaysia, which has a lot of private companies and headquarters of all government sectors. Selangor, on the other hand, has a lot of business and trading activities as this state has the busiest port in Malaysia and is the main gateway by sea for traders. The urbanization of the place had caused a great impact on the health condition of the population in Selangor and W.P. Kuala Lumpur/ Putrajaya. These hotspots are the center of human attention that involve high ingress and egress activities with different culture and sanitary management. Hence, there are a lot of chances for the physical contacts to happen between the human-mosquitoes due to the lack of sanitary cleanliness and also due to the high density of population-centered at a place. These factors are able to promote the virus spreading, causing dengue outbreaks.

For other states in Figure 6, low percentage infected area were plotted such as Johor, Kedah, Kelantan, Perlis, Terengganu, Sabah, Sarawak and W.P. Labuan. These states have several factors that result in the low percentage of infection on dengue virus. Some of the states, such as Perlis is the smallest state in Malaysia. Hence, the population density in Perlis is not high that gives more harmony and clean condition for the residence of the states. However, some of the states that have low percentage of infection are the big states such as Sabah, Sarawak, Terengganu and Johor. These states have less dengue outbreaks due to high quality of cleanliness and the least number of attraction points that lead towards the low frequencies of ingress and egress.

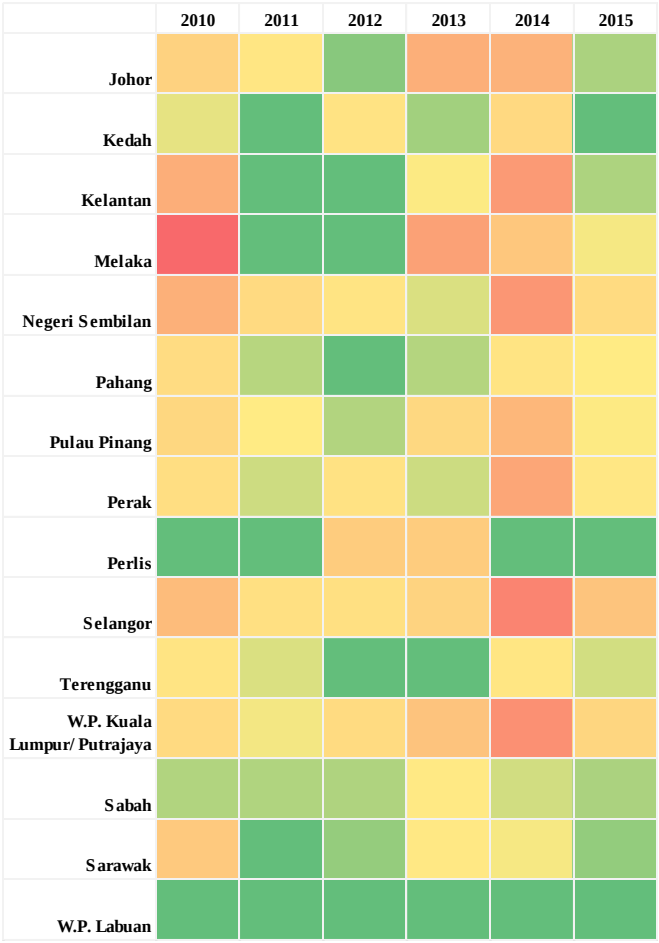

Figure 6: The heat map on the infected area in Malaysia due to the outbreaks of dengue virus from 2010 until 2015

VI. CONCLUSION

The epidemic disease is the most dangerous disease for the 21st Century, if the infectious disease is not managed despite of the modern medical treatment. The classical model of modern medical treatment has turned far side due to the epidemic disease dissemination factors such as the increasing of population density and the speedy outbreaks of new infectious diseases. Hence, due to the recent development of computer technology that promotes high compute intensive, the epidemic disease factors have become the selected features over the unsupervised learning on the data collected from EHR, social media record, and etc. for the predictive analysis purposes.

The output determined in this research paper will be used as the training data for the future output expectation determination. Based on the correlation of the health predictive analysis and epidemic disease dissemination factors, the backpropagation method had shown that the predictive analysis is able to predict the infected area of epidemic diseases outbreak. Based on the case study, Selangor and W.P. Kuala Lumpur/ Putrajaya were predicted as the most infected areas of dengue virus in Malaysia.

ACKNOWLEDGMENT

Research reported here is pursued under the Short Term Grant Scheme by Universiti Sains Malaysia for “Pedestrian Simulator and Heuristic Search Methods for Spatial Layout Design” [304/PKOMP/6313169] and supported by Fundamental Research Grant Scheme (FRGS) by Ministry

of Education Malaysia for "Spatial Layout Design Optimization with Pedestrian Simulation in a Panic Situation using Genetic Algorithm" [203.PKOMP.6711534].

REFERENCES

[1] Boivin, G., et al., *Predicting Influenza Infections during Epidemics with Use of a Clinical Case Definition.* Clinical Infectious Diseases, 2000. 31(5): p. 1166-1169.

[2] Ravì, D., et al., *Deep Learning for Health Informatics.* IEEE Journal of Biomedical and Health Informatics, 2017. 21(1): p. 4-21.

[3] Shope, R., *Global Climate Change and Infectious Diseases.* Environmental Health Perspectives, 1991. 96: p. 171-174.

[4] Shuman , E.K., *Global Climate Change and Infectious Diseases.* New England Journal of Medicine, 2010. 362(12): p. 1061-1063.

[5] Sadilek, A., H. Kautz, and V. Silenzio, *Predicting Disease Transmission from Geo-Tagged Micro-Blog Data*, in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*. 2012, AAAI Press: Toronto, Ontario, Canada. p. 136-142.

[6] Andrick, B., et al. *Infectious Disease and Climate Change: Detecting Contributing Factors and Predicting Future Outbreaks*. in *Geoscience and Remote Sensing, 1997. IGARSS '97. Remote Sensing - A Scientific Vision for Sustainable Development., 1997 IEEE International*. 1997.

[7] Masuda, N. and P. Holme, *Predicting and Controlling Infectious Disease Epidemics Using Temporal Networks.* F1000Prime Reports, 2013. 5: p. 6.

[8] Kimura, Y., et al., *Geodemographics Profiling of Influenza A and B Virus Infections in Community Neighborhoods in Japan.* BMC Infectious Diseases, 2011. 11(1): p. 36.

[9] Abas, A., *Five diphtheria deaths in Malaysia so far*, in *New Straits Times*. 2016, New Straits Times Press (M) Berhad (4485-H).