



Prediction Model of Leptospirosis Occurrence for Seremban (Malaysia) using Meteorological Data

Fariq Rahmat, Asnor Juraiza Ishak*, Zed Zulkafli, Hazlina Yahaya, Afiqah Masrani

Department of Electrical and Electronic Engineering, Faculty Engineering, University Putra Malaysia, Selangor, Malaysia

*Corresponding Author

DOI: <https://doi.org/10.30880/ijie.2019.11.04.007>

Received 29 April 2019; Accepted 8 July 2019; Available online 5 September 2019

Abstract: Leptospirosis is one of the waterborne diseases that widespread in Asia Pacific regions, especially developed countries. Over the past few years, the clinical data have shown Seremban experienced a significant number of leptospirosis patient. To minimize the impact of this disease, this study has set one objective which is to develop one prediction model to predict the leptospirosis diseases confirmed-cases by using Back-Propagation Neural Network (BPNN). A growing number of studies has shown the climate can be a predictor in outbreak incidence. Likewise, climate variable such as rainfall, temperature, and relative humidity affect in many ways especially for the transmission of vector and pathogens. Thus, these 3 parameters will be the main input for this model. Technically, this study will focus on the accuracy and the sensitivity of the model by finding the relationship between the meteorological data and clinical data. The clinical data was provided from the ministry of health Negeri Sembilan, while the meteorological data was provided from the Drainage and Irrigation Department and the Malaysian Meteorological Department. This study acknowledges that the amount of rainfall was correlated with the leptospirosis cases in all region of Seremban states such as Mantin, Seremban, Perentian, and Sikamat. In this study, preliminary exploration was performed by finding the best time for the meteorological data to correlate with clinical data (1 until 5-month lag). The model achieved 70% accuracy in prediction by combining the sum of rainfall, relative humidity, and temperature with 3-month lag as an input of the BPNN model. In conclusion, the authors believe this achievement of the model is an early stage for the prediction model. This model can achieve more than 70% accuracy by adapting some exploratory data analysis for every single variable or predictor.

Keywords: Leptospirosis prediction; Artificial Neural Network; Meteorological data; Prediction model

1. Introduction

Waterborne disease has a worldwide distribution and it was frequently happening in developing countries but rarely happen in developed countries [1]. The waterborne disease belongs to the top five common diseases that cause of death. While leptospirosis is one of the top killers of water-borne diseases because more than 500 000 cases were recorded every year [2]. Malaysia is one of the developing countries and is one of the countries that face this disaster [3]. Generally, human will be infected with leptospirosis when they have direct contacts with the product of infected animals such as urine. Besides, they also can be infected by the indirect way by contact with the contaminated water or soil which consist of the leptospira species. This disease also can be infected through human to human transmission but based on the number of cases it is very rare [4]. This pathogen can survive in tropical and subtropical environments

*Corresponding author: asnorji@upm.edu.my

2019 UTHM Publisher. All rights reserved.

penerbit.uthm.edu.my/ojs/index.php/ijie

which is warm and wet for long period [5]. Based on the previous study have proved this pathogenic able to survive in a robust condition such as low temperature and acidic for at least 20 months and remain harmful [6], [7]. *Leptospira* can be found in animals such as cats, rodents, dog, pigs and cattle [8].

There are a few factors that will cause the leptospirosis to happen. One of the factors is lack of awareness and lack of clinical diagnosis on this disease [9]. This is because most of the infected the patients have symptom-less and show undifferentiated fever with normal fever. In addition, endemic common in a rural area compared to an urban area due to the lack of infrastructure. Secondly, this disease is very sensitive to the weather. Heavy precipitation and flooding become the catalyst for this problem since it will bring the bacteria closer contact with humans. Flooding may carry the urine of infected animal to the area of human habitation [4].

Negeri Sembilan is a state in Malaysia which lies on the western coast of Peninsular Malaysia. The monsoon rains caused variability of rainfall distribution across Negeri Sembilan to form two significant features which are the area on the west of the highlands up to the coast indicates an increase in annual rainfall while the area on the east of the highlands (including Jelebu and Kuala Pilah) experienced the decreasing rate of rainfall. Thus, this state also has a quite high number of rural people based on Statistical Department of Malaysia due to the dry and wet environment [Appendix A].

Artificial neural network (ANN) is very widespread in its application on predicting any diseases. However, this model most popular in the prediction of vector-borne disease such as malaria and dengue [10], [11]. Thus, in this study very interested to test and analyses the accuracy of the ANN model to predict the occurrence of Leptospirosis disease.



Fig. 1 - Location of study sites

2. Methods

2.1 Predictive Modeling

Briefly, the prediction model is a tool that is used to combine several mathematical formulas to understand and appreciate the relationship between predictor data and predicted element. This tool will analyze the historical data and do prediction about the future. This analysis will be statistically done and interpreted the pattern of each predictor data. Besides that, the predictive model will help to find the association between each predictor as well as to get a better prediction. This model will be tested and validated very well before it was used to forecast the future event. Previous studies have proven the predictive model has managed to get a place and very effective in certain types of sectors, including marketing and merchandising, engineering, finance, psychology, medicine, and healthcare [12]. In the health field, it will be used for clustering the ranks of the group in which each group will represent the risk of the specific outcome [13]. Division of this rank can help health center to decide the level of severity some disease on a certain time and at once will help in controlling the healthcare cost. In other words, a prediction model can give early warning to the health center and communities to prepare for the endemic before it happens.

2.2 Clinical data

This dataset was collected from the Department of Health of Negeri Sembilan for leptospirosis disease. This dataset consists of all patient's medical record such as notification date, offset date, patient's date of treatment, symptom, diagnosis status and patient address. Regarding clinical data, from 2011 until 2017 Seremban is the highest

number of cases was recorded for Leptospirosis disease which is 564 cases. This is the main reason why authors choose Seremban as a study site. Besides that, during the cleaning process, we realize that Seremban state is the smaller number of missing data compared the other state. This process required this data to link with meteorological data. This is because our model was to be training to cannot handle the missing data. During the pre-processing stage, the data was divided into 4 regions which are Seremban city, Sikamat, Perentian and Mantin. All this region was divided according to the patient’s residence which we suspect as the last location before the patient got infected by the leptospira virus. We calculate the Euclidean distance for every single patient’s residence so that we can classify all these residences belong to which regions. As an improvement, we are suggesting using Thiessen Polygon as a process to divide all this patient’s residence into these four regions.

2.3 Meteorological data

This study uses the meteorological data provided by 2 different sources which are the Department of Irrigation and Drainage (DID) and Malaysia Meteorological Department (MMD). In this study, 3 types of climate variable were selected which is total rainfall, temperature, and relative humidity. Both sources have their own pros and cons. For example, MMD very helpful to provide the data for all kind of variable, but we are facing some problem with the location where the data was collected. MMD only can provide one station to collect all these 3 data. The location of this station is Seremban City. While, it differs from DID which only provide rainfall data, but they can provide this data at a different location. In other words, this department provides many stations to collect the rain reading data per day. 4 stations were selected which very close with 4 regions that have been divided during pre-processing clinical data. During the early stage, all this data was given in daily format. Based on a few previous studies, it almost impossible to predict accurately in daily the occurrence of disease. Thus, a few things must be done with this data before we proceed with data analysis. Authors were decided to convert the data from daily to weekly. Thus, they sum up the value of rainfall every 7 days while for temperature and humidity, they take the mean value for every 7 days. Since all this data are weekly, temperature data can be processed again to create another 2 variable which a maximum and minimum temperature for the given week. Thus, for the temperature variable, we successfully create 3 covariates as input for our model. As mentioned before 7% of raw meteorological data was missing. All this missing data may be erased as a solution to solve this problem, but it may also affect missing out on very important information which would interfere with the accuracy of this model. After doing a few analyses, authors have found rainfall in Seremban have a constant pattern for every year (from 2011 until 2017) or in other words, it was seasonally. Thus, it will help in handle missing data problem. Linear regression was applied to predict both of this variable. Authors also found temperature and rainfall have a strong correlation by doing the linear regression analysis. For example, if rainfall data is to be predicted, the temperature will become as a predictor (x-axis) while vice versa if temperature data need to be predicted as shown in Figure 2. This regression was considered as successful because the error percentage is below than 20% as shown in Figure 3.

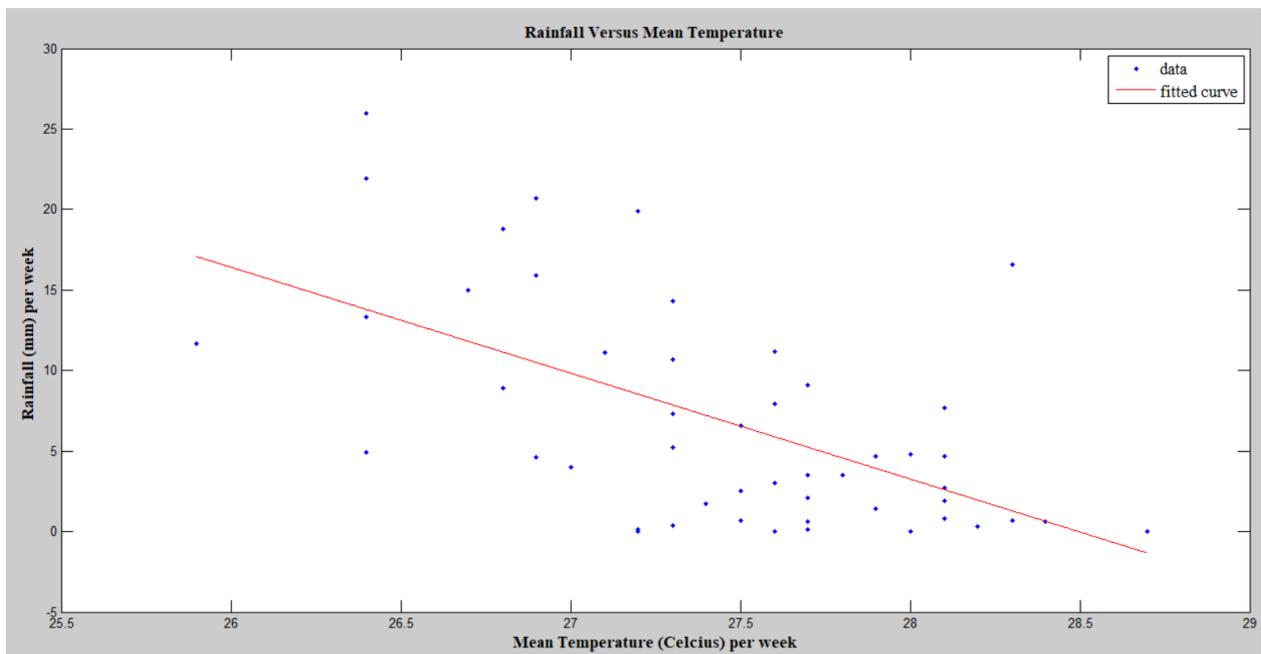


Fig. 2 - Linear regression for rainfall prediction.

```

Linear regression model:
y ~ 1 + x1

Estimated Coefficients:

```

	Estimate	SE	tStat	pValue
(Intercept)	178.1	31.787	5.603	8.5275e-07
x1	-6.233	1.1674	-5.3394	2.1749e-06

```

Number of observations: 53, Error degrees of freedom: 51
Root Mean Squared Error: 8.13
R-squared: 0.359, Adjusted R-Squared 0.346
F-statistic vs. constant model: 28.5, p-value = 2.17e-06

```

Fig. 3 - Result for linear regression for rainfall prediction.

$$RH = \frac{100 \left[e^{1.8096 + \left(\frac{17.2694 + T_w}{237.3 + T_w} \right)} - 7.866 \times 10^{-4} P(T - T_w) \left(1 + \frac{T_w}{610} \right) \right]}{e^{1.8096 + \left(\frac{17.2694 + T}{237.3 + T} \right)}} \quad (1)$$

For humidity data, the time series model (regression) cannot be implemented to handle missing data due to irregular pattern and it required implementation of calculation. This calculation will be related to precipitation and the temperature for the same day. However, the equation as shown in (1) cannot be applied because of a few variable values cannot be identified. The authors did some checks on the humidity data and they found out most of these data have a similar value with the humidity readings at 6 weeks ago. Finally, they decide to replace the missing value with the humidity value at the lag of 6 weeks. Mean Square Error (MSE) was calculated and we realize this method very helpful to implement when we are missing a few attributes for relative humidity data.

2.4 Preliminary Exploration

Since this model was developed to monthly forecasts for 365 weeks, we managed to collect the 365 attributes based on these 3 variables. This stage has been done in a pre-processing step. It required us to study every single variable to select the best input for our model. This exploration will help to get better accuracy in our prediction. The objective for this preliminary study is to find the best lag time for each independent variable which can help the distribution of leptospirosis disease. We decide to use Partial autocorrelation function (PACF) and Autocorrelation function (ACF) analysis for each variable. ACF was used to find the correlation between a number of a series of numbers arranged in time. In other words, the correlation between each number with different lag. While PACF little bit different with ACF which help to find the correlation between time series and each of its intermediate lagged values. Most of the time, ACF and PACF are used to identify the number of orders for the ARMA model or ARIMA model [14]. However, in this study used the application of PACF and ACF to see which lag have better correlation for this prediction model.

Before ACF and PACF analysis were started, one thing must be done which is make sure the data is stationarity. In other words, the data must have constant mean and variance over time and does not have trends. Rainfall and temperature data showed both data is stationary while humidity data showed a decreasing trend over time. Figure 4 shows the PACF and ACF graph for rainfall data from Seremban City region respectively. Both Figures shows there is not much difference between the PACF and ACF graph. Thus, we decide to use the PACF graph as our reference for Seremban City. This graph showed there is 2 significant spike which is at lag 2 and lag 4 weeks. This graph also shows there is one negative spike at lag 11. A positive correlation can be considered as the future observation might be the same state with the previous value, while a negative value is vice versa. For Example, the likelihood of this month for the disease to occur if last month (lag 4 weeks) have experiencing rainy day more than 150 mm per month. Then the same procedure happens on the temperature data. Based on Figure 5, it was clearly showing the significant spike

happen during 1-week lag. However, this method cannot be applied to humidity data because of this data not stationary. This might give issues because we cannot identify the best time lag which will affect the accuracy of the predictive model. Thus, we decide to use humidity data in the same week (No lag) as a third input for the predictive model. We choose humidity data without lag time series because based on one previous study have mentioned the effect of humidity in the distribution of water-borne disease [15]. In this study, the author just found out by increasing 10% in humidity, the malaria cases increased by 10.4% in the same week. Finally, we repeat the same manner to gain the best time lag for rainfall data on 3 other regions. Based on this analysis, we got the significant spike to happen between lag with 1 week until 23 weeks or from 1-month lag until 6-month lag approximately. Based on Figure 4 and 5 which represent the graph for PACF and ACF, there is one very significant spike at no lag (lag 0) compared with others time lag. This spike can be ignored since it does not give any meaning for this analysis.

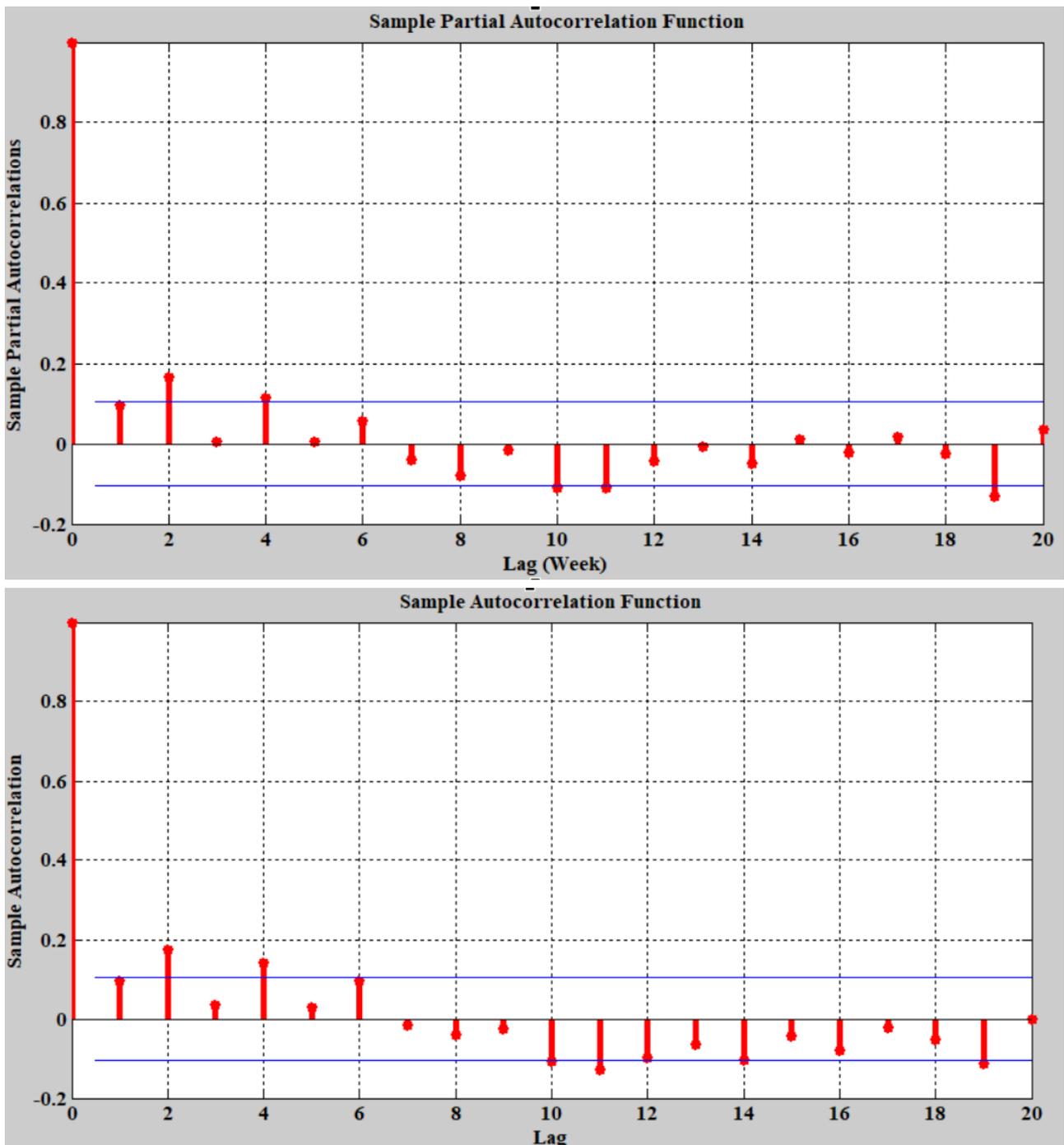


Fig. 4 - PACF (above) and ACF (below) graph for rainfall data on Seremban City region.

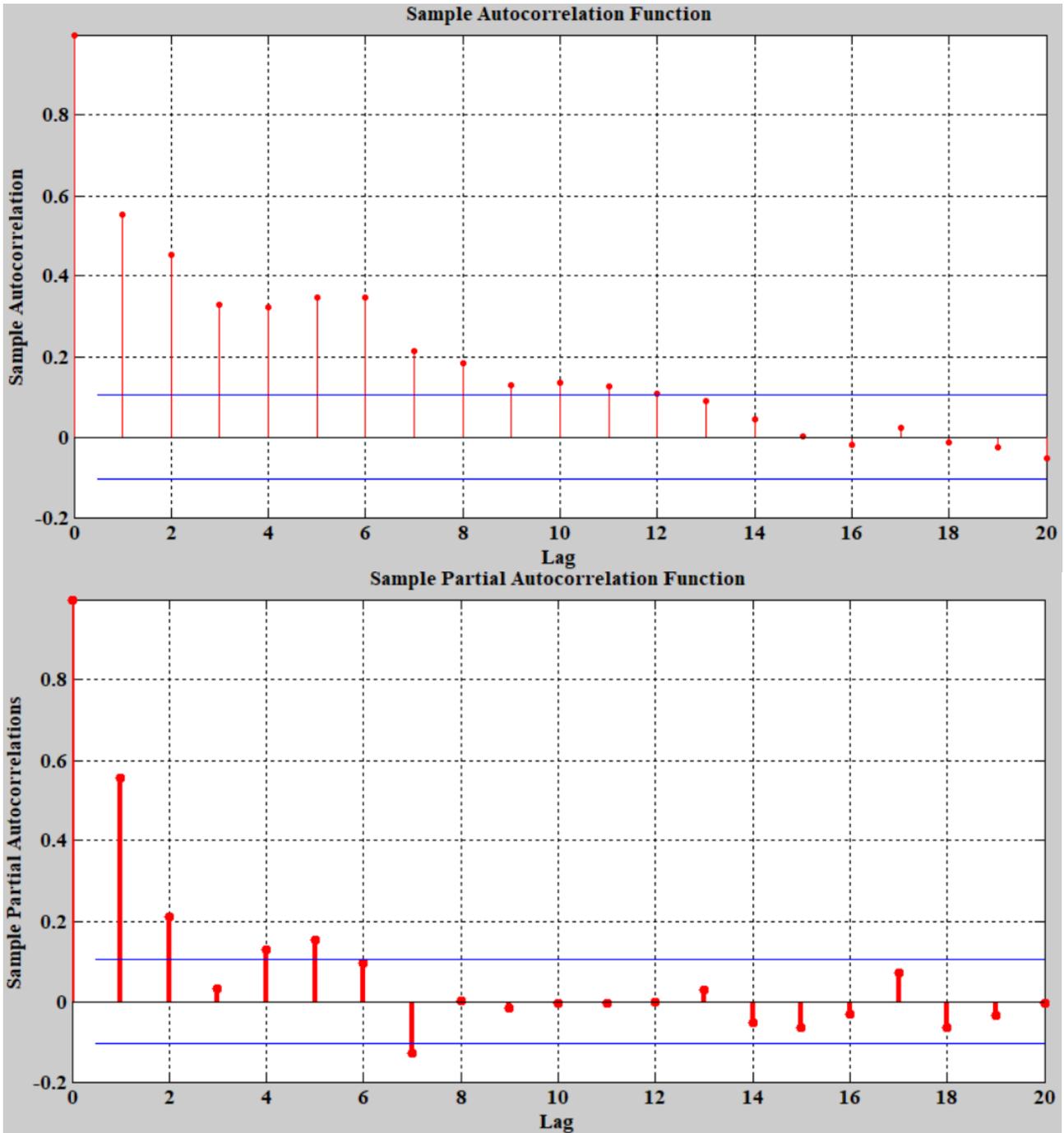


Fig. 5 - ACF (above) and PACF (below) graph for temperature data on Seremban City region.

2.5 Artificial Neural Network

To the best knowledge of authors, there is a lot of application of Artificial Neural Network (ANN) such as forecasting, data mining, traffic, image processing and so on. There are various types of ANNs including, feedforward neural network, backpropagation neural network, radial basis function neural network, recurrent neural network, neuro-fuzzy and many other types [16]. In this study, authors have applied the same model of the neural network in recent work which used to predict the abundance of mosquito by using simple backpropagation method [17]. It contains 3 important element which is input, hidden neuron, and output nodes. The output nodes can be more than one depending on the situation and cases that want to be predicted. This model consists 9 input data including mean temperature, maximum temperature, minimum temperature, mean of relative humidity, and 5 datasets for the sum of rainfall from a different location (2 datasets fro Seremban city, 1 from sikamat, perentian and mantin). Authors decide to use only one

output nodes with the output range is 0 until 1 which represent disease not occur (output value approach to 0) and disease occur (output value approach to 1). This model also applied 10 neurons for the hidden layer. In each hidden neuron, there is one activation function which will help the model to process the information from input and bring to the output nodes. The sigmoid function was selected to become an activation function due to their ability to handle the complex data and very suitable to be implemented in the classification process. In order to predict disease occurrence without noise and overfitting, we decide to divide the data into 3 group which is training data, testing data and validate data. Based on the previous study, there is 2 way to validation technique which is a percentage split and cross-validation [18]. Then we decide to use percentage split which 70% from dataset will be applied as training, 15% for training and another 15% for validation.

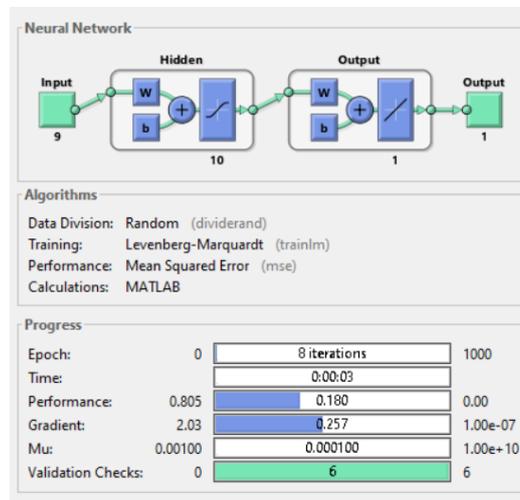


Fig. 6 - Visualization of back propagation neural network by using the MATLAB tool.

2.6 Measurement

Regarding the previous study in [19], the evaluation process has been done by using the root mean square error (RMSE) and error percentage. This method is very practical if and only the objective of the study wants to predict the number of disease cases. This is because it can measure how big the difference between observation and prediction reading. However, in this study, we more focussing to predict the occurrence of the disease. Thus, the output must be only into 2 class which is Yes and No or 1 and 0 (binary output). The application of RMSE in this situation not so practical because the aim of this study is to know how many weeks or month that this model successfully forecast the occurrence of the disease. Then authors decide to use Confusion Matrix to evaluate the accuracy, sensitivity and specificity. Table 1 shows an example of the confusion matrix or contingency table. Accuracy is the value of percentage of successful for the model to predict correctly either the event (disease) occurred or not occurred (Accuracy = $[h + q] / [h + f + m + q]$) in a given week.

Sensitivity is a measured value which represents the effectiveness of the prediction model to predict correctly the week with the disease among another confirmed week with the disease was reported. (Sensitivity = $h / [h + m]$). It also can be known as the Probability of Detection (POD). Lastly, specificity is the measured value for successfully predict a week where does not have the leptospirosis disease reported (specificity = $q / [q + f]$) [20].

Table 1 - The contingency table for the yes-or-no forecast

		Event occurred?	
		Yes	No
Event Forecast?	Yes	Hit (h)	False Alarm (f)
	No	Miss (m)	Correct rejection (q)

3. Result and Discussion

All meteorological data such as mean temperature, maximum temperature, minimum temperature, relative humidity and the sum of rainfall became the input of BPNN model while the clinical data which is number confirmed

leptospirosis cases became the target of the model. The meteorological data were collected with 1 month until 5 months before the week of cases were reported. As previously stated, this study wants to predict the occurrence of disease, this analysis is very dedicated to verifying the probability of detection (POD) and the fraction of positive forecast that turned out to be incorrect (false alarm rate). Besides that in this part, the author will discuss based on the result that has been collected. The higher the number of POD, there is high chance to predict the disease to occur, while the lower false alarm rate, the possibility the model to give a wrong or incorrect prediction. Besides that, the value of accuracy is also very important to verify either this is good enough or not. Based on Figure 7, meteorological data with 3-month lag is the best input to forecast the occurrence of leptospirosis because the POD is optimum which is 86% and the rate of false alarm is 0.3 which lower than another month. That means the model will achieve 70% accuracy to predict the occurrence of disease. This study showed lag 3 months give enough time for *Leptospira* virus distribute among this area and able to survive for a longer period. Besides that, this result can satisfy preliminary study, which also showed the lag with 11 weeks (3 months) have significant spike which mean have very good correlation and good to be a predictor. Regarding this study, we have proved the selection of lag time for meteorological data very important to help this variable to deliver the correct information. We can see in table 1, lag 4 week will help the model to have very high sensitivity because this model successfully predicts individual with the disease at 206 is the highest compared with another model at a different lag. However, this model cannot be the best model because of this model very less specificity and very high false alarm rate. In other words, this model very hard to predict the day or week not have leptospirosis. Overall, this model is very suitable to predict when the disease will occur regarding sensitivity value. But if we refer to specificity value, we might not suggest this model predict the day where the disease not occur.

After doing some analysis, there are a few reasons why this model not so specificity and become very sensitive. First reason, this is because of a characteristic of leptospirosis disease itself. This disease is very sensitive to hygiene status [21]. Thus, by using meteorological data only cannot be satisfied enough in the prediction of leptospirosis disease. Besides that, to improve the accuracy of this model, the authors suggest adding another input which can explain human activity. But this feature might be hard to implement because the researcher needs to do some additional work to study danger level of work types and human activity which can trigger disease infection. The second reason is the meteorological data. We can say that, the meteorological data in Seremban very easy to trigger and very sensitive to this disease to occur. The rainfall and temperature contribute to each other in the distribution of this disease. For example, this model senses the disease occurs during the rainy day and no rain. So, it might give a challenge to this model to distinguish the same condition when the day or week the disease not occur. Since this study is using ANN which has the ability to find the non-linearity of the data, this model can interpret and find the difference between the condition of disease occurrence and not occur by correlating each of input and starts to tune the weight of the input nodes. Besides that, geographical types in a certain area also can become the effect of the occurrence of this disease. If the residence is located on high ground, it has a high chance to this place not infected by leptospirosis disease because this place will not be flooded. While this place at lower ground, possibility to flood is very high and at the same time will invite the disease to distribute. Thus, if include all this additional covariate, we might improve the specificity and optimize the accuracy of the model.

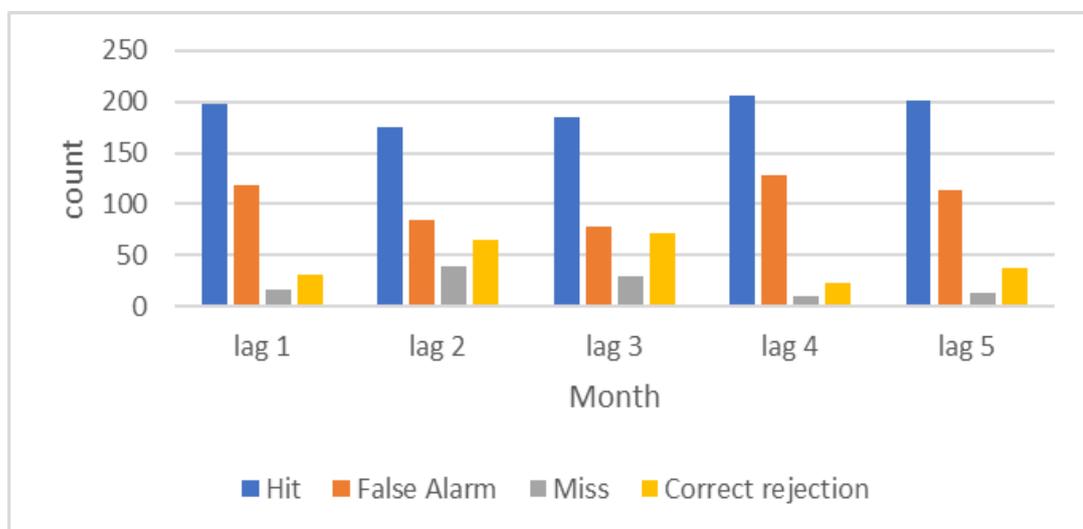


Fig. 7 - The number of counts for 4 different types of possibilities with 1 until 5-month lag.

Table 2 - Accuracy, Sensitivity and Specificity value for each month lag

Lag Time (month)	Accuracy (%)	Sensitivity (%)	Specificity (%)
Lag 1	62.74	92	20.66
Lag 2	66	81.86	43.33
Lag 3	70.41	86	48
Lag 4	62.46	95.81	14.67
Lag 5	65.48	94	24.67

4. Conclusion

In conclusion, this paper employed the BPNN algorithm as the model to predict the occurrence of leptospirosis. This study was different from other the previous study which using the same model to predict another type of vector-borne diseases such as malaria and dengue. Besides, the authors were using a different method and different topic in the disease prediction. Based on the previous study, they were using the ANN to predict the abundance of the diseases by predicting the number of cases while in this paper, the author decides to use yes-or-no forecast to predict the occurrence of leptospirosis. Regarding this paper, the author only achieves 70% accuracy to predict the occurrence of leptospirosis disease. The author acknowledges that to predict when the disease will occur is more challenging compared to predict the number of cases every week. In the future, the study will focus to improvise the accuracy of the model by doing the exploratory data analysis on each meteorological data.

Acknowledgement

This research project is supported by Putra Grant-Putra Graduate Initiative (IPS), Universiti Putra Malaysia. Appreciation is extended toward Newton Fund which one of the agencies that have supplied some expense to buy the meteorological data from Malaysian Meteorological Department (MetMalaysia).

Appendix A: Population number of rural and urban area for every state in Malaysia until

Jadual 1.3 : Taburan penduduk Malaysia mengikut negeri dan strata, 2016⁵

Negeri	Bandar ('000)	% Bandar	Luar bandar ('000)	% Luar bandar	Jumlah ('000)	Jumlah %
Johor	2,787.8	76.3	867.3	23.7	3,655.1	100.0
Kedah	1,455.2	68.6	665.5	31.4	2,120.7	100.0
Kelantan	860.9	47.9	936.3	52.1	1,797.2	100.0
Melaka	836.1	92.7	65.6	7.3	901.7	100.0
N. Sembilan	803.1	73.0	296.6	27.0	1,099.7	100.0
Pahang	912.1	56.0	716.0	44.0	1,628.1	100.0
P. Pinang	1,623.2	94.4	96.1	5.6	1,719.3	100.0
Perak	1,868.6	75.3	614.4	24.7	2,483.0	100.0
Perlis	154.5	61.5	96.6	38.5	251.0	100.0
Selangor	5,875.7	93.3	422.7	6.7	6,298.4	100.0
Terengganu	761.2	64.3	422.7	35.7	1,183.9	100.0
W.P. Kuala Lumpur	1,787.2	100.0	0.0	0.0	1,787.2	100.0
W.P. Putrajaya	83.3	100.0	0.0	0.0	83.3	100.0
Sabah	2,217.9	58.2	1,595.4	41.8	3,813.2	100.0
Sarawak	1,584.7	57.8	1,156.2	42.2	2,741.0	100.0
W.P. Labuan	83.1	84.9	14.7	15.1	97.8	100.0
Malaysia	23,694.7		7,966.0		31,660.7	

References

- [1] A. D. G. R. J. B. R. C. D. C. G. C. R. F. V. G. J.A. Cotruva, "identification, causes and control," in *Waterborne zoonoses*, World Health Organisation, IWA, 2004.
- [2] Fraga TR, Barbosa AS, Isaac L. Leptospirosis: Aspects of innate immunity, immunopathogenesis and immune evasion from the complement system. *Scand J Immunol* 2011; 73(5): 408-419.
- [3] S. G. J. Wright, "Household characteristics associated with home water treatment: an analysis of the Egyptian demographic and health survey," *Journal of Water and Health*, vol. 7(1), pp. 21-29, 2009.
- [4] Chadsuthi, S., Modchang, C., Lenbury, Y., Iamsirithaworn, S., & Triampo, W. (2012). Modeling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time-series and ARIMAX analyses. *Asian Pacific journal of tropical medicine*, 5(7), 539-546.
- [5] Ridzlan, F. R., Bahaman, A. R., Khairani-Bejo, S., & Mutalib, A. R. (2010). Detection of pathogenic *Leptospira* from selected environment in Kelantan and Terengganu, Malaysia. *Trop Biomed*, 27(3), 632-638.
- [6] Andre-Fontaine, G., Aviat, F., & Thorin, C. (2015). Waterborne leptospirosis: survival and preservation of the virulence of pathogenic *Leptospira* spp. in fresh water. *Current microbiology*, 71(1), 136-142.
- [7] Trueba, G., Zapata, S., Madrid, K., Cullen, P., & Haake, D. (2004). Cell aggregation: a mechanism of pathogenic *Leptospira* to survive in fresh water. *International microbiology*, 7(1), 35-40.
- [8] Mgode, G. F., Machang'u, R. S., Mhamphi, G. G., Katakweba, A., Mulungu, L. S., Durnez, L., ... & Belmain, S. R. (2015). *Leptospira* serovars for diagnosis of leptospirosis in humans and animals in Africa: common *Leptospira* isolates and reservoir hosts. *PLoS neglected tropical diseases*, 9(12), e0004251.
- [9] Bharti, A. R., Nally, J. E., Ricaldi, J. N., Matthias, M. A., Diaz, M. M., Lovett, M. A., ... & Vinetz, J. M. (2003). Leptospirosis: a zoonotic disease of global importance. *The Lancet infectious diseases*, 3(12), 757-771.
- [10] Thakur, S., & Dharavath, R. (2018). Artificial neural network based prediction of malaria abundances using big data: A knowledge capturing approach. *Clinical Epidemiology and Global Health*.
- [11] Aburas, H. M., Cetiner, B. G., & Sari, M. (2010). Dengue confirmed-cases prediction: A neural network model. *Expert Systems with Applications*, 37(6), 4256-4260.
- [12] Warner, B., & Misra, M. (1996). Understanding neural networks as statistical tools. *The American Statistician*, 50(4), 284-293.
- [13] Cousins, M. S., Shickle, L. M., & Bander, J. A. (2002). An introduction to predictive modeling for disease management risk stratification. *Disease Management*, 5(3), 157-167.
- [14] Chadsuthi, S., Modchang, C., Lenbury, Y., Iamsirithaworn, S., & Triampo, W. (2012). Modeling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time-series and ARIMAX analyses. *Asian Pacific journal of tropical medicine*, 5(7), 539-546.
- [15] Mohammadkhani, M., Khanjani, N., Bakhtiari, B., & Sheikhzadeh, K. (2016). The relation between climatic factors and malaria incidence in Kerman, South East of Iran. *Parasite epidemiology and control*, 1(3), 205-210.
- [16] Krenker, A., Bešter, J., & Kos, A. (2011). Introduction to the artificial neural networks. *Artificial Neural Networks: Methodological Advances and Biomedical Applications*. InTech, 1-18.
- [17] Lee, K. Y., Chung, N., & Hwang, S. (2016). Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas. *Ecological informatics*, 36, 172-180.
- [18] Nery Jr, N. R. R., Claro, D. B., & Lindow, J. C. (2017). Prediction of leptospirosis cases using classification algorithms. *IET Software*, 11(3), 93-99.
- [19] Thakur, S., & Dharavath, R. (2018). Artificial neural network-based prediction of malaria abundances using big data: a knowledge capturing approach. *Clin. Epidemiol. Glob. Health*.
- [20] Lalkhen, A. G., & McCluskey, A. (2008). Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care & Pain*, 8(6), 221-223.
- [21] Wasinski, B., & Dutkiewicz, J. (2013). Leptospirosis-current risk factors connected with human activity and the environment. *Annals of Agricultural and Environmental Medicine*, 20(2).