UNIVERSITI PUTRA MALAYSIA

FEATURE SELECTION METHODS BASED ON METEOROLOGICAL DATA FOR PREDICTION OF LEPTOSPIROSIS OCCURRENCE IN SEREMBAN, MALAYSIA

MOHAMAD FARIQ BIN RAHMAT

FK 2020 113

FEATURE SELECTION METHODS BASED ON METEOROLOGICAL DATA FOR PREDICTION OF LEPTOSPIROSIS OCCURRENCE IN SEREMBAN, MALAYSIA

By

MOHAMAD FARIQ BIN RAHMAT

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of the Master of Science

November 2019

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

FEATURE SELECTION METHODS BASED ON METEOROLOGICAL DATA FOR PREDICTION OF LEPTOSPIROSIS OCCURRENCE IN SEREMBAN, MALAYSIA

By

MOHAMAD FARIQ BIN RAHMAT

November 2019

Chair : Asnor Juraiza Bt. Ishak, PhD
Faculty : Engineering

The use of predictive model is useful for preventing and controlling disease out-break. This can be done by analysing weather behavior in relation to disease occurrence. In Malaysia, leptospirosis disease is the one of the higher number of cases that reported for past 7 years, and the absence of understanding and modelling studies that allows development of an early warning system. In this study, predictive model is developed using machine learning to capture the relation between weather variables such as temperature, sum of rainfall, and relative humidity, and Leptospira occurrence. The aim of this study is to predict the occurrence of Leptospirosis in Seremban district using a machine learning and meteorological data as input. The first objective of the study is to investigate the best time lags for each weather variable using feature selection methods. The second objective is to develop, train and test a neural network model for disease prediction based on the selected features. Feature selection was conducted using two methods: firstly, though correlation analysis, and secondly through graphical and non-graphical Exploratory Data Analysis (EDA). The neural network model is developed using Backpropagation training, optimizing the number of hidden layers and hidden nodes. The success is measured using accuracy, sensitivity, and specificity of the model. Correlation analysis has shown that Seremban district has higher correlation with disease occurrence when sum of rainfall at lag 4 until 16 weeks and temperature at lag 1 week, while by using EDA has shown Seremban can have high correlation with leptospirosis occurrence when the temperature at lag 16 weeks and sum of rainfall at lag 12 until 20 weeks. This study also shown the predictive model can achieve high accuracy between 80% to 84% when the input variables were following the feature selection that have been made by EDA and the number of hidden neurons is 10. In conclusion, this study is able to show

the trend of the environmental variable in predicting the leptospirosis occurrence at different time lag. Besides, by having this predictive model, it helps the public health not only to predict the occurrence of the disease, but it can prevent from the outbreak start to spread to the community by giving the early warning based on the weather status in future.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains

KAEDAH PEMILIHAN CIRI BERDASARKAN DATA METEOROLOGI DALAM MERAMAL LEPTOSPIROSIS DI SEREMBAN, MALAYSIA

Oleh

MOHAMAD FARIQ BIN RAHMAT

November 2019

Pengerusi : Asnor Juraiza Bt. Ishak, PhD
Fakulti : Kejuruteraan

Penggunaan model ramalan berguna untuk mencegah dan mengawal wabak penyakit. Ini boleh dilakukan dengan menganalisis perubahan dan keadaan cuaca yang berkait rapat dengan kejadian penyakit. Di Malaysia, penyakit leptospirosis adalah salah satu daripada penyakit yang mempunyai bilangan kes tertinggi yang dilaporkan selama 7 tahun yang lalu, dan ketiadaan pemahaman dan kajian pemodelan terhadap penyakit ini telah mendorong penciptaan sistem amaran awal. Dalam kajian ini, model ramalan dibangunkan menggunakan pembelajaran mesin (machine learning) untuk mencari hubungan diantara pembolehubah cuaca seperti suhu, jumlah hujan, dan kelembapan relatif, dan kejadian Leptospirosis. Tujuan kajian ini adalah untuk meramalkan berlakunya Leptospirosis di daerah Seremban menggunakan data pembelajaran mesin dan meteorologi sebagai input. Objektif pertama kajian ini adalah untuk menyiasat tempoh masa terbaik bagi setiap pembolehubah cuaca menggunakan kaedah pemilihan ciri. Objektif kedua adalah untuk membangun, melatih dan menguji model rangkaian neural untuk ramalan penyakit berdasarkan ciri-ciri yang dipilih. Pemilihan ciri dijalankan menggunakan dua kaedah: pertama, analisis korelasi, dan kedua melalui Analisis Data Eksplorasi grafik dan bukan grafik (EDA). Model rangkaian saraf dibangunkan menggunakan latihan Backpropagation, mengoptimumkan jumlah lapisan tersembunyi dan simpul tersembunyi. Kejayaan diukur menggunakan ketepatan, kepekaan dan kekhususan model. Analisis korelasi menunjukkan bahawa daerah Seremban mempunyai korelasi yang lebih tinggi dengan kejadian penyakit apabila jumlah hujan pada 4 hingga 16 minggu sebelum kejadian leptospirosis,manakala suhu pada 1 minggu sebelum kejadian, sedangkan dengan menggunakan EDA menunjukkan Seremban dapat mempunyai korelasi tinggi dengan kejadian leptospirosis ketika suhu pada 16 minggu sebelumnya dan jumlah hujan pada 12 hingga 20 minggu sebelum kejadian penyakit. Kajian ini juga menunjukkan model ramalan dapat mencapai ketepatan yang tinggi antara 80% hingga 84% apabila pembolehubah input mengikuti pemilihan ciri yang telah dibuat oleh EDA dan bilangan neuron tersembunyi adalah 10. Kesimpulannya, kajian ini mampu untuk menunjukkan

corak pembolehubah cuaca dalam meramalkan kejadian leptospirosis pada waktu yang berbeza. Selain itu, dengan menggunakan model ramalan ini, ia bukan sahaja dapat membantu pusat kesihatan untuk meramalkan berlakunya penyakit itu, tetapi ia dapat mencegah daripada wabak mula merebak didalam masyarakat setempat dengan cara memberi amaran awal berdasarkan status cuaca pada masa akan datang.

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor Dr Asnor Juraiza Ishak from Department of Electric and Electronic Engineering of the Faculty of Engineering at Universiti Putra Malaysia. Her office door was always open whenever I ran into a trouble spot or had a question about my research or writing. She consistently allowed this paper to be my own work but steered me in the right the direction whenever she thought I needed it.

I would also like to thank the experts who were involved in the validation of clinical data for this research project: Dr Hazlina Yahaya and Dr Afiqah Masrani from Department of Health Negeri Sembilan. Without their passionate participation and input, the validation process could not have been successfully conducted.

I would also like to acknowledge Dr Zed Diyana Zulkafli from Department of Civil Engineering of the Faculty of Engineering at Universiti Putra Malaysia as the second reader of this thesis, and I am gratefully indebted to her for her very valuable comments on this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EDA	Exploratory Data Analysis
ACF	Auto-Correlation Function
PACF	Passive Auto Correlation Function
ARIMAX	Autoregressive Integrated Moving Average with eXplanatory
DoH	Department of Health
DID	Department of Irrigation and Drainage
MMD	Malaysian Meteorological Department
PDF	Probability Distribution Function
NBR	Negative Binomial Regression
RMSE	Root Mean Squared Error
MSE	Mean Squared Error
ROC	Receiver Operating Characteristic
MA	Moving Average

CHAPTER 1

INTRODUCTION

1.1 Introduction

Waterborne disease has a worldwide distribution and it was frequently happens in developing countries causing human suffering (Cotruvo et al., 2004). In 2009, 4 billion cases of diarrhoea were reported that brought 1.6 million to a death and 62.5 million Disability Adjusted Life Years (DALYs) (Wright and Gundry, 2009). Waterborne disease belongs to top five common disease that causes of death. Malaysia is one of the developing countries that face this disaster. Generally, these diseases that able to spread rapidly in contaminated water. Small worms and parasitic protozoa can live in water naturally and most of protozoa are harmful. Because they cannot be seen, they are hard to be avoided. Most of the time, developed countries have a small number of cases that relate to this disease because they have sophisticated and updated water system including the filter and chlorinate water to kill all worms and protozoan that can cause disease. In other words, waterborne disease closely related to the water management such as inadequate water supply, improper sewage disposal, poor personal hygiene and unsatisfactory environmental sanitation. However, drinking water quality in developed countries is also not assured. In France, when drinking water was tested, it was uncovered that 3 million people were drinking water that have not meet the World Health Organization (WHO) standard, and 97% of groundwater sample did not meet standards for nitrate in the same study (World Water Assessment Programme, 2003).

The first waterborne disease was reported occur in Malaysia during the first half of the twentieth century. The infection starts at a northern state in Peninsular Malaysia which is Kedah. The first infection that was recorded is Cholera disease (FEDERATION et al., 1954). Then, the disease was continue spread to another state such as Sarawak, Kelantan, Perak and Malacca. Based on the analysis and the statistic that was recorded from years 1970 to 1997, this disease was fluctuating and become peak every five years, and mostly happen May, June and July where it a dry season (Kin, 2007). There are others waterborne disease that occur in Malaysia such as dengue, malaria, leptospirosis and Hepatitis B. All this disease can be spread by water easily. This is because, during severe drought, this condition force many people to use water directly from river (Kin, 2007). Based on all these studies, instead of condition of water itself, the climate also can be a cause to the spreading of the waterborne disease. Many do not realize that this water-borne disease can cost many lives if not prevented from the beginning. This is because the disease is very sensitive to weather (National Research Council (US) Committee on Climate, Ecosystems, Infectious Diseases and Health, 2001). According to Department of Statistic Malaysia, leptospirosis is one of waterborne diseases that have ranked in the top 5 most cases and mortality among others killer disease from 2012 until 2015 (Department of Statistics Malaysia, 2015).

One previous study also shown, number of deaths that caused by leptospirosis have significant increase since 2007 until 2010 and Malaysia recorded the highest number of deaths due to the disease in 2014 (Garba et al., 2017). There are 5 states in Malaysia contribute the highest cases for leptospirosis such as Melaka, Selangor, Kuala Lumpur, Negeri Sembilan and Sarawak (Tan et al., 2016), and in 2015, Negeri Sembilan become second larger number of outbreaks of leptospirosis after Kuala Lumpur (Garba et al., 2017). These two records have made leptospirosis a dangerous disease in Malaysia and have received attention from various parties including the Department of Health Malaysia over the past several years (Tan et al., 2016; Garba et al., 2017).

There are previous studies that have investigate the factors that can contribute in spreading the leptospirosis disease. In tropical country like Malaysia, environmental conditions are one of the variables that associated to a survival of specific bacteria especially leptospirosis disease. Extreme weather event such as floods and cyclones that occur in recent year may have potential to increase the disease incidence as well as the magnitude of leptospirosis outbreaks (Lau et al., 2010; Vijayachari et al., 2008). The findings of these studies have been a turning point for many researchers to develop a model that can predict the number of cases of leptospirosis in the future. In Thailand, few researchers have used to investigate the impact of time variation of meteorological variables in the number of leptospirosis cases.

Autoregressive Integrated Moving Average with Exogenous Inputs (ARIMAX) was used as a predictive model to predict the number of cases for leptospirosis from 2003 until 2010(Chadsuthi et al., 2012). In this study, they have used correlation analysis and found rainfall with 10-month lag time and temperature with an 8-month lag time can show the trend in leptospirosis cases and indirectly may increase the prediction of the number of leptospirosis cases. Their research shown the positive impact on the performance of the model and this study has also strengthened the theory of the relationship between meteorological factors and the transmission of leptospirosis. Besides, another retrospective study was undertaken to describe the meteorological impact on the patterns of human leptospirosis cases that recorded in Reunion Island (Indian Ocean) (Desvars et al., 2011). In this study, the researchers used the same type of predictive model which is ARIMAX to find the correlation between leptospirosis cases and meteorological variable. However, this study has found the rainfall and temperature 2-month prior is the most effective variable to predict the number of cases of leptospirosis. To the best of our knowledge, there are no studies in Malaysia that have focused on developing a predictive model that can help public health officially to predict the occurrence of the leptospirosis using meteorological variables.

Understanding extreme weather events and how this can explain the occurrence of leptospirosis diseases is a necessary first step at improving predictability and ultimately the community response to the epi- demic. However, due to the complexity of the physical and microbiological interactions that lead to conditions favouring disease occurrences, a mathematical model development can be laborious because it required several processes characterizing of physical and microbiological fundamental (Tompkins and Di Giuseppe, 2015). In contrast, a data mining approach can be more resource effective as models can be trained to learn patterns from historical records and be blind to the modeller's prior knowledge. Data mining models have long gained popularity in the fields of hydrology, agriculture, ecology, as well as health, yet limited work has been done for a couple hydro-meteorological-health systems and specifically for improving the understanding and forecasting of water-borne diseases (Babovic, 2005; Debeljak et al., 2009; Lucas, 2004; Mucherino et al., 2009).

1.2 Problem Statement

Malaysia is heavily influenced by the monsoon rains. The monsoon will cause a rain cycle based on southwest monsoon, northeast monsoon and two transition periods. This phenomenon indirectly will cause the whole of Peninsular Malaysia to be particularly humid in the east coast in the beginning of the northeast monsoon season and to dry at the end of the season. Negeri Sembilan is a state in Malaysia which lies on the western coast of Peninsular Malaysia. The monsoon rains cause variability of rainfall distribution across Negeri Sembilan and form two significant features which is a wetter region west of the highlands (including Seremban) up to the coast indicates an increase in annual rainfall while the area on the east of the highlands (including Jelebu and Kuala Pilah) were experience the decreasing rate of rainfall (Wong et al., 2016). The variability on rainfall distribution has influenced on the survival period, growth, transmission of leptospira in the external environment.

When conditions are optimal, pathogenic leptospires can survive in water and wet soil for weeks to months (Vijayachari et al., 2008). Besides, it also can influence the rodent behaviour because rodent activity increase during raining (Kraus et al., 2005). Rodent has tendency to moving indoors to seek shelter during raining or winter season or during colder ambient temperatures (Ng, 2016). In other words, the rodent would move to residential area where it would increase the chance of human to contact with rodent or rat dropping. Furthermore, growth rate of rodent may increase during this season because they would reduce their reproduction, thus they not facing the competition for access to food (Ng, 2016). These statements have been proved by two previous study which have been undertaken at two different country which is Thailand and Reunion Island (Indian Ocean) (Chadsuthi et al., 2012; Desvars et al., 2011). Both studies shown different finding where study at Thailand shown rainfall 10-month prior give positive correlation to increasing number of leptospirosis cases while study on Reunion Island found 2 months prior of rainfall give positive influence. Thus, in Negeri Sembilan (Malaysia) might have different correlation result due to different geographical and climate zone.

Referring these two previous studies, both used correlation analysis to find the best time lag for meteorological variable which have higher correlation with leptospirosis cases (Chadsuthi et al., 2012; Desvars et al., 2011). Cross-correlation might be the best and fast solution to identify the correlation between 2 independent variables, but it also provides meaningless correlation that exist between time series. For example, if the values of x series does not give any information to the y series at any times, then it still possible for cross-correlation to appear significant non-zero when the measurement against the stan- dard criteria (Dean and Dunsmuir, 2016). This is because cross-correlation analysis only investigates the change of y series when the x series changed at any times without give good expositions of pre- conception between these two variables. Besides, cross-correlation analysis also not promising a good performance for predictive model. Regarding to previous study that develop predictive model by using ARIMAX, the model was achieve better performance when the ARIMAX model combine with single input variable such as rainfall (low RMSE) compare ARIMAX with both input variables (rainfall and temperature) (Chadsuthi et al., 2012). This finding has proven that cross-correlation only find the best correlation between one input and one output variable without giving strong preconception reason for that correlation. Thus, when it come together with another variable, the model cannot fit because it has 2 difference correlation for 2 difference input variables.

ARIMAX model is one of mathematical model which very famous among researchers that involve in predictive modelling (Dhewantara et al., 2019). This model become popular due to ability to have solid underlying theory, stable estimation of time-varying trends (due to stationary characteristic) and can give advantages on simplify a complex situation (Li et al., 2012). However, implementation mathematical model required few assumptions or estimation in their equation. In early model development, it may seem that the problem is very complex to make any progress. Thus, it very necessary to assume to help in simplifying the problem and focus on the model's objective. The assumption may include the number of factors affecting the model, thereby deciding which factors are most important. Thus, this might cause simplification on the real problem and does not include all aspects of the problems. The model output might gives very precise result. But it does not mean the model have very accurate. The model was built with statistical technique based on the specific range that has been covered by input data. But if the model faced with unseen data, model need to have few changes on the parameter to keep the model to perform well. In other words, mathematical model cannot generalize the real problem and less reliable (Richardson, 1979).

In conclusion, to overcome all these drawbacks and improve the predictive model of disease prediction, we are proposing one method for feature selection that can see through the data and select which data that associated with the leptospirosis occurrence. Besides, exploring the ability of modern mathematical models may help in improving disease prediction as well.

1.3 Objective

The overall aim is predicting the occurrence of Leptospirosis in Seremban district using a machine learning and meteorological data as input. Specific objectives:

I. To design and analyze feature selection methods which are correlation analysis and Exploratory Data Analysis to find the best time lag of temperature and rainfall data.
II. To develop a predictive model using backpropagation neural network for the direct and indirect impacts of environmental variables on the occurrence of Leptospirosis

1.4 Research Contributions

To correlate between the meteorological variables and occurrence of leptospirosis disease by using Exploratory Data Analysis is new. This study designed and investigate the suitable approach by using this method to perform better selection on the time lag of temperature and rainfall data.

1.5 Research Scope

This study has set some limitation as guidance and reference for the researchers. First, this study used secondary data for both meteorological and clinical data. The secondary data is the data that was obtained by collection from the government departments including Department of Health Negeri Sembilan, Department of Meteorological Malaysia (MetMalaysia) and Department of Irrigation and Drainage and it was not retrieved by measurement of a rain gauge or thermometer. Secondly, this study does not in- volve any scientific experiment that will use any laboratory equipment. Lastly, this study only done on simulation which only require uses of few software and does not planning in developing the hardware.

1.6 Thesis Outline

1.6.1 Introduction

This section discussed the general topic of transmission of leptospirosis in Malaysia. Besides, emphasis on the purpose of this study in disease prediction also has been discussed. Furthermore, the research question as well as the objective and research scope study also filled in this section.

1.6.2 Literature Review

This section would describe all relevant topics that fall in this study subject. Topics that would be reviewed are meteorological factors that may affect the transmission of leptospirosis, previous and current studies in feature selection and implementation of a mathematical model in disease prediction. At the end of this chapter, the conclusion has been made based on the review of previous studies and finally, the research gap identified.

1.6.3 Methodology

This section gives more detail on how the study gets access to the selected methodological approach including data retrieved, data processing, data analysis and model development. Analyzed data emphasize more to the feature selection technique while modelling development more to the parameter selection for the predictive model.

1.6.4 Result and Discussion

This chapter was divided into 3 sections. The first section has presented the result of the based-line model as well as the preliminary study in this research. The second section presents and discusses the result of the time lag of rainfall and temperature data based on the type of feature selection techniques. The final section presents the performance of the proposed model in terms of accuracy, specificity, and sensitivity during the training and testing phase. Besides, this section also discussed how different time lags of rainfall and temperature data may affect the performance of the model.

1.6.5 Conclusion

This is the final section of this thesis. Thus, the overall conclusion including the methodological approach and performance of the model has been made. Besides, this section also included a recommendation that may improve disease prediction in future studies.

REFERENCES

Abhishek, K., Kumar, A., Ranjan, R. and Kumar, S. 2012. A rainfall prediction model using artificial neural network. In 2012 IEEE Control and System Graduate Research Colloquium, 82–87. IEEE.

Agostinelli, F., Hoffman, M., Sadowski, P. and Baldi, P. 2014. Learning activation functions to improve deep neural networks. arXiv preprint arXiv:1412.6830 .

Ahumada, J. A., Laoointe, D. and Samuel, M. D. 2004. Modeling the population dynamics of Culex quinquefasciatus (Diptera: Culicidae), along an elevational gradient in Hawaii. Journal of medical entomology 41 (6): 1157–1170.

Aieb, A., Madani, K., Scarpa, M., Bonacorso, B. and Lefsih, K. 2019, A new approach for processing climate missing databases applied to daily rainfall data in Soummam watershed, Algeria.

Andre-Fontaine, G., Aviat, F. and Thorin, C. 2015. Waterborne Leptospirosis: Survival and Preservation of the Virulence of Pathogenic Leptospira spp. in Fresh Water. Current Microbiology 71 (1): 136–142.

Babovic, V. 2005. Data mining in hydrology. Hydrological Processes 19 (7): 1511–1515.

Baraldi, A. and Blonda, P. 1999. A survey of fuzzy clustering algorithms for pattern recognition. II. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 29 (6): 786–801.

Barrett, M. A., Humblet, O., Hiatt, R. A. and Adler, N. E. 2013. Big data and disease prevention: from quantified self to quantified communities. Big data 1 (3): 168–175.

Borovicka, T., Jirina, M., Kordik, P. and Jiri, M. 2012. Selecting Representative Data Sets. Advances in Data Mining Knowledge Discovery and Applications .

Box, G. E., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M. 2015. Time series analysis: forecasting and control. John Wiley & Sons.

Chadsuthi, S., Modchang, C., Lenbury, Y., Iamsirithaworn, S. and 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.

Coelho, M. S. and Massad, E. 2012. The impact of climate on Leptospirosis in São Paulo, Brazil. Inter- national journal of biometeorology 56 (2): 233–241.

Cook, A., Watson, J., van Buynder, P., Robertson, A. and Weinstein, P. 2008. 10th Anniversary Review: Natural disasters and their long-term impacts on the health of communities. Journal of Environmental Monitoring 10 (2): 167–175.

Cotruvo, J., Dufour, A., Rees, G., Bartram, J., Carr, R., Cliver, D. O., Craun, G. F., Fayer, R. and Gannon,

V. P. 2004. Waterborne zoonoses: identification, causes, and control. World Health Organization.

Dean, R. T. and Dunsmuir, W. T. 2016. Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. Behavior Research Methods 48 (2): 783–802.

Debeljak, M., Dzeroski, S., Jørgensen, S., T-Chon, S. and Recknagel, F. 2009. Applications of data mining in ecological modelling. Handbook of ecological modelling and informatics. WIT Press, Southampton, UK 409–423.

Department of Statistics Malaysia. 2015, Statistik Penyakit Berjangkit yang Didaftarkan Mengikut Kate- gori Umur Dan Negeri Di Malaysia.

Deshmukh, P., Narang, R., Jain, J., Jain, M., Pote, K., Narang, P., Raj, R., Kumar, P. and Vijayachari, P. 2019. Leptospirosis in Wardha District, Central India—Analysis of hospital based surveillance data. Clinical Epidemiology and Global Health 7 (1): 102–106.

Desvars, A., Jégo, S., Chiroleu, F., Bourhy, P., Cardinale, E. and Michault, A. 2011. Seasonality of human leptospirosis in Reunion Island (Indian Ocean) and its association with meteorological data. PLoS ONE 6 (5).

Dhewantara, P. W., Lau, C. L., Allan, K. J., Hu, W., Zhang, W., Mamun, A. A. and Soares Magalhães, R. J. 2019. Spatial epidemiological approaches to inform leptospirosis surveillance and control: A systematic review and critical appraisal of methods. Zoonoses and Public Health 66 (2): 185–206.

Dufour, B., Moutou, F., Hattenberger, A., Rodhain, F. et al. 2008. Global change: impact, management, risk approach and health measures–the case of Europe. Rev. Sci. Tech. Off. Int. Epizoot. 27: 529–550.

Dutta, P. S. and Tahbilder, H. 2014. Prediction of Rainfall Using Datamining Technique Over Assam 5 (2): 85–90.

Ellis, W. A. 2015, 99–137, 99–137.

Evangelista, K. V. and Coburn, J. 2010. Leptospira as an emerging pathogen: a review of its biology, pathogenesis and host immune responses. Future Microbiology 5 (9): 1413–1425.

Fan, H., Zheng, L., Yan, C. and Yang, Y. 2018. Unsupervised person re-identification: Clustering and fine- tuning. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 14 (4): 83.

FEDERATION, O. M. et al. 1954. Annual Report of the Institute for Medical Research, Kuala Lumpur, for the Year 1954. Annual Report of the Institute for Medical Research, Kuala Lumpur, for the Year 1954. .

Garba, B., Bahaman, A. R., Khairani-Bejo, S., Zakaria, Z. and Mutalib, A. R. 2017. Retrospective Study of Leptospirosis in Malaysia. EcoHealth 14 (2): 389–398.

Gatto, M., Mari, L., Bertuzzo, E., Casagrandi, R., Righetto, L., Rodriguez-Iturbe, I. and Rinaldo, A. 2012. Generalized reproduction numbers and the prediction of patterns in waterborne disease. Proceedings of the National Academy of Sciences of the United States of America 109 (48): 19703–19708.

Gazzaz, N. M., Yusoff, M. K., Ramli, M. F., Juahir, H. and Aris, A. Z. 2015. Artificial Neural Network Modeling of the Water Quality Index Using Land Use Areas as Predictors. Water Environment Research 87 (2): 99–112.

Gharbi, M., Quenel, P., Gustave, J., Cassadou, S., Ruche, G. L., Girdary, L. and Marrama, L. 2011. Time series analysis of dengue incidence in guadeloupe, french west indies: Forecasting models using climate variables as predictors. BMC Infectious Diseases 11 (May 2014).

Grassmann, A. A., da Cunha, C. E. P., Bettin, E. B. and McBride, A. J. A. 2017, 245–275, 245–275.

Ho Yu, C. 2010. Exploratory data analysis in the context of data mining and resampling. International Journal of Psychological Research 3 (1): 9.

Jalalkamali, A., Moradi, M. and Moradi, N. 2015. Application of several artificial intelligence models and ARIMAX model for forecasting drought using the Standardized Precipitation Index. International journal of environmental science and technology 12 (4): 1201–1210.

Joshi, Y. P., Kim, E.-H. and Cheong, H.-K. 2017. The influence of climatic factors on the development of hemorrhagic fever with renal syndrome and leptospirosis during the peak season in Korea: an ecologic study. BMC Infectious Diseases 17 (1): 406.

Kalton, G. and Kish, L. 1984. Some efficient random imputation methods. Communications in Statistics - Theory and Methods 13 (16): 1919–1939.

Karsoliya, S. 2012. Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture. International Journal of Engineering Trends and Technology 3 (6): 714–717.

Kin, F. 2007, In A World of Water, In A World of Water, 279–295, Brill, 279–295.

Kraus, A. A., Priemer, C., Heider, H., Krüger, D. H. and Ulrich, R. 2005. Inactivation of hantaan virus-containing samples for subsequent investigations outside biosafety level 3 facilities. Intervirology 48 (4): 255–261.

Lalkhen, A. G. and McCluskey, A. 2008. Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia Critical Care & Pain 8 (6): 221–223.

Lau, C. L., Smythe, L. D., Craig, S. B. and Weinstein, P. 2010. Climate change, flooding, urbanisation and leptospirosis: Fuelling the fire? Transactions of the Royal Society of Tropical Medicine and Hygiene 104 (10): 631–638.

Li, Q., Guo, N. N., Han, Z. Y., Zhang, Y. B., Qi, S. X., Xu, Y. G., Wei, Y. M., Han, X. and Liu, Y. Y.

2012. Application of an autoregressive integrated moving average model for predicting the incidence of hemorrhagic fever with renal syndrome. American Journal of Tropical Medicine and Hygiene 87 (2): 364–370.

Lobitz, B., Beck, L., Huq, A., Wood, B., Fuchs, G., Faruque, A. S. and Colwell, R. 2000. Climate and infectious disease: Use of remote sensing for detection of Vibrio cholerae by indirect measurement. Proceedings of the National Academy of Sciences of the United States of America 97 (4): 1438–1443.

Louangrath, P. 2015. Normal Distribution and Common Tests Used to Verify Normality.

Lucas, P. 2004. Bayesian analysis, pattern analysis, and data mining in health care. Current Opinion in Critical Care 10 (5): 399–403.

M. Jones, Z. and J. Linder, F. 2016. edarf: Exploratory Data Analysis using Random Forests. The Journal of Open Source Software 1 (6): 92.

Maciel, E. A., de Carvalho, A. L. F., Nascimento, S. F., de Matos, R. B., Gouveia, E. L., Reis, M. G. and Ko, A. I. 2008. Household transmission of Leptospira infection in urban slum communities. PLoS neglected tropical diseases 2 (1): e154.

Madsen, H. 2007. Time series analysis. Time Series Analysis 1–373.

Mahesh, C., Kiruthika, K. and Dhilsathfathima, M. 2014. Diagnosing hepatitis B using artificial neural network based expert system. In International Conference on Information Communication and Em- bedded Systems (ICICES2014), 1–7. IEEE.

Manap, R. 2015. LEPTOSPIRAL INFECTION. In Proceeding of the 2nd International Conference on Management and Muamalah.

Mgode, G. F., Machang'u, R. S., Mhamphi, G. G., Katakweba, A., Mulungu, L. S., Durnez, L., Leirs, H., Hartskeerl, R. A. and 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).

Misra, A. K. and Singh, V. 2012. A delay mathematical model for the spread and control of water borne diseases. Journal of Theoretical Biology 301: 49–56.

Mohammadinia, A., Saeidian, B., Pradhan, B. and Ghaemi, Z. 2019. Prediction mapping of human lep- tospirosis using ANN, GWR, SVM and GLM approaches. BMC infectious diseases 19 (1): 1–18.

Mucherino, A., Papajorgji, P. and Pardalos, P. M. 2009. A survey of data mining techniques applied to agriculture. Operational Research 9 (2): 121–140.

Mueez, A., Islam, K. A. T. and Iqbal, W. 2018. Exploratory Data Analysis and Success Prediction of Google Play Store Apps Authors (December).

Mustafidah, H., Hartati, S., Wardoyo, R. and Harjoko, A. 2014. Selection of Most Appropriate Backprop- agation. Internasional Journal of Computer Trends and Technology (IJCTT) 14 (2): 92–95.

Mutalip, M. H. A., Mahmud, M. A. F., Yoep, N., Muhammad, E. N., Ahmad, A., Hashim, M. H. and Muhamad, N. A. 2019. Environmental risk factors of leptospirosis in urban settings: a systematic review protocol. BMJ open 9 (1): e023359.

Muthulakshmi, A. and BaghavathiPriya, S. 2015. A survey on weather forecasting to predict rainfall using big data analytics. International Journal of Innovative Science, Engineering & Technology 2 (10).

National Research Council (US) Committee on Climate, Ecosystems, Infectious Diseases and Health,

H. 2001, In Under the Weather: Climate, Ecosystems, and Infectious Disease. Washington (DC): Na- tional Academies Press (US), In Under the Weather: Climate, Ecosystems, and Infectious Disease. Washington (DC): National Academies Press (US).

Navid, M. 2018. Multiple Linear Regressions for Predicting Rainfall for Bangladesh. Communications 6 (1): 1.

Nery, N. R. R., Claro, D. B. and Lindow, J. C. 2017. Prediction of leptospirosis cases using classification algorithms. IET Software 11 (3): 93–99.

Ng, M. 2016. Environmental factors associated with increased rat populations: A Focused Practice Question (November).

Panchal, F. S. and Panchal, M. 2014. International Journal of Computer Science and Mobile Computing Review on Methods of Selecting Number of Hidden Nodes in Artificial Neural Network. International Journal of Computer Science and Mobile Computing 3 (11): 455–464.

Parker, J. and Walker, M. 2011. Survival of a pathogenic Leptospira serovar in response to combined in vitro pH and temperature stresses. Veterinary microbiology 152 (1-2): 146–150.

Penna, M. L. F. 2004. Use of an artificial neural network for detecting excess deaths due to cholera in Ceará, Brazil. Revista de saude publica 38 (3): 351–357.

Pezeshki, Z., Tafazzoli-Shadpour, M., Nejadgholi, I., Mansourian, A. and Rahbar, M. 2016. Model of cholera forecasting using artificial neural network in Chabahar City, Iran. Int J Enteric Pathog 4 (1): 1–8.

Radford, P. J., Velleman, P. F. and Hoaglin, D. C. 1983. Applications, Basics, and Computing of Ex- ploratory Data Analysis. Biometrics 39 (3): 815.

Rahmat, F., Ishak, A. J., Zulkafli, Z., Yahaya, H. and Masrani, A. 2019. Prediction model of leptospirosis occurrence for Seremban (Malaysia) using meteorological data. International Journal of Integrated Engineering 11 (4): 61–69.

Richardson, B. 1979. Limitations on the use of mathematical models in transportation policy analysis. Motor Vehicle Manufacturers Association 1–13.

Ridzlan, F. R., Bahaman, A. R., Khairani-Bejo, S. and Mutalib, A. R. 2010. Detection of pathogenic Leptospira from selected environment in Kelantan and Terengganu, Malaysia. Tropical Biomedicine 27 (3): 632–638.

Schneider, M., Nájera, P., Aldighieri, S., Bacallao, J., Soto, A., Marquiño, W., Altamirano, L., Saenz, C., Marin, J., Jimenez, E. et al. 2012. Leptospirosis outbreaks in Nicaragua: identifying critical areas and exploring drivers for evidence-based planning. International journal of environmental research and public health 9 (11): 3883–3910.

Schneider, M. C., Jancloes, M., Buss, D. F., Aldighieri, S., Bertherat, E., Najera, P., Galan, D. I., Durski,

K. and Espinal, M. A. 2013. Leptospirosis: A silent epidemic disease. International Journal of Envi- ronmental Research and Public Health 10 (12): 7229–7234.

Song, Y., Wang, F., Wang, B., Tao, S., Zhang, H., Liu, S., Ramirez, O. and Zeng, Q. 2015. Time series analyses of hand, foot and mouth disease integrating weather variables. PLoS ONE 10 (3): 1–18.

Tan, W. L., Soelar, S. A., Suan, M. A. M., Hussin, N., Cheah, W. K., Verasahib, K. and Goh, P. P. 2016. Leptospirosis incidence and mortality in Malaysia. Southeast Asian Journal of Tropical Medicine and Public Health 47 (3): 434–440.

Thakur, S. and Dharavath, R. 2019. Artificial neural network based prediction of malaria abundances using big data: A knowledge capturing approach. Clinical Epidemiology and Global Health 7 (1): 121–126.

Thibeaux, R., Geroult, S., Benezech, C., Chabaud, S., Soupé-Gilbert, M.-E., Girault, D., Bierque, E. and Goarant, C. 2017. Seeking the environmental source of Leptospirosis reveals durable bacterial viability in river soils. PLOS Neglected Tropical Diseases 11 (2): e0005414.

Tompkins, A. M. and Di Giuseppe, F. 2015. Potential Predictability of Malaria in Africa Using ECMWF Monthly and Seasonal Climate Forecasts. Journal of Applied Meteorology and Climatology 54 (3): 521–540.

Triampo, W., Baowan, D., Tang, I. M., Nuttavut, N. and Doungchawee, G. 2007. A Simple Deterministic Model for the Spread of Leptospirosis in Thailand. International Journal of Biological and Medial Sciences 2 (1): 22–26.

Trueba, G., Zapata, S., Madrid, K., Cullen, P. and Haake, D. 2004. Cell aggregation: A mechanism of pathogenic Leptospira to survive in fresh water. International Microbiology 7 (1): 35–40.

Vanasco, N. B., Schmeling, M., Lottersberger, J., Costa, F., Ko, A. I. and Tarabla, H. D. 2008. Clinical characteristics and risk factors of human leptospirosis in Argentina (1999–2005). Acta Tropica 107 (3): 255–258.

Victoriano, A. F., Smythe, L. D., Gloriani-Barzaga, N., Cavinta, L. L., Kasai, T., Limpakarnjanarat, K., Ong, B. L., Gongal, G., Hall, J., Coulombe, C. A., Yanagihara, Y., Yoshida, S. I. and Adler, B. 2009. Leptospirosis in the Asia Pacific region. BMC Infectious Diseases 9 (October): 147.

Vijayachari, P., Sugunan, A. P. and Shriram, A. N. 2008. Leptospirosis: an emerging global public health problem. (Special issue. Emerging and re-emerging infections in India). Journal of Biosciences 33 (November): 557–569.

Wang, W. 2006. Stochasticity, nonlinearity and forecasting of streamflow processes. Ios Press.

Wasin´ski, B. and Dutkiewicz, J. 2013. Leptospirosis - Current risk factors connected with human activity and the environment. Annals of Agricultural and Environmental Medicine 20 (2): 239–244.

Weinberger, D., Baroux, N., Grangeon, J.-P., Ko, A. I. and Goarant, C. 2014. El Niño Southern Oscillation and Leptospirosis Outbreaks in New Caledonia. PLoS Neglected Tropical Diseases 8 (4): e2798.

Wong, C. L., Liew, J., Yusop, Z., Ismail, T., Venneker, R. and Uhlenbrook, S. 2016. Rainfall charac- teristics and regionalization in peninsular malaysia based on a high resolution gridded data set. Water (Switzerland) 8 (11).

World Water Assessment Programme, U. 2003. United Nations world water assessment Programme. The world water development report 1: water for people, water for life. UNESCO: Paris, France.

Wright, J. and Gundry, S. W. 2009. Household characteristics associated with home water treatment: An analysis of the Egyptian demographic and health survey. Journal of Water and Health 7 (1): 21–29.

Xue, F., Dong, H., Wu, J., Wu, Z., Hu, W., Sun, A., Troxell, B., Yang, X. F. and Yan, J. 2010. Transcrip- tional responses of Leptospira interrogans to host innate immunity: significant changes in metabolism, oxygen tolerance, and outer membrane. PLoS neglected tropical diseases 4 (10): e857.

Yes¸ilova, A., Kaya, Y. and Almali, M. N. 2011. A comparison of hot deck imputation and substitution methods in the estimation of missing data. Gazi University Journal of Science 24 (1): 69–75.

Zhao, F. and Li, W. 2019. A Combined Model Based on Feature Selection and WOA for PM2.5 Concen- tration Forecasting. Atmosphere 10 (4): 223.

Zinszer, K., Kigozi, R., Charland, K., Dorsey, G., Brewer, T. F., Brownstein, J. S., Kamya, M. R. and Buckeridge, D. L. 2015. Forecasting malaria in a highly endemic country using environmental and clinical predictors. Malaria journal 14 (1): 245.

Zitek, K. and Benes, C. 2005. Longitudinal epidemiology of leptospirosis in the Czech Republic (1963- 2003). Epidemiologie, mikrobiologie, imunologie: casopis Spolecnosti pro epidemiologii a mikrobiologii Ceske lekarske spolecnosti J.E. Purkyne 54 (1): 21–6.

Zupan, J. 1994. Introduction to artificial neural network (ANN) methods: what they are and how to use them. Acta Chimica Slovenica 41 (September): 327–327.