LEPTOSPIROSIS MODELLING USING HYDROMETEOROLOGICAL INDICES AND RANDOM FOREST MACHINE LEARNING FOR HUMID TROPICAL NORTH-EAST PENINSULAR MALAYSIA

By

VEIANTHAN A/L JAYARAMU

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

June 2022

FK 2022 126

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

LEPTOSPIROSIS MODELLING USING HYDROMETEOROLOGICAL INDICES AND RANDOM FOREST MACHINE LEARNING FOR HUMID TROPICAL NORTH-EAST PENINSULAR MALAYSIA

By

VEIANTHAN A/L JAYARAMU

June 2022

Chair : Zed Diyana binti Zulkafli, PhD
Faculty : Engineering

Leptospirosis is a zoonotic tropical disease caused by pathogenic *Leptospira* sp. whose transmission has been linked to extreme hydrometeorological phenomena. Hydrometeorological variability in the form of averages and extremes indices have been used before as drivers in statistical prediction of disease occurrence; however, their importance and predictive capacity are still little known. Random forest classification models of leptospirosis occurrence were developed to identify the important hydrometeorological indices and models' prediction accuracy, sensitivity, and specificity based on the sets of indices used, using case data from three districts in Kelantan, Malaysia. This region experiences annual monsoonal rainfall and flooding, and that record high leptospirosis incidence rates. First, hydrometeorological data including rainfall, streamflow, water level, relative humidity and temperature were derived into 164 weekly average and extreme indices in accordance with the Expert Team on Climate Change Detection and Indices (ETCCDI). Then, the weekly number of cases were classified into binary classes 'high' and 'low' based on an average threshold. 17 models based on 'average', 'extreme' and 'mixed' sets of indices – based on the type of indices used as input – were trained by optimizing the feature subsets using the embedded approach that utilized the mean decrease Gini (MDG) scores. The variable importance was assessed through cross correlation analysis and the MDG scores. The results showed that the average and extreme models showed similar prediction accuracy ranges while the mixed models showed some improvement. An extreme model was the most sensitive while and average model was the most specific. The time lag associated with the driving indices agreed with the seasonality of the monsoon. The variable importance analysis based on the MDG scores indicated that overall, the rainfall (extreme) factor dominated, suggesting its strong influence on Leptospirosis incidence while the streamflow variable was the least important to the model development despite showing higher cross correlations with leptospirosis.

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

PEMODELAN LEPTOSPIROSIS DENGAN MENGGUNAKAN INDEKS HIDROMETEOROLOGI DAN PEMBELAJARAN MESIN HUTAN RAWAK UNTUK KAWASAN TROPIK LEMBAP TIMUR LAUT SEMENANJUNG MALAYSIA

Oleh

VEIANTHAN A/L JAYARAMU

Jun 2022

Pengerusi : Zed Diyana binti Zulkafli, PhD
Fakulti : Kejuruteraan

Leptospirosis merupakan penyakit tropikal zoonotik yang disebabkan oleh *Leptospira sp.* patogenik di mana penularannya sering dikaitkan dengan fenomena hidrometeorologi yang ekstrim. Kebolehubahan hidrometeorologi dalam bentuk purata dan ekstrim pernah digunakan sebagai faktor dalam ramalan statistik kejadian penyakit; walaubagaimanapun, kepentingan dan keupayaannya masih kurang diketahui. Model pengelasan hutan rawak kejadian leptospirosis dibangunkan untuk mengenalpasti indeks-indeks hidrometeorologi yang penting dan ketepatan ramalan, sensitiviti dan kekhususan model yang bergantung pada set-set indeks tersebut, dengan menggunakan data kes dari tiga daerah di Kelantan, Malaysia. Kawasan ini mengalami hujan monsun tahunan dan banjir serta kejadian leptospirosis yang tinggi. Pertama, data hidrometeorologi termasuk hujan, aliran sungai, paras air, kelembapan dan suhu udara diagregatkan kepada 164 indeks purata dan ekstrim mingguan selaras dengan Pasukan Pakar dalam Pengesanan Perubahan Iklim dan Indeks (*ETCCDI*). Kemudian, bilangan kes mingguan dikelaskan kepada kelas binari 'tinggi' dan 'rendah' mengikut ambang purata. Model berdasarkan 'purata', 'ekstrim' dan 'campuran' – berdasarkan kelas indeks sebagai input – dilatih dengan mengoptimumkan subset ciri menggunakan cara *embedded* yang memanfaatkan skor purata penurunan Gini (*MDG*). Kepentingan pemboleh ubah telah dinilai melalui *cross correlation analysis* (*CCA*) dan skor *MDG*. Hasil kajian menunjukkan bahawa model purata dan ekstrim menunjukkan ketepatan ramalan yang serupa manakala model campuran menunjukkan penambahbaikan. Model ekstrim adalah yang paling sensitif manakala model purata adalah yang paling khusus. Skor MDG menunjukkan hujan ekstrim sebagai faktor dominan, mencadangkan pengaruh kuatnya terhadap kejadian leptospirosis, manakala aliran sungai, walaupun berkorelasi tinggi dengan leptospirosis, kurang penting untuk pembangunan model.

ACKNOWLEDGEMENTS

The completion of my master's study could not have been possible without the support of my loving parents, Jayaramu and Paremeswari. Their countless sacrifices and unconditional love have always kept my momentum going and pushed me to the next levels in my life. I am forever grateful for having such great parents in my life.

Next, I feel so blessed and grateful for conducting my master's research under the supervision of Assoc. Prof. Dr. Zed Diyana Zulkafli. All the guidance provided my supervisor throughout my master's journey are extremely invaluable for my research as well for my life. She has always been encouraging me to pursue new things in the research so that I can come up with new ideas and methodology. She has given very useful comments throughout my master's journey, which helped me to clearly define my research direction and refine the methodology.

Furthermore, I wish to thank my co-supervisors, Dr. Simon de Strecke, from the Department of Civil and Environmental Engineering, Faculty of Engineering, Imperial College London, Assoc. Prof. Dr. Asnor Juraiza Ishak and Dr. Ribhan Zafira Abdul Rahman, from the Department of Electric and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia. They have always posed critical questions regarding my research, which helped me to explore deeper into the methods and gain a better understanding about them.

Not to forget, special thanks to my friend, Mohamad Fariq Rahmat for being a supportive colleague during my entire master's program. He has always been keen to engage discussions with me regarding my research and give his point of views. This prepared myself to present the research findings more confidently to my supervisory committee.

Lastly, I want to also thank the health officer, Dr. Nik Mohd Hafiz Mohd Fuzi from the Department of Health, Kelantan, who was involved in the provision and validation of leptospirosis e-notification data. He was willing and eager to help with the verification of the information contained in the health data.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee were as follows:

Zed Diyana binti Zulkafli, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

Asnor Juraiza binti Dato Hj Ishak, PhD
Associate Professor Ts.
Faculty of Engineering
Universiti Putra Malaysia
(Internal Member)

Ribhan Zafira binti Abdul Rahman, PhD
Senior Lecturer
Faculty of Engineering
Universiti Putra Malaysia
(Internal Member)

Simon De Stercke, PhD
Postdoctoral Researcher
Faculty of Engineering
Imperial College London
(External Member)

ZALILAH MOHD SHARIF, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 11 May 2023

Declaration by Graduate Student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any institutions;
- intellectual property from the thesis and the copyright of the thesis are fully-owned by Universiti Putra Malaysia, as stipulated in the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from the supervisor and the office of the Deputy Vice-Chancellor (Research and innovation) before the thesis is published in any written, printed or electronic form (including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials) as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld in accordance with the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2015-2016) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software

Signature: _____ Date: _____

Name and Matric No.: Veianthan a/l Jayaramu

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

API	Application Programming Interface
CCA	Cross Correlation Analysis
DID	Drainage and Irrigation Department
ELISA	Enzyme-Linked Immunosorbent Assay
ETCCDI	Expert Team on Climate Change Detection and Indices
MAT	Microscopic Agglutination Test
MDA	Mean Decrease Accuracy
MDG	Mean Decrease Gini
MetMalaysia	Malaysian Meteorological Department
PCR	Polymerase Chain Reaction
QGIS	Quantum Geographic Information System
ROC	Receiver Operating Characteristic
WHO	World Health Organization

GLOSSARY OF TERMS

Terms	Definition
Feature	Independent variable that is used as a model input
Gini	Probability of misclassification of output classes due to the splitting of the features at the nodes of the single tree (decision tree)
Hydrometeorological index	Weekly indices derived/aggregated from the daily hydrometeorological data, i.e., weekly mean rainfall, weekly maximum temperature
Hydrometeorological variable	Type of hydrometeorological data used in the study, e.g., rainfall, temperature
Lag	Difference in the number of weeks between hydrometeorological indices and leptospirosis occurrence
Modelling	A process of representing the interaction between dependent and independent variables by using a set of equations or algorithms
Prediction	A process of a model generating outcomes based on an unseen data.
Predictor	An independent variable that is used to predict the dependent variable.
Training	A process of developing model by learning data using algorithm. This term is similar to the term 'calibration' in water resource modelling.
Testing	A process of assessing the model's predictability using a set of unseen data. This term is similar to the term 'validation' in in water resource modelling.

CHAPTER 1

INTRODUCTION

1.1 Background

Leptospirosis is a zoonotic disease caused by a pathogenic spiral-shaped bacteria of genus *Leptospira sp.* that gets transmitted from animals to humans (Levett, 2001). The infection takes place in two ways either directly or indirectly. The direct infection occurs when humans contract the disease by having a contact with the urine of a host that has been infected by the bacteria (Ansdell, 2017). On the other hand, the indirect infection happens when humans get the disease through a contact with the environment (soil and water bodies) that has been contaminated by the bacterial community (Ansdell, 2017). The bacteria reach the human blood circulation system via an injured skin and/or mucous membrane (Ansdell, 2017). The disease has been described in the ancient literature, but its modern history began a century ago attributing it to jaundice with enlarged spleen, kidney failure, pink eye and skin rashes (Adler, 2015). The disease exists in both temperate and tropical countries, but usually higher incidence rates are recorded in tropics (10 – 100 per 100,000 population per year) compared with the temperate climates (0.1 – 1 per 100,000) (WHO, 2003). Tropical regions experience rainfall and hot temperature all year-round, which keeps the environment warm and humid. Such an environment is suitable for rodent proliferation as well as for the longevity of leptospires (Garba et al., 2018). The processes involved in the hydrologic cycle circulate the bacterial community among the maintenance (rodents), accidental (humans) hosts, soil and water bodies ensuring the endemicity of leptospirosis.

A recent systematic review study estimated that 1.03 million cases and 58,900 deaths have occurred annually due to leptospirosis worldwide (Costa et al., 2015) indicating its global public health importance. However, this is expected to be an underestimated number as leptospirosis cases are normally misdiagnosed and underreported due to its similar manifestations with other febrile illnesses and lack of laboratory diagnostic facilities (WHO, 2001). The actual number of leptospirosis cases remains unknown as many probably endemic countries do not have a proper notification system (Pappas et al., 2008). The current situation of leptospirosis can become even worse as the driving forces are expected to be more vigorous with the recent severe climate changes, major flooding events and rapid urbanisation (Lau et al., 2010).

Initially, leptospirosis was seen as an occupational disease as it mostly affects farmers, sanitation workers, veterinarians and those who conduct activities in soil and water bodies that may be contaminated with infected animal urine (Picardeau, 2013). However, recently, the disease incidence has increased and been linked with hydrometeorological catastrophes (Guerra, 2013). Hydrometeorological catastrophes such as flooding events tend to quicken and broaden the transmission of leptospirosis by bringing the bacteria closer to

humans and causing outbreaks (Ansdell, 2017). As part of investigating the disease occurrence, modelling studies have long been undertaken at different spatial and temporal resolutions to understand the roles of risk factors in driving leptospirosis (Dhewantara et al., 2019). Modelling is the process of representing the interaction between dependent and independent variables by using a set of equations or algorithms. Traditional statistical models have established the relationships between the risk factors and leptospirosis. Since leptospirosis occurs under complex ecological settings involving diverse predictors, machine learning models have gained attention as they are highly capable of handling a large number of predictors and nonlinear patterns (Ahangarcani et al., 2019).

1.2 Problem Statement

Although leptospirosis has been ubiquitous globally due to a vast array of driving factors, rapid urbanisation, climate change and hydrometeorological extreme events are likely to aggravate the current situation of leptospirosis (Lau et al., 2010; Picardeau, 2013). Flooding events have contributed to a higher rate of, more widespread, and longer lagged infections (Barcellos & Sabroza, 2001; Ding et al., 2019; Radi et al., 2018; Sehgal et al., 2002; Togami et al., 2018). This indicated that extreme hydrometeorological events including heavy rainfall and the consecutive flood disperse leptospires broadly.

To better understand the mechanism behind leptospirosis transmission, several different approaches to statistical prediction have been explored. While many have considered the spatial dependency (Lau et al., 2012; Mayfield et al., 2018; Mohammadinia et al., 2017; Sánchez-Montes et al., 2015; Schneider et al., 2012; Suwanpakdee et al., 2015; Vega-Corredor & Opadeyi, 2014; Zhao et al., 2016), fewer studies have analysed the drivers behind the occurrence, transmission and outbreak using the time series (Chadsuthi et al., 2012; Desvars et al., 2011; Joshi et al., 2017; Weinberger et al., 2014). Additionally, most have employed conventional statistical modelling techniques, which inadequately handled the non-linearity present in the relationship of leptospirosis and its risk factors (Dhewantara et al., 2019). The complex mechanism of leptospirosis transmission due to the involvement of multiple variables impedes the models' ability of explaining the disease trends (WHO, 2011). Machine learning models, in contrast, can capture the complex patterns in disease occurrences, and therefore predict the output with a higher accuracy (Ahangarcani et al., 2019; Carvajal et al., 2018; Guo et al., 2017; Hu et al., 2018). Machine learning algorithms do not assume linearity between the independent and dependent variables as how the traditional statistical models do. Rather, they use non-linear functions, which will be implemented individually to each aspect of data to learn the complex patterns present in it. Since the relationship between the hydrometeorological variables and leptospirosis are highly non-linear, machine learning could be a better approach since it can handle non-linearity. For example, artificial neural network uses non-linear activation functions, i.e., sigmoid and relu, to make non-linear transformations to the input making it capable of learning and performing more complex tasks (Grossberg, 1988).

However, they are often treated as black boxes when the objective is to optimise predictive performance, rather than to gain process insight.

Nevertheless, knowledge extraction is possible with the use of interpretable machine learning algorithms. Random forest machine learning (Breiman, 2001) is one that allows insight into feature (input variable) importance. Unlike the neural network and support vector machine, the random forest algorithm uses tree-based decision making, and can rank the features involved during the model training based on how well they contribute to the classification of output classes (Hastie et al., 2009). This indicates that random forest is a better model to use when it comes to understanding the contribution of features in the model development. The learning is well known for its diversity as the algorithm includes data bootstrapping and random selection of predictor subsets at each splitting node (Breiman, 2001). Bootstrapping creates multiple training sets with shuffled and duplicated records, which makes the models robust since they get trained on a variety of data points. This also eventually prevents the model from overfitting since the results are aggregated across the different datasets (Zhao et al., 2020). Besides that, random forest is an ensemble technique that is developed by creating individual weak learners (decision trees), and grouped together to achieve a better prediction (Hastie et al., 2009). Prediction is a process of a model generating outcomes based on an unseen data. Moreover, the random selection of predictors allows for a less biased selection of features at each splitting node. This decorrelates the trees present in the forest and reduces the effect of multicollinearity. All these advantages could have contributed to an equal or better performance of random forest models in the past studies (Carvajal et al., 2018; Zhao et al., 2020).

Random forest machine learning has been applied for predicting water-borne and vector-borne diseases, e.g., cholera (Campbell et al., 2020), dengue (Carvajal et al., 2018; Khan et al., 2017; Zhao et al., 2020), malaria (Didier Barradas-Bautista, 2020), and tick-borne encephalitis (Uusitalo et al., 2020). It has also been used to model animal leptospirosis based on annual precipitation and temperature as well as socio-economic and landscape factors (Zakharova et al., 2021). The random forest model developed in Zakharova et al., (2021) ranked the independent variables based on the importance metric of Gini that reflects the variable's responsibility in splitting the output. However, the study did not further optimise the model by removing the less important (lower in the rank) variables. Eliminating less important or irrelevant variables can reduce the complexity of the model, which improves its run time, comprehensibility and performance (Kumar, 2014).

Besides that, hydrometeorological variability in the form of average and extreme indices have been used as drivers in past modelling studies. For example, simple average and extreme hydrometeorological indices, i.e., mean, sum, minimum and maximum, have been investigated (Chadsuthi et al., 2012; Cunha et al., 2019; Desvars et al., 2011; Gómez et al., 2021; Kupek et al., 2000; Radi et al., 2018; Rahmat et al., 2019; Schneider et al., 2012; Sumi et al., 2017; Weinberger et al., 2014), while more elaborate covariates that represented extreme dry and

wet conditions have also been used (Dhewantara et al., 2017; Ding et al., 2019; Ehelepola et al., 2019; Rahayu et al., 2018; Sánchez-Montes et al., 2015; Tassinari et al., 2008). However, none of the above studies have systematically analysed and compared the effects of different variables and their average and extreme indices on case predictions.

1.3 Research Questions

In this study, cross correlation analysis and the capabilities of the random forest algorithm were leveraged to answer the following research questions:

i. What hydrometeorological variables are highly cross correlated with leptospirosis and important in classifying the disease occurrence?

ii. Does prediction capacity change according to the type of index used as model features, whether in the form of average or extreme indices or their combination?

1.4 Aim and Objectives

The research aim is to compare the variable importance and predictive capacity of average and extreme hydrometeorological indices in the leptospirosis occurrence of three districts of Kelantan, which are subject to higher incidence rate and flooding events, through cross correlation analysis and random forest modelling technique. There are several objectives derived as below to achieve the aim of the study:

i. To analyse the importance of hydrometeorological indices through the lagged correlation analysis and mean decrease Gini (MDG) score.

ii. To cross-evaluate the accuracy, sensitivity and specificity of random forest classification models for leptospirosis that are built using average and extreme hydrometeorological indices with the feature subsets optimised mean decrease Gini (MDG).

1.5 Research Contribution

The research is a systematic comparison between the derived extreme and corresponding average hydrometeorological indices that correlate and classify leptospirosis. Apart from that, the random forest model for predicting human leptospirosis is a new application. Additionally, this research contributes to developing machine learning leptospirosis prediction models for Kelantan.

1.6 Research Scope

The study area only covers those which are susceptible to flooding events as the research focuses on extreme hydrometeorological events as well. Only three districts with higher leptospirosis incidence rates are selected to reduce complexity in a lumped model. Although leptospirosis cases occur among both humans and animals, the research develops prediction models for human leptospirosis only. Moreover, the research uses five hydrometeorological variables only including rainfall, streamflow, water level, relative humidity and temperature. These are secondary data, which are collected from the Drainage and Irrigation Department (DID) and Malaysian Meteorological Department (MetMalaysia). Lastly, the variable importance of the hydrometeorological indices is measured based on mean decrease Gini (MDG).

1.7 Thesis Outline

1.7.1 Introduction

This chapter provides a general overview on the research that covers the history and background and current information of the thesis problem. The problem to be addressed in the study is concisely described. The aim and objectives to address the problem are listed in this section. The extent of which the research is conducted, and its contributions are also discussed.

1.7.2 Literature Review

This chapter synthesises the information gained from the previously published works that are related to the field of study. This section presents (1) leptospirosis and (2) its relation to hydrometeorological variables, (3) the index representation and (4) temporal lags of hydrometeorological variables, (5) the leptospirosis modelling and prediction of former studies, (6) machine learning techniques used to perform leptospirosis predictions, (7) random forest machine learning and (7) the feature selection approaches and measures used in machine learning are presented in this section. Conclusion is drawn by reviewing the former studies and the research gap is identified.

1.7.3 Methodology

This chapter presents all the methods approached to conduct the research including case study selection, data collection, processing and analysis, model development and optimization. This section explains in detail on how the input features are generated and models are configured according to the types of hydrometeorological indices and settings.

1.7.4 Result and Discussion

This chapter highlights the main and important findings from the result. The research draws the findings from three main stages of the research i.e., cross correlation analysis, model development, model optimization using selected feature subsets based on an independent criterion. This chapter also interprets the meaning of the results with the support of literature at the last of each section. The possible reasons for obtaining such results and their importance to the study are discussed in this section. Unexpected results and the limitation of the findings are also discussed in this section.

1.7.5 Conclusion and Recommendation

This chapter summarises the conclusions drawn from the research findings. The recommendations on bringing the research forward have also been included.

REFERENCES

Adler, B., & de la Peña Moctezuma, A. (2010). Leptospira and leptospirosis. *Veterinary Microbiology*, *140*(3–4), 287–296. https://doi.org/10.1016/j.vetmic.2009.03.012

Adler B. (2015). History of leptospirosis and leptospira. *Current topics in microbiology and immunology*, *387*, 1–9. https://doi.org/10.1007/978-3-662-45059-8_1

Agampodi, S. B., Dahanayaka, N. J., Bandaranayaka, A. K., Perera, M., Priyankara, S., Weerawansa, P., Matthias, M. A., & Vinetz, J. M. (2014). Regional Differences of Leptospirosis in Sri Lanka: Observations from a Flood-Associated Outbreak in 2011. *PLoS Neglected Tropical Diseases*, *8*(1), e2626. https://doi.org/10.1371/journal.pntd.0002626

Ahangarcani, M., Farnaghi, M., Shirzadi, M. R., Pilesjö, P., & Mansourian, A. (2019). Predictive risk mapping of human leptospirosis using support vector machine classification and multilayer perceptron neural network. *Geospatial Health*, *14*(1). https://doi.org/10.4081/gh.2019.711

Ansdell, V. E. (2017). Chapter 23 - Leptospirosis. In *The Travel and Tropical Medicine Manual (Fifth Edition)*. Elsevier Inc. https://doi.org/10.1016/B978-0-323-37506-1.00023-4

Auret, L., & Aldrich, C. (2011). Empirical comparison of tree ensemble variable importance measures. *Chemometrics and Intelligent Laboratory Systems*, *105*(2), 157–170. https://doi.org/10.1016/j.chemolab.2010.12.004

Azimullah, A. Z., Aziah, B. D., & Fauziah, M. N. (2016). The rise of Leptospirosis in Kelantan 2014: Characteristics geographical pattern and associated factors. *International Journal of Public Health and Clinical Sciences*, *3*(4), 2289–7577. https://pdfs.semanticscholar.org/cbcc/72aa5c7f77b37dd477966bb90a5a51d85f4f.pdf

Babudieri, B. (1958). Animal reservoirs of leptospires. *Annals of the New York Academy of Sciences*, *70(3)*, 393-413. https://doi.org/10.1111/j.1749-6632.1966.tb45498.x

Barcellos, C., & Sabroza, P. C. (2001). The place behind the case: leptospirosis risks and associated environmental conditions in a flood-related outbreak in Rio de Janeiro. *Cadernos de Saúde Pública*, *17 Suppl*, 59–67. https://doi.org/10.1590/s0102-311x2001000700014

Batchelor, T. W. K., Stephenson, T. S., Brown, P. D., Amarakoon, D., & Taylor, M. A. (2012). Influence of climate variability on human leptospirosis cases in Jamaica. *Climate Research*, *55*(1), 79–90. https://doi.org/10.3354/cr01120

Benacer, D., Thong, K. L., Min, N. C., Verasahib, K. B, Galloway, R. L., Hartskeerl, R. A., Souris, M., & Zain, S. N. M. (2016). Epidemiology of human leptospirosis in Malaysia, 2004-2012. *Acta Tropica*, *157*, 162–168. https://doi.org/10.1016/j.actatropica.2016.01.031

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Bierque, E., Thibeaux, R., Girault, D., Soupé-Gilbert, M. E., & Goarant, C. (2020). A systematic review of Leptospira in water and soil environments. *PLoS ONE*, *15*(1), e0227055. https://doi.org/10.1371/journal.pone.0227055

Breiman, L. (2001). Random Forests. *Machine learning, 45(1), 5-32*. https://doi.org/10.1007/978-3-030-62008-0_35

Campbell, A. M., Racault, M. F., Goult, S., & Laurenson, A. (2020). Cholera risk: A machine learning approach applied to essential climate variables. *International Journal of Environmental Research and Public Health*, *17*(24), 9378. https://doi.org/10.3390/ijerph17249378

Cann, K. F., Thomas, D. R., Salmon, R. L., Wyn-Jones, A. P., & Kay, D. (2012). Extreme water-related weather events and waterborne disease. *Epidemiology and Infection*, *141*(4), 671–686. https://doi.org/10.1017/S0950268812001653

Carvajal, T. M., Viacrusis, K. M., Hernandez, L. F. T., Ho, H. T., Amalin, D. M., & Watanabe, K. (2018). Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. *BMC Infectious Diseases*, *18*(1), 1–15. https://doi.org/10.1186/s12879-018-3066-0

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. https://doi.org/10.1016/S1995-7645(12)60095-9

Chang, S., Cohen, T., & Ostdiek, B. (2018). What is the machine learning? *Physical Review D*, *97*(5), 56009.

https://doi.org/10.1103/PhysRevD.97.056009

Ciceroni, L., Pinto, A., Benedetti, E., Pizzocaro, P., Lupidi, R., Cinco, M., Gelosa, L., Grillo, R., Rondinella, V., Marcuccio, L., Mansueto, S., Ioli, A., Franzin, L., Giannico, F., & Cacciapuoti, B. (1995). Human leptospirosis in Italy, 1986-1993. *European Journal of Epidemiology*, *11*(6), 707–710.

Coelho, M. S., & Massad, E. (2012). The impact of climate on Leptospirosis in São Paulo, Brazil. *International Journal of Biometeorology*, *56*(2), 233–241. https://doi.org/10.1007/s00484-011-0419-4

Costa, F., Hagan, J. E., Calcagno, J., Kane, M., Torgerson, P., Martinez-Silveira, M. S., Stein, C., Abela-Ridder, B., & Ko, A. I. (2015). Global Morbidity and Mortality of Leptospirosis: A Systematic Review. *PLoS Neglected Tropical Diseases*, *9*(9), e0003898. https://doi.org/10.1371/journal.pntd.0003898

Cucchi, K., Liu, R., Collender, P. A., Cheng, Q., Li, C., Hoover, C. M., Chang, H. H., Liang, S., Yang, C., & Remais, J. V. (2019). Hydroclimatic drivers of highly seasonal leptospirosis incidence suggest prominent soil reservoir of pathogenic Leptospira spp. in rural western China. *PLoS Neglected Tropical Diseases, 13*(12), e0007968.

Cunha, M., Costa, F., Ribeiro, G. S., Carvalho, M. S., Reis, R. B., Nery Jr, N., Pischel, L., Gouveia, E. L., Santos, A. C., Queiroz, A., Wunder Jr, E. A., Reis, M. G, Diggle, P. J, & Ko, A. I. (2022). Rainfall and other meteorological factors as drivers of urban transmission of leptospirosis. *PLoS Neglected Tropical Diseases, 16*(4), e0007507. https://doi.org/10.1101/658872

Dechet, A. M., Parsons, M., Rambaran, M., Mohamed-Rambaran, P., Florendo-Cumbermack, A., Persaud, S., Baboolal, S., Ari, M. D., Shadomy, S. V., Zaki, S. R., Paddock, C. D., Clark, T. A., Harris, L., Lyon, D., & Mintz, E. D. (2012). Leptospirosis outbreak following severe flooding: A rapid assessment and mass prophylaxis campaign; Guyana, January-February 2005. *PLoS ONE*, *7*(7). https://doi.org/10.1371/journal.pone.0039672

Department of Statistics, Malaysia. (2010). Population distribution by local authority areas and Mukims.

Department of Irrigation and Drainage, Malaysia. (2017). Flood Management - Programme and Activities: Flood Prone Area in Malaysia.

Department of Irrigation and Drainage, Malaysia. (2020). Laman Web Rasmi Public InfoBanjir.

Derrick, T. R., & Thomas, J. M. (2004). Time-Series Analysis: The cross-correlation function. *Innovative Analyses of Human Movement*, 189–205.

Desvars, A., Jégo, S., Chiroleu, F., Bourhy, P., Cardinale, E., & Michault, A. (2011). Seasonality of human leptospirosis in Reunion Island (Indian Ocean) and its association with meteorological data. *PLoS ONE*, *6*(5), e20377. https://doi.org/10.1371/journal.pone.0020377

Dhewantara, P. W., Ipa, M., Riandi, M. U., Sadali, M., & Djati, A. P. (2017). Ecological niche model as tools to predict current and future distribution of Leptospirosis occurrence in western Java, Indonesia. Poster presented at the Conference of International Leptospirosis Society 2017, Indonesia.

Dhewantara, P. W., Lau, C. L., Allan, K. J., Hu, W., Zhang, W., Mamun, A. A., & 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. https://doi.org/10.1111/zph.12549

Dhewantara, P. W., Zhang, W., Al Mamun, A., Yin, W. W., Ding, F., Guo, D., Hu, W. & Magalhães, R. J. S. (2020). Spatial distribution of leptospirosis incidence in the Upper Yangtze and Pearl River Basin, China: Tools to support intervention and elimination. *Science of the Total Environment*, *725*, 138251.

Didier Barradas-Bautista. (2020). Random Forest and Deep Learning Performance on the Malaria DREAM Sub Challenge One Random Forest and Deep Learning Performance. *Research in Computing Science, 149*(5), 163-170.

Ding, G., Li, X., Li, X., Zhang, B., Jiang, B., Li, D., Xing, W., Liu, Q., Liu, X., & Hou, H. (2019). A time-trend ecological study for identifying flood-sensitive infectious diseases in Guangxi, China from 2005 to 2012. *Environmental Research*, *176*, 108577. https://doi.org/10.1016/j.envres.2019.108577

Dufour, B., Moutou, F., Hattenberger, A. M., & Rodhain, F. (2008). Global change: Impact, management risk approach and health measures - The case of Europe. *Revue Scientifique et Technique (International Office of Epizootics)*, *27*(2), 529–550. https://doi.org/10.20506/rst.27.2.1817

Dzulaikha, K., Nurul Yuziana, M. Y., Maizatulriah, J. J., & Marfiah, A. W. (2017, December). Association of Rainfall and the Occurrence of Pathogenic Leptospira spp. in Recreational Stream Water, Hulu Langat, Selangor. In *International Conference for Innovation in Biomedical Engineering and Life Sciences,* 119–124. Springer, Singapore. https://doi.org/10.1007/978-981-

10-7554-4_20

Ehelepola, N. D. B., Ariyaratne, K., & Dissanayake, W. P. (2019). The correlation between local weather and leptospirosis incidence in Kandy district, Sri Lanka from 2006 to 2015. *Global Health Action*, *12*(1), 1553283. https://doi.org/10.1080/16549716.2018.1553283

Fan, J., Upadhye, S., & Worster, A. (2006). Understanding receiver operating characteristic (ROC) curves. *Canadian Journal of Emergency Medicine*, *8*(1), 19-20.

Fefferman, N. H., O'Neil, E. A., & Naumova, E. N. (2005). Confidentiality and confidence: Is data aggregation a means to achieve both? *Journal of Public Health Policy*, *26*(4), 430–449. https://doi.org/10.1057/palgrave.jphp.3200029

Ganoza, C. A., Matthias, M. A., Collins-Richards, D., Brouwer, K. C., Cunningham, C. B., Segura, E. R., Gilman, R. H., Gotuzzo, E., & Vinetz, J. M. (2006). Determining risk for severe leptospirosis by molecular analysis of environmental surface waters for pathogenic Leptospira. *PLoS Medicine*, *3*(8), e308. https://doi.org/10.1371/journal.pmed.0030308

Garba, B., Bahaman, A. R., Bejo, S. K., Zakaria, Z., Mutalib, A. R., & Bande, F. (2018). Major epidemiological factors associated with leptospirosis in Malaysia. *Acta Tropica*, *178*, 242–247. https://doi.org/10.1016/j.actatropica.2017.12.010

Ghizzo Filho, J., Nazàrio, N. O., Freitas, P. F., Pinto, G. D. A., G., & Schlindwein, A. D. (2018). Temporal analysis of the relationship between leptospirosis, rainfall levels and seasonality, Santa Catarina, Brazil, 2005-2015. *Revista Do Instituto de Medicina Tropical de São Paulo*, *60*.

Gini, C. (1921). Measurement of inequality of incomes. *The economic journal*, *31*(121), 124-126.

Gómez, A. A., López, M. S., Müller, G. V., López, L. R., Sione, W., & Giovanini, L. (2022). Modeling of leptospirosis outbreaks in relation to hydroclimatic variables in the northeast of Argentina. *Heliyon, 8*(6), e09758. https://doi.org/10.1101/2021.07.06.21260095

Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural networks*, *1*(1), 17-61.

Guerra, M. A. (2013). Leptospirosis: Public health perspectives. *Biologicals*, *41*(5), 295–297. https://doi.org/10.1016/j.biologicals.2013.06.010

Guo, P., Liu, T., Zhang, Q., Wang, L., Xiao, J., Zhang, Q., Luo, G., Li, Z., He, J., Zhang, Y., & Ma, W. (2017). Developing a dengue forecast model using machinelearning: A case study in China. *PLoS Neglected Tropical Diseases, 11(10), e0005973.*

Gutierrez, J. D. (2021). Effects of meteorological factors on human leptospirosis in Colombia. *International Journal of Biometeorology*, *65*(2), 257–263. https://doi.org/10.1007/s00484-020-02028-2

Haake, D. A., & Levett, P. N. (2015). Leptospira in Humans. In *Journal of Biological Education,* *25*(3). https://doi.org/10.1080/00219266.1991.9655201

Hacker, K. P., Sacramento, G. A., Cruz, J. S., de Oliveira, D., Nery Jr, N., Lindow, J. C., Carvalho, M., Hagan, J., Diggle, P. J., Begon, M., Reis, M. G., Wunder, E. A., Ko, A. I., & Costa, F. (2020). Influence of rainfall on leptospira infection and disease in a tropical urban setting, Brazil. *Emerging Infectious Diseases*, *26*(2), 311–314. https://doi.org/10.3201/eid2602.190102

Han, H., Guo, X., & Yu, H. (2016, August). Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest. *Proceedings of 7th IEEE International Conference on Software Engineering and Service Sciences (ICSESS)*, 219–224. https://doi.org/10.1109/ICSESS.2016.7883053

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Springer Series in Statistics: The Elements of Statistical Learning. *The Mathematical Intelligencer*, *27*(2), 83–85. http://www.springerlink.com/index/D7X7KX6772HQ2135.pdf

Hayati, K. S., Sharifah, N., Salmiah, M. S., Edre, M. A., & Khin, T. D. (2018). Hot-spot and cluster analysis on legal and illegal dumping sites as the contributors of leptospirosis in a flood hazard area in Pahang, Malaysia. *Asian Journal of Agriculture and Biology*, (Special Issue), 78-82.

Hu, H., Wang, H., Wang, F., Langley, D., Avram, A., & Liu, M. (2018). Prediction of influenza-like illness based on the improved artificial tree algorithm and artificial neural network. *Scientific Reports*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-23075-1

Ismail, W. R., & Haghroosta, T. (2018). Extreme weather and floods in Kelantan

state, Malaysia in December 2014. *Research in Marine Sciences*, *3*(1), 231–244.

Jamaludin, N., Mohammed, N. I., Khamidi, M. F., & Wahab, S. N. A. (2015). Thermal Comfort of Residential Building in Malaysia at Different Micro-climates. *Procedia-Social and Behavioral Sciences*, *170*, 613–623. https://doi.org/10.1016/j.sbspro.2015.01.063

Joshi, Y. P., Kim, E. H., & 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), 1-11. https://doi.org/10.1186/s12879-017-2506-6

Khan, S., Ullah, R., Khan, A., Sohail, A., Wahab, N., Bilal, M., & Ahmed, M. (2017). Random Forest-Based Evaluation of Raman Spectroscopy for Dengue Fever Analysis. *Applied Spectroscopy*, *71*(9), 2111–2117. https://doi.org/10.1177/0003702817695571

Kira, R., Bilung, L. M., Ngui, R., Apun, K., & Su'ut, L. (2021). Spatially varying correlation between environmental conditions and human leptospirosis in Sarawak, Malaysia. *Tropical Biomedicine*, *38*(2), 31–39. https://doi.org/10.47665/TB.38.2.034

Kumar, V., & Minz, S. (2014). Feature Selection: A literature Review. *The Smart Computing Review*, *4*(3), 211-229. https://doi.org/10.6029/smartcr.2014.03.007

Kupek, E., de Sousa Santos Faversani, M. C., & de Souza Philippi, J. M. (2000). The relationship between rainfall and human leptospirosis in Florianópolis, Brazil, 1991-1996. *The Brazilian Journal of Infectious Diseases : An Official Publication of the Brazilian Society of Infectious Diseases*, *4*(3), 131–134.

Lau, C. L., Smythe, L. D., Craig, S. B., & 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. https://doi.org/10.1016/j.trstmh.2010.07.002

Levett, P. N. (2001). Leptospirosis. *Clinical Microbiology Reviews*, *14*(2), 296–326. https://doi.org/10.1128/CMR.14.2.296

Lin, W. Y., Hong, Q. J., Cai, Z. C., Kui, G. X., Gao, J. X., & Ping, H. E. (2014). An outbreak of leptospirosis in Lezhi county, China in 2010 may possibly be linked to rainfall. *Biomedical and Environmental Sciences*, *27*(1), 56–

59. https://doi.org/10.3967/bes2014.016

López, M. S., Müller, G. V., Lovino, M. A., Gómez, A. A., Sione, W. F., & Pomares, L. A. (2019). Spatio-temporal analysis of leptospirosis incidence and its relationship with hydroclimatic indicators in northeastern Argentina. *Science of the Total Environment*, *694,* 133651. https://doi.org/10.1016/j.scitotenv.2019.133651

Mason, M. R., Encina, C., Sreevatsan, S., & Muñoz-Zanzi, C. (2016). Distribution and Diversity of Pathogenic Leptospira Species in Peri-domestic Surface Waters from South Central Chile. *PLoS Neglected Tropical Diseases*, *10*(8), e0004895. https://doi.org/10.1371/journal.pntd.0004895

Matsushita, N., Ng, C. F. S., Kim, Y., Suzuki, M., Saito, N., Ariyoshi, K., Salva, E. P., Dimaano, E. M., Villarama, J. B., Go, W. S., & Hashizume, M. (2018). The non-linear and lagged short-term relationship between rainfall and leptospirosis and the intermediate role of floods in the Philippines. *PLoS Neglected Tropical Diseases*, *12*(4), e0006331. https://doi.org/10.1371/journal.pntd.0006331

Mawonike, R., & Mandonga, G. (2017). The Effect of Temperature and Relative Humidity on Rainfall in Gokwe Region, Zimbabwe: A Factorial Design Perspective. *International Journal of Multidisciplinary Academic Research*, *5*(2), 36–46.

Mayfield, H. J., Lowry, J. H., Watson, C. H., Kama, M., Nilles, E. J., & Lau, C. L. (2018). Use of geographically weighted logistic regression to quantify spatial variation in the environmental and sociodemographic drivers of leptospirosis in Fiji: a modelling study. *The Lancet Planetary Health*, *2*(5), e223–e232. https://doi.org/10.1016/S2542-5196(18)30066-4

Metcalf, C. J. E., Walter, K. S., Wesolowski, A., Buckee, C. O., Shevliakova, E., Tatem, A. J., Boos, W. R., Weinberger, D. M., & Pitzer, V. E. (2017). Identifying climate drivers of infectious disease dynamics: Recent advances and challenges ahead. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1860), 20170901. https://doi.org/10.1098/rspb.2017.0901

Mohammadinia, A., Alimohammadi, A., & Saeidian, B. (2017). Efficiency of geographically weighted regression in modeling human leptospirosis based on environmental factors in Gilan province, Iran. *Geosciences*, *7*(4), 136. https://doi.org/10.3390/geosciences7040136

Mohammed, H., Nozha, C., Hakim, K., & Abdelaziz, F. (2011). LEPTOSPIRA: Morphology, Classification and Pathogenesis. *Journal of Bacteriology &*

Parasitology, *2*(06). https://doi.org/10.4172/2155-9597.1000120

Radi, M. F. M., Hashim, J. H., Jaafar, M. H., Hod, R., Ahmad, N., Nawi, A. M., Baloch, G. M., Ismail, R., & Ayub, N. I. F. (2018). Leptospirosis outbreak after the 2014 major flooding event in Kelantan, Malaysia: a spatial-temporal analysis. *The American Journal of Tropical Medicine and Hygiene*, *98*(5), 1281–1295. https://doi.org/10.4269/ajtmh.16-0922

Naing, C., Reid, S. A., Aye, S. N., Htet, N. H., & Ambu, S. (2019). Risk factors for human leptospirosis following flooding: A meta-analysis of observational studies. *PLoS ONE*, *14*(5), e0217643. https://doi.org/10.1371/journal.pone.0217643

Nicodemus, K. K. (2011). Letter to the editor: on the stability and ranking of predictors from random forest variable importance measures. *Briefings in Bioinformatics*, *12*(4), 369–373. https://doi.org/10.1093/bib/bbr016

Pappachan, M. J., Sheela, M., & Aravindan, K. P. (2004). Relation of rainfall pattern and epidemic leptospirosis in the Indian state of Kerala. *Journal of Epidemiology and Community Health*, *58*(12), 1054. https://doi.org/10.1136/jech.2003.018556

Pappas, G., Papadimitriou, P., Siozopoulou, V., Christou, L., & Akritidis, N. (2008). The globalization of leptospirosis: worldwide incidence trends. *International Journal of Infectious Diseases*, *12*(4), 351–357. https://doi.org/10.1016/j.ijid.2007.09.011

Peterson, T. C., Folland, C., Gruza, G., Hogg, W., Mokssit, A., & Plummer, N. (2001). Report on the activities of the Working Group on Climate Change Detection and Related Rapporteurs 1998–2001. *Rep. WCDMP-47, WMO-TD 1071*, *March*, 143. http://etccdi.pacificclimate.org/docs/wgccd.2001.pdf

Picardeau, M. (2013). Diagnosis and epidemiology of leptospirosis. *Medecine et Maladies Infectieuses*, *43*(1), 1–9. https://doi.org/10.1016/j.medmal.2012.11.005

Rahayu, S., Adi, M. S., & Saraswati, L. D. (2018). Mapping of Leptospirosis Environmental Risk Factors and Determining the Level of Leptospirosis Vulnerable Zone in Demak District Using Remote Sensing Image. In *E3S Web of Conferences* (Vol. 31, p. 06003). EDP Sciences. https://doi.org/10.1051/e3sconf/20183106003

Rahmat, F., Ishak, A. J., Zulkafli, Z., Yahaya, H., & Masrani, A. (2019). Prediction model of leptospirosis occurrence for Seremban (Malaysia) using

meteorological data. *International Journal of Integrated Engineering*, *11*(4), 61–69. https://doi.org/10.30880/ijie.2019.11.04.007

Rahmat, F., Zulkafli, Z., Juraiza Ishak, A., Mohd Noor, S. B., Yahaya, H., & Masrani, A. (2020). Exploratory Data Analysis and Artificial Neural Network for Prediction of Leptospirosis Occurrence in Seremban, Malaysia Based on Meteorological Data. *Frontiers in Earth Science*, *8*, 377. https://doi.org/10.3389/feart.2020.00377

Robertson, C., Nelson, T. A., & Stephen, C. (2012). Spatial epidemiology of suspected clinical leptospirosis in Sri Lanka. *Epidemiology and Infection*, *140*(4), 731–743. https://doi.org/10.1017/S0950268811001014

Sánchez-Montes, S., Espinosa-Martínez, D. V., Ríos-Muñoz, C. A., Berzunza-Cruz, M., & Becker, I. (2015). Leptospirosis in Mexico: Epidemiology and potential distribution of human cases. *PLoS ONE*, *10*(7), e0133720. https://doi.org/10.1371/journal.pone.0133720

Schneider, M. C., Nájera, P., Aldighieri, S., Bacallao, J., Soto, A., Marquiño, W., Altamirano, L., Saenz, C., Marin, J., Jimenez, E., Moynihan, M., & Espinal, M. (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. https://doi.org/10.3390/ijerph9113883

Schneider, M. C., Najera, P., Pereira, M. M., Machado, G., dos Anjos, C. B., Rodrigues, R. O., Cavagni, G. M., Muñoz-Zanzi, C., Corbellini, L. G., Leone, M., Buss, D. F., Aldighieri, S., & Espinal, M. A. (2015). Leptospirosis in Rio Grande do Sul, Brazil: An Ecosystem Approach in the Animal-Human Interface. *PLoS Neglected Tropical Diseases*, *9*(11), e0004095. https://doi.org/10.1371/journal.pntd.0004095

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Sehgal, S. C., Sugunan, A. P., & Vijayachari, P. (2002). Outbreak of leptospirosis after the cyclone in Orissa. *The National Medical Journal of India*, *15*(1), 22–23.

Settouti, N., Daho, M. E. H., & Chikh, M. A. (2017). Using conditional inference forest to identify variable importance. *International Journal of Bioinformatics Research and Applications*, *13*(2), 95–108. https://doi.org/10.1504/IJBRA.2017.083129

Simundic, A. M. (2009). Measures of diagnostic accuracy: basic definitions. *ejifcc, 19*(4), 203.

Socolovschi, C., Angelakis, E., Renvoisé, A., Fournier, P. E., Marié, J. L., Davoust, B., Stein, A., & Raoult, D. (2011). Strikes, flooding, rats, and leptospirosis in Marseille, France. *International Journal of Infectious Diseases*, *15*(10), e710-e715. https://doi.org/10.1016/j.ijid.2011.05.017

Sohail, M. L., Khan, M. S., Ijaz, M., Naseer, O., Fatima, Z., Ahmad, A. S., & Ahmad, W. (2018). Seroprevalence and risk factor analysis of human leptospirosis in distinct climatic regions of Pakistan. *Acta Tropica*, *181*, 79–83. https://doi.org/10.1016/j.actatropica.2018.01.021

Soo, Z. M. P., Khan, N. A., & Siddiqui, R. (2020). Leptospirosis: Increasing importance in developing countries. *Acta Tropica*, *201*, 105183. https://doi.org/10.1016/j.actatropica.2019.105183

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*(1), 1-21. https://doi.org/10.1186/1471-2105-8-25

Sumi, A., Telan, E. F. O., Chagan-Yasutan, H., Piolo, M. B., Hattori, T., & Kobayashi, N. (2017). Effect of temperature, relative humidity and rainfall on dengue fever and leptospirosis infections in Manila, the Philippines. *Epidemiology and Infection*, *145*(1), 78–86. https://doi.org/10.1017/S095026881600203X

Suwanpakdee, S., Kaewkungwal, J., White, L. J., Asensio, N., Ratanakorn, P., Singhasivanon, P., Day, N. P. J., & Pan-Ngum, W. (2015). Spatio-temporal patterns of leptospirosis in Thailand: Is flooding a risk factor? *Epidemiology and Infection*, *143*(10), 2106–2115. https://doi.org/10.1017/S0950268815000205

Tassinari, W. S., Pellegrini, D. C., Sá, C. B., Reis, R. B., Ko, A. I., & Carvalho, M. S. (2008). Detection and modelling of case clusters for urban leptospirosis. *Tropical Medicine and International Health*, *13*(4), 503–512. https://doi.org/10.1111/j.1365-3156.2008.02028.x

Togami, E., Kama, M., Goarant, C., Craig, S. B., Lau, C., Ritter, J. M., Imrie, A., Ko, A. I., & Nilles, E. J. (2018). A large leptospirosis outbreak following successive severe floods in Fiji, 2012. *American Journal of Tropical Medicine and Hygiene*, *99*(4), 849. https://doi.org/10.4269/ajtmh.18-0335

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. https://doi.org/10.2436/im.v7i1.9442

Ucar, M. K., Nour, M., Sindi, H., & Polat, K. (2020). The Effect of Training and Testing Processon Machine Learning in Biomedical Datasets. *Mathematical Problems in Engineering*, *2020*. https://downloads.hindawi.com/journals/mpe/2020/2836236.pdf

Uusitalo, R., Siljander, M., Dub, T., Sane, J., Sormunen, J. J., Pellikka, P., & Vapalahti, O. (2020). Modelling habitat suitability for occurrence of human tick-borne encephalitis (TBE) cases in Finland. *Ticks and tick-borne diseases, 11*(5), 101457.

Van Stralen, K. J., Stel, V. S., Reitsma, J. B., Dekker, F. W., Zoccali, C., & Jager, K. J. (2009). Diagnostic methods I: sensitivity, specificity, and other measures of accuracy. *Kidney international*, *75*(12), 1257-1263.

Vega-Corredor, M. C., & Opadeyi, J. (2014). Hydrology and public health: linking human leptospirosis and local hydrological dynamics in Trinidad, West Indies. *Earth Perspectives*, *1*(1), 1-14. https://doi.org/10.1186/2194-6434-1-3

Weinberger, D., Baroux, N., Grangeon, J. P., Ko, A. I., & Goarant, C. (2014). El Niño Southern Oscillation and Leptospirosis Outbreaks in New Caledonia. *PLoS Neglected Tropical Diseases*, *8*(4), e2798. https://doi.org/10.1371/journal.pntd.0002798

Wijerathne, K. B. P. C. A., & Senevirathna, E. M. T. K. (2018). Identify the risk for leptospirosis disease during flooding periods (Special reference to Medirigiriya Divisional Secretariat Division in Polonnaruwa district). In *Procedia Engineering, 212*, 101-108. https://doi.org/10.1016/j.proeng.2018.01.014

World Health Organization. (2001). *WHO recommended standards and strategies for surveillance, prevention and control of communicable diseases* (No. WHO/CDS/CPE/SMT/2001.13). World Health Organization. https://www.who.int/zoonoses/diseases/Leptospirosissurveillance.pdf

World Health Organization. (2003). *Human leptospirosis: guidance for diagnosis, surveillance and control* (No. WHO/CDS/CSR/EPH 2002.23). World Health Organization.

World Health Organization. (2011). Report of the Second Meeting of the

Leptospirosis Burden Epidemiology Reference Group.

Zakharova, O. I., Korennoy, F. I., Iashin, I. V., Toropova, N. N., Gogin, A. E., Kolbasov, D. V., Surkova, G. V., Malkhazova, S. M., & Blokhin, A. A. (2021). Ecological and Socio-Economic Determinants of Livestock Animal Leptospirosis in the Russian Arctic. *Frontiers in Veterinary Science*, *8*, 658675. https://doi.org/10.3389/fvets.2021.658675

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B., & Zwiers, F. W. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(6), 851–870. https://doi.org/10.1002/wcc.147

Zhao, J., Liao, J., Huang, X., Zhao, J., Wang, Y., Ren, J., Wang, X., & Ding, F. (2016). Mapping risk of leptospirosis in China using environmental and socioeconomic data. *BMC Infectious Diseases*, *16*(1), 1–10. https://doi.org/10.1186/s12879-016-1653-5

Zhao, N., Charland, K., Carabali, M., Nsoesie, E. O., Maheu-Giroux, M., Rees, E., Yuan, M., Garcia Balaguera, C., Jaramillo Ramirez, G., & Zinszer, K. (2020). Machine learning and dengue forecasting: Comparing random forest and artificial neural networks for predicting dengue burden at national and sub-national scales in Colombia. *PLoS Neglected Tropical Diseases, 14*(9), e0008056.