SPATIO TEMPORAL FORECASTING OF DENGUE OUTRBREAKS USING MACHINE LEARNING

Manju Lasantha Fernando

168061F

Degree of Master of Science

Department of Computer Science & Engineering

University of Moratuwa

Sri Lanka

July 2019

SPATIO TEMPORAL FORECASTING OF DENGUE OUTRBREAKS USING MACHINE LEARNING

Manju Lasantha Fernando

168061F

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree Master of Science in Computer Science and Engineering

Department of Computer Science & Engineering

University of Moratuwa

Sri Lanka

July 2019

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The above candidate has carried out research for the Masters Thesis/Dissertation under my supervision.

Signature of the Supervisor: Date:

ACKNOWLEDGEMENTS

I am sincerely grateful for the advice and guidance of my supervisor Dr. Amal Shehan Perera. Without his insight, encouragement, and his willingness to be available at all times for advice and direction, this project would have not been completed. His consistent positive outlook and resourcefulness guided me through each stumbling block and made sure I carried this study to its conclusion. It would be amiss of me if I did not give similar thanks to Sriganesh Lokanathan, my team lead of the big data team at LIRNEasia, for his advice, guidance, his words of encouragement and the insights provided by him regarding this project work.

I would like to extend my gratitude to my progress review committee, Prof. Nalin Wickremasinghe and Dr. Surangika Ranathunga. Their valuable insights and guidance helped me immensely in shaping the direction of this research.

I would like to thank the entire staff of the Department of Computer Science and Engineering and my colleagues at LIRNEasia for providing their help in various ways during the course of this work, as well as for providing me with the adequate resources to conduct this research.

This work was partially funded by LIRNEasia through a grant from IDRC-Canada as well as from a research grant awarded by the Senate Research Committee, University of Moratuwa, for which I am extremely grateful.

I would also like to express my gratitude to all my friends who lent their support in numerous ways by engaging in discussion, questioning my assumptions and providing knowledge on different domains related to this work. Finally, I would like to thank my family, and especially my parents, for the unconditional support and encouragement provided by them to complete this study to the best of my ability.

ABSTRACT

Spatio Temporal Forecasting of Dengue Outbreaks using Machine Learning

Dengue is one of the most critical public health concerns in Sri Lanka which imposes a severe economic and welfare burden on the nation annually. Prior work has shown that there are multiple factors that contribute to propagation of dengue, including sociological factors such as human mobility. Therefore, it is a non-trivial task to model the propagation of this disease accurately at a regional level. However, accurate quantitative modeling approaches that can predict dengue incidence for a public health administrative division would be invaluable in allocating valuable public health resources and preventing sudden disease outbreaks.

In this study, we make use of large-scale pseudonymized call detail records of approximately 10 million mobile phone subscribers to derive human mobility patterns that can contribute towards disease propagation. We develop 3 distinct proxy indicators for human mobility based on different assumptions and evaluate the suitability of each indicator to accurately model the disease transmission dynamics of dengue. Using the proxy measures developed by us, we go on to show that human mobility has a significant impact on the disease incidence at a regional level, even if the disease is already endemic to a given region.

Combining these proxy mobility indicators with other climatic factors that is known to affect dengue incidence, we build multiple predictive models using different machine learning methods to predict dengue incidence 2 weeks ahead of time for a given MOH division. By introducing an automated input feature selection method based on genetic algorithms, we show that we are able to improve the predictive accuracy of our models significantly, with predictive models based on XGBoost yielding the best performance, with an R^2 of 0.935 and RMSE of 7.688.

Keywords: disease outbreak forecasting; human mobility models; mobile network big data; machine learning applications;

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
ARIMA	Autoregressive Integrated Moving Average
BTS	Base Transceiver Station
CDR	Call Detail Record
DALY	Disability-Adjusted Life Years
DHF	Dengue Hemorrhaegic Fever
DSS	Dengue Shock Syndrome
GA	Genetic Algorithm
LASSO	Least Absolute Shrinkage and Selection Operator
LS-SVM	Least Squares - Support Vector Machines
MC	Municipal Council
MOH	Medical Officer of Health
NDVI	Normalized Difference Vegetation Index
NN	Neural Networks
RF	Random Forests
RMSE	Root Mean Squared Error
RNA	Ribonucleic Acid
SEI	Susceptiple-Exposed-Infected
SEIR	Susceptiple-Exposed-Infected-Recovered
SIR	Susceptiple-Infected-Recovered
SVM	Support Vector Machines
SVR	Support Vector Regression
WHO	World Health Organization

TABLE OF CONTENTS