

# ANNEX 19: PREDICTIVE MODEL FOR THE DENGUE INCIDENCES IN SRI LANKA USING MOBILE NETWORK BIG DATA

Dharmawardana, K.G.S. ;Lokuge, K. S.;Dassanayake, P. S. B. ;Sirisena, M. L.;Fernando, Lasantha;Perera, Amal Shehan;Lokanathan, Sriganesh;  
;

© 2018, LIRNEASIA



This work is licensed under the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted use, distribution, and reproduction, provided the original work is properly credited.

Cette œuvre est mise à disposition selon les termes de la licence Creative Commons Attribution (<https://creativecommons.org/licenses/by/4.0/legalcode>), qui permet l'utilisation, la distribution et la reproduction sans restriction, pourvu que le mérite de la création originale soit adéquatement reconnu.

*IDRC Grant/ Subvention du CRDI: 108008-001-Leveraging Mobile Network Big Data for Developmental Policy*

## Annex 19: Predictive Model for the Dengue Incidences in Sri Lanka Using Mobile Network Big Data.

K.G.S. Dharmawardana<sup>1</sup>, J.N. Lokuge<sup>1</sup>, P.S.B. Dassanayake<sup>1</sup>, M.L. Sirisena<sup>1</sup>, M.L. Fernando<sup>1,2</sup>, A.S. Perera<sup>1</sup>, and S. Lokanathan<sup>2</sup>

1 Department of Computer Science and Engineering University of Moratuwa Katubedda, Sri Lanka

Email: {suchira.12, janitha.12, pramod.12, menaka.12, lasantha.06, shehan}@cse.mrt.ac.lk

2 LIRNEasia 12, Balcombe Place Colombo 08, Sri Lanka

Email: {lasantha,sriganesh}@lirneasia.net

**Abstract**—Multiple studies in recent literature have shown that human mobility has a considerable effect on the importation of dengue to an immunologically dengue 'naive' regions. The purpose of this study is to validate whether this fact is applicable to dengue endemic regions as well as assess the impact of human mobility when developing predictive models for dengue incidence of a particular geographical region. The information regarding human population mobility of Sri Lanka was derived using Mobile Network Big Data (MNBD) due to its better representativity and accuracy over other conventional data sources such as census data or surveys. The derived mobility values for each region of the country is then weighted using reported past dengue cases of the relevant region. We used Artificial Neural Networks and XGBoost to predict the dengue incidence for selected administrative regions within Western Province of Sri Lanka. Accuracy of the prediction for dengue incidence improved on average when considering all the administrative regions included in our study by incorporating weighted human mobility. Our study shows that there is an impact of human mobility on the spread of Dengue fever in Sri Lanka while introducing a generalizable methodology to fuse big data sources with traditional data sources in developing predictive models using machine learning techniques. **Index Terms**—Mobile Network Big Data, CDR, Human Mobility, Dengue, Machine Learning Applications

### I. INTRODUCTION

Dengue fever is a mosquito-borne tropical disease caused by the dengue virus, transmitted by the mosquitoes of the genus *Aedes*, mainly *Aedes aegypti* and *Aedes albopictus* [1]. In Sri Lanka, dengue has become a growing menace that causes considerable financial and social impact to the society as a whole each year. In 2015 there were total of 29,777 reported dengue cases, but in 2016 that count was nearly doubled with a total of 55,150 reported dengue cases [2]. Currently there is no effective vaccine for dengue accepted by the National Medicines Regulatory Authority of Sri Lanka and it would take several years to get an approval for such vaccine [3]. Therefore the government and public health sector officials invest heavily on prevention techniques as well as on the treatment of dengue patients. The value addition of our implementation is to provide the ability to model the spread of disease beforehand and predict the areas at risk, which will provide a better lead time for the National Dengue Control Unit to apply preventive measures in advance and reduce the number of patients who fall victim to acute Dengue fever. The Mobile Network Big Data (MNBD) consists of different types of data categories such as Call Detail Records (CDR), Internet access usage records, airtime recharge records etc. Out of these subcategories, CDRs include an identifier for the Base Transceiver Station (BTS) associated with a particular call along with other meta information that is needed for billing purposes of a mobile

network operator. A mobile subscriber network with 10 million subscribers can generate billions of records monthly which would account for terabytes of data. Until the significant advances and the research interest that has been developing around the field of MNBD lately, these enormous volumes of data had not been utilized for any other purpose other than to be collected and archived mainly as a record of proof for billing activities of mobile operators. It is stated that Sri Lanka has a mobile penetration of over 90% and coverage of nearly 100% of the landmass (Telecommunication Regulatory Commission of Sri Lanka, 2014) and this percentage keeps increasing every year. Therefore CDR data can be taken as a proxy which would provide insight into human mobility patterns within the country, with better accuracy and representativity when compared to conventional models used in transportation theory and data collected from surveys.

In our study, the objective is to construct a predictive model for any given Medical Officer of Health (MOH) division, which is the smallest medical administrative district in Sri Lanka, by taking human mobility into account. This decision was based on literature which shows a significant link between the spread of epidemics and human mobility when an infectious disease is propagated across geographical regions [4], [5], [6]. Most of the previous studies done in Sri Lanka related to modeling spatial propagation of Dengue have compared the impact of human mobility as a contributing factor [7], [8]. In our work, several MOH divisions nearer to Colombo were chosen as focus regions because of the high population density and the relatively high number of dengue cases reported. Furthermore, based on the fact that the number of reported dengue cases in Colombo is highest when compared to other areas, Colombo can be considered as a driver city for the spread of dengue in Sri Lanka. Therefore the selected MOH divisions are good candidates to study the impact of mobility due to its high mobility and proximity to Colombo.

## II. RELATED WORK

### A. *Mobile Network Big Data*

A CDR contains different data fields such as a BTS identifier, the calling party number, the called party number, call time and the call duration. The BTS identifier for a particular record is used to obtain the spatial coordinates of the BTS, which is indicative of the general region or location of a mobile subscriber at a particular time. This information can then be used to derive aggregated human mobility information at a much larger scale. The extraction of human mobility information from CDR data has significant applications in transport and traffic controlling [9], predicting the spatial spread of infectious diseases, planning of infrastructure projects and detecting crimes [10].

### B. *Human mobility models constructed using CDR data*

In epidemiological mobility models there is significant motivation to identify the most frequented MOH divisions of a mobile subscriber as the Home or Work location, due to the fact that transmission dynamics of the disease would get affected based on this location type. The significance of this fact is reflected from the different epidemiological human mobility models discussed below. The research done in Kenya to find the quantitative impact of human mobility on malaria used CDR data to find human mobility [11]. Their human mobility model is able to identify the dynamics of human carriers that drive disease importation between regions. Wesolowski et. al have also used CDR data for the purpose of studying the impact of human mobility on spread of dengue in Pakistan [6]. The mobility model [6] used in this study can be considered a much improved model compared to previously discussed mobility model [11], which did not incorporate detailed epidemiological

reasoning. In a research done in Senegal, researchers were able to identify that the major cause for a cholera outbreak in the country was due to a mass gathering which took place during the initial phase of the outbreak by modelling human mobility using mobile phone data [5]. Mobility model of this research has considered not only the number of trips by subscribers from one region to another, but also the amount of time which a subscriber spends in each region. Moreover, this work shows that the assumptions that they made to calculate the time spent by a user in a region is feasible and justifiable, and we have made use of these same assumptions in our work as well.

### *C. Machine Learning approaches*

Recent research studies have adopted machine learning models for dengue incidence analysis. In contrast to mathematical models which only consider the entomology and prespecified and pre-identified factors that affect the spread of dengue arbovirus, the machine learning models are able to capture the effect of many factors as well as their unseen variations which cannot be captured by mathematical models. However, the accuracy of results will vary with the approach the researchers have taken for the model construction. For example, rather than using raw weather data, Y. Wu et al [12] have done a wavelet transformation for data normalization and selected the important features through a genetic algorithm. Artificial Neural Networks(ANN) are a very powerful computational methodology that can be used to predict an output value, classify an object, recognize patterns, and complete a known pattern [13]. ANNs have been extensively used in the field of predicting Dengue outbreaks. Rachata et al researched on the use of ANNs with an entropy technique to build a predictive model for Dengue outbreaks in Thailand [14]. When we consider research studies done in Sri Lanka, there are two recent studies that use ANN for dengue predictions. One of those research studies have introduced an ANN which is capable of predicting the dengue outbreak in Kandy district, Sri Lanka, using past weather patterns and past dengue cases reported in the region as inputs to the network[7]. The other research study attempts to predict dengue outbreaks with a model consisting of three main components and one of the components is an ANN[15].

## **III. HUMAN MOBILITY MODEL IMPLEMENTATION**

Human mobility values generated from CDRs at cell tower level had to be aggregated to MOH level since our epidemiological predictive model predicts at an MOH level. Subscribers connected to a cell towers were distributed to the MOH division according to the proportion of the area of the MOH division that overlaps with a voronoi cell of the BTS distribution of Sri Lanka. The mobility model was developed under the assumption that the number of calls taken or received by a subscriber in a particular region is proportional to the amount of time spent in that region, which has been used in similar studies in contemporary literature [5]. This mobility model has not only considered the number of trips from one MOH division to another, but also the amount of time that a subscriber spends in each MOH. The goal of our mobility model is to find a mobility value for each MOH division for a given week which can be taken as an indicator for the aggregate time fraction spent by a potentially infected subscriber in a particular MOH division who came from another MOH.

**Identifying the home location of mobile subscriber:** A phone users home MOH is identified from the most frequently communicated cell tower during nights over the study period. In our study, we select the time period 21h00 to 05h00 as the home hours of users since it is the period of time with minimal variation in average mobility in Sri Lanka according to the research done by Danaja et al [9].

**Probabilistic mobility value for a subscriber:** Once the home MOHs are derived a mobility value for a subscriber is calculated by the total CDRs made by a subscriber who is not a resident of that particular MOH in a given week. The value is divided by the total CDRs produced by the subscriber during the week which represents a probabilistic value of the presence of the subscriber in that MOH.  $S$  is the set of all the subscribers.  $s_p$  is any subscriber from the set of all subscribers  $S$ .  $M$  is the set of all MOHs in the country.  $m_q$  is any MOH from the MOH set  $M$ .  $W$  is the set of all the weeks.  $w_r$  is any week from the week set  $W$ . Probability of subscriber  $s_p$  at MOH division  $m_q$  is defined as,

$$P(s_p, m_q, w_r) = \frac{CDR(s_p, m_q, w_r)}{\sum_i^M CDR(s_p, m_i, w_r)} \quad (1)$$

where  $\forall s_p \in S, \forall m_q \in \{M - Home(s_p)\}$

**Mobility value between two MOH divisions:** Probabilistic value per subscriber is aggregated to get the total time fraction of all subscribers travelling from home MOH to destination MOH.  $Q_{ij}(t)$  is the aggregation of all time fractions of each subscriber living in MOH  $i$  spent in MOH  $j$  during week  $t$ .

$$Q_{ij}(t) = \sum_k^{N_i} P(s_k, m_i, w_t) \quad (2)$$

**Non weighted incoming mobility to MOH  $m_j$  :** Weekly incoming mobility value for a particular MOH is generated by aggregating the total probabilistic mobility from a all other MOH division to a particular MOH.  $Mob(m_j, t)$  is the non weighted incoming mobility to MOH  $m_j$  during week  $t$ . This is calculated by aggregating  $Q_{ij}(t)$  from every other MOH  $i$  to particular MOH  $m_j$ .

$$Mob(m_j, t) = \sum_i^M Q_{ij}(t) \quad (3)$$

Figure 1 shows the non weighted weekly incoming mobility variation for three MOH areas of Colombo district (urban region and capital of Sri Lanka) throughout the year 2013. Figure 2 shows the non weighted weekly incoming mobility variation for three MOH areas of Badulla district (remote rural area) throughout the year 2013. 15th and 16th weeks of the year fall on the new year festival season of Sri Lanka. One can see there is a significant drop in incoming mobility value in Colombo (city) while significant rise in incoming mobility in Badulla (remote area) during this period. This is due to people traveling to their home towns during the new year festival season. This seasonal variation of human mobility has been captured by our mobility model.

**Weighted incoming mobility to MOH  $m_j$  :**  $Q_{ij}(t)$  value is a mobility value which represent the total time fraction spent at MOH  $j$  by subscribers traveled from MOH  $i$  to MOH  $j$ . But this does not represent the effect of the mobility of the dengue infected people. Therefore, in

the weighted mobility, we are multiplying the mobility value from a source MOH to destination MOH by the number of infected humans ( $I_h$ ) in the source MOH.

$$Mob_{weighted}(m_j, t) = \sum_i^M \frac{Q_{ij}(t) * I_h(i, t)}{N_h(i)} \quad (4)$$

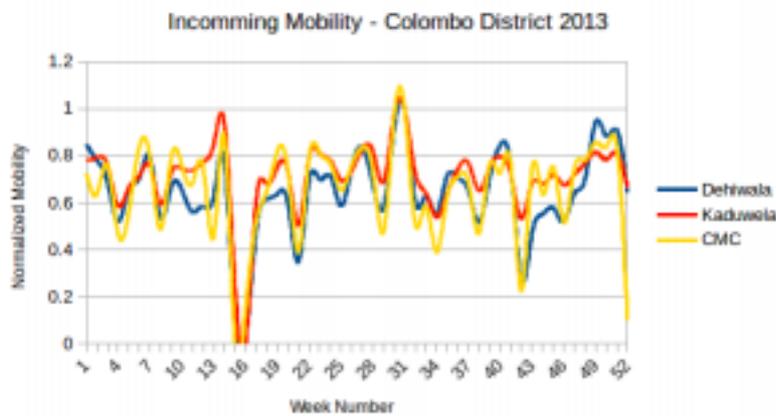


Fig. 1. Incoming Mobility - Colombo District 2013

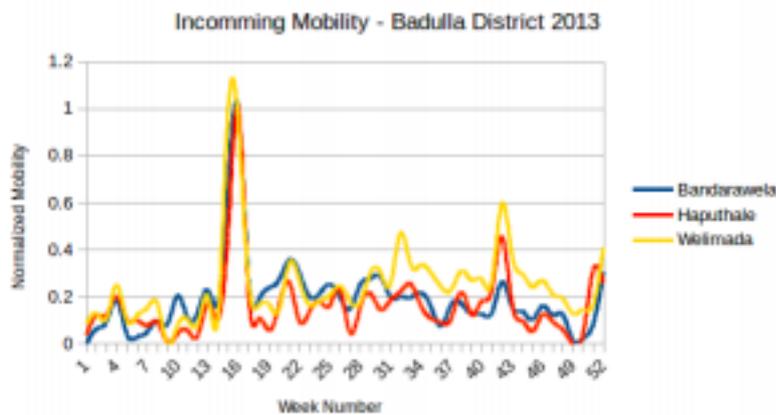


Fig. 2. Incoming Mobility - Badulla District 2013

$Mob_{weighted}(m_j, t)$  is the dengue cases weighted incoming mobility value calculated for the MOH  $j$  for the week  $t$ . Mobility value from each MOH  $i$  to  $j$  ( $Q_{ij}(t)$ ) is multiplied by the number infected host in each MOH  $i$  ( $I_h(i, t)$ ) divided by the population of that MOH ( $N_h(i)$ ). Then we aggregate this value from every MOH  $i$  to MOH  $j$  to get the  $Mob_{weighted}(m_j, t)$ . Population estimates from CDR data is used rather than from census data to avoid having a conflict due to different scales in census data vs. CDR data, since all previous calculations were carried out on values derived from CDR data.

## IV. MATERIALS AND METHODS

### A. CDR

We analysed CDR data of more than 10 million mobile phone subscribers, belonging to multiple mobile phone operators in Sri Lanka for a period spanning more than 1 year. The dataset used in our research consists of a pseudonymized identifier in place of the calling party number as well as the called party number, so that no privately identifiable information is available. CDR data was initially mapped to a 1km voronoi cell area. There were approximately 2500 cell regions after this mapping was done. Since we have data spanning for only just over 1 year, we made an assumption that the aggregate human mobility patterns will not drastically change between consecutive years. Apache Hadoop HDFS file system was used as the data storage framework to store the CDR data while Apache Spark was used to analyse the data.

### B. Dengue Data

The past dengue cases are the reported dengue cases in each MOH area in Sri Lanka, for the time period from January 2012 to December 2014 - for a total of 156 weeks, on a weekly basis. This data was acquired from the National Dengue Control Unit, Sri Lanka.

### C. Meteorological Data

Weather data includes the daily rainfall and temperature data for year 2013. This data was acquired from the Department of Meteorology, Sri Lanka. Rainfall data was obtained from 112 weather stations around the country, and the temperature data was from 22 stations.

### D. Vegetation Index

For our research, we also consider ground vegetation cover as a predictor for dengue incidence [16]. Vegetation indices are calculated for different MOHs separately from 2013 to 2016 using MOD13Q1 data set provided by NASA Land Data Products and Services.

### E. Feature Selection and Feature Reduction

All models developed for our work have a lead time of four weeks. In order to achieve this lead time all the features that were included had to be considered with a minimum lag of four weeks up to a maximum lag of 12 weeks. Feature reduction process was carried out by filtering in the features with correlation values higher than the median of correlation values of all features. Among the features that were selected as input data, cases and mobility values which are lagging back to 4 weeks showed the highest correlation.

### F. Artificial Neural Network and TensorFlow

Although Artificial Neural Networks (ANN) have been used extensively in the field of predicting dengue outbreaks deep neural networks have not been used often. [14], [7]. In this research we have used TensorFlow which is an open source deep learning library developed by the Google Brain team [17]. Separate ANNs were trained for several MOH areas after input features were normalized by removing the mean and scaling to unit variance before been fed into the neural network. When designing the ANN, the number of input layer neurons is chosen according to the number of attributes in the optimum data set after feature reduction. Since this is a regression ANN which predicts the number of dengue

cases in four weeks lead time it has only one neuron in the output layer. Two hidden layers were used and a range for the number of neurons in each hidden layer was evaluated based on recommendations in literature [18] to achieve optimality. Rectifier functions are used as the activation function for the hidden layers, with back propagation as the learning algorithm which was used in similar related work[7] and Adagrad [19] was used as the optimizing algorithm. The dropout technique [20] was used to avoid over fitting which is a serious problem in ANN. ANN for each MOH was trained using 104 weeks (2012 and 2013) and tested using 52 weeks (2014).

### *G. XGBoost*

Extreme gradient boosting is a supervised machine learning technique that is scalable, fast and has become highly popular for many machine learning problems. The technique was published in 2016[21], although gradient boosting was already a well known technique and the XGBoost library was available and in use widely well before that. The model is based on classification and regression trees (CART) , where a real score is taken into consideration with the ensemble approach of multiple trees. Low latency and high accuracy makes XGBoost a significant predictive learning technique for dengue forecasting when compared to other predictive learning techniques. The same training set and testing set as used in ANN model above was used. There are several parameters which holds a significant importance when it comes to XGBoost. Learning rate, number of rounds, depth and the objective function are some of them. For the purpose of this study, the optimum results were obtained by comparing the outcome from multiple different parameter settings. The learning rate was tuned in the range from 0.006 to 0.01. The number of rounds were looped with a incremental value of 100 from 1000 to 1600. The depth was varied between 2-10 to avoid over fitting.

## V. RESULTS

We used R2 value as the metric to measure the performance of our predictive models. Figure 3 and 4 shows the prediction graph for Maharagama MOH area without any incorporation of human mobility. The best results when we used mobility values are illustrated in Figure 5 and 6. The R2 value has improved when mobility is used as an input to machine learning methodology. The best results for five MOH areas viz, MC - Colombo, Dehiwala, Maharagama, Moratuwa, Kalutara(North) are shown in the Table I.

## VI. DISCUSSION

There are key benefits of using MNBD to construct human mobility models. Namely, the greater spatio-temporal granularity when compared with traditional survey information, negligible cost associated with the collection of data and most importantly it acts as a single source for understanding different aspects of mobility. While there are many advantages of MNBD, we cannot neglect the inherent limitations of these datasources as well. These limitations vary with respect to the context of a particular country. But in general, the main limitations of MNBD would be sampling biases and sparsity of data in some regions. Sampling biases arise due to high/low activity users and varying mobile phone penetration in different regions. When considering Sri Lanka, there is less than 25 records per day on average for 90% of the users, which has a significant effect on the sparsity of data [9]. However, when considering the granularity and density of the CDR data when compared to traditional survey based methods, it can be argued that CDR based models are able to provide much more accurate and interesting insights into human movement patterns which we have not being able to replicate with other data sources.

Various mobility models can be obtained using MNBD as presented in this paper. Yet the logical reasoning behind the mobility model construction should align well with the context of the problem. The human mobility model employed in our study has only considered the Home location of the mobile phone subscribers. According to a recent survey conducted to find about habitats of dengue vectors, state schools and construction sites had been identified as the major breeding places for dengue mosquitoes in Sri Lanka [22]. Therefore, special consideration should be given for defining Work locations of the mobile phone subscribers as well, which is an important location type according to the existing entomological and epidemiological knowledge available to us. We plan to examine the significance of defining the work location and make relevant adjustments to the human mobility model based on this evaluation in the future.

A main characteristic of machine learning techniques is that accuracy of the results increases with a larger and qualitatively better input dataset. The current dataset used in this study covers only data for 3 years with 156 data points for an MOH division. This is insufficient for machine learning techniques to capture seasonal variations. It can be anticipated that results of our models will further improve with continuous training of the models as new data is received.

Analysis of the total dengue incidences of each year in Sri Lanka distinctly shows that if the total dengue incidence for an year is lower than the previous year, then there will be higher number of total dengue incidence in next year and vice versa [2]. This fact exhibits that the predictive models must be trained with at least two consecutive years of dengue cases. We did not have two consecutive years of data on MNBD as well as meteorological data. Since the predictive models need more than one year of data, we assumed that mobility,

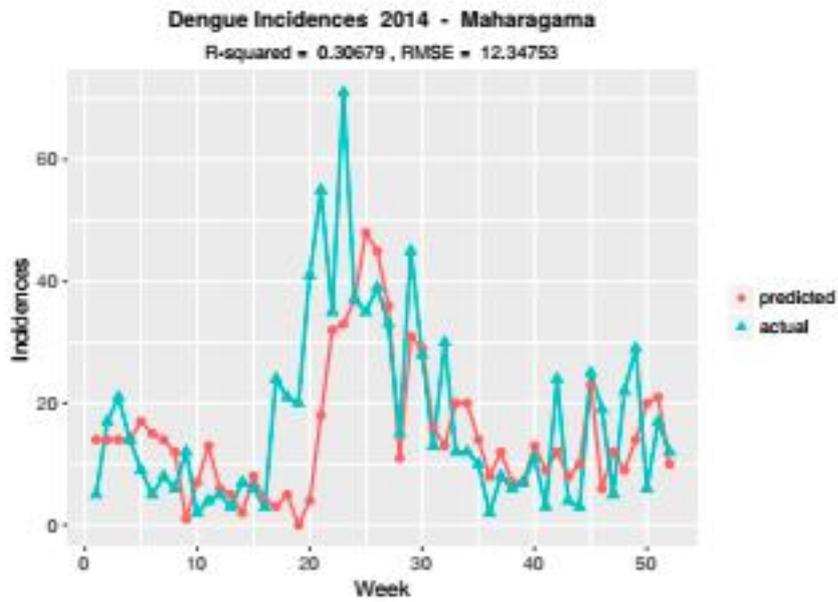


Fig. 4. Without Mobility (ANN) - Maharagama MOH 2014

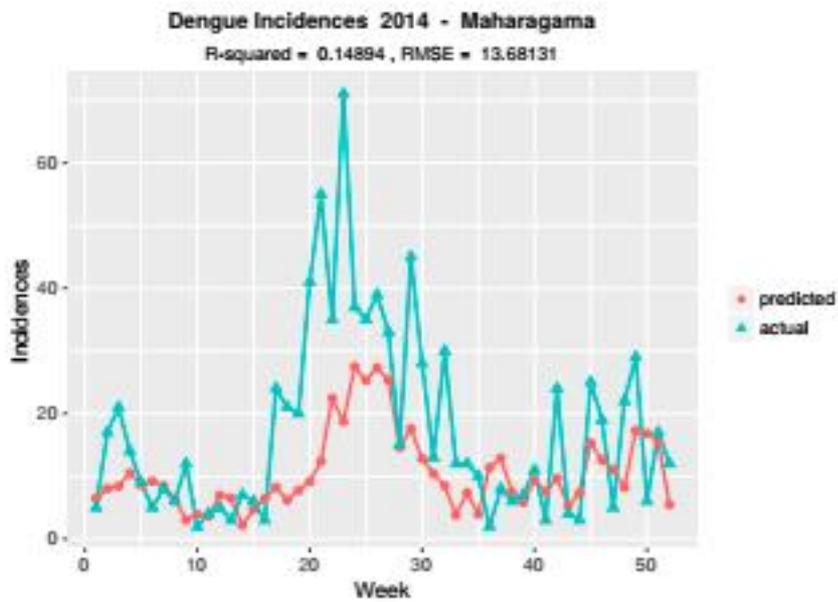


Fig. 5. With Mobility (XGBoost)-Maharagama MOH 2014

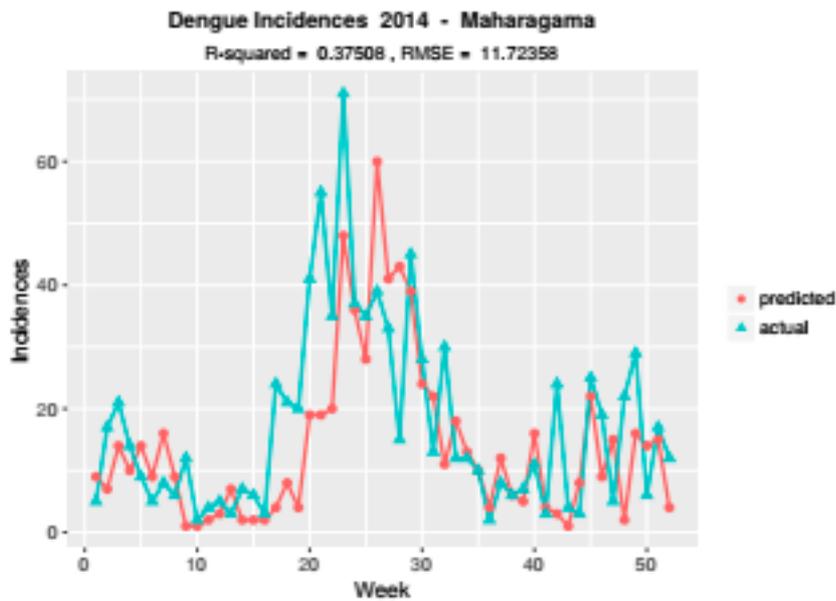


Fig. 6. With Mobility (ANN)-Maharagama MOH 2014

MOH	Without Mobility		With Mobility	
	XGBoost	ANN	XGBoost	ANN
MC - Colombo	0.54	0.67	0.44	0.58
Dehiwala	0.33	0.43	0.18	0.30
Maharagama	0.06	0.30	0.15	0.37
Moratuwa	0.10	0.31	0.16	0.35
Kalutara(North)	0.13	0.36	0.18	0.43

TABLE I  
R SQUARED VALUE MOH LEVEL

temperature and rainfall have similar seasonal variation each year. However, when computing the factorized mobility values, we incorporated weekly reported dengue incidences with the corresponding weekly mobility values of the respective MOH area. This may reduce the total error caused by the assumption that mobility patterns are identical each year.

The results from our models show that for some MOH areas, the predictions have better performance when mobility was used as an input value. MOH areas with high population density such as Colombo depict a different pattern when compared to less populous areas like Maharagama, Moratuwa and Kalutara. The predictions from our models show a decrease in accuracy for Colombo when human mobility was introduced as an input factor. Wesolowski et al considered Karachchi as a driver city for dengue [6]. Similarly, being the commercial capital of Sri Lanka and densely populated MOH divisions, Colombo can be considered as a driver city in Sri Lanka, which can attribute for the deviation from the expected impact from human mobility. However, other MOH areas do show a positive impact from human mobility when predicting dengue incidences.

## VII. CONCLUSION

Use of MNBD to model disease propagation is still an evolving area of study and not many studies have been conducted to explore its applications in day-to-day problems. The extraction of human mobility models or any other mobility variables from CDR datasets is a significant application of MNBD with multiple potential benefits. The viability and usefulness of these mobility models have been validated mostly in a limited number of domains such as traffic controlling and urban policy development [9], [23]. Our study shows that information derived from these human mobility models can be used to predict dengue incidence. The results obtained in this study corroborate this fact taking into consideration the high feature importance shown for human mobility in our predictive models, the high correlation shown between dengue incidence and human mobility and the increase in R2 value between predicted dengue incidence and actual reported dengue incidence in multiple MOH divisions when human mobility is taken as an input. This work can be extended further to model the spatial spread of other infectious diseases as well which has a strong correlation with human mobility.

## VIII. ACKNOWLEDGEMENT

The authors would like to thank Dr. Hasitha A. Tissera and the staff of National Dengue Control Unit, Sri Lanka for providing statistics related to past dengue reported cases and geographical boundary information of MOH divisions. Also authors like to extend their gratitude towards Danaja Maldeniya and Madhushi Bandara who are researchers at LIRNEasia, Sri Lanka, for their valuable opinions and suggestions on this research study. The weekly vegetation index values used in this study were calculated by Amila De Silva, a research student at Department of Computer Science and Engineering, University of Moratuwa.

## REFERENCES

- [1] Dengue and severe dengue, 2016. [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs117/en/>
- [2] Distribution of Notification Dengue Cases by Month, 2017. [Online]. Available: <http://www.epid.gov.lk>
- [3] G. Lucas, "A fresh look at dengue vaccines," Sri Lanka Journal of Child Health, vol. 45(3), pp. 149–150, 2016.
- [4] J. Lourenc\_o and M. Recker, "The 2012 Madeira Dengue Outbreak: Epidemiological Determinants and Future Epidemic Potential," PLoS Neglected Tropical Diseases, vol. 8, no. 8, 2014.
- [5] F. Finger, T. Genolet, L. Mari, G. C. de Magny, N. M. Manga, A. Rinaldo, and E. Bertuzzo, "Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks," Proceedings of the National Academy of Sciences, vol. 113, no. 23, p. 201522305, 2016.
- [6] A. Wesolowski, T. Qureshi, M. F. Boni, P. R. Sundsøy, M. A. Johansson, S. B. Rasheed, K. Engø-Monsen, and C. O. Buckee, "Impact of human mobility on the emergence of dengue epidemics in Pakistan," Proceedings of the National Academy of Sciences, vol. 112, no. 38, pp. 11 887–11 892, 2015.
- [7] P. H. M. N. Herath, A. A. I. Perera, and H. P. Wijekoon, "Prediction of dengue outbreaks in Sri Lanka using artificial neural networks," International Journal of Computer Applications, vol. 101, no. 15, pp. 1–5, 2014.
- [8] P. Liyanage, H. Tissera, M. Sewe, and M. Quam, "A Spatial Hierarchical Analysis of the Temporal Influences of the El Nino-Southern Oscillation ~ and Weather on Dengue in Kalutara District , Sri Lanka," International Journal of Environmental Research and Public Health, vol. 13, Nov. 2016.

- [9] D. Maldeniya, A. Kumarage, S. Lokanathan, G. Kriendler, and K. Madhawa, "Where did you come from ? Where did you go ? Robust policy relevant evidence from mobile network big data," no. September, 2015.
- [10] Area: Predicting human behavior. [Online]. Available: <http://www.tid.es/long-term-innovation/predicting-human-behavior>.
- [11] A. Wesolowski, N. Eagle, A. J. Tatem, D. L. Smith, A. M. Noor, R. W. Snow, and C. O. Buckee, "Quantifying the impact of human mobility on malaria." *Science (New York, N.Y.)*, vol. 338, no. 6104, pp. 267–70, oct 2012.
- [12] Y. Wu, G. Lee, X. J. Fu, and T. Hung, "Detect climatic factors contributing to dengue outbreak based on wavelet, support vector machines and genetic algorithm," vol. 1, 2008, pp. 303–307.
- [13] P. D. Wasserman, *Advanced methods in neural computing*. John Wiley & Sons, Inc., 1993.
- [14] N. Rachata, P. Charoenkwan, T. Yooyativong, K. Chamnongthai, C. Lursinsap, and K. Higuchi, "Automatic prediction system of dengue haemorrhagic-fever outbreak risk by using entropy and artificial neural network," no. Iscit, 2008, pp. 210–214.
- [15] A. Munasinghe, H. Premaratne, and M. Fernando, "Towards an Early Warning System to Combat Dengue," *International Journal of Computer Science and Electronics Engineering*, vol. 1, no. 2, 2013.
- [16] D. Fuller, A. Troyo, and J. C. Beier, "El nino southern oscillation and vegetation dynamics as predictors of dengue fever cases in costa rica," *Environmental Research Letters*, vol. 4, no. 1, p. 014011, 2009.
- [17] TensorFlow: An open-source software library for Machine Intelligence, 2017. [Online]. Available: <https://www.tensorflow.org/>
- [18] S. Karsoliya, "Approximating number of hidden layer neurons in multiple hidden layer bpnn architecture," *International Journal of Engineering Trends and Technology*, vol. 3, 2012.
- [19] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, 2011.
- [20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [21] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 785–794.
- [22] K. Hettiarachchi, *Destroying breeding grounds vital to curb dengue: Study*, 2016. [Online]. Available: <http://www.sundaytimes.lk/160626/news/destroying-breedinggrounds-vital-to-curb-dengue-study-198819.html>
- [23] R. Samarajiva, *Mobile Network Big Data for Urban Development - LIRNEasia*, 2016. [Online]. Available: [http://lirneasia.net/wpcontent/uploads/2016/09/Samarajiva\\_UoM\\_Sep2016-v2.pdf](http://lirneasia.net/wpcontent/uploads/2016/09/Samarajiva_UoM_Sep2016-v2.pdf)