MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231091

Artificial Intelligence Approach for Severe Dengue Early Warning System

Dina Nur ANGGRAINI NINGRUM[a,1], Yu-Chuan (Jack) LI[b,c,d], Chien-YEH HSU[e,f], Muhammad SOLIHUDDIN MUHTAR[g] and Hanif PANDU SUHITO[h]

[a] *Public Health Department, Universitas Negeri Semarang, Semarang City, Indonesia*
[b] *Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan*
[c] *Department of Dermatology, Wan Fang Hospital, Taipei, Taiwan*
[d] *International Center for Health Information Technology (ICHIT), Taipei Medical University, Taipei, Taiwan*
[e] *Department of Information Management, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan*
[f] *College of Public Halth, Taipei Medical University, Taipei, Taiwan*
[g] *Graduate Institute of Data Science, Taipei Medical University, Taiwan*
[h] *Semarang City Regional Health Offfice, Semarang City, Taiwan*
ORCiD ID: Dina Nur Anggraini Ningrum: https://orcid.org/0000-0002-8557-2862

Abstract. Dengue fever is a viral infectious disease transmitted through mosquito bites, and has symptoms ranging from mild flu-like symptoms to deadly complications. Dengue fever is one of the global burden diseases which annually have 50-100 million cases with 500,000 cases of severe dengue fever, of which 22,000 deaths occur mostly in children. Despite the discovery of vaccines, vector control is still the main approach for prevention efforts. Early detection and accessibility to medical care can reduce severe Dengue mortality rate from 50% to 2%. In the previous study, both statistical and machine learning methods have the potential for predicting a Dengue outbreak, but the study is still fragmented and limited on implementing the generated model into an early warning system application. In this study, we developed an artificial intelligence model with spatiotemporal to predict Dengue outbreak and Dengue incidence case which is ready to be implemented into an early warning system application. Indonesia, especially Semarang City, has experienced an endemic Dengue. We used Semarang City spatiotemporal, meteorological, climatological, and Dengue surveillance epidemiology data from January 2014 to December 2021 in 16 districts of Semarang City. We reviewed 7208 samples from 16 districts and 1 city per week during 8 years. The entire dataset was divided into training (80%) and testing (20%) to develop a prediction model. We used machine learning and Long Short Term Memory (LSTM) to predict Dengue outbreak 1 week before the event for each district. and machine learning to predict Dengue incident cases 1 week before the event for each district. Accuracy, area under the receiver operating characteristic curve (AUROC), precision, recall, and F1 score were considered to evaluate the Dengue outbreak prediction model. The Dengue incidence cases prediction model will evaluate using Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-Squared (R^2). Extra Trees Classifier model shown outperform in Dengue outbreak prediction, with accuracy 0.8925, AUROC 0. 9529, Recall 0.6117, precision 0.8880, and F1 score 0.7238. CatBoost Regressor model is shown to outperform in Dengue incidence cases prediction, with R^2 0.5621,

[1] Corresponding Author: Dina Nur Anggraini Ningrum, email: dinanan@mail.unnes.ac.id.

MAE 0.6304, MSE 1.1997, and RMSE 1.0891. The study proves that Artificial Intelligence (AI) with a spatiotemporal approach can give higher performance in Dengue outbreak and incidence cases prediction. Utilization of AI approaches that are sensitive with spatiotemporal feasibility to implement in Dengue early warning system application may contribute to increase the policy makers and community attention to do accurate community-based vector control.

Keywords. artificial intelligence, dengue outbreak prediction, dengue incidence cases prediction, early warning system.

1. Introduction

Dengue fever is an infection from a virus that is transmitted through mosquito bites (DEN 1-4) which has flu-like symptoms until lethal complication [1]. Dengue spread throughout the tropical world especially developing countries and now affects also in sub-tropical areas. Dengue fever threatens more than half of the world's population. Due to climate change and urbanization, the geographic reach of Dengue is expected to expand. Dengue fever is one of the global burden diseases with annually 50-100 million cases and 500,000 cases of severe Dengue, in which 22,000 deaths mainly belong to children. Dengue vector also transmits Chikungunya, Yellow Fever, and Zika infection. Despite the discovery of vaccines, vector control is still the main approach for prevention efforts. Early detection and accessibility to medical care can reduce severe Dengue mortality rate from 50% to 2% [2].

Severe Dengue consists of Dengue Hemorrhagic Fever (DHF) and Dengue Shock Syndrome (DSS) cases. Unexpected surges in Severe Dengue cases and influence by climate change and urbanization are some challenges in prevention and control intervention [3]. Dengue transmission modeling allows determining the main influencing factors and helps policy makers to determine interventions [4]. The ability to detect early emerging outbreaks can help policy makers to manage outbreaks effectively [5]. With the acceleration of the technology, Artificial Intelligence (AI) could help analyze massive surveillance data effectively [6].

Overall purpose of this study is using an AI model with the spatiotemporal data to predict Dengue outbreak and incidence cases that are feasible to implement in Dengue early warning system application.

2. Methods

2.1. Data Source

We use an open dataset from Semarang City, Indonesia as input data. The spatiotemporal and Dengue surveillance epidemiology open data was collected from Semarang City Regional Health Office (http://116.254.113.136:8080/tunggaldara/). The Meteorological and climatological open data was collected from Indonesian Meteorological, Climatological, and Geophysical Agency (http://dataonline.bmkg.go.id). We collected the data from January 2014 until December 2021. Outcome of prediction are Dengue outbreak events (outbreak or non-outbreak) and incidence Dengue cases.

2.2. Data Preprocessing

We reviewed 7208 samples from 16 districts and 1 city per week (53 weeks) during 8 years. We divided data from 1st week January 2004 until 53rd week 2021 into 80:20 for training (5766 samples) and testing (1442 samples). We use weekly (53 weeks) cases collected for an 8 years dataset to predict 1 week before the event. The event duration for outbreak and incidence Dengue cases are 1 week.

2.3. Model Development

Model development and performance evaluation was conducted using a Python library, PyCaret, to simplify the comparison pipeline among many different machine learning algorithms.

3. Results

3.1. Dengue Outbreak Prediction Model Performance Evaluation

The top 4 best performance machine learning models are Extra Trees Classifier (ETC), CatBoost Classifier, Extreme Gradient boosting, and Light Gradient Boosting Machine. We also compare the performance with LSTM as a deep learning approach. The Extra Trees Classifier outperforms compared with other machine learning and LSTM as a deep learning approach.

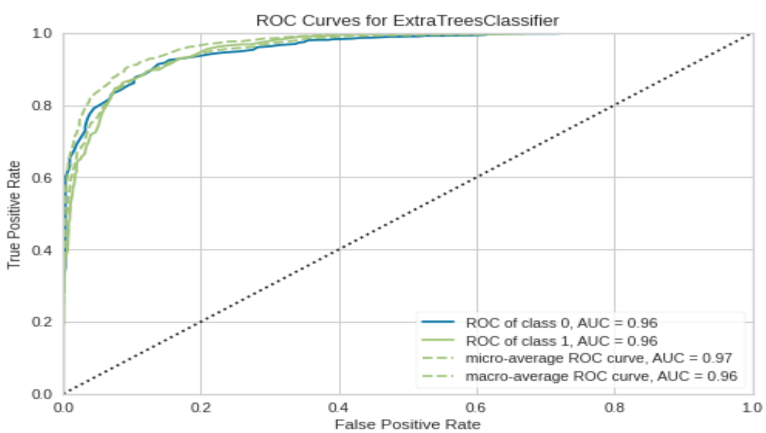

Figure 1. ROC Curves for ETC Dengue Outbreak Prediction Model.

Based on Figure 1, AUROC of class 0 (non-outbreak), AUROC of class 1 (outbreak), micro-average AUROC, and macro-average AUROC shown similar between 0.96-0.97. The recall tends to be low since the outbreak events are imbalanced.

3.2. Dengue Incidence Cases Prediction Model Performance Evaluation

The top 4 best performance machine learning models are CatBoost Regressor (CBR), Gradient Boosting Regressor, Orthogonal Matching Pursuit, and Huber Regressor. The

CBR outperforms compared with other machine learning.

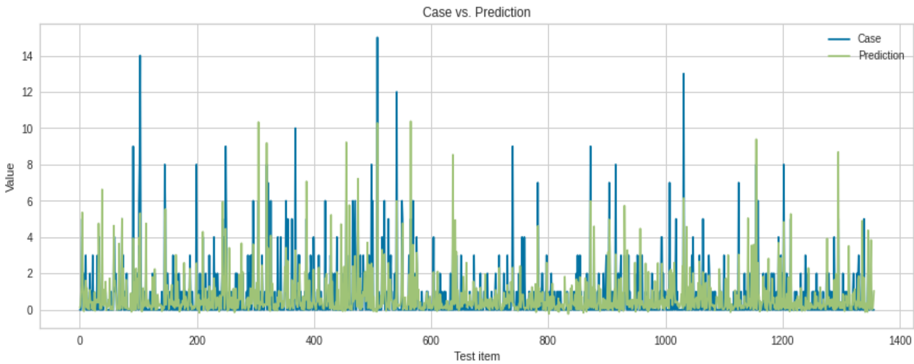

Figure 2. Visualization Between Dengue Incidence Case Actual and Prediction Result from CBR Dengue Incidence Cases Prediction Model

Figure 2 visualizes which the Dengue incidence cases prediction result tend to be lower than the actual Dengue incidence cases.

4. Discussion

Many studies related Dengue outbreak prediction using statistical [7], machine learning methods [8] and still limited using deep learning [9]. Since our data was limited to run in a deep learning method, the result in the deep learning method was not satisfied, so we chose using a machine learning approach. In the previous studies related Dengue outbreak prediction still fragmented among variables [10] and limited implementing the model into an early warning system application [11].

In this study, we used comprehensive variables consisting of meteorological, climatological, and Dengue epidemiology surveillance data during 2014-2021 from many government institutions and open access data sources. This model is feasible to be implemented in a Dengue early warning system. Compared to the previous study which used an AI (SVM Linear) and spatiotemporal approach to predict Dengue outbreak per district in Malaysia, with accuracy 70.12%, Recall 14.40%, and precision 56.25%, this study (ETC Dengue outbreak prediction model) promised better performance. Compared to the previous study which used statistical approaches to predict Dengue incidence cases in Semarang City, this study with AI (CBR Dengue incidence cases model) and spatiotemporal approach promises more detailed prediction for high-risk location (district) and feasibility to deploy AI model in the Dengue early warning system than statistical model.

Implementation of the prediction model into open access application, detailed geographical unit, and user-friendliness dashboard also recommended by Dengue Early Warning and Response System (EWARS) participants [12]. For further research, we will deploy our model into Dengue EWARS for the Semarang City Regional Health Office. The granularity of climatological and meteorological data limited to the city region, there are no district climatological and meteorological stations in Semarang City. Further research needs to use variables which represent per region and possibly collect the data from wearable devices/IoT smart homes to get detailed climatological or meteorological data in each district/sub district.

5. Conclusions

AI model with the spatiotemporal data could predict Dengue outbreak with accuracy 89.25 % (AUROC 95.29%) and Dengue incidence cases with MAE 0.6304 and RMSE 1.0891 that are feasible to be implemented in Semarang City Dengue early warning system application. Further research needs to use deep learning such that facilitates transfer learning.

References

[1] World Health Organization. (2019, April). Facts sheet: Dengue and Severe Dengue. Retrieved June 26, 2019, from https://www.who.int/en/news-room/fact-sheets/detail/dengue-and-severe-dengue

[2] Messina JP, Brady OJ, Golding N, Kraemer MUG, Wint GRW, Ray SE, Pigott DM, Shearer FM, Johnson K, Earl L, Marczak LB, Shirude S, Davis Weaver N, Gilbert M, Velayudhan R, Jones P, Jaenisch T, Scott TW, Reiner RC Jr, Hay SI. The current and future global distribution and population at risk of dengue. Nat Microbiol. 2019 Sep;4(9):1508-15, doi: 10.1038/s41564-019-0476-8.

[3] World Health Organization. (2012). Global strategy for dengue prevention and control, 2012-2020. World Health Organization.

[4] Kong L, Wang J, Li Z, Lai S, Liu Q, Wu H, Yang W. Modeling the Heterogeneity of Dengue Transmission in a City. Int J Environ Res Public Health. 2018 May;15(6):1128, doi: 10.3390/ijerph15061128.

[5] Tissera H, Pannila-Hetti N, Samaraweera P, Weeraman J, Palihawadana P, Amarasinghe A. Sustainable dengue prevention and control through a comprehensive integrated approach: the Sri Lankan perspective. WHO South East Asia J Public Health. 2016 Sep;5(2):106-12, doi: 10.4103/2224-3151.206246.

[6] Wong ZSY, Zhou J, Zhang Q. Artificial Intelligence for infectious disease Big Data Analytics. Infect Dis Health. 2019 Feb;24(1):44-8, doi: 10.1016/j.idh.2018.10.002.

[7] Racloz V, Ramsey R, Tong S, Hu W. Surveillance of dengue fever virus: a review of epidemiological models and early warning systems. PLoS Negl Trop Dis. 2012;6(5):e1648, doi: 10.1371/journal.pntd.0001648.

[8] Guo P, Liu T, Zhang Q, Wang L, Xiao J, Zhang Q, Luo G, Li Z, He J, Zhang Y, Ma W. Developing a dengue forecast model using machine learning: A case study in China. PLoS Negl Trop Dis. 2017 Oct;11(10):e0005973, doi: 10.1371/journal.pntd.0005973.

[9] Laureano-Rosario AE, Duncan AP, Mendez-Lazaro PA, Garcia-Rejon JE, Gomez-Carro S, Farfan-Ale J, Savic DA, Muller-Karger FE. Application of Artificial Neural Networks for Dengue Fever Outbreak Predictions in the Northwest Coast of Yucatan, Mexico and San Juan, Puerto Rico. Trop Med Infect Dis. 2018 Jan;3(1):5, doi: 10.3390/tropicalmed3010005.

[10] Li C, Lu Y, Liu J, Wu X. Climate change and dengue fever transmission in China: Evidences and challenges. Sci Total Environ. 2018 May;622-3:493-501, doi: 10.1016/j.scitotenv.2017.11.326.

[11] McKerr C, Lo YC, Edeghere O, Bracebridge S. Evaluation of the national Notifiable Diseases Surveillance System for dengue fever in Taiwan, 2010-2012. PLoS Negl Trop Dis. 2015 Mar;9(3):e0003639, doi: 10.1371/journal.pntd.0003639.

[12] Hussain-Alkhateeb L, Kroeger A, Olliaro P, Rocklöv J, Sewe MO, Tejeda G, Benitez D, Gill B, Hakim SL, Gomes Carvalho R, Bowman L, Petzold M. Early warning and response system (EWARS) for dengue outbreaks: Recent advancements towards widespread applications in critical settings. PLoS One. 2018 May;13(5):e0196811, doi: 10.1371/journal.pone.0196811.