Recent Developments in Electronics and Communication Systems
KVS Ramachandra Murthy et al. (Eds.)

doi:10.3233/ATDE221302

Comparative Analysis of Machine Learning Classifiers on Forecasting Dengue Fever Infection

Supreet Kaur [a,1] and Dr Sandeep Sharma [b]
[a]*Department of Computer Engineering & Technology, GNDU, ASR.*
[b]*Department of Computer Engineering & Technology, GNDU, ASR.*

Abstract. Each year Dengue infection causes havoc almost across the entire globe raising the death toll and thus becoming a global burden. With the absence of Global approved vaccine and its scope not limiting itself to tropical regions anymore, it has become a rapidly growing epidemic. Researchers have comprehensively explored mechanisms to predict and diagnose infectious diseases and machine learning has revolutionized the medical field by reducing the computing time in analyzing complex data and finding hidden patterns for accurate predictions. In this study, a comparative machine learning-based analysis is achieved using python based classification to predict dengue fever infection in a person. This paper helps interested researchers choose the most efficient classification technique among the selected machine learning classifiers by not only focusing on accuracy achieved but on overall classification metrics analysis to develop a better Dengue predictive model.

Keywords. *Classification algorithms, Dengue* infection, Expert system, Predictive approach, Machine Learning

1. Introduction

The burden of Vector-borne diseases is highest in tropical and subtropical areas. It accounts for more than 17% of all infectious diseases globally resulting in a rise in the death toll of more than 700000 annually[1, 2]. The infectious disease spread by mosquitoes can be caused by any of 3 i.e. a parasite (as in the case of malaria), the virus (as in the case of zika virus), and bacteria (as in the case of plague) [3]. For their ability in spreading deadly diseases, mosquitoes are considered one of the most dangerous creatures on the planet. Most people have been bitten by a mosquito at some time in their lives. Usually, this causes no more than an extreme itch or swelling which produces significant discomfort for a day or so and no one cares about that one mosquito bite[4, 5]. But the deadliest diseases: yellow fever, Zika fever, chikungunya, and dengue fever are transmitted mostly by the mosquito's genus- Aedes Aegypti bite[5]. Although we can never tell by the naked eye, which disease strain it is carrying the only sure way of combating mosquito-borne diseases is the complete eradication of mosquitoes which may sound feasible and effortless theoretically but in practical seems murky[6,7].

[1] Corresponding Author.

Dengue is often labeled as the most ignored tropical disease irrespective of the medical advancements achieved; to date, no vaccine treatment is available[8]. For this reason, researchers have comprehensively explored mechanisms to predict and diagnose this disease with a special focus on designing intelligent medical expert systems that could accurately and timely predict Dengue which could save many lives in an area having minimal or no medical facility available[9, 10]. This paper not only summarizes the various works done in the literature for devising expert systems but also deploys various machine learning algorithms to effectively predict dengue-infected persons.

The summary of the main contribution of this study can be broadly categorized as follows: The summary of the main contribution of this study can be broadly categorized as follows:

- To emphasize the need to predict Dengue infection effectively and precisely.
- To perform a deep analysis on various aspects of Dengue epidemiology and highlight the key areas of study.
- To study the role of different machine prediction models in combating Dengue disease.
- To detect critical Dengue infected patients based on symptoms observed.

The rest of the article is structured as follows, "Related work" summarizing the existing related work. The model to predict Dengue-infected patients is proposed in "The Proposed Model". The experimentation results and performance analysis of the said model are conducted in "Results". "Conclusion" gives the conclusions derived from this study's proposed model.

2. Related Work

Based on physical symptoms alone, a dengue infection prediction model can be achieved by the machine learning algorithm, also acting as an early patient diagnosis system [11]. Though the author also emphasized the importance of the medical test parameter inclusion alongside physical symptoms for properly finding from which stage of Dengue, the person is affected. During this study, the author incorporated some symptoms i.e. fever, gastrointestinal findings, headache, pain, skin rash, eye-ball pain, and clinical factors like WBC, AST/ALT, PLT, CRP, and PTT to offer results comprising three possible diagnoses- no Dengue, probable Dengue, and Dengue confirm[12]. The technique incorporated was the fuzzy toolbox of MATLAB. Similarly in [13], a study on 60 Dengue patients data was conducted in which predictors like systolic /diastolic blood pressure, pulse pressure, HR, BT/CT, HCT, WBC, and PLT to classify patients based on severity index as Dengue fever without warning than with warning and lastly severe Dengue category. The author suggests an effective dengue prediction model by employing IoT sensors, data gathered from sensors and mobiles will be sent to the FOG environment, and Attributes such as name, location, mobile number, age, and sex along with water quality, humidity, and rainfall. The temperature range and degree of mosquitoes are being considered for study [14, 15]. Another study targeted to propose an early detection as well as a precise dengue diagnosis model and evaluated three ML i.e. machine learning techniques namely ANN, NB & DT. The dengue disease set was collected from different hospitals located in the

dengue region and ANN technique to produce better outcomes with larger computational time when compared to NB and DT [16, 17]. The researchers presented an early dengue detection expert system by using FIS i.e. fuzzy inference system. The main was to help patients combat this disease in early-stage by using symptoms and reports of medical tests as input to the model [18, 19, 20].

3. The Proposed Model

The proposed Dengue prediction model is further classified into sub-sections to urge a better understanding of every stage's operations by starting with an overall working idea which can eventually progress into more detailed information. So that Dengue prediction can be analyzed in a better and easy way which can further help me in designing tools for epidemic aid. In the first subsection proposed model's architecture is shown to offer a generalized view and within the proceeding subsections each stage is being discussed highlighting the work wiped out of them.

3.*1. Model Architecture*

The working design of the proposed model is divided into major three phases

- *Phase1-Data Collection*: The Dengue patient's information screened and retrieved that were admitted/referred to the government hospital for dengue monitoring as well as those who although was Dengue positive but hospitalization was not required. The positive patients were confirmed to have Dengue infection through NS1 antigen test or IgM/IgG tests. The vital signs, as well as clinical test parameters, were then recorded in a Microsoft Excel sheet for further pre-processing in which missing values records were dropped.
- *Phase2-Prediction Model*: The classifier algorithms in the machine learning library of Python are trained by using training data. The predicting algorithms are saved in Jupyter Notebook for further work.
- *Phase3-Validation and Improvement*: The predictive models are tested using a testing dataset, generated in Phase2, and accuracy for each is calculated. The confusion metrics are then generated to verify the models' performance and upon comparison, the best algorithm is selected.

3.2. Data Collection

The population for this study comprises subjects with Dengue infection which is being collected manually from Government Medical College (GMC), Amritsar ensuring all legal and ethical formalities were taken care of during this study. The physical signs/symptoms along with clinical tests result were collected from the Dengue patient's file records that are managed and stored in the Dengue ward, GMC. Excel sheet was used as a data entering platform in which further data were grouped into two main classes i.e. Dengue Fever (DF) and No Dengue Fever (NDF). The dataset attributes are age, sex, address (city), fever, loss of appetite, nausea, rash, extreme

weakness, retro-orbital headache, joint/muscle pain, mucosal bleed, blood in stool, abdomen pain, unusually heavy menstrual bleeding, loose motion, large BP drop along with result class as DF (Dengue fever) and DFW(No Dengue fever) [22, 23, 24].

4. Results

Dengue data containing vital signs/ symptoms of patients suffering from Dengue were trained using machine learning classifiers. In this study, more than one machine classifier was selected and compared based on the accuracy achieved. The results in Figure1 below show each model type along with the accuracy achieved by it. Among the selected models: DT, CART, and NB showed high accuracy compared to LDA and KNN. Figure2 shows the confusion matrix for the DT analysis in which prediction reveals the highest accuracy of 99.8% with 0.86 precision and 1.00 recall.

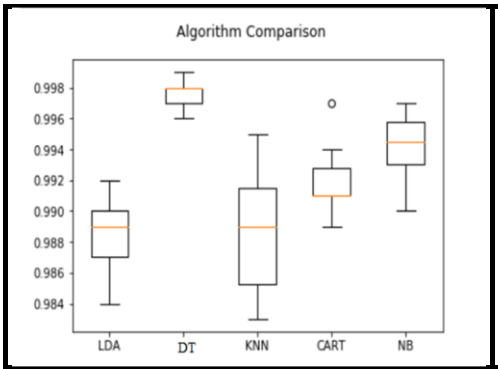

Figure1: Comparative Accuracy of Selected Machine Classifier Algorithms

	precision	recall	f1-score	support
0	0.86	1.00	0.92	6
1	1.00	0.75	0.86	4
accuracy			0.90	10
macro avg	0.93	0.88	0.89	10
weighted avg	0.91	0.90	0.90	10

Figure2: Classification Matrix of Decision Tree

The accuracy of the testing data as shown in the above figures was found to be more than 90 % which successfully correctly classified 9 patients out of test 10 Dengue-infected patients in almost all the classifier algorithms but DT, CART, and NB outperformed all others. After the vital signs input, a classification would be provided depicting either one of two classes i.e. Dengue fever (DF) or No Dengue fever (NDF). However, on closer look, DT outperforms others not only in accuracy but also in sensitivity, specificity, and recall though in initial computation it took little more minutes than others.

5. Conclusion

The purpose of this study was to see how machine learning algorithms work on the dengue dataset and mine hidden patterns in it for an effective detection process for this disease. Among machine learning algorithms, different classification techniques were applied, and based on the accuracy achieved five classifiers LDA, DT, KNN, CART, and NB were selected for this study to further choose a suitable prediction model for Dengue fever that can further assist in improving the current Dengue disease management practice, especially in areas having minimal or no medical facilities or in

pandemics-like scenarios which are already achieved. Among the selected models: DT, CART, and NB showed high accuracy compared to LDA and KNN. However, DT was found to be the best with not only accuracy, recall, sensitivity, and specificity which is calculated by addressing its classification matrix. Though DT was found to be the best learner yet we assume that since the dataset was not very large and that was the reason that DT didn't show its limitation of over-fitting here. Yet this can't be ruled out unless tested by taking a considerably large dataset. Overall, promising results were observed from the proposed model to classify the level of Dengue patients – from where we can also emphasize on patients showing warning signs to consult a doctor for timely treatment if found any. The proposed model was able to achieve good classification parameters such as 99.8% accuracy, 0.86 precision, and 1.0 recall which gave promising results among the selected pool of classifiers. From the study, it is found that there are a set of warning signs in the symptomatic set that also indicate internal bleeding manifestations which are often a sign of thrombocytopenia in Dengue patients which could also be worked on to get better predictions. If data about DHF and DSS patients can also be included in the study it could facilitate dengue forecasting by alerting disease prognosis into many deadly stages which could help on-time medical aid to save a human life. However, Deep learning can also enhance the overall disease prediction efficiency if sufficient big data volume can be collected and hence experimented with. Although we have achieved all the aims stated in the Introduction section, we would like to further investigate by trying to correctly calculate the severity index of Dengue patients based on the classification provided by WHO i.e. DHF and DSS to enhance the system to withstand the pressure of this infectious disease and also further test all other classifiers under decision tree algorithm to ensure among tree classifiers, if any could outperformed this proposed learner algorithm to help medical sector against dengue-the rapidly growing epidemic.

References

[1] Anggraeni, W., Abdillah, A., Pujiadi, Trikoratno, L. T., Wibowo, R. P., Purnomo, M. H., & Sudiarti, Y. (2019). Modelling and Forecasting the Dengue Hemorrhagic Fever Cases Number Using Hybrid Fuzzy-ARIMA. *2019 IEEE 7th International Conference on Serious Games and Applications for Health, SeGAH 2019*, 1–8. https://doi.org/10.1109/SeGAH.2019.8882433

[2] Binti Mohd Zainee, N., & Chellappan, K. (2016). A preliminary dengue fever prediction model based on vital signs and blood profile. *IECBES 2016 - IEEE-EMBS Conference on Biomedical Engineering and Sciences*, 652–656. https://doi.org/10.1109/IECBES.2016.7843530

[3] Gambhir, S., Malik, S. K., & Kumar, Y. (2017). PSO-ANN based diagnostic model for the early detection of dengue disease. *New Horizons in Translational Medicine*, *4*(1–4), 1–8. https://doi.org/10.1016/j.nhtm.2017.10.001

[4] Gambhir, S., Malik, S. K., & Kumar, Y. (2018). The Diagnosis of Dengue Disease: An Evaluation of Three Machine Learning Approaches. *International Journal of Healthcare Information Systems and Informatics*, *13*(3), 1–19. https://doi.org/10.4018/IJHISI.2018070101

[5] Hair, G. M. E., Nobre, F. F., & Brasil, P. (2019). Characterization of clinical patterns of dengue patients using an unsupervised machine learning approach. *BMC Infectious Diseases*, *19*(1), 1–11. https://doi.org/10.1186/s12879-019-4282-y

[6] https://www.cdc.gov/dengue/index.html. (n.d.). *Dengue*.

[7] https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue. (n.d.). *Dengue and severe Dengue*.

[8] Nayak, S. D. P., & Narayan, K. A. (2019). Forecasting Dengue Fever Incidence Using ARIMA Analysis. *International Journal of Collaborative Research on Internal Medicine & Public Health*, *11*(3), 924–

932. https://www.iomcworld.org/abstract/forecasting-dengue-fever-incidence-using-arima-analysis-44475.html

[9] Pandiyarajan, P., & Thangairulappan, K. (2018). Classification of dengue serotypes using protein sequence based on rule extraction from neural network. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11308 LNAI*, 127–137. https://doi.org/10.1007/978-3-030-05918-7_12

[10] Potts, J. A., Gibbons, R. V., Rothman, A. L., Srikiatkhachorn, A., Thomas, S. J., Supradish, P. on, Lemon, S. C., Libraty, D. H., Green, S., & Kalayanarooj, S. (2010). Prediction of dengue disease severity among pediatric Thai patients using early clinical laboratory indicators. *PLoS Neglected Tropical Diseases*, *4*(8), 2–8. https://doi.org/10.1371/journal.pntd.0000769

[11] Saikia, D., & Dutta, J. C. (2016). Early diagnosis of dengue disease using fuzzy inference system. *International Conference on Microelectronics, Computing and Communication, MicroCom 2016*. https://doi.org/10.1109/MicroCom.2016.7522513

[12] Sreenivasan, P., Geetha, S., & Sasikala, K. (2018). Development of a Prognostic Prediction Model to Determine Severe Dengue in Children. *Indian Journal of Pediatrics*, *85*(6), 433–439. https://doi.org/10.1007/s12098-017-2591-y

[13] Idris, M. F. I. M., Abdullah, A., & Fauzi, S. S. M. (2018). Prediction of Dengue Outbreak in Selangor Using Fuzzy Logic. In Proceedings of the Second International Conference on the Future of ASEAN (ICoFA) 2017–Volume 2 (pp. 593-603). Springer, Singapore.

[14] Divya, A., & Lavanya, S. (2020). Real Time Dengue Prediction Using Machine Learning. Indian Journal of Public Health Research & Development, 11(2).

[15] Chakraborty, A., & Chandru, V. (2020). A robust and non-parametric model for prediction of dengue incidence. Journal of the Indian Institute of Science, 1-7.

[16] Pravin, A., Jacob, T. P., & Nagarajan, G. (2019). An intelligent and secure healthcare framework for the prediction and prevention of Dengue virus outbreak using fog computing. Health and Technology, 1-9.

[17] Mohapatra, C., Rautray, S. S., & Pandey, M. (2017, February). Prevention of infectious disease based on big data analytics and map-reduce. In Electrical, Computer and Communication Technologies (ICECCT), 2017 Second International Conference on (pp. 1-4). IEEE.

[18] Gambhir, S., Malik, S. K., & Kumar, Y. (2018). The Diagnosis of Dengue Disease: An Evaluation of Three Machine Learning Approaches. International Journal of Healthcare Information Systems and Informatics (IJHISI), 13(3), 1-19.

[19] Ooi, J. Y. L., & Thomas, J. J. (2017, November). DengueViz: A Knowledge-Based Expert System Integrated with Parallel Coordinates Visualization in the Dengue Diagnosis. In International Visual Informatics Conference (pp. 50-61). Springer, Cham.

[20] Chellappan, K. (2016, December). A preliminary dengue fever prediction model based on vital signs and blood profile. In *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)* (pp. 652-656). IEEE.

[21] Saikia, D., & Dutta, J. C. (2016, January). Early diagnosis of Dengue disease using fuzzy inference system. In Microelectronics, Computing and Communications (MicroCom), 2016 International Conference on (pp. 1-6). IEEE.

[22] Mondal, M. R. H., Bharati, S., Podder, P., & Podder, P. (2020). Data analytics for novel coronavirus disease. *Informatics in Medicine Unlocked*, *20*, 100374.

[23] Bharati, S., Podder, P., Mondal, M., & Prasath, V. B. (2021). CO-ResNet: Optimized ResNet model for COVID-19 diagnosis from X-ray images. *International Journal of Hybrid Intelligent Systems*, (Preprint), 1-15.

[24] Bharati, S., Podder, P., & Mondal, M. R. H. (2020). Hybrid deep learning for detecting lung diseases from X-ray images. *Informatics in Medicine Unlocked*, *20*, 100391.

[25] Bharati, S., & Mondal, M. R. H. 12 Applications and challenges of AI-driven IoHT for combating pandemics: a review. *Intelligence for Managing Pandemics*, 213.