Ensemble machine learning based prediction of dengue disease with performance and accuracy elevation patterns

Rekha Gangula [a], Lingala Thirupathi [b], Rajashekar Parupati [c], K. Sreeveda [d], Saritha Gattoju [b]

[a] CSE Department, Kakatiya Institute of Technology and Science, Warangal 506015, Telangana, India
[b] CSE Department, Methodist College of Engineering & Technology, Abids, Hyderabad 500001, Telangana, India
[c] CSE Department, Vidya Jyothi Institute of Technology(Autonomous), Aziz nagar, Moinabad, RangaReddy 500075, Telangana, India
[d] CSE Department, Institute of Aeronautical Engineering, Dundigal, Hyderabad 500043, Telangana, India

ARTICLE INFO

ABSTRACT

Mosquitoes have numerous illnesses and are one of the deadliest animals in the planet. Including Zika, dengue, palaria, West Niles, chikungunya, yellow fever, and more, mosquitoborne illnesses. Various areas suffer from various climate-induced mosquito-borne illnesses, kinds of mosquitoes widespread across the region and access to preventive measures and medicines. Dengue fever is a mosquito-borne disease that is transferred to the dengue virus via the bite of an Aedes mosquito. The bits of the infected female Aedes mosquito, which spreads the virus to others as it feeds on the infected people's blood. Transmission of dengue is susceptible to climate due to many causes, such as temperature, humidity, precipitation, etc. Areas with higher vapor pressure and precipitation rates are more prone to dengue illness transmission. We utilized the classification algorithms to discover the essential characteristics that spread the dengue. Machine learning is one of the most important approaches of current analysis. For medical applications, many algorithms were employed. Dengue disease is one of the worst infectious diseases that require a high level machine to develop good models in order to learn. We employed the Ensemble Machine Learning technique in hybrid integrations to identify characteristics associated with the spread of the Dengue illness and achieve improved performance.

© 2021 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Nanoelectronics, Nanophotonics, Nanomaterials, Nanobioscience & Nanotechnology.

1. Introduction

Dengue is an arboviral illness of one of four related viral dengue serotypes caused by infection. Today, it's the largest mosquito-borne viral infection infecting people and is a serious worldwide health hazard. The worldwide range of individuals at risk of infection has been estimated at roughly 3 billion and around 40 million symptom episodes and approximately 20,000 fatalities are reported annually [1] (Fig. 1).

It is reported currently from over 100 countries worldwide, with an impact of roughly 70 percent on world dengue morbidity and death in Southeast Asia. It has an impact on Southeast Asia, Latin America and the Caribbean. However, the real incidence of the disease is likely to be several-fold greater since many clinically apparent cases are still undetected in the community by the hospital based surveillance systems conventionally used for disease

reports [2]. Fig. 2 shows dramatic increase in case numbers reporting to the WHO over the past 50 years.

Dengue fever is a tropical viral illness which is caused by and passed on by Dengue viruses [3]. Dengue fever symptoms are headache, muscular acha, fever, and a measleslike rash, which is also known as breakbone fever. According to the IDAMS, the incidence of dengue fever has grown substantially since the 1960s, infecting 100 million people each year, making it "the leading mosquito-borne virus disease in humans today." IDAMS is a joint research effort on the complete control of dengae fever (Small or Medium-Scale Focused Research Project) (Fig. 3).
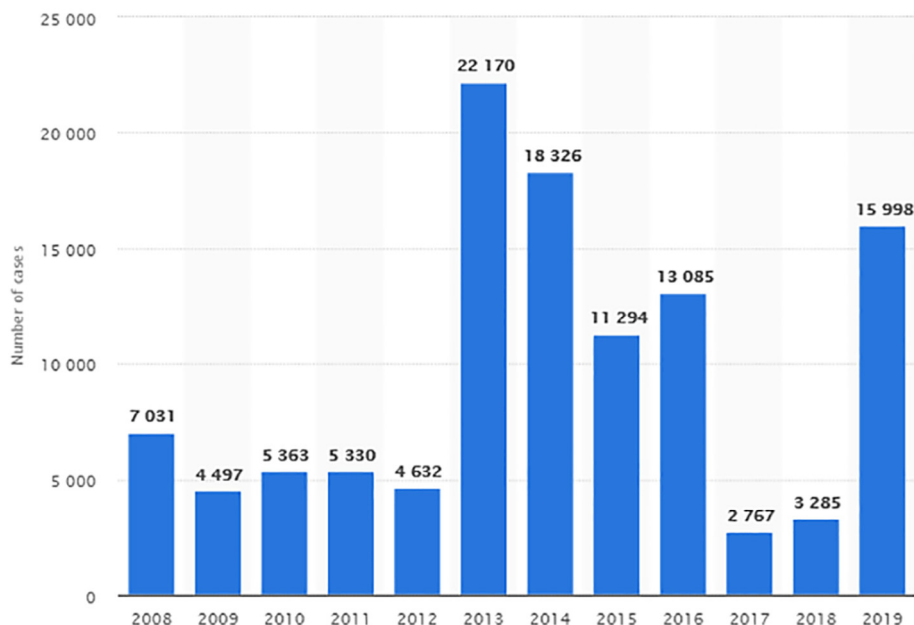
Dengue is an arboviral illness of one of four related viral dengue serotypes caused by infection. Currently, it is the leading mosquito-borne viral infection in humans and a serious threat to world health [5] is arising from the illness. Best estimates show that around 3 trillion individuals are at risk of infection in regions of the world, and that about 40 million symptomatic episodes and over 20,000 fatalities occur every year. It is reported currently from over 100 countries worldwide, with an impact of roughly 70

Fig. 1. Number of dengue fever and dengue hemorrhagic fever cases in Singapore from 2010 to 2019.

Fig. 2. Deadliest creatures worldwide by annual number of human deaths as of 2018.

percent on world dengue morbidity and death in Southeast Asia. It has an impact on Southeast Asia, Latin America and the Caribbean. However, the actual incidence of the disease is probably multiplied by many, as many clinically apparent cases in the community remain uncovered through hospital-based monitoring systems conventionally relied on for disease reporting [6,7]. Fig. 2 shows the dramatic increase in cases reported in the last 50 years to WHO (Fig. 4).

Dengue fever is a tropical illness caused by and spread by dengue viruses. Dengue fever symptoms are headache, muscular acha, fever, and a measleslike rash, which is also known as breakbone fever. According to the IDAMS, the incidence of dengue fever has grown substantially since the 1960s, infecting 100 million people each year, making it "the leading mosquito-borne virus disease in humans today." IDAMS is a joint initiative on the Comprehensive
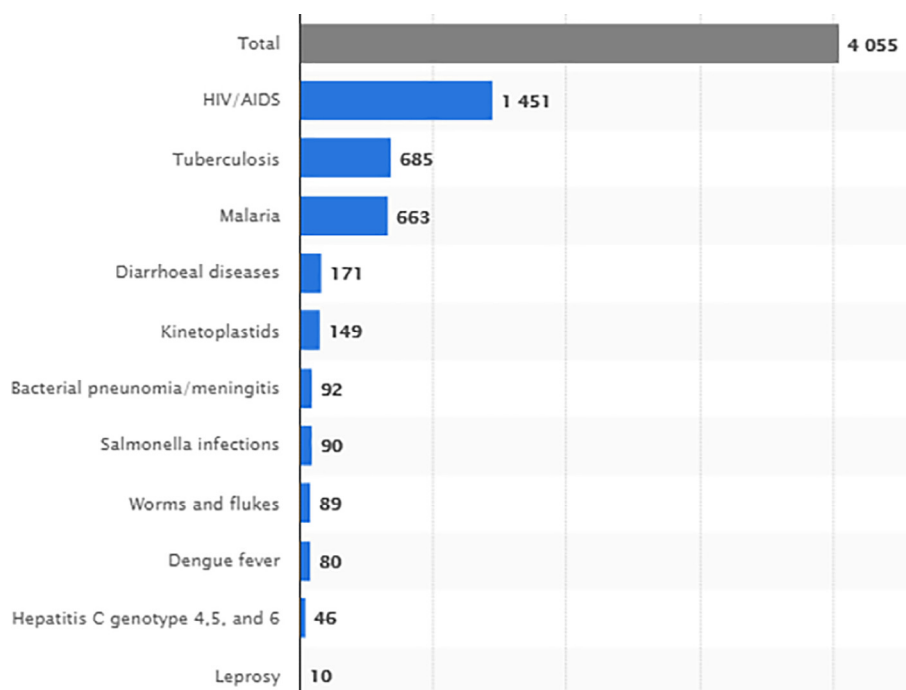
Fig. 3. Donated funds for research and development by disease in 2018 (in million U.S. dollars)*
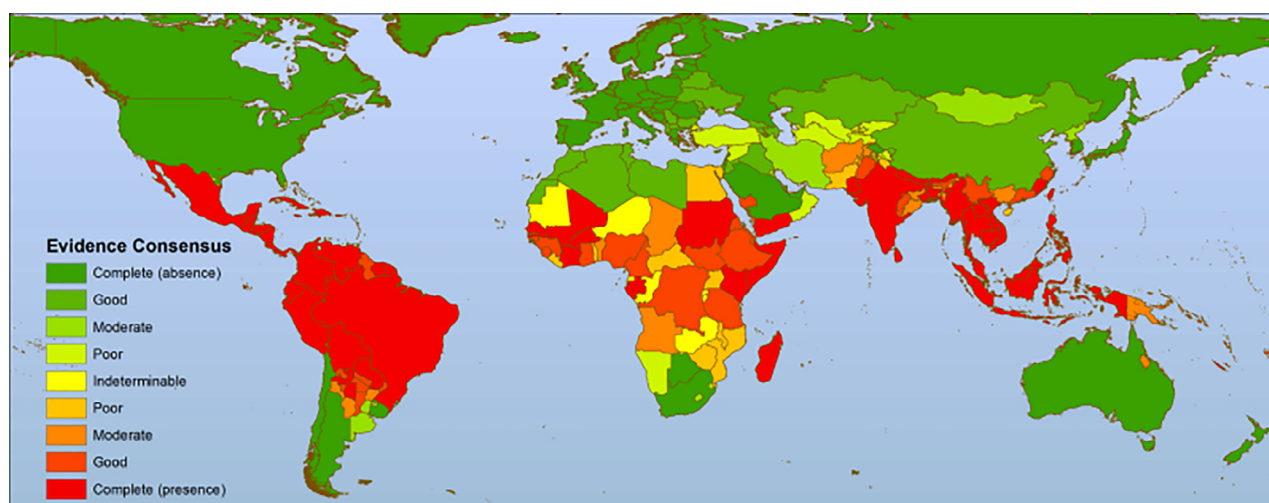

Fig. 4. International Scenario of Dengue.

Control for Dengue Fever (Small or Medium Scale Focused Research Projects) [8].

In many nations, communicable illnesses remain among the top causes of mortality, despite improvements in health conditions worldwide. Control, preventative measures, outbreak research, and creation of control mechanisms are crucial to the fight against infectious illness. Detriments are motivating new attempts on short to medium-term data to collect Country-level statistical data on confirming communicable diseases cases, such as dengue fever [9].

Dengue has expanded quickly in all areas of the WHO in recent years, as a mosquito virus illness. Dengue virus predominantly from the Aedes aegypti and, to a lesser degree, the Ae. albopictus are spread by female mosquitos. These mosquitoes also have chikungunya, yellow fever and Zika. Dengue is common in the tropics and local risk differences are driven by rainfall, temperature, relative moisture and unexpected fast urbanization [10].

Dengue is responsible for a wide range of illnesses. The spectrum of symptoms in persons who are sick may be from subclinical diseases, (people may not realize they're even afflicted). Though less common, certain persons have severely dengue, which can be accompanied with a variety of serious bleeding, organ damage and/or plasma leaking problems. When not properly handled, severe dengue is more susceptible to mortality.

In the 1950s, during dengue outbreaks in the Philippines and Thailand, severe dengue was first observed. Today, sèvère dengue is a major cause of illness and mortality among children and adults in most Asian and Latin American countries [11].

Dengue is caused by Flaviviridae virus, and the virus is causing dengue in four different, but closely related, serotypes (DENV-1, DENV-2, DENV-3 and DENV-4). Infection recovery is supposed to confer lifetime immunity to this serotype. Cross-immunity is only partial and only transient to the other serotypes following recov-
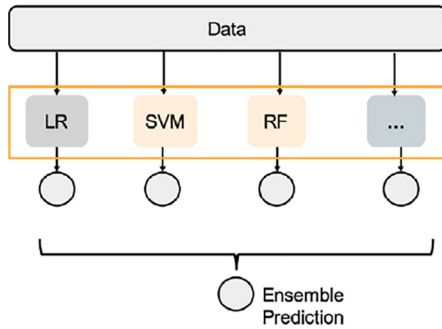
R. Gangula, L. Thirupathi, R. Parupati et al.

Fig. 5. Ensemble Machine Learning.

ery. The risk of getting severe dengue is increased following infections (secondary infection) by different serotypes.

Dengue is related with four viral serotypes in a distinction of epidemiological patterns. These can co-circulate in an area, and for all four serotypes many nations are hyperendemic. The effect of dengue on human health as well as on the world and national economy is worrying. DENV is often transferred by sick travelers from one site to another; if sensitive vectors are available in these new sites, local transmission is likely to develop [12].

2. Global scenario of dengue disease

In recent decades, the prevalence of dengue worldwide has increased significantly. A large majority of cases are self-managed asymptomatic or moderate and hence the true number of cases of dengue is underreported. In many situations other febrile diseases are also misdiagnosed [13].

An estimate model indicating 390 million dengue virus infections each year (95% credible 284–528 million interval), 96 million of which are clinically (67–136 million with illness severity) [2]. In another study on the incidence of dengue the risk of becoming infected with dengue viruses is estimated at 3.9 billion people. Despite a risk of infection in 129 countries [3], Asia is now responsible for 70 percent of the burden [14].

WHO has received 8 reports of dengue cases, from 505 430 recorded cases in 2000, to 2.4 million reported in 2010 and 4.2 million reported in 2019 during the past two decades. Deaths reported rose from 960 to 4032 between 2000 and 2015.

3. Ensemble Machine Learning for performance based predictions

Ensemble based on performance (Collective) Learning [16] incorporates a concept of machine-based higher performance learning. The following portions are included

- Bagging. Multiple (usually identical) models from distinct training dataset sub-samples are created.
- Boosting. Multiple (usually identical) models are constructed which each learns to repair the prediction mistakes of a previous chain model.
- Voting. The combination of forecasts is based on many (usually distinct) models using straightforward statistics (such computing the average).

Bootstrap Aggregate or bag many samples (with replacement) from your training dataset and a pattern for each sample is required. The final output forecast is averaged in all sub-models' predictions (Fig. 5).

In this section, the three bagging models are as follows:

- Random Forest
- Extra Trees
- Bagged Decision Trees

Boost Algorithms construct a succession of models which try to repair the errors of the models in the sequence before them.

Once established, the models predict that their demonstrated accuracy may be weighed and the findings combined to generate a final output forecast.

Algorithms are commonly used for boosting

- AdaBoost
- Boosting of the stochastic gradient

4. Research statement

The high-performance technique of ensemble-level learning forecasts must be developed since Dengue is a major worldwide developing illness and a massive burden on endemic nations' health systems. It is difficult, prior to complications, to differentiate dengue from other common febrile diseases - simple and unpriced strategy is urgently needed in the promotion of early diagnosis, in order both to enhance case management and to facilitate the appropriate use of limited resources, and to identify patients who are at high risk of complications. In order to create a strong case-definition for dengue in the creation of prognostic algorithms it may also be beneficial to evaluate early clinical characteristics with easily available laboratory tests in a large cohort of patients with a range of dengue disorders seen in the environment. In order to reduce the number of dengue infections worldwide, early detection of outbreaks with enhanced surveillance scheme and an early reaction to impending epidemics may be extremely beneficial. Together with identifying regions that are susceptible to dengue outbreaks, as identified by riskmapping, these policies may have significant health advantages by avoiding outbreaks and lowering cases, therefore decreasing overall morbidity and dengue death.

5. Research methodology

Dengue illness dataset is used to measure fever, bleeding, myalgia, glandular influenza, fatigue, results (positive/negative). The hybrid ensemble model is constructed using these data set features and then trained to become more precise.

6. Hybrid ensemble model

For this job, weak students are utilized for creating a hybrid ensemble model of learning of several sorts of machine models (Fig. 6).

These models might include the Logistic Regression Model, Decision Tree, Support Vector Machine, Naive Baye Model or others.

The word hybrid is employed here because a homogenous group of weak students is utilized in previous ensemble models, while a heterogeneous collection of weak students is used in this assignment.

7. Results and outcomes

7.1. Hybrid ensemble model

function Return Solution:
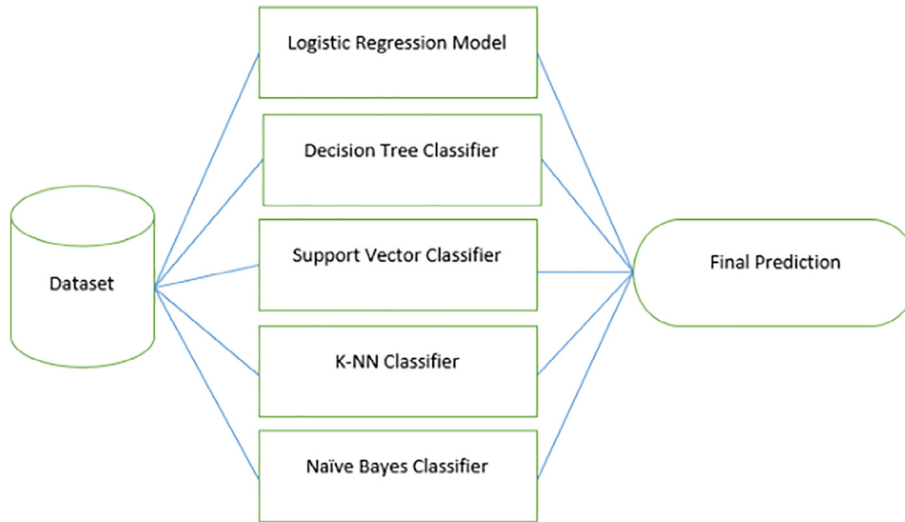Ensemble Model(S,F): S --> Set of (Xi,Yi)
F training —> Function No.

R. Gangula, L. Thirupathi, R. Parupati et al.

Fig. 6. Ensemble Based Hybrid Machine Learning.

Table 1
Evaluation of Assorted Algorithms.

Execution Scenario	Random Forest Approach	Support Vector Machine	Ensemble Learning
1	67	71	91
2	76	81	93
3	69	72	93
4	78	82	96
5	74	82	95

Table 2
Evaluation of Performance

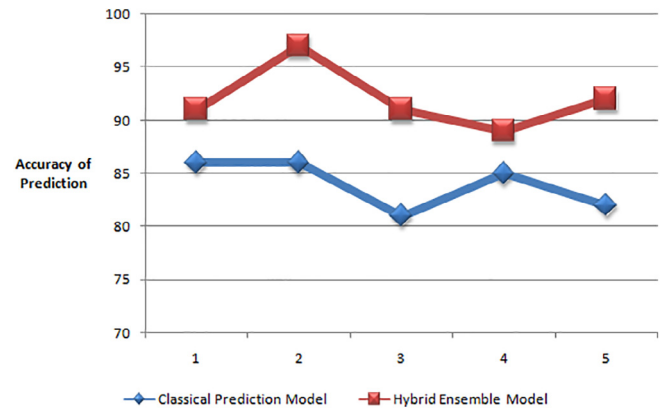| Execution Scenario	Classical Prediction Model	Hybrid Ensemble Model
1	86	91
2	86	97
3	81	91
4	85	89
5	82	92

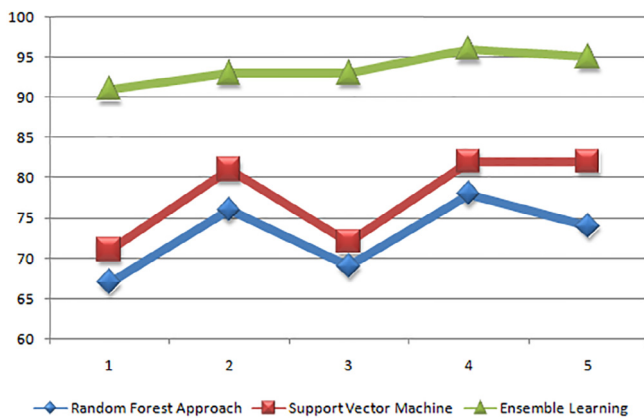**Fig. 8.** Accuracy Elevation in Projected Ensemble Approach.

Fig. 7. Accuracy Elevation in Assorted Approaches.

Key characteristics Dengue Data Extraction
Current Solution <— zero
for each I 1,..(No.fr) do
H(i)-> Post Cleaning Data H (Current Solution)
SPASSOL(s(i),F) SPECIAL(s)

Return H <- H U {h(i)
}
From the entire "m" characteristics, randomly choose "k."
Where k less than m
Calculate "d" node with the optimum split point among the "k" characteristics.
Divide the node in the best split nodes into daughter nodes.
Repeat 1 to 3 stages until nodes "l" have achieved their number.
Construct forests by repeating steps 1 to 4 for n numbers of trees to build "n."
function Set Model(S,F)
return learnt tree:
do f<—Randomized subset of F for each node in tree
Split and select the best function in f
There are several execution situations or tries in the following results table.
Each attempt indicates if the code is executing time or slot.
The code is performed 5 times, and the results are recorded to ensure that the outgoings in each iteration are overall consistent. The findings show that the suggested technique to ensemble learning at a substantial level improves the implementation time. The highlighted numbers indicate maximum and minimum timings (in seconds) for both methods to the algorithm.
Table shows five distinct execution scenarios for calculating both techniques' cost factor. Ensemble learning is therefore a superior

method than conventional prediction approach (Tables 1 and 2, Figs. 7 and 8).

In figure it is shown that the overall performance of hybrid method is more sensitive to random forest technique compared with the traditional strategy in 5 execution situations. The technique based on the hybrid ensemble produces findings as in the graph and the tabular format.

8. Conclusion

Dengue fever is a dengue virus-borne tropical illness. Symptoms usually start 3 to 14 days after infection. High fever, headaches, vomiting, muscular and joint problems and a typical skin rash may be included. It usually takes 2–7 days for recovery. The condition becomes severe dengue in a tiny proportion of patients, also known as dengue hemorrhagic fever, resulting in bleeding, low blood platelet and bleeding, and dengue shock syndrome, which can be caused by dangerously low blood pressure. Dengue is distributed by various Aedes-genus women's mosquitoes, mainly Aedes aegypti. The virus contains five serotypes; one kind of infection generally provides that type lifetime immunity, but only the other short-term protection. The risk of serious consequences is increased by subsequent infections of a different kind. A range of assays, including the detection of antibodies to the virus or its RNA, can be used to confirm diagnoses. This paper focuses on the integration of ensemble-based learning for dengue illness prediction with a greater degree of accuracy and results.

CRediT authorship contribution statement

Rekha Gangula: Conceptualization, Methodology, Data curation. **Lingala Thirupathi:** Writing - review & editing, Visualization. **Rajashekar Parupati:** Investigation. **K. Sreeveda:** Writing – original draft. **Saritha Gattoju:** Validation, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] N. Iqbal, M. Islam, Machine learning for dengue outbreak prediction: a performance evaluation of different prominent classifiers, Informatica 43 (3) (2019).

[2] M. Fatima, M. Pasha, Survey of machine learning algorithms for disease diagnostic, J. Intell. Learn. Syst. Appl. 09 (01) (2017).

[3] D. Salami, C.A. Sousa, M.D.R.O. Martins, C. Capinha, Predicting dengue importation into Europe, using machine learning and model-agnostic methods, Sci. Rep. 10 (1) (2020) 1–13.

[5] S. Anno, T. Hara, H. Kai, M.-A. Lee, Y. Chang, K. Oyoshi, Y. Mizukami, T. Tadono, Spatiotemporal dengue fever hotspots associated with climatic factors in Taiwan including outbreak predictions based on machine-learning, Geospat. Health 14 (2) (2019).

[6] G. Rekha, C.H. Sudha, S. Rekha, M. Nirmala, A Conceptual framework for understanding the role of machine learning in artificial intelligence, Int. J. Adv. Sci. Technol. 29 (4s).

[7] D.S.R. Sanjudevi, D. Savitha, Dengue fever prediction using classification techniques, Int. Res. J. Eng. Technol. (IRJET) 6 (02) (2019) 558–563.

[8] V. Janani, N. Maadhuryaa, D. Pavithra, S.R. Sree, Dengue Prediction Using (MLP) Multilayer Perceptron-A Machine Learning Approach (No. 2444). EasyChair, 2020.

[9] K. Balasaravanan, M. Prakash, Detection of dengue disease using artificial neural network based classification technique, Int. J. Eng. Technol. 7 (1.3) (2018) 13–15.

[10] G. Rekha, B. SridharaMurthy, D. ShashiRekha, K. Sreeveda, Optimal machine learning classifier in mobile malware detection, Int. J. Adv. Res. Eng. Technol. . . . IAEME, 11(10).

[11] T.-S. Ho, T.-C. Weng, J.-D. Wang, H.-C. Han, H.-C. Cheng, C.-C. Yang, C.-H. Yu, Y.-J. Liu, C.H. Hu, C.-Y. Huang, M.-H. Chen, C.-C. King, Y.-J. Oyang, C.-C. Liu, A.L. Rothman, Comparing machine learning with case-control models to identify confirmed dengue cases, PLoS Negl.Trop. Dis. 14 (11) (2020) e0008843.

[12] S. Gambhir, S.K. Malik, Y. Kumar, PSO-ANN based diagnostic model for the early detection of dengue disease, New Horiz. Transl. Med. 4 (1-4) (2017) 1–8.

[13] C. Davi, A. Pastor, T. Oliveira, F.B. de Lima Neto, U. Braga-Neto, A.W. Bigham, M. Bamshad, E.T.A. Marques, B. Acioli-Santos, Severe dengue prognosis using human genome data and machine learning, IEEE Trans. Biomed. Eng. 66 (10) (2019) 2861–2868.

[14] A. Appice, Y.R. Gel, I. Iliev, V. Lyubchich, D. Malerba, A multi-stage machine learning approach to predict dengue incidence: a case study in Mexico, IEEE Access 8 (2020) 52713–52725.

[16] S. Gambhir, Y. Kumar, S. Malik, G. Yadav, A. Malik, Early Diagnostics Model for Dengue Disease Using Decision Tree-Based Approaches, in: Pre-Screening Systems for Early Disease Prediction, Detection, and Prevention (pp. 69–87). IGI Global, 2019.

Further reading

[4] S. Kannimuthu, K.S. Bhuvaneshwari, D. Bhanu, A. Vaishnavi, S. Ahalya, Performance evaluation of machine learning algorithms for dengue disease prediction, J. Comput. Theor. Nanosci. 16 (12) (2019) 5105–5110.

[15] T.M. Carvajal, K.M. Viacrusis, L.F.T. Hernandez, H.T. Ho, D.M. Amalin, K. Watanabe, Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines, BMC Infect. Dis. 18 (1) (2018) 1–15.