EPH- International Journal of Medical and Health Sciences
Volume 12 Issue 4 October 2026
ISSN- 2456 - 6063
DOI: https://doi.org/10.67772/qxhwcz89

INTERPRETABLE MACHINE LEARNING FOR DENGUE RISK STRATIFICATION: INTEGRATING CLINICAL PROFILES, DIAGNOSTIC MARKERS, AND CROSS-MODEL PREDICTOR SIGNATURES

Suzan Mohamed Naeem[1]

[1]Department of Biomedical Engineering – Medical Research Institute, Egypt. Orcid id- https://orcid.org/0009-0009-9702-5928, Email Id: suzannaim7@alexu.edu.eg

Received:- 05/06/2026	Revised:- 12/09/2026	Accepted:- 08/10/2026	Published:- 28/10/2026

Abstract
Despite the substantial clinical heterogeneity of dengue infection, its early recognition can still be difficult, particularly when its clinical presentation overlaps with other febrile illnesses, which are also borne by mosquitoes. Interpretable prediction models can be used to facilitate the process of risk stratification by combining routinely available clinical and diagnostic data. This study aimed to compare Logistic Regression and Random Forest models for dengue risk stratification and evaluate their discrimination, calibration, robustness, and cross-model predictor importance. A retrospective secondary-data analysis included 299 participants with valid dengue outcomes, comprising 56 dengue-positive and 243 dengue-negative cases. Following data-quality screening, 56 predictors were retained. Models were developed using a stratified training–test design, optimized through cross-validation, and evaluated using ROC-AUC, PR-AUC, bootstrap confidence intervals, repeated cross-validation, calibration measures, and permutation importance. Logistic Regression and Random Forest achieved held-out ROC-AUCs of 0.9443 and 0.9425 and PR-AUCs of 0.7947 and 0.7606, respectively. Repeated cross-validation produced mean ROC-AUCs of 0.7148 and 0.7525. Random Forest demonstrated better calibration, with a lower Brier score (0.1334 vs. 0.1697) and expected calibration error (0.1977 vs. 0.2818). Common influential factors were health centre, thick blood smear, weight, joint pain, rapid diagnostic testing, muscle pain, and hepatomegaly. The models were both helpful for discrimination, and Random Forest demonstrated better repeated-validation ROC-AUC and calibration. There is a need for external validation for wider clinical use.

Keywords: Dengue, interpretable machine learning, risk stratification, Random Forest, Logistic Regression

Introduction

Dengue remains a significant disease burden for health care systems in tropical and subtropical countries and is transmitted by mosquitoes. All four major serotypes of the dengue virus cause a spectrum of clinical disease from asymptomatic or mild febrile illnesses to severe disease with plasma leakage, haemorrhage, organ dysfunction and shock. With increasing geographical spread of dengue and the frequent outbreaks of large numbers of cases, early detection is more critical for clinical management and public health response (Paz-Bailey et al., 2024). Viral properties, human susceptibility, mosquito vectors, environment and population movements all influence the way in which dengue is transmitted. Its clinical course may also evolve rapidly, and close clinical observation of the patient with acute febrile illness is warranted (Wilder-Smith et al., 2019). The current understanding of the pathogenesis of dengue suggests that there is a mosaic of clinical manifestation and severity, which involves multiple mechanisms of the virus, immunity and the host, making the identification of one single clinical profile applicable to all patients difficult (Halstead, 2019). Dengue has also been found to have high societal costs as it impacts healthcare utilization, productivity, on the household and national economic levels in addition to morbidity and mortality (Castro et al., 2017).

The number of people exposed to dengue transmission is also evolving. The combined influence of population growth and climate has led to a rise in the number of individuals exposed to environments that facilitate the

transmission of the disease-causing dengue virus, further fueling concerns about the potential for increased disease burden in the future (Nakase et al., 2024). Another critical component of this transmission system is vector abundance – the adult Aedes mosquito population is linked to the fluctuation of dengue transmission risk (Ong et al., 2021). The relationship between meteorological conditions and transmission patterns may also vary and may be modified by the effects of meteorological conditions on the development, survival, and biting activity of mosquitoes, and the dynamics of virus transmission (Gui et al., 2021). The risk of dengue can not be considered a purely virological phenomenon. Its spatial distribution is modulated by socioeconomic and demographic factors and by climate and geography, demonstrating the non-uniform environments where suspected infections are found (Watts et al., 2020).

Dengue is difficult to clinically recognize due to similar symptoms to many other acute febrile diseases. Fever, headache, myalgia, arthralgia, nausea, rash and other systemic symptoms may be helpful for diagnosis; however, these are not all specific to dengue. Laboratory confirmation is thus a cornerstone in the diagnosis, and includes direct detection of viral components and serological evidence of infection; the latter is partly dependent on the stage of the illness and timings of the specimens (Muller et al., 2017). The clinical spectrum is also broader, making the distinction between uncomplicated febrile illness and dengue more complex since there are clinical symptoms that are not usually considered part of the disease. While neurological complications are less common than systemic symptoms, there are a few central and peripheral nervous system complications, which can affect clinical assessment and treatment (Trivedi & Chakravarty, 2022). This highlights the importance of keeping to the right diagnostic suspicion of atypical presentations of dengue while still maintaining appropriate differentials to other causes of acute illness (Patel et al., 2024).

In recent years, the role of combinations of routinely available variables to enhance recognition and risk assessment of dengue has been taken up and discussed in the literature. The updated evidence details dengue as a multifactorial disease in which epidemiological exposure, host factors, clinical presentation, laboratory findings and progression all contribute to the diagnosis and management of the disease (Khan et al., 2023). Predictive approaches are relevant because many symptoms or measurements can be informative independently, but multiple variables can be useful when observing patterns in relationships among several variables. Dengue early-warning systems research has expanded to include an increasing interest in statistical and computational approaches to predicting dengue outbreaks, but significant variability in the choice of predictors, modelling methods, validation techniques, and assessment of the predictive performance exists (Baharom et al., 2022). From a patient standpoint, this will provide a justification for investigating the possibility of creating interpretable statistical and machine learning models that can identify the presence of dengue in the patient's presentation based on information provided in the clinical record.

A crucial need exists to demonstrate predictive discrimination and to determine if a model offers stable, interpretable and properly calibrated risk estimates. The results of studies focused mainly on classification performance can give limited information on either the uniformity of the model's performance over validation samples or the appropriateness of the predicted probabilities in relation to the observed results. Comparisons of linear and nonlinear models are also useful because similar problems of discrimination may occur with different models for the relationship among the predictors. Predictor importance analysis comparing modelling approaches could thus be useful to differentiate clinical signals that are consistently informative from those specific to the modelling. These factors are particularly important when the positive event occurs less often than the negative event, and when the sample size is relatively small; these factors can lead to optimistic measures of performance if not carefully validated within the sample.

The objectives of the present study were to describe differences in demographic, clinical, physiological, and diagnostic features between individuals with and without dengue; to develop and compare dengue risk-stratification models using regularized Logistic Regression and Random Forest; and to evaluate model discrimination, calibration, robustness, and cross-model predictor importance. The analytical framework was designed to assess predictive performance while retaining interpretability and accounting for uncertainty in model evaluation. This clinically oriented approach enabled multivariable dengue risk stratification while positioning predictive modelling as a complementary decision-support tool rather than a replacement for confirmatory diagnosis.

2. Methodology

2.1 Research Design

The study design employed was a retrospective secondary data analysis, which enabled the development of models that can be interpreted in the context of the risk of dengue. Demographic, clinical, physiological, laboratory, and diagnostic variables were assessed and compared with binary dengue status. Logistic regression and random forest were chosen to compare the linear and nonlinear patterns of prediction.

2.2 Data Source

The data collected were public data reported by Ouedraogo et al. (2025). The source had individual-level demographic, clinical, laboratory and diagnostic data for the assessment of vector-borne diseases. The data were limited to those with valid dengue outcomes, and 299 observations were included in this analysis.

2.3 Data Cleaning and Feature Selection

Any extremely implausible values or scale inconsistencies of continuous variables were checked. Errors in decimal scale were corrected conservatively, and four variables in laboratory scale were removed due to their unclear measurement scale. To avoid instability caused by a low number of observations in each class, binary predictors with fewer than 10 observations in each of the two classes were rejected. The final set of pre-imputed predictors was a set of 56, six of which were continuous, 45 were binary, and five were categorical.

2.4 Descriptive and Comparative Analysis

The continuous variables were presented as median and interquartile range, and compared between the dengue-positive and dengue-negative groups by Mann–Whitney U test. Categorical variables were presented as frequencies and percentages and compared with a suitable test of contingency tables. A $p<0.05$ was considered statistically significant. There were 56 dengue-positive and 243 dengue-negative participants in the analytical cohort.

2.5 Data Preprocessing

The cohort was split into a training set of 239 individuals and a held-out test set of 60 individuals, including 45 and 11 dengue-positive cases. The missing values were taken care of in the preprocessing pipeline; missingness indicators were maintained if applicable. All continuous variables were standardized, binary variables were coded numerically and categorical variables were coded. The parameters of the preprocessing were determined from the training set and then used for the test set.

2.6 Model Development

The classifiers, namely regularized Logistic Regression and Random Forest, were fitted using the training data. Five-fold stratified cross-validation was used to optimize the hyperparameters of the Logistic Regression and Random Forest. PR-AUC was highlighted for model selection due to the small number of cases for the dengue-positive class. Only predictions during the training phase were used to guide the classification thresholds to be applied to the held-out test set.

2.7 Model Evaluation and Validation

ROC-AUC, PR-AUC, as well as accuracy, balanced accuracy, sensitivity, specificity, precision and F1 score were used to measure model discrimination. The test ROC-AUC and PR-AUC and their corresponding difference were estimated using bootstrap resampling with 5,000 resamples and 95% confidence intervals. A robustness test was also conducted with five-fold stratified cross-validation repeated five times, which resulted in 25 different validation folds per model.

2.8 Calibration and Predictor Importance

The Brier score, log loss, expected calibration error, and the dengue probabilities observed versus predicted were used to evaluate calibration. The permutation importance for each predictor was calculated as the reduction in ROC-AUC after the permutation of the predictor being assessed. Logistic Regression and Random Forest were separately ranked and compared to determine consistency in the predictors that were important.

3. Results

3.1 Study Population and Characteristics According to Dengue Status

There was a total of 299 patients whose outcomes of dengue were available for analysis. Of these, 56 (18.73%) were dengue-positive, and 243 (81.27%) were dengue-negative. The dengue-positive group had a median age of 25.50 years (IQR: 16.75–40.25), compared with 8.00 years (IQR: 2.80–30.00) among dengue-negative participants ($p<0.001$). Finally, the median weight was also greater for those who were positive for dengue (61.50 kg vs. 15.50 kg; $p<0.001$). There was no significant difference between groups for axillary temperature ($p=0.233$), but respiratory rate ($p=0.024$), pulse rate ($p=0.020$) and creatinine ($p=0.005$) had significant differences between groups. Of the participants who had been confirmed as positive for dengue, 71.4% reported joint pain, as 39.5% of those who were negative for dengue reported the same. For muscle pain, the numbers were 78.6% and 49.6%, respectively. Dizziness was also more common in the dengue-positive group (48.2% compared to 22.2%). There were differences found in the health centre, type of fever, rapid diagnostic test and thick blood smear results. Table 1 shows that multiple demographic, clinical, physiological, and diagnostic characteristics differed significantly between dengue-positive and dengue-negative participants, whereas axillary temperature and gender showed no statistically significant differences.

Table 1. Clinical and demographic characteristics according to dengue status

Characteristic	Dengue positive	Dengue negative	P value
Age	25.50 (16.75–40.25)	8.00 (2.80–30.00)	<0.001
Weight	61.50 (45.00–68.75)	15.50 (10.00–43.50)	<0.001
Axillary temperature (°C)	38.20 (37.50–39.00)	38.00 (37.00–39.00)	0.233
Respiratory rate	28.00 (26.00–34.00)	32.00 (28.00–38.00)	0.024
Pulse rate	101.00 (92.00–113.50)	110.00 (92.00–135.00)	0.020
Elevated creatinine	99.00 (90.50–112.00)	83.50 (61.00–103.50)	0.005
Mucosal skin pallor or anemia	11 (19.6%)	125 (52.3%)	<0.001
Joint pain	40 (71.4%)	96 (39.5%)	<0.001
Oliguria	3 (5.4%)	76 (31.4%)	<0.001
Dizziness	27 (48.2%)	54 (22.2%)	<0.001
Muscle pain	44 (78.6%)	120 (49.6%)	<0.001
Hemoconcentration	4 (7.3%)	77 (32.2%)	<0.001
Positive tourniquet test	5 (9.1%)	87 (36.0%)	<0.001
Hepatomegaly	3 (5.4%)	68 (28.1%)	<0.001
Moderate splenomegaly	4 (7.1%)	73 (30.2%)	<0.001
Back pain	1 (1.8%)	44 (18.2%)	<0.001
Health centre: CMA de DAFRA	13 (23.2%)	139 (57.2%)	<0.001
Health centre: CMA de DO	43 (76.8%)	104 (42.8%)	<0.001
Gender: Female	29 (51.8%)	122 (50.6%)	0.875
Gender: Male	27 (48.2%)	119 (49.4%)	0.875
Fever type: Intermittent	12 (21.8%)	122 (51.0%)	<0.001
Fever type: Recurrent	43 (78.2%)	117 (49.0%)	<0.001
TDR: Negative	21 (40.4%)	38 (17.4%)	<0.001
TDR: Positive	31 (59.6%)	181 (82.6%)	<0.001
Thick blood smear: Negative	20 (36.4%)	36 (16.7%)	0.001
Thick blood smear: Positive	35 (63.6%)	179 (83.3%)	0.001

Note: Continuous variables are presented as median (Q1–Q3); categorical variables are presented as n (%).

3.2 Model Discrimination

The case distribution was preserved, and 239 of the observations were included in the training set, with 60 held out for testing. The training set included 45 dengue-positive and 194 dengue-negative participants, while the held-out test set included 11 dengue-positive and 49 dengue-negative participants. Five-fold cross-validation was used to determine the regularization parameter C using logistic regression, which is 0.0886. The out-of-fold ROC-AUC and PR-AUC for it were 0.7143 and 0.4386, respectively. The same values were obtained with the Random Forest model, which were 0.7483 and 0.4302. Logistic regression had an ROC-AUC value of 0.9443 and a PR-AUC value of 0.7947 on the held-out test set. Random Forest had an ROC-AUC of 0.9425 and a PR-AUC of 0.7606. The 95% CI of logistic regression ROC-AUC and Random Forest were between 0.8757–0.9919 and 0.8727–0.9896, respectively. The corresponding PR-AUC confidence intervals were 0.5492–0.9697 and 0.5011–0.9611. Figure 1A shows the ROC curves for Logistic Regression and Random Forest, withheld-out ROC-AUC values of 0.944 and 0.942, respectively. Figure 1B shows the corresponding precision–recall curves, with PR-AUC values of 0.795 and 0.761, respectively. The precision–recall curves are presented in Figure 1B; logistic regression PR-AUC is 0.795, and Random Forest PR-AUC is 0.761.

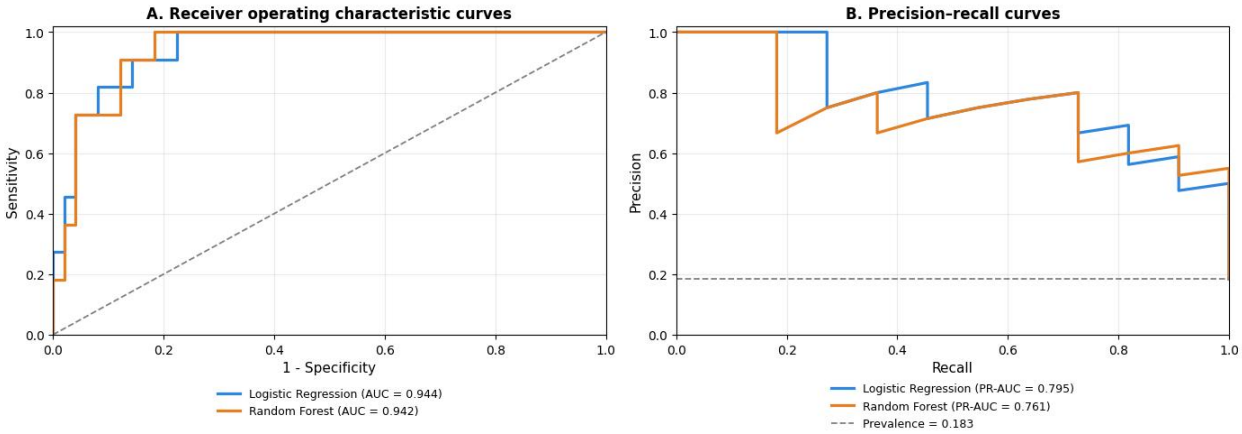


Figure 1. Discrimination performance of the dengue prediction models. (A) Receiver operating characteristic curves for Logistic Regression and Random Forest. (B) Precision–recall curves for Logistic Regression and Random Forest.

3.3 Comparative and Repeated Cross-Validation Performance

Stratified cross-validation consisted of 25 folds and 5 repeats for each model. Random Forest produced a mean ROC-AUC of 0.7525 ± 0.0694 compared with 0.7148 ± 0.0712 for logistic regression. Random Forest had the largest ROC-AUC in 21 out of 25 validation folds. Mean PR-AUC was 0.4777 ± 0.1180 for logistic regression and 0.4711 ± 0.0933 for Random Forest. The balanced accuracy for logistic regression was 0.6603 and 0.6863 for Random Forest, and the mean accuracy for repeated validation was 0.6812 and 0.7539, respectively. Logistic regression had an average of 0.6267 in sensitivity, while Random Forest had an average of 0.5778 in sensitivity. Mean F1 scores were 0.4271 and 0.4719, respectively. Both algorithms yielded an ROC-AUC value above 0.94 on an independent test set, however. The comparative repeated cross-validation, held-out discrimination and calibration metrics of the two models are presented in Table 2.

Table 2. Comparative performance of Logistic Regression and Random Forest

Model	Repeated CV ROC-AUC, mean ± SD	Repeated CV PR-AUC, mean ± SD	Test ROC-AUC	Test PR-AUC	Brier score	ECE
Logistic Regression	0.7148 ± 0.0712	0.4777 ± 0.1180	0.9443	0.7947	0.1697	0.2818
Random Forest	0.7525 ± 0.0694	0.4711 ± 0.0933	0.9425	0.7606	0.1334	0.1977

Note: CV, cross-validation; ROC-AUC, area under the receiver operating characteristic curve; PR-AUC, area under the precision–recall curve; ECE, expected calibration error.

3.4 Calibration of Predicted Dengue Probabilities

The Brier scores for calibration analysis were 0.1697 for logistic regression and 0.1334 for Random Forest. The expected calibration errors were 0.2818 and 0.1977 for logistic regression and Random Forest, respectively. Logistic regression predicted a mean of 0.4651, while the Random Forest model predicted a mean of 0.3805,

in comparison to the observed mean dengue prevalence in the test sample not used for fitting the models (held out), which was 0.1833. For Logistic Regression, observed dengue proportions of 0.000, 0.000, 0.000, 0.250, and 0.667 corresponded to mean predicted probabilities of 0.170, 0.302, 0.457, 0.597, and 0.800, respectively, across grouped predicted probabilities for logistic regression. Random Forest produced observed proportions of 0.000, 0.000, 0.000, 0.250, and 0.667 at mean predicted probabilities of 0.122, 0.214, 0.350, 0.551, and 0.665. The plot of the observed versus predicted probabilities for logistic regression is shown in Figure 2A along with the Brier score of 0.170 and ECE of 0.282. The calibration curve using the Random Forest model is shown in Figure 2B, with Brier score = 0.133 and ECE = 0.198.

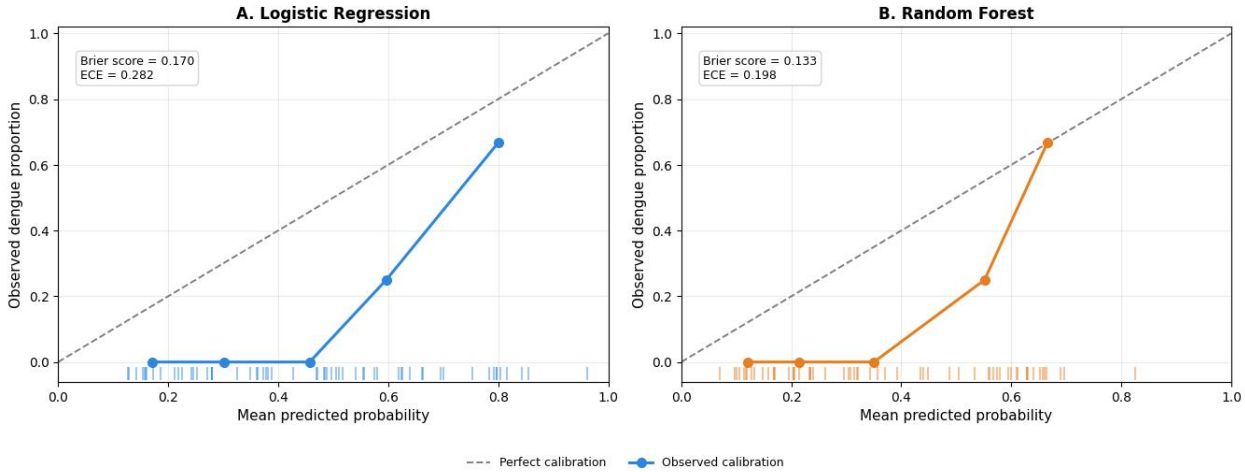


Figure 2. Calibration performance of the dengue prediction models. (A) Calibration curve for Logistic Regression. (B) Calibration curve for Random Forest.

3.5 Predictor Importance

The top parameters for both algorithms were health centre, thick blood smear, weight, joint pain, rapid diagnostic test, muscle pain, and hepatomegaly, which were identified by cross-model permutation analysis. Health centre ranked first for logistic regression with a mean permutation importance of 0.0337 ± 0.0139, followed by weight (0.0311 ± 0.0233), joint pain (0.0199 ± 0.0170), and muscle pain (0.0191 ± 0.0082). Thick blood smear ranked first for Random Forest (0.0365 ± 0.0125), followed by rapid diagnostic test (0.0344 ± 0.0141) and health centre (0.0307 ± 0.0158). There were seven common predictors in both of the top-ten lists. The lowest combined mean rank was for health centre (2.0); the thick blood smear (3.0), joint pain (3.5), and weight (3.5) were next; the rapid diagnostic test (4.5) and muscle pain (6.0) were next. Cross-model permutation importance indicates the relative importance of the 15 top variables (Table 3).

Table 3. Leading predictors of dengue based on cross-model permutation importance

Rank	Variable	Logistic mean importance	Logistic SD	Logistic rank	RF mean importance	RF SD	RF rank	Mean rank
1	Health centre	0.0337	0.0139	1	0.0307	0.0158	3	2.0
2	Thick blood smear	0.0144	0.0119	5	0.0365	0.0125	1	3.0
3	Joint pain	0.0199	0.0170	3	0.0192	0.0215	4	3.5
4	Weight	0.0311	0.0233	2	0.0190	0.0177	5	3.5
5	Rapid diagnostic test	0.0105	0.0133	7	0.0344	0.0141	2	4.5
6	Muscle pain	0.0191	0.0082	4	0.0100	0.0124	8	6.0
7	Hepatomegaly	0.0094	0.0084	8	0.0089	0.0103	10	9.0
8	Type of fever	0.0043	0.0022	14	0.0142	0.0094	6	10.0
9	Positive tourniquet test	0.0049	0.0066	12	0.0079	0.0057	12	12.0
10	Oliguria	0.0093	0.0104	9	0.0058	0.0075	17	13.0
11	Age	0.0025	0.0016	19	0.0137	0.0184	7	13.0
12	Headache	0.0053	0.0099	11	0.0067	0.0044	16	13.5
13	Hemoconcentration	0.0032	0.0066	17	0.0089	0.0105	11	14.0
14	Back pain	0.0058	0.0081	10	0.0031	0.0060	21	15.5
15	Cough	0.0109	0.0098	6	0.0019	0.0036	26	16.0

Note: RF, Random Forest. Permutation importance represents the mean decrease in ROC-AUC after permutation of the corresponding predictor.

Health centre and weight are the most important variables for logistic regression, as seen in Figure 3A. The Ranking of the Random Forest model is shown in Fig. 3B, with the top three being thick blood smear, rapid diagnostic test and health centre, all with the highest mean decreases in ROC-AUC.

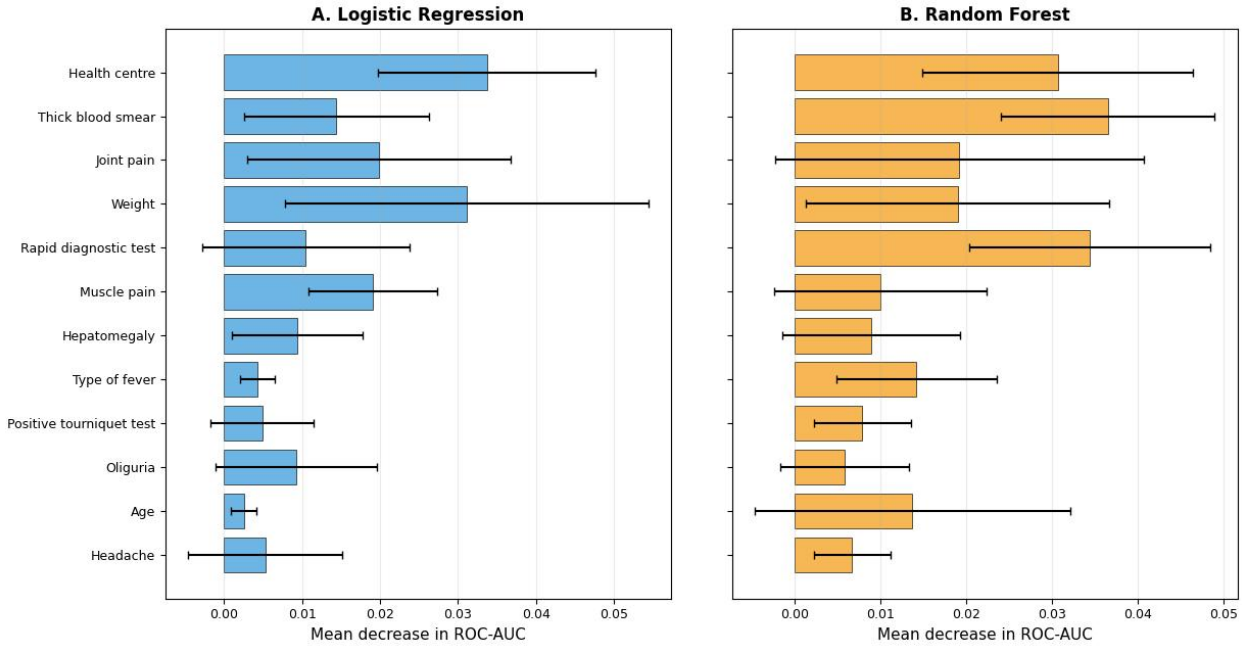


Figure 3. Leading dengue predictors based on permutation importance. (A) Permutation importance of leading predictors in Logistic Regression. (B) Permutation importance of the corresponding predictors in Random Forest.

4. Discussion

The results showed that there was a clinical difference between the dengue-positive and dengue-negative cases. The age, body weight, respiratory rate, pulse rate and creatinine were significant in distinguishing between the two groups of participants with and without dengue. Increased incidence of joint pain, muscle pain, and dizziness was seen in the dengue-positive group, while pallor or anaemia, oliguria, hepatomegaly, and moderate splenomegaly were less common. Similarly, the results of fever type, health centre, rapid diagnostic test, and thick blood smear were also different based on the dengue status. In sum, these patterns suggest that classification was based on information from various demographic, symptomatic, physiological, and diagnostic dimensions rather than on any single dimension of the classification. Logistic Regression and Random Forest demonstrated a very similar discrimination in terms of their ROC-AUC (0.9443 vs. 0.9425) and PR-AUC (0.7947 vs. 0.7606) on the held-out test set. They showed no convincing differences in held-out discrimination, with their bootstrap confidence intervals for the difference in discrimination overlapping. The estimates were more conservative when the cross-validation was repeated. The mean ROC-AUC values were 0.7525 for Random Forest and 0.7148 for Logistic Regression, and the mean PR-AUC values were 0.4711 and 0.4777, respectively. Random Forest also had better calibration properties, as evidenced by a lower Brier score and expected calibration error. Cross-model permutation analysis revealed that certain factors, including health centre, thick blood smear, weight, joint pain, rapid diagnosis, muscle pain and hepatomegaly, were shared influential predictors.

The relationships observed are congruent with evidence based on multivariable approaches for identifying dengue. The work by Tuhin et al. (2025) exemplifies the power of interpretable machine learning techniques, which show that patient data, routinely collected data from patients, can contain meaningful discriminatory signals by the use of clinical and hematological features. Similarly, using machine learning, Madewell et al. (2025) showed the importance of using various patient factors for the prediction of severe dengue. This is not the first time that Logistic Regression has been compared to and performed as well as Random Forest. Ho et al. (2020) reported that machine-learning methods can be used to assist in the identification of confirmed dengue cases, although the impact of the use of such methods depends on the characteristics used and the modelling framework. The current results further illustrate this finding by demonstrating that the more flexible random forest algorithm did not necessarily outperform the logistic regression algorithm on each of the

measures. Random Forest performed better on repeated cross-validation ROC-AUC, accuracy, BAC and F1 score, while the Logistic Regression had slightly better PR-AUC and sensitivity. This variation emphasizes the importance of evaluating complementary measures in addition to the ROC-AUC for clinical classifiers. The clinical prediction factors indicated by the models are also known clinical features of dengue. The study by Flórez et al. (2024) revealed that musculoskeletal and systemic symptoms could differ within the different age groups of dengue patients, supporting the importance of joint pain, muscle pain, and other associated symptoms. Goweda and Faisal (2020) claimed that Dengue presentation is a combination of clinical and laboratory features, and that each clinical/laboratory feature has a greater diagnostic significance when placed in the context of the total presentation.

The weight to be given to the healthcare context needs to be interpreted in the context. Darshan et al. (2024) outlined the clinical and seasonal differences for the patients visiting the clinic for Dengue, where the patient's characteristics might vary from one clinic to another and from one season to another. Therefore, the health centres may be indicative of variations in case-mix, referral patterns, diagnostic practices or local transmission conditions, and not directly indicative of the presence of dengue. Multivariable modelling is further supported by the use of machine-learning applications at an epidemiological level. In a study, Guo et al. (2017) showed that machine-learning techniques could describe complex interactions among the predictors when predicting the activity of the disease dengue. Moreover, Siriyasatien et al. (2018) highlighted data quality, predictor selection, the validation strategy, and the choice of model as key factors in predicting dengue. These concepts are applied in the present analysis by repeatedly stratifying the validation test set, independently evaluating the test set, assessing calibration, estimating the uncertainty in the analysis using bootstrapping, and ranking the predictors using permutations.

Results of the study corroborate the idea of a combined approach of dengue risk stratification based on routinely available patient characteristics. Using a combination of demographic, symptom, physiological and diagnostic variables could give a better picture of the risk of dengue than using each one of these variables alone. Having several influential predictors that had similar agreement between Logistic Regression and Random Forest also suggests that the signals were not solely reliant on any specific modelling architecture. This is especially important for discrimination and calibration. A model can successfully rank patients, but make probabilities that are not very similar to the observed risk. Random Forest had relatively good probability calibration, while Logistic Regression had competitive discrimination and was more transparent. It is important to consider that the models are more complementary risk-stratification techniques than alternatives to confirmatory diagnosis procedures.

Some restrictions are noted. The analytical sample was moderate with 56 dengue-positive subjects, but the held-out test set had only 11 dengue-positive subjects. The small number of events leads to increased uncertainty about the seemingly good test set performance and helps to explain why the more conservative repeated cross-validation estimates are so important. Some predictors were missing, and four lab variables were dropped due to being measured on different scales that could not be reliably resolved. The characteristics of the participating sites may also be captured in health centres, which might reduce the extent to which these can be transferred to other settings.

The models should be validated on larger, more geographically representative populations of participants and the models' predictive rankings should be compared to see if they are consistent across health-care settings. Independent external validation would give a more robust assessment of discrimination and calibration, and prospective evaluation could be used to set performance in the context of routine clinical practice. The models might also need to be recalibrated for other populations with different dengue prevalence. This would help determine if the clinical patterns identified do provide reliable enough information to support widespread dengue risk stratification.

5. Conclusion

Using commonly available clinical and diagnostic data, interpretable prediction of dengue was shown to be possible, and meaningful classification without relying on a single dominant marker was possible. Dengue-positive participants showed distinctive socio-demographic and clinical characteristics, and health centre, thick blood smear, weight, joint pain, rapid diagnostic testing, muscle pain, and hepatomegaly were influential across both models. Using Logistic Regression and Random Forest yielded similar discriminatory estimates on the test set, while repeated validation gave more moderate and stable estimates. Random Forest had the highest mean AUC (AUC on repeated CV) and had the best probability calibration, while Logistic Regression had comparable PR-AUC and sensitivity. The results indicate that the use of discrimination, calibration, periodical validation, and a predictor interpretation approach in the construction of dengue risk-stratification models is valid. The models are to be used as a support tool for analysis and not a substitute for confirmatory diagnosis. A larger external and prospective validation study is needed to confirm the observed patterns of predictors and

the performance of the model in different clinical settings and infection prevalence to confirm the stability of the patterns and performance of the model.

References

1. Baharom, M., Ahmad, N., Hod, R., & Abdul Manaf, M. R. (2022). Dengue early warning system as outbreak prediction tool: a systematic review. Risk Management and Healthcare Policy, 871-886.
2. Castro, M. C., Wilson, M. E., & Bloom, D. E. (2017). Disease and economic burdens of dengue. The Lancet Infectious Diseases, 17(3), e70-e78.
3. Darshan, B. B., Holla, R., Unnikrishnan, B., Achappa, B., Poovattil, R., Sharma, A., ... & Prateek, S. (2024). Clinical and seasonal pattern of dengue in a tertiary care hospital of South West India. F1000Research, 12, 817.
4. Flórez, J. E. S., Velasquez, K. M., Cardona, Á. M. S., Jaramillo, B. N. R., Díaz, Y. E. O., Cardona, L. S. G., & Naranjo, M. A. (2024). Clinical manifestations of dengue in children and adults in a hyperendemic region of Colombia. The American journal of tropical medicine and hygiene, 110(5), 971.
5. Goweda, R., & Faisal, A. (2020). A study of clinical features and laboratory profile of dengue fever in outpatient setting. Malaysian Journal of Public Health Medicine, 20(2), 94-100.
6. Gui, H., Gwee, S., Koh, J., & Pang, J. (2021). Weather factors associated with reduced risk of dengue transmission in an urbanized tropical city. International journal of environmental research and public health, 19(1), 339.
7. Guo, P., Liu, T., Zhang, Q., Wang, L., Xiao, J., Zhang, Q., ... & Ma, W. (2017). Developing a dengue forecast model using machine learning: A case study in China. PLoS neglected tropical diseases, 11(10), e0005973.
8. Halstead, S. (2019). Recent advances in understanding dengue. F1000Research, 8, F1000-Faculty.
9. Ho, T. S., Weng, T. C., Wang, J. D., Han, H. C., Cheng, H. C., Yang, C. C., ... & Liu, C. C. (2020). Comparing machine learning with case-control models to identify confirmed dengue cases. PLoS neglected tropical diseases, 14(11), e0008843.
10. Khan, M. B., Yang, Z. S., Lin, C. Y., Hsu, M. C., Urbina, A. N., Assavalapsakul, W., ... & Wang, S. F. (2023). Dengue overview: An updated systemic review. Journal of infection and public health, 16(10), 1625-1642.
11. Madewell, Z. J., Rodriguez, D. M., Thayer, M. B., Rivera-Amill, V., Paz-Bailey, G., Adams, L. E., & Wong, J. M. (2025). Machine learning for predicting severe dengue in Puerto Rico. Infectious diseases of poverty, 14(1), 5.
12. Muller, D. A., Depelsenaire, A. C., & Young, P. R. (2017). Clinical and laboratory diagnosis of dengue virus infection. The Journal of infectious diseases, 215(suppl_2), S89-S95.
13. Nakase, T., Giovanetti, M., Obolski, U., & Lourenço, J. (2024). Population at risk of dengue virus transmission has increased due to coupled climate factors and population growth. Communications Earth & Environment, 5(1), 475.
14. Ong, J., Aik, J., & Ng, L. C. (2021). Adult Aedes abundance and risk of dengue transmission. PLoS Neglected Tropical Diseases, 15(6), e0009475.
15. Ouedraogo, I., Some, B., & Ouedraogo, I. (2025). Tabular dataset for AI-based vector-borne disease prediction (Version 2) [Data set]. Mendeley Data. https://doi.org/10.17632/cf49v47z4c.2
16. Patel, J. P., Saiyed, F., & Hardaswani, D. (2024). Dengue fever accompanied by neurological manifestations: challenges and treatment. Cureus, 16(5).
17. Paz-Bailey, G., Adams, L. E., Deen, J., Anderson, K. B., & Katzelnick, L. C. (2024). Dengue. The Lancet, 403(10427), 667-682.
18. Siriyasatien, P., Chadsuthi, S., Jampachaisri, K., & Kesorn, K. (2018). Dengue epidemics prediction: A survey of the state-of-the-art based on data science processes. IEEE Access, 6, 53757-53795.
19. Trivedi, S., & Chakravarty, A. (2022). Neurological complications of dengue fever. Current neurology and neuroscience reports, 22(8), 515-529.
20. Tuhin, I. A., Siam, A. F. K., Shanto, M. M. R., Mia, M. R., Mahmud, I., & Ghosh, A. (2025). An interpretable machine learning model for dengue detection with clinical hematological data. Healthcare Analytics, 100430.
21. Watts, M. J., Kotsila, P., Mortyn, P. G., Sarto i Monteys, V., & Urzi Brancati, C. (2020). Influence of socio-economic, demographic and climate factors on the regional distribution of dengue in the United States and Mexico. International journal of health geographics, 19(1), 44.
22. Wilder-Smith, A., Ooi, E. E., Horstick, O., & Wills, B. (2019). Dengue. The Lancet, 393(10169), 350-363.