

1 **Machine learning for predicting severe dengue, Puerto Rico**

2
3 Zachary J. Madewell, PhD^{1,*}; Dania M. Rodriguez, PhD¹; Maile B. Thayer, PhD¹; Vanessa Rivera-Amill,
4 PhD²; Gabriela Paz-Bailey, MD, PhD¹; Laura E. Adams, DVM¹; Joshua M. Wong, MD¹

5
6 ¹ Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, San Juan, Puerto Rico.

7 ² Ponce Health Sciences University/Ponce Research Institute, Ponce, Puerto Rico.

8
9 * Correspondence to: Zachary J. Madewell, PhD; ock0@cdc.gov

10 **Abstract**

11 **Background:** Distinguishing between non-severe and severe dengue is crucial for timely intervention and
12 reducing morbidity and mortality. Traditional warning signs recommended by the World Health
13 Organization (WHO) offer a practical approach for clinicians but have limitations in sensitivity and
14 specificity. This study evaluates the performance of machine learning (ML) models compared to WHO-
15 recommended warning signs in predicting severe dengue among laboratory-confirmed cases in Puerto
16 Rico.

17 **Methods:** We analyzed data from Puerto Rico's Sentinel Enhanced Dengue Surveillance System (May
18 2012–August 2024), using 40 clinical, demographic, and laboratory variables. Nine ML models,
19 including Decision Trees, K-Nearest Neighbors, Naïve Bayes, Support Vector Machines, Artificial
20 Neural Networks, AdaBoost, CatBoost, LightGBM, and XGBoost, were trained using 5-fold cross-
21 validation and evaluated with area under the receiver operating characteristic curve (AUC-ROC),
22 sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). A
23 subanalysis excluded hemoconcentration and leukopenia to assess performance in resource-limited
24 settings. An AUC-ROC value of 0.5 indicates no discriminative power, while a value closer to 1.0 reflects
25 better performance.

26 **Results:** Among the 1,708 laboratory-confirmed dengue cases, 24.3% were classified as severe. Gradient
27 boosting algorithms achieved the highest predictive performance, with AUC-ROC values exceeding 94%
28 for CatBoost, LightGBM, and XGBoost. Feature importance analysis identified hemoconcentration
29 ($\geq 20\%$ increase during illness or $\geq 20\%$ above baseline for age and sex), leukopenia (white blood cell
30 count $< 4,000/\text{mm}^3$), and timing of presentation to a healthcare facility at 4–6 days post-symptom onset as
31 key predictors. Excluding hemoconcentration and leukopenia did not significantly affect model
32 performance. Individual warning signs like abdominal pain and restlessness had sensitivities of 79.0%
33 and 64.6%, but lower specificities of 48.4% and 59.1%, respectively. Combining ≥ 3 warning signs
34 improved specificity (80.9%) while maintaining moderate sensitivity (78.6%), resulting in an AUC-ROC
35 of 74.0%.

36 **Conclusions:** ML models, especially gradient boosting algorithms, outperformed traditional warning
37 signs in predicting severe dengue. Integrating these models into clinical decision-support tools could help
38 clinicians better identify high-risk patients, guiding timely interventions like hospitalization, closer
39 monitoring, or the administration of intravenous fluids. The subanalysis excluding hemoconcentration
40 confirmed the models' applicability in resource-limited settings, where access to laboratory data may be
41 limited.

42 **Keywords:** ensemble learning, gradient boosting, feature importance, clinical decision support,
43 Caribbean

44 **Background**

45 Dengue is a significant public health concern worldwide, with approximately 390 million
46 infections annually, of which 96 million manifest clinically [1, 2]. In Puerto Rico, dengue has been
47 associated with nearly 30,000 confirmed and probable cases from 2010 to 2020, including 584 severe
48 cases, 10,000 hospitalizations, and 68 deaths [3]. A surge in dengue cases on the island in 2024 prompted
49 a public health emergency declaration by Puerto Rico’s Department of Health, highlighting the ongoing
50 threat of dengue to the island [4]. These regular outbreaks strain healthcare resources and pose substantial
51 morbidity and mortality risks. A critical aspect of managing dengue is distinguishing between non-severe
52 and severe cases, as the latter require intensive medical intervention to prevent complications and
53 fatalities. Early identification of patients at risk of severe dengue is important for timely intervention and
54 improved patient outcomes. However, predicting which patients will progress to severe dengue remains a
55 challenge, often leading to delayed treatment and increased healthcare burden.

56 The World Health Organization (WHO) has recommended identifying severe dengue through
57 clinical assessment of warning signs such as persistent vomiting, abdominal pain, mucosal bleeding,
58 restlessness, and hepatomegaly [5]. Although these warning signs offer a practical approach for clinicians,
59 their specificity and sensitivity in accurately predicting severe dengue are limited. Studies have shown
60 that relying solely on these warning signs can result in both false positives and negatives, potentially
61 leading to over- or under-treatment of patients [6-10]. The substantial burden of dengue on the healthcare
62 system, both in terms of economic cost and human suffering, underscores the need for innovative
63 approaches to disease diagnosis. A more accurate and efficient method for risk stratification could lead to
64 substantial improvements in patient care and resource allocation.

65 In recent years, machine learning (ML) has emerged as a powerful tool for analyzing complex
66 datasets and uncovering patterns not easily discernible by traditional methods. In the context of dengue,
67 ML models can analyze a multitude of factors beyond the established warning signs, including patient
68 demographics, laboratory results, clinical symptoms, and epidemiological data, to enhance the prediction
69 of severe disease [11, 12]. By leveraging ML, we aim to improve the accuracy of severe dengue

70 predictions, offering a more robust and data-driven approach to risk stratification. If successful, these
71 models could transform dengue management by enabling early, accurate identification of high-risk
72 patients, ultimately leading to improved patient outcomes and reduced mortality rates. Additionally,
73 targeted intervention for high-risk patients can optimize resource allocation, ensuring critical care is
74 available to those who need it most.

75 This project specifically leverages data from Puerto Rico's Sentinel Enhanced Dengue
76 Surveillance System (SEDSS), which has detailed clinical and laboratory information on dengue cases,
77 allowing for the exploration of potential predictors of disease severity beyond the conventional warning
78 signs. In addition to exploring ML approaches, this project also aims to evaluate the performance of
79 WHO-recommended warning signs in predicting severe dengue among laboratory-confirmed cases,
80 which include both molecular (RT-PCR) and serologic (IgM ELISA) testing. By doing so, we seek to
81 bridge the gap between research and clinical practice by demonstrating the practical applications of both
82 traditional and advanced computational tools in identifying severe dengue. If our findings demonstrate
83 that ML models offer improved prediction of severe dengue compared to traditional methods, this could
84 highlight the potential for integrating advanced computational tools into public health strategies and
85 clinical protocols. For example, predictive ML models could be incorporated into clinical decision
86 support systems used in emergency departments or outpatient clinics, enabling real-time risk stratification
87 for severe dengue. This integration could help healthcare providers prioritize patients for hospitalization,
88 allocate medical resources more efficiently, and guide timely interventions to prevent complications and
89 fatalities. These tools could enable early and accurate identification of high-risk patients, improve patient
90 outcomes, and optimize resource allocation.

91

92 **Methods**

93 *Study population*

94 In this analysis, we used data from SEDSS, an ongoing facility-based study in Puerto Rico that
95 tracks the frequency and causes of acute febrile illness [13, 14]. Our study included data from SEDSS

96 from May 2012 to August 2024. SEDSS has included five sites: 1) Centro Médico Episcopal San Lucas
97 (CMESL) in Ponce, a tertiary acute care facility (2012-present), 2) Hospital Episcopal San Lucas (HESL)
98 - Guayama, a secondary acute care hospital (2013–2015), 3) Hospital de La Universidad de Puerto Rico
99 in Carolina, another secondary acute care teaching hospital (2013–2015), 4) Centro de Emergencia y
100 Medicina Integrada (CEMI), an outpatient acute care clinic in Ponce (2016-present), and 5) Auxilio
101 Mutuo Hospital, a tertiary care facility in the San Juan Metro Area (2018–present).

102

103 *Study enrollment and data collection*

104 SEDSS enrolls participants using convenience sampling. Potential participants are identified by
105 triage nurses as any patient with an acute febrile illness (AFI) defined by the presence of fever ($\geq 38.0^{\circ}\text{C}$
106 for temperatures measured orally, $\geq 37.5^{\circ}\text{C}$ for temperatures measured rectally, and $\geq 38.5^{\circ}\text{C}$ for
107 temperatures measured axillarily for both children and adults) at the time of triage or chief complaint of
108 having a fever within the past seven days. During the Zika virus epidemic in Puerto Rico (June 2016–
109 June 2018), patients were eligible if they presented with either rash and conjunctivitis, rash and arthralgia,
110 or fever [15]. Starting in April 2020, patients with cough or dyspnea within the last 14 days (with or
111 without fever) were also eligible to better capture respiratory viruses [16]. No age groups were excluded,
112 although infants were only eligible for enrollment if they presented to the hospital after their initial
113 discharge after birth. After meeting the inclusion criteria and being informed about the study, participants
114 provided written informed consent. In cases where patients were incapacitated at the time of triage due to
115 acute illness, consent was sought after their stabilization.

116 SEDSS collects data via patient interviews and medical record reviews at enrollment and
117 convalescence (~7–14 days later). The case investigation form (CIF) gathers information about patient
118 demographics, comorbidities, and clinical features. The convalescent sample processing form (CSPF)
119 echoes CIF data, adding the second specimen collection date and AFI severity indicators
120 (hospitalizations, clinic visits). Inpatient medical data for participants with AFIs who were admitted to the
121 hospital from CMESL, HESL-Guayama, and Auxilio Mutuo Hospital also were collected using a separate

122 form (Hospital Admitted Abstraction Form) to collect key clinical indicators of disease severity and
123 progression. For admitted patients, these data included information on extent and nature of hemorrhage,
124 plasma leakage (e.g., ascites and pleural and cardiac effusions), hematologic indicators of increased
125 intravascular permeability (e.g., hematocrit and serum albumin levels), additional blood pressure and
126 heart rate measures to assess shock, and indicators of severe organ involvement (e.g., liver impairment,
127 meningitis, and encephalitis) [14].

128 Dengue warning signs and severe dengue were defined by the World Health Organization [17],
129 incorporating available clinical indicators from SEDSS intake and follow-up forms and abstracted
130 inpatient medical records. Dengue warning signs were defined by abdominal pain or tenderness,
131 persistent vomiting, plasma leakage (pleural or pericardial effusion or ascites), mucosal bleeding,
132 restlessness, hemoconcentration (defined as either a hematocrit increase of $\geq 20\%$ during illness or a
133 hematocrit value $\geq 20\%$ above baseline for age and sex), or hepatomegaly. Detailed definitions for these
134 variables have been provided previously [18]. Severe dengue was defined as severe plasma leakage or
135 shock, severe bleeding, or severe organ impairment [18]. The presence and overlap of warning signs
136 among severe dengue cases were visualized using an Euler plot via the *eulerr* R package [19].

137

138 *Sample collection and laboratory procedures*

139 Blood, nasopharyngeal (NP), and oropharyngeal (OP) specimens were collected at enrollment
140 from eligible participants. Additional blood samples (serum and whole blood) were also collected during
141 the convalescent phase. Participation required providing at least one sample (blood or OP/NP swab). All
142 patients had molecular testing for dengue virus for specimens collected within 7 days of symptom onset.
143 Serologic testing was done by Immunoglobulin M (IgM) antibody capture enzyme-linked immunosorbent
144 assay (ELISA) for anti-DENV antibodies for specimens collected >3 days after symptom onset [20].

145

146 *Variables*

147 A total of 40 variables were selected based on the WHO's criteria for severe dengue, physicians'
148 clinical experience, and a review of current literature to potentially differentiate between severe and non-
149 severe dengue cases [17, 18, 21-24]. These features included age group, days post onset of symptoms,
150 clinical signs and symptoms, laboratory findings, pre-existing health conditions, and dengue virus
151 serotype. Clinical symptoms included report of fever, rash, headache, myalgia, abdominal pain, chills,
152 itchy skin, eye pain, nasal discharge, cough, sore throat, persistent vomiting, diarrhea, arthralgia, arthritis,
153 back pain, calf pain, nausea, no appetite, and restlessness. Clinical signs, as observed by healthcare
154 providers, included objective fever at the time of enrollment, yellow skin (jaundice), observed bruising,
155 conjunctivitis, hepatomegaly, mucosal bleeding, pale skin, and blue lips (cyanosis). Clinical laboratory
156 findings comprised leukopenia (defined as white blood cell count $<4,000/\text{mm}^3$), the calculated value of
157 hemoconcentration (an increase in the concentration of red blood cells due to plasma loss), and dengue
158 immune status (primary or post-primary) as measured from results for anti-dengue virus immunoglobulin
159 G (IgG) on or before day 5 of illness. Pre-existing health conditions like obesity ($\text{BMI} \geq 30$), gastritis,
160 chronic arthritis, hypertension, chronic kidney disease, diabetes, thyroid disease, and high cholesterol
161 were also considered. Additionally, dengue virus serotype was included, coded as "unknown" for
162 probable cases identified through IgM ELISA, as serotype data was only available for confirmed reverse
163 transcription polymerase chain reaction (RT-PCR) cases. Thrombocytopenia and clinical fluid
164 accumulation were excluded from the analysis, as they generally manifest after the onset of severe disease
165 or are components of its definition, making them less useful as predictive features. This comprehensive
166 set of variables was intended to represent the multifaceted nature of factors influencing dengue infection
167 severity.

168

169 *Sensitivity analyses*

170 To further explore the performance of the models in resource-constrained settings, where
171 complete blood counts (CBCs), dengue immune status, and serotype information might not be readily
172 available, we performed sub-analyses to evaluate model robustness and applicability. The sub-analyses

173 included: 1) excluding results found on a CBC (i.e., leukopenia and hemoconcentration), 2) excluding
174 IgG and serotype results, and 3) excluding leukopenia, hemoconcentration, IgG, and serotype results.
175 These analyses were designed to assess how well the models could predict progression to severe dengue
176 in various clinical scenarios, particularly where access to comprehensive clinical laboratory results or
177 pathogen-specific testing might be limited.

178 To ensure the robustness of our models and account for potential confounding from co-circulating
179 arboviruses, we performed a sensitivity analysis with the highest-performing individual ML model,
180 excluding cases that tested positive for chikungunya virus (CHIKV) by either IgM or RT-PCR. This
181 analysis aimed to confirm that the predictive features for severe dengue remain consistent even in the
182 absence of CHIKV, given the potential overlap in clinical presentations between the two viruses.

183

184 *Sampling*

185 Our analysis included laboratory-confirmed dengue cases, confirmed by either molecular or
186 serologic testing, focusing on differentiating between severe and non-severe cases. Due to an imbalance
187 in the dataset, where non-severe cases were more prevalent, we used upsampling to balance the class
188 distribution. Upsampling involves increasing the number of minority class samples (severe dengue cases)
189 through duplication, which helps to prevent model bias towards the majority class and improves the
190 model's ability to accurately predict severe dengue cases [25]. Upsampling was done using the upSample
191 function from the caret package in R [26]. Following upsampling, the dataset was divided into training
192 and testing sets using a 70/30 split: 70% of the data was allocated for training the models, while the
193 remaining 30% was reserved for testing. This partitioning ensured that the models were trained on a
194 substantial portion of the data while retaining a sufficient amount for unbiased evaluation.

195

196 *Machine learning models*

197 An initial logistic regression (LR) model served as a baseline simple model to explore the
198 relationship between potential predictors and the outcome of severe dengue. Stepwise selection,

199 implemented using the stepAIC function from the MASS package in R [27], was used to iteratively add
200 or remove variables to identify the optimal model with the lowest Akaike Information Criterion. This
201 approach balances model complexity and goodness-of-fit by selecting variables that contribute
202 significantly to the model. The final logistic regression model, derived from stepwise selection, was
203 evaluated on both the training and testing sets.

204 In addition, we used nine ML methods to predict severe dengue and analyze feature importance.
205 A comprehensive selection of models was used to leverage different strengths, enhance predictive
206 performance, and provide a nuanced understanding of the factors contributing to severe dengue. The
207 algorithms used include Decision Trees (DT), K-Nearest Neighbors (KNN), Naïve Bayes, Support Vector
208 Machines (SVM), Artificial Neural Networks (ANN), Adaptive Boosting (AdaBoost), Categorical
209 Boosting (CatBoost), Light Gradient Boosting Machine (LightGBM), and eXtreme Gradient Boosting
210 (XGBoost). DTs create a tree-like structure to make predictions by recursively splitting the data based on
211 feature values [28]. KNNs predict the class of a data point by considering the majority class of its closest
212 neighbors in the feature space [29]. Naïve Bayes is a probabilistic classifier that applies Bayes' theorem,
213 assuming independence between predictors [30]. SVMs identify the optimal hyperplane to separate
214 classes, making them effective for high-dimensional data [31]. ANNs are inspired by biological neural
215 networks and consist of interconnected nodes that can capture complex patterns [32]. AdaBoost,
216 CatBoost, LightGBM, and XGBoost are ensemble methods that combine multiple weak learners to
217 improve predictive performance [33, 34]. AdaBoost adjusts weights to focus on difficult-to-predict
218 instances, CatBoost handles categorical features effectively, LightGBM is efficient with large datasets
219 due to its leaf-wise tree growth, and XGBoost uses regularization techniques to prevent overfitting,
220 enhancing accuracy and robustness [33-35].

221 The hyperparameters (model-specific settings, such as learning rate, maximum tree depth, or
222 number of estimators) for each ML model were carefully tuned using a grid search strategy to optimize
223 performance. The area under the receiver operating characteristic curve (AUC-ROC) was used as the
224 optimization metric, ensuring a focus on maximizing classification performance. We used 5-fold cross-

225 validation during model training to enhance robustness and mitigate overfitting. Specific details of the
226 grid search strategy and parameters included in each model are provided in Table S1.

227 The following R packages were used for model implementation: rpart [36] for DT, e1071 [37] for
228 Naïve Bayes and SVM, caret [26] for KNN, nnet [38] for ANN, keras [39] for DNN, ada [40] for
229 AdaBoost, catboost [41] for CatBoost, lightgbm [42] for LightGBM, and xgboost [43] for XGBoost. All
230 analyses were done using R version 4.4.0 [44].

231

232 *Ensemble model*

233 To leverage the predictive power of multiple ML algorithms, we used an ensemble learning
234 approach with a stacked generalization framework. This method combines the strengths of various
235 individual models to improve overall predictive performance and robustness. We used predictions from
236 LR and the nine different ML models as base learners in our ensemble. Specifically, we used a logistic
237 regression model as the meta-learner to combine the outputs of the base models. This approach allows the
238 meta-model to learn the optimal combination of base models' predictions. To improve the performance of
239 the meta-model, we again used stepwise selection with the stepAIC function from the MASS package
240 [27].

241 Pearson correlation coefficients were calculated to measure the linear correlation between the
242 predictions of the ML models. This analysis helps determine whether the models are making similar
243 predictions for severe dengue, potentially reflecting the selection of similar variables and patterns across
244 the models. The results were visualized in a heatmap using ggplot2 [45].

245

246 *Performance evaluation*

247 Model performance for each ML model and the meta-model was evaluated on both the training
248 and testing sets using AUC-ROC as the primary performance metric. AUC-ROC is an aggregate measure
249 of performance across all possible classification thresholds, providing a comprehensive assessment of the

250 model's ability to distinguish between classes. We used the DeLong method to calculate the confidence
251 intervals for the AUC-ROC to ensure accurate estimation of the model's performance [46].

252 The performance of the ensemble model was further evaluated using several metrics, including
253 accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1
254 score, and Cohen's kappa. These metrics provide a comprehensive view of the model's performance,
255 capturing both the ability to correctly classify severe dengue cases and the overall agreement between
256 predicted and actual classifications.

257

258 *Feature importance*

259 Feature importance was calculated for each ML model to quantify the contribution of each
260 variable to the model's predictive accuracy. This approach enables the identification of the most
261 influential features, which enhances our understanding of the factors driving the predictions for severe
262 dengue. Feature importance was assessed for both the 40-variable feature set and a subset excluding
263 CBCs, IgG, and serotype results. Different methods were applied across the ML algorithms to determine
264 feature importance. For ensemble-based methods, including XGBoost, LightGBM, and CatBoost, feature
265 importance was calculated using the Gain metric, which measures the contribution of each feature to the
266 model's decision-making process. Gain represents the improvement in the model's accuracy brought by a
267 feature, with higher values indicating greater importance. For XGBoost, LightGBM, and CatBoost, we
268 used the `xgb.importance`, `lgb.importance`, and `catboost.get_feature_importance` functions from the
269 `xgboost` [43], `lightgbm` [42], and `catboost` [41] packages, respectively.

270 Permutation importance was applied to assess feature importance for KNN, Naive Bayes, and
271 ANN. This method involves randomly shuffling feature values and measuring the subsequent decline in
272 model performance. A substantial decrease in accuracy indicates a highly influential feature. For DT,
273 feature importance was determined by the reduction in impurity (Gini index or entropy) achieved by
274 splitting data based on that feature. AdaBoost assigned importance to features based on their contribution
275 to correcting errors in subsequent models, with higher weights indicating greater influence. For SVM,

276 feature importance was derived from the absolute value of the model coefficients. The magnitude of these
277 coefficients reflects the influence of each feature on the decision boundary, with larger coefficients
278 indicating greater importance.

279 We also calculated and plotted SHapley Additive exPlanations (SHAP) values for the top three
280 performing models based on AUC-ROC. SHAP values provide a nuanced measure of each feature's
281 contribution to the prediction of severe dengue cases, enabling a deeper understanding of model decision-
282 making. Positive SHAP values indicate a higher likelihood of severe dengue, whereas negative values
283 suggest a protective effect. The SHAP approach is particularly valuable as it allows for the decomposition
284 of the prediction into individual feature contributions, offering a clear interpretation of how different
285 variables influence the model's predictions.

286

287 *Post-hoc variable reduction analysis*

288 To assess the predictive performance of a simplified variable set, we conducted a post-hoc
289 analysis using the ML model that achieved the highest AUC. This analysis aimed to identify the
290 minimum number of features needed to maintain high accuracy. We began with the top features identified
291 through SHAP values in the original 40-variable analysis, adding one feature at a time, starting with the
292 highest-ranking. At each step, we evaluated the AUC-ROC to determine the impact of including
293 additional features. The goal was to develop a more streamlined model that remains feasible and
294 interpretable, especially in clinical settings with limited diagnostic resource.

295

296 *Diagnostic accuracy of warning signs*

297 In addition to ML, we evaluated the diagnostic accuracy of individual warning signs for
298 identifying severe dengue cases. The performance of each warning sign was assessed using sensitivity,
299 specificity, PPV, NPV, and AUC-ROC. Sub-analyses assessed the performance of warning signs by
300 dengue serotype and immune status. Dengue serotype was determined via RT-PCR, whereas immune
301 status was classified based on IgG antibody results in the first 5 days after illness onset (primary: IgG;

302 post-primary: positive IgG). Cases lacking serotype or immune status data were excluded from sub-
303 analyses. This approach aimed to identify potential clinical differences in the presentation and predictive
304 capacity of warning signs for severe dengue across serotypes and infection statuses.

305

306 *Ethics statement*

307 The Institutional Review Boards at the Centers for Disease Control and Prevention (CDC),
308 Auxilio Mutuo, and Ponce Medical School Foundation approved the SEDSS study protocols 6214, and
309 120308-VR/2311173707, respectively. Written consent to participate was obtained from all adult
310 participants and emancipated minors. For minors aged 14 to 20 years, written consent was obtained, and
311 for those aged 7 to 13 years, parental written consent and participant assent were obtained.

312

313 **Results**

314 *Characteristics of dengue cases*

315 From May 2012 to August 2024, there were 51,877 unique AFI visits from 41,647 participants
316 enrolled in SEDSS, including 8,404 hospitalizations or transfers and 75 deaths. Of these visits, there were
317 50,189 AFI visits from 40,495 participants tested for DENV. From these, 1,708 (3.4%) had dengue (1,218
318 confirmed, 490 probable). The majority of the 1,206 serotyped dengue cases were DENV-1 (n=905,
319 75.0%), followed by DENV-3 (n=149, 12.4%), DENV-2 (n=102, 8.5%), and DENV-4 (n=50, 4.1%). Of
320 1,708 dengue cases, 759 (44.4%) were hospitalized or transferred, and two (0.1%) died. The median
321 duration from symptom onset to presentation at the emergency room was 3 days [IQR: 2, 5]. Of the 730
322 participants assessed for immune status using DENV IgG, 577 (79.0%) were positive, indicating post-
323 primary dengue, whereas the remaining 153 (21.0%) were negative, suggesting primary dengue
324 infections.

325 Among the 1,708 laboratory-confirmed dengue cases, 24.3% (n=415) were classified as severe
326 dengue. Compared to those without severe dengue, participants with severe dengue were more likely to
327 present between 4 to 6 days post-symptom onset (52.4% vs. 32.0%, $p < 0.001$) and be aged 10-19 years

328 (51.6% vs. 36.3%, $p < 0.001$) (Table 1). Among dengue cases tested, a higher proportion of severe cases
329 were post-primary DENV infections (85.7% vs. 76.3%, $p = 0.007$). Participants with severe dengue had a
330 higher prevalence of warning signs such as persistent vomiting (37.6% vs. 20.4%), abdominal pain
331 (79.0% vs. 51.6%), restlessness (64.6% vs. 40.9%), mucosal bleeding (22.9% vs. 13.8%), and
332 hemoconcentration (20.7% vs. 3.3%) compared to lab-confirmed non-severe cases (all $p < 0.001$) (Figure
333 1). All 30 dengue cases with seizures were classified as severe dengue. Leukopenia (77.1% vs. 53.5%)
334 was more prevalent among participants with severe dengue ($p < 0.001$).

335

336 *Performance of warning signs for predicting severe dengue*

337 Among warning signs, abdominal pain and restlessness had the highest sensitivities for predicting
338 severe dengue at 79.0% and 64.6%, respectively, but the lowest specificities of 48.4% and 59.1% (Table
339 2). In contrast, hepatomegaly and hemoconcentration demonstrated the highest specificities at 97.8% and
340 96.7%, respectively, but were less sensitive at 4.6% and 20.7%, respectively. The presence of any
341 warning sign yielded the highest sensitivity (92.8%) but a low specificity (29.2%), with an AUC-ROC of
342 61.1%. Combining three or more warning signs increased the specificity to 65.1% while maintaining
343 moderate sensitivity (87.2%), resulting in the highest AUC-ROC (71.3%) among the combinations tested.

344 Performance of warning signs for predicting severe dengue demonstrated some variability across
345 serotypes and immune status, though the interpretation is constrained by limited sample sizes and
346 overlapping confidence intervals (Tables S2-S3).

347

348 *Performance evaluation of machine learning models*

349 The ensemble model demonstrated a strong correlation between predictions from the CatBoost,
350 XGBoost, LightGBM, and AdaBoost models, with Pearson correlation coefficients of 0.91, 0.89, 0.89,
351 and 0.84, respectively, indicating that these gradient boosting models had substantial influence on the
352 ensemble's predictions (Figure 2). This high correlation suggests that the models may be selecting and
353 emphasizing similar variables in their predictive processes. In contrast, weaker correlations were observed

354 between the ensemble model and simpler models like KNN (0.48), Naïve Bayes (0.48), and DT (0.51),
355 indicating different prediction patterns and potential differences in variable selection. Additionally, high
356 inter-model correlations among gradient boosting models, particularly between LightGBM and XGBoost
357 (0.98), further support the idea that these models capture similar patterns in the data and rely on
358 comparable sets of variables.

359 AUC values for the 40-variable feature set across various ML models demonstrated varying
360 levels of predictive performance. Gradient boosting algorithms achieved the highest AUC values of
361 97.1% for CatBoost, 95.5% for XGBoost, and 94.5% for LightGBM, indicating strong discriminatory
362 power (Figure 3). ANN showed moderate performance (AUC = 88.4%), whereas LR and SVM had lower
363 discrimination (AUC = 79.4% and 78.9%, respectively). KNN, Naïve Bayes, and DT had the lowest AUC
364 values of 74.1%, 75.9%, and 76.2%, respectively, indicating limited predictive ability. The ensemble
365 meta-model provided a slight improvement over CatBoost with an AUC of 97.7%.

366 Exclusion of immune status and serotype data minimally affected model performance across all
367 ML algorithms (Figure 3). Conversely, removing leukopenia and hemoconcentration significantly
368 reduced predictive power for Naïve Bayes, LR, SVM, and ANN (AUC decreased by 3.2% ~ 5.2%).
369 CatBoost, XGBoost, LightGBM, and the ensemble model consistently maintained high performance,
370 showing minimal to no change in AUC, regardless of the inclusion or exclusion of leukopenia and
371 hemoconcentration. Excluding CHIKV-positive cases in the sensitivity analysis resulted in minimal
372 changes to the AUC-ROC scores for CatBoost, confirming that the model's predictive performance for
373 severe dengue remains robust even in the presence of co-circulating arboviruses (Table S4).

374 The ensemble model with 40 variables achieved the highest overall AUC of 97.7% with
375 corresponding sensitivity and specificity of 95.6% and 93.3%, respectively (Table S5). The F1 score was
376 94.5% and Kappa was 88.9%, indicating a high level of agreement and balanced performance between
377 precision and recall in the model's classification of severe dengue cases.

378

379 *Feature importance*

380 For the 40-variable feature set, SHAP values identified hemoconcentration, days post symptom
381 onset, and leukopenia as most important features for CatBoost, XGBoost, and LightGBM (Figure 4,
382 Figures S1-S2). Similarly, LR highlighted these variables as having the highest adjusted odds ratios for
383 severe dengue (hemoconcentration: aOR 7.02; leukopenia: aOR 2.24; days post onset 4–6 days: aOR
384 1.96) (Table S6). Additionally, these models highlighted pale skin, age group, and the clinical warning
385 signs of restlessness, abdominal pain, and persistent vomiting as key predictors of severe dengue
386 progression. Hemoconcentration also stood out as a top feature for Naïve Bayes and SVM (Figure S3-S4).
387 AdaBoost, which focuses on correcting errors from previous classifiers, assigned greater importance to
388 chronic conditions such as high cholesterol, chronic arthritis, and hypertension. Although hepatomegaly is
389 a recognized warning sign, it had a lower importance score in our analysis, suggesting it may play a more
390 limited role in predicting severe dengue in this context.

391

392 *Post-hoc variable reduction analysis*

393 To explore a more streamlined predictive model, we conducted a post-hoc variable reduction
394 analysis using the ML model with the highest AUC, CatBoost. Starting with the top feature identified by
395 SHAP values (days post onset), we sequentially added variables, assessing AUC-ROC at each step. The
396 AUC improved consistently with each additional variable, though the gains diminished over time. By
397 including just 20 variables—compared to the original 40-variable set—the model achieved an AUC of
398 96.5% (Figure 5). The optimal reduced feature set included days post onset, hemoconcentration,
399 leukopenia, restlessness, pale skin, abdominal pain, age group, diarrhea, rash, persistent vomiting, cough,
400 calf pain, sore throat, arthralgia, itchy skin, eye pain, back pain, mucosal bleeding, myalgia, and arthritis.
401 This reduced model offers a more practical and interpretable approach while maintaining high predictive
402 accuracy, making it feasible for use in clinical settings, especially where diagnostic resources are limited.

403

404 **Discussion**

405 Our study underscores the potential of ML models, particularly gradient boosting algorithms, to
406 outperform traditional warning signs in predicting severe dengue. This improved predictive ability could
407 transform clinical decision-making, enabling earlier and more accurate identification of high-risk patients,
408 thereby improving outcomes in dengue-endemic regions like Puerto Rico.

409 Hemoconcentration, days post symptom onset, and leukopenia emerged as the most important
410 features across multiple ML models, aligning with their known relevance in dengue prognosis [47-49].
411 Hemoconcentration, which reflects plasma leakage through an increased red blood cell concentration, was
412 consistently highlighted as a top predictor by CatBoost, XGBoost, LightGBM, and several other
413 algorithms. Days post symptom onset is a crucial temporal marker, likely capturing the dynamic nature of
414 disease as cases often progress to the critical phase of dengue (when severe disease occurs) 3–7 days after
415 symptom onset. Leukopenia, or low white blood cell count, often reflects the body’s response to viral
416 infections, including dengue. In addition, pale skin, age group, and clinical warning signs such as
417 restlessness, abdominal pain, and persistent vomiting were identified as key predictors of severe dengue
418 progression. The variability in feature importance across models emphasizes the complexity of severe
419 dengue prediction, highlighting the need for tailored approaches that account for both individual patient
420 characteristics and disease progression.

421 Our analysis highlights the strong predictive performance of gradient boosting algorithms—
422 CatBoost, XGBoost, and LightGBM—with AUC values above 94%, reflecting their ability to capture
423 complex, non-linear patterns in clinical data [35, 50-52]. Despite this high predictive accuracy, the
424 interpretability of these models remains a limitation in clinical settings, where transparency in decision-
425 making is critical for trust and practical use [53-55]. Compared to simpler models like LR, which offers
426 straightforward interpretations of how each variable influences severe dengue risk, gradient boosting
427 algorithms can be challenging to interpret. This trade-off between high performance and interpretability
428 suggests that ML models may be most useful as supplementary tools for alerting clinicians to high-risk
429 cases, rather than as standalone decision aids.

430 In contrast, simpler models like Naïve Bayes, Decision Trees, and KNN showed weaker
431 correlations and lower AUC values, reflecting their limitations in capturing data complexity. The
432 ensemble model, achieving the highest AUC of 97.7%, demonstrates the added value of combining
433 multiple ML algorithms to enhance predictive accuracy, with high sensitivity and specificity, making it
434 particularly useful in resource-limited settings [56, 57]. Additionally, the exclusion of immune status and
435 serotype data had minimal impact on model performance, indicating these variables are not essential for
436 accurate prediction in this context. The high NPV across models suggests that ML tools can still be
437 valuable for identifying low-risk patients who may not require intensive monitoring. In these cases, the
438 ML model's recommendation could serve as an early discharge or outpatient management decision-
439 support mechanism, further optimizing healthcare resource allocation.

440 The post-hoc variable reduction analysis demonstrated that a streamlined set of 20 variables
441 achieved strong predictive accuracy (AUC of 96.5%), close to the full 40-variable model. This reduced
442 set offers a balance between interpretability and performance, making it more practical for clinical
443 application, particularly in settings with limited diagnostic resources. By focusing on essential
444 predictors—such as hemoconcentration, days post onset, leukopenia, and key symptoms—this approach
445 prioritizes feasibility and interpretability, even if it means a modest sacrifice in predictive power. For
446 added clinical utility, LR could complement ML approaches by further refining and validating the
447 reduced variable set with interpretable odds ratios, enabling clinicians to apply these findings more
448 confidently in practice [58].

449 Traditional warning signs showed both strengths and limitations. Abdominal pain and restlessness
450 were the most sensitive indicators, consistent with other studies [6, 9], yet their low specificities limit
451 their utility. Conversely, markers like hepatomegaly and hemoconcentration had high specificity but low
452 sensitivity. Combining multiple warning signs improved specificity while maintaining high sensitivity,
453 yielding the highest AUC among tested combinations. In contrast, gradient boosting ML algorithms
454 offered a more balanced approach with high sensitivity and specificity, crucial for accurate risk
455 stratification in clinical settings. Our findings align with a recent study where an 8-gene XGBoost model

456 outperformed clinical warning signs, significantly improving negative predictive power and
457 demonstrating strong generalizability across patient cohorts [54]. Although our models focus on
458 demographic and clinical features, the integration of gene expression data presents an intriguing avenue
459 for future exploration [54, 55].

460 This study has several limitations. First, these ML models need to be re-fitted to different
461 variables and populations to ensure accuracy across various settings. Second, the relatively small dataset
462 increases the risk of overfitting, potentially affecting the models' robustness and generalizability. Third,
463 the prevalence of DENV-1 cases from the 2012-2013 outbreak may limit the models' applicability to
464 other periods, regions, populations, age groups, or serotypes. Fourth, the models were developed using
465 data from the SEDSS, where inclusion criteria required febrile illness, potentially limiting generalizability
466 to broader populations. Fifth, due to limited sample sizes for serotypes and immune statuses, we could not
467 conduct ML analyses for these subgroups; future studies with larger datasets are needed to validate
468 findings. Sixth, there is a potential limitation related to the inclusion of false negatives—SEDSS cases
469 who may have later presented to a non-SEDSS facility with severe disease—although this is considered
470 unlikely given typical healthcare-seeking behaviors. Seventh, the SEDSS data's robustness may not
471 accurately reflect real-world conditions, where datasets are often sparse, contain free-text fields, or have
472 incomplete information, potentially affecting model performance. Eighth, dengue and severe dengue are
473 often underdiagnosed and underreported, which could impact model results; however, our findings likely
474 represent a conservative estimate when accounting for underreporting. Finally, implementing ML models
475 in clinical practice may require computational resources and infrastructure not available in all settings,
476 particularly in low-resource environments where dengue is endemic.

477

478 **Conclusions**

479 Although traditional warning signs are essential in clinical practice, their low specificity often
480 leads to high hospitalization rates, potentially overwhelming healthcare systems. Our findings suggest
481 that ML models, particularly gradient boosting algorithms, offer a more effective approach by integrating

482 multiple variables and capturing complex interactions, thereby improving specificity while maintaining
483 sensitivity. Implementing these models in clinical decision-making could help identify patients at highest
484 risk for progression to severe dengue, reducing unnecessary hospitalizations and easing healthcare
485 burdens. Although resource constraints may limit direct ML implementation in some settings, platforms
486 like SEDSS can still leverage ML techniques to identify key predictors of severe disease. This approach
487 can optimize patient care by prioritizing the most critical predictors, even in low-resource environments
488 where advanced ML algorithms may not be feasible.

489

490 **Abbreviations**

491 AdaBoost: Adaptive Boosting

492 AFI: acute febrile illness

493 ANN: Artificial Neural Networks

494 AUC-ROC: area under the receiver operating characteristic curve

495 CatBoost: Categorical Boosting

496 CBC: complete blood count

497 CHIKV: chikungunya virus

498 CSPF: convalescent sample processing form

499 CEMI: Centro de Emergencia y Medicina Integrada

500 CIF: case investigation form

501 CMESL: Centro Médico Episcopal San Lucas

502 DT: Decision Trees

503 ELISA: enzyme-linked immunosorbent assay

504 HESL: Hospital Episcopal San Lucas

505 KNN: K-Nearest Neighbors

506 LightGBM: Light Gradient Boosting Machine

507 LR: logistic regression

- 508 ML: machine learning
- 509 NPV: negative predictive value
- 510 PPV: positive predictive value
- 511 RT-PCR: reverse transcription polymerase chain reaction
- 512 SEDSS: Sentinel Enhanced Dengue Surveillance System
- 513 SHAP: SHapley Additive exPlanations
- 514 SVM: Support Vector Machines
- 515 XGBoost: eXtreme Gradient Boosting
- 516 WHO: World Health Organization

517 **Declarations**

518 **Ethics approval and consent to participate:** The Institutional Review Boards at the Centers for Disease
519 Control and Prevention (CDC), Auxilio Mutuo, and Ponce Medical School Foundation approved the
520 SEDSS study protocols 6214, and 120308-VR/2311173707, respectively. Written consent to participate
521 was obtained from all adult participants and emancipated minors. For minors aged 14 to 20 years, written
522 consent was obtained, and for those aged 7 to 13 years, parental written consent and participant assent
523 were obtained.

524 **Consent for publication:** Not applicable.

525 **Availability of data and materials:** Data cannot be shared publicly because data cannot be deidentified
526 at the granular level of analyses performed. Data are available from the CDC management team (contact:
527 dengue@cdc.gov) for researchers who meet the criteria for access to confidential data.

528 **Competing interests:** The authors declare no conflict of interests.

529 **Funding:** This research was funded by Centers for Disease Control and Prevention, grant
530 numbers U01CK000473 and U01CK000580 (VRA).

531 **Authors' contributions:** Conception and design of the study: ZJM, MBT, and JMW. Acquisition of data:
532 VR. Analysis and interpretation of data: ZJM. Drafting the article: ZJM. Revising the article critically for
533 important intellectual content: ZJM, DR, MBT, VR, GP, LEA, and JMW. All authors have made
534 substantial contributions to the study. All authors read and approved the final manuscript.

535 **Acknowledgements:** Not applicable.

536 **Disclaimer:** The findings and conclusions in this report are those of the authors and do not necessarily
537 represent the official position of the US Centers for Disease Control and Prevention.

538 **References**

- 539 1. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS,
540 Hoen AG, Sankoh O *et al*: **The global distribution and burden of dengue.** *Nature* 2013,
541 **496**(7446):504-507.
- 542 2. **Global, regional, and national age-sex-specific mortality for 282 causes of death in 195**
543 **countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease**
544 **Study 2017.** *Lancet* 2018, **392**(10159):1736-1788.
- 545 3. Ryff KR, Rivera A, Rodriguez DM, Santiago GA, Medina FA, Ellis EM, Torres J, Pobutsky A,
546 Munoz-Jordan J, Paz-Bailey G, Adams LE: **Epidemiologic Trends of Dengue in U.S.**
547 **Territories, 2010-2020.** *MMWR Surveill Summ* 2023, **72**(4):1-12.
- 548 4. Thayer MB, Marzan-Rodriguez M, Torres Aponte J, Rivera A, Rodriguez DM, Madewell ZJ,
549 Rysava K, Paz-Bailey G, Adams LE, Johansson MA: **Dengue epidemic alert thresholds: A tool**
550 **for surveillance and epidemic detection.** *medRxiv* 2024:2024.2010.2022.24315684.
- 551 5. **WHO Guidelines Approved by the Guidelines Review Committee.** In: *Dengue: Guidelines for*
552 *Diagnosis, Treatment, Prevention and Control: New Edition.* edn. Geneva: World Health
553 Organization, Copyright © 2009, World Health Organization.; 2009.
- 554 6. Ahmad MH, Ibrahim MI, Mohamed Z, Ismail N, Abdullah MA, Shueb RH, Shafei MN: **The**
555 **sensitivity, specificity and accuracy of warning signs in predicting severe dengue, the severe**
556 **dengue prevalence and its associated factors.** *International Journal of Environmental Research*
557 *and Public Health* 2018, **15**(9):2018.
- 558 7. Narvaez F, Gutierrez G, Pérez MA, Elizondo D, Nuñez A, Balmaseda A, Harris E: **Evaluation of**
559 **the traditional and revised WHO classifications of dengue disease severity.** *PLoS neglected*
560 *tropical diseases* 2011, **5**(11):e1397.
- 561 8. Macedo GA, Gonin MLC, Pone SM, Cruz OG, Nobre FF, Brasil P: **Sensitivity and specificity of**
562 **the World Health Organization dengue classification schemes for severe dengue assessment**
563 **in children in Rio de Janeiro.** *PloS one* 2014, **9**(4):e96314.

- 564 9. Jayaratne SD, Atukorale V, Gomes L, Chang T, Wijesinghe T, Fernando S, Ogg GS, Malavige
565 GN: **Evaluation of the WHO revised criteria for classification of clinical disease severity in**
566 **acute adult dengue infection.** *BMC Res Notes* 2012, **5**:645.
- 567 10. Thein TL, Gan VC, Lye DC, Yung CF, Leo YS: **Utilities and limitations of the World Health**
568 **Organization 2009 warning signs for adult dengue severity.** *PLoS Negl Trop Dis* 2013,
569 **7**(1):e2023.
- 570 11. Hoyos W, Aguilar J, Toro M: **Dengue models based on machine learning techniques: A**
571 **systematic literature review.** *Artificial Intelligence in Medicine* 2021, **119**:102157.
- 572 12. Gupta G, Khan S, Guleria V, Almjally A, Alabduallah BI, Siddiqui T, Albahlal BM, Alajlan SA,
573 Al-Subaie M: **DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment**
574 **Analysis and Machine Learning Algorithms.** *Diagnostics (Basel)* 2023, **13**(6).
- 575 13. Tomashek KM, Rivera A, Torres-Velasquez B, Hunsperger EA, Munoz-Jordan JL, Sharp TM,
576 Rivera I, Sanabria D, Blau DM, Galloway R *et al*: **Enhanced Surveillance for Fatal Dengue-**
577 **Like Acute Febrile Illness in Puerto Rico, 2010-2012.** *PLoS Negl Trop Dis* 2016,
578 **10**(10):e0005025.
- 579 14. Madewell ZJ, Hernandez-Romieu AC, Wong JM, Zambrano LD, Volkman HR, Perez-Padilla J,
580 Rodriguez DM, Lorenzi O, Espinet C, Munoz-Jordan J *et al*: **Sentinel Enhanced Dengue**
581 **Surveillance System - Puerto Rico, 2012-2022.** *MMWR Surveill Summ* 2024, **73**(3):1-29.
- 582 15. Read JS, Torres-Velasquez B, Lorenzi O, Rivera Sanchez A, Torres-Torres S, Rivera LV, Capre-
583 Franceschi SM, Garcia-Gubern C, Munoz-Jordan J, Santiago GA, Alvarado LI: **Symptomatic**
584 **Zika Virus Infection in Infants, Children, and Adolescents Living in Puerto Rico.** *JAMA*
585 *Pediatr* 2018, **172**(7):686-693.
- 586 16. Wong JM, Volkman HR, Adams LE, García CO, Martínez-Quiñones A, Perez-Padilla J, Bertrán-
587 Pasarell J, de la Pena DS, Tosado-Acevedo R, Santiago GA: **Clinical Features of COVID-19,**
588 **Dengue, and Influenza among Adults Presenting to Emergency Departments and Urgent**

- 589 **Care Clinics—Puerto Rico, 2012–2021.** *The American Journal of Tropical Medicine and*
590 *Hygiene* 2023, **108**(1):107.
- 591 17. **Handbook for clinical management of dengue**
592 [\[https://www.who.int/publications/i/item/9789241504713\]](https://www.who.int/publications/i/item/9789241504713)
- 593 18. Paz-Bailey G, Sánchez-González L, Torres-Velasquez B, Jones ES, Perez-Padilla J, Sharp TM,
594 Lorenzi O, Delorey M, Munoz-Jordan JL, Tomashek KM *et al*: **Predominance of Severe Plasma**
595 **Leakage in Pediatric Patients With Severe Dengue in Puerto Rico.** *The Journal of Infectious*
596 *Diseases* 2022, **226**(11):1949-1958.
- 597 19. **eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses** [\[https://cran.r-](https://cran.r-project.org/web/packages/eulerr/index.html)
598 [project.org/web/packages/eulerr/index.html\]](https://cran.r-project.org/web/packages/eulerr/index.html)
- 599 20. Tomashek KM, Lorenzi OD, Andújar-Pérez DA, Torres-Velásquez BC, Hunsperger EA, Munoz-
600 Jordan JL, Perez-Padilla J, Rivera A, Gonzalez-Zeno GE, Sharp TM *et al*: **Clinical and**
601 **epidemiologic characteristics of dengue and other etiologic agents among patients with**
602 **acute febrile illness, Puerto Rico, 2012-2015.** *PLoS Negl Trop Dis* 2017, **11**(9):e0005859.
- 603 21. Tsheten T, Clements ACA, Gray DJ, Adhikary RK, Furuya-Kanamori L, Wangdi K: **Clinical**
604 **predictors of severe dengue: a systematic review and meta-analysis.** *Infect Dis Poverty* 2021,
605 **10**(1):123.
- 606 22. Sangkaew S, Ming D, Boonyasiri A, Honeyford K, Kalayanaroj S, Yacoub S, Dorigatti I,
607 Holmes A: **Risk predictors of progression to severe disease during the febrile phase of**
608 **dengue: a systematic review and meta-analysis.** *Lancet Infect Dis* 2021, **21**(7):1014-1026.
- 609 23. Khan MAS, Al Mosabbir A, Raheem E, Ahmed A, Rouf RR, Hasan M, Alam FB, Hannan N,
610 Yesmin S, Amin R *et al*: **Clinical spectrum and predictors of severity of dengue among**
611 **children in 2019 outbreak: a multicenter hospital-based study in Bangladesh.** *BMC Pediatr*
612 2021, **21**(1):478.
- 613 24. Yuan K, Chen Y, Zhong M, Lin Y, Liu L: **Risk and predictive factors for severe dengue**
614 **infection: A systematic review and meta-analysis.** *PLoS One* 2022, **17**(4):e0267186.

- 615 25. Werner de Vargas V, Schneider Aranda JA, Dos Santos Costa R, da Silva Pereira PR, Victória
616 Barbosa JL: **Imbalanced data preprocessing techniques for machine learning: a systematic**
617 **mapping study**. *Knowl Inf Syst* 2023, **65**(1):31-57.
- 618 26. **caret: Classification and Regression Training** [[https://cran.r-](https://cran.r-project.org/web/packages/caret/index.html)
619 [project.org/web/packages/caret/index.html](https://cran.r-project.org/web/packages/caret/index.html)]
- 620 27. **MASS: Support Functions and Datasets for Venables and Ripley's MASS** [[https://cran.r-](https://cran.r-project.org/web/packages/MASS/index.html)
621 [project.org/web/packages/MASS/index.html](https://cran.r-project.org/web/packages/MASS/index.html)]
- 622 28. Kotsiantis SB: **Decision trees: a recent overview**. *Artificial Intelligence Review* 2013, **39**:261-
623 283.
- 624 29. Peterson LE: **K-nearest neighbor**. *Scholarpedia* 2009, **4**(2):1883.
- 625 30. Webb GI, Keogh E, Miikkulainen R: **Naïve Bayes**. *Encyclopedia of machine learning* 2010,
626 **15**(1):713-714.
- 627 31. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B: **Support vector machines**. *IEEE*
628 *Intelligent Systems and their applications* 1998, **13**(4):18-28.
- 629 32. Aggarwal CC: **Neural networks and deep learning**, vol. 10: Springer; 2018.
- 630 33. Schapire RE: **The boosting approach to machine learning: An overview**. *Nonlinear estimation*
631 *and classification* 2003:149-171.
- 632 34. Bentéjac C, Csörgő A, Martínez-Muñoz G: **A comparative analysis of gradient boosting**
633 **algorithms**. *Artificial Intelligence Review* 2021, **54**:1937-1967.
- 634 35. Khan AA, Chaudhari O, Chandra R: **A review of ensemble learning and data augmentation**
635 **models for class imbalanced problems: Combination, implementation and evaluation**.
636 *Expert Systems with Applications* 2024, **244**:122778.
- 637 36. **rpart: Recursive Partitioning and Regression Trees** [[https://cran.r-](https://cran.r-project.org/web/packages/rpart/index.html)
638 [project.org/web/packages/rpart/index.html](https://cran.r-project.org/web/packages/rpart/index.html)]
- 639 37. **e1071: Misc Functions of the Department of Statistics, Probability Theory Group**
640 **(Formerly: E1071), TU Wien** [<https://cran.r-project.org/web/packages/e1071/index.html>]

- 641 38. **nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models** [[https://cran.r-](https://cran.r-project.org/web/packages/nnet/index.html)
642 [project.org/web/packages/nnet/index.html](https://cran.r-project.org/web/packages/nnet/index.html)]
- 643 39. **keras: R Interface to 'Keras'** [[https://cran.r-](https://cran.r-project.org/web/packages/keras/index.html)
644 [project.org/web/packages/keras/index.html](https://cran.r-project.org/web/packages/keras/index.html)]
- 645 40. **ada: The R Package Ada for Stochastic Boosting** [[https://cran.r-](https://cran.r-project.org/web/packages/ada/index.html)
646 [project.org/web/packages/ada/index.html](https://cran.r-project.org/web/packages/ada/index.html)]
- 647 41. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A: **CatBoost: unbiased boosting with categorical features**. *Advances in neural information processing systems* 2018, **31**.
- 648 42. **lightgbm: Light Gradient Boosting Machine** [[https://cran.r-](https://cran.r-project.org/web/packages/lightgbm/index.html)
649 [project.org/web/packages/lightgbm/index.html](https://cran.r-project.org/web/packages/lightgbm/index.html)]
- 650 43. **xgboost: Extreme Gradient Boosting** [[https://cran.r-](https://cran.r-project.org/web/packages/xgboost/index.html)
651 [project.org/web/packages/xgboost/index.html](https://cran.r-project.org/web/packages/xgboost/index.html)]
- 652 44. **R: A language and environment for statistical computing.**
- 653 45. Wickham H: **ggplot2**. *Wiley interdisciplinary reviews: computational statistics* 2011, **3(2)**:180-
654 185.
- 655 46. DeLong ER, DeLong DM, Clarke-Pearson DL: **Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach**. *Biometrics*
656 1988, **44(3)**:837-845.
- 657 47. Lee IK, Huang CH, Huang WC, Chen YC, Tsai CY, Chang K, Chen YH: **Prognostic Factors in Adult Patients with Dengue: Developing Risk Scoring Models and Emphasizing Factors Associated with Death ≤ 7 Days after Illness Onset and ≤ 3 Days after Presentation**. *J Clin Med* 2018, **7(11)**.
- 662 48. Wakimoto MD, Camacho LA, Guaraldo L, Damasceno LS, Brasil P: **Dengue in children: a systematic review of clinical and laboratory factors associated with severity**. *Expert Rev Anti Infect Ther* 2015, **13(12)**:1441-1456.

- 665 49. Sami CA, Tasnim R, Hassan SS, Khan AH, Yasmin R, Monir-Uz-Zaman M, Sarker MAS, Arafat
666 SM: **Clinical profile and early severity predictors of dengue fever: Current trends for the**
667 **deadliest dengue infection in Bangladesh in 2022.** *IJID Reg* 2023, **9**:42-48.
- 668 50. Ming DK, Hernandez B, Sangkaew S, Vuong NL, Lam PK, Nguyet NM, Tam DTH, Trung DT,
669 Tien NTH, Tuan NM: **Applied machine learning for the risk-stratification and clinical**
670 **decision support of hospitalised patients with dengue in Vietnam.** *PLOS digital health* 2022,
671 **1**(1):e0000005.
- 672 51. Ong SQ, Isawasan P, Ngesom AMM, Shahar H, Lasim AmM, Nair G: **Predicting dengue**
673 **transmission rates by comparing different machine learning models with vector indices and**
674 **meteorological data.** *Scientific reports* 2023, **13**(1):19129.
- 675 52. Madewell ZJ, Rodriguez DM, Thayer MB, Rivera-Amill V, Torres Aponte J, Marzan-Rodriguez
676 M, Paz-Bailey G, Adams LE, Wong JM: **Machine learning for improved dengue diagnosis,**
677 **Puerto Rico.** *medRxiv* 2024:2024.2011.2013.24317272.
- 678 53. Chaw JK, Chaw SH, Quah CH, Sahrani S, Ang MC, Zhao Y, Ting TT: **A predictive analytics**
679 **model using machine learning algorithms to estimate the risk of shock development among**
680 **dengue patients.** *Healthcare Analytics* 2024, **5**:100290.
- 681 54. Liu YE, Saul S, Rao AM, Robinson ML, Agudelo Rojas OL, Sanz AM, Verghese M, Solis D,
682 Sibai M, Huang CH *et al*: **An 8-gene machine learning model improves clinical prediction of**
683 **severe dengue progression.** *Genome Med* 2022, **14**(1):33.
- 684 55. Hung SJ, Tsai HP, Wang YF, Ko WC, Wang JR, Huang SW: **Assessment of the Risk of Severe**
685 **Dengue Using Intra-host Viral Population in Dengue Virus Serotype 2 Patients via Machine**
686 **Learning.** *Front Cell Infect Microbiol* 2022, **12**:831281.
- 687 56. Cramer EY, Ray EL, Lopez VK, Bracher J, Brennen A, Castro Rivadeneira AJ, Gerding A,
688 Gneiting T, House KH, Huang Y *et al*: **Evaluation of individual and ensemble probabilistic**
689 **forecasts of COVID-19 mortality in the United States.** *Proc Natl Acad Sci U S A* 2022,
690 **119**(15):e2113561119.

- 691 57. Dean NE, Pastore YPA, Madewell ZJ, Cummings DAT, Hitchings MDT, Joshi K, Kahn R,
692 Vespignani A, Halloran ME, Longini IM, Jr.: **Ensemble forecast modeling for the design of**
693 **COVID-19 vaccine efficacy trials.** *Vaccine* 2020, **38**(46):7213-7216.
- 694 58. Ho TS, Weng TC, Wang JD, Han HC, Cheng HC, Yang CC, Yu CH, Liu YJ, Hu CH, Huang CY
695 *et al*: **Comparing machine learning with case-control models to identify confirmed dengue**
696 **cases.** *PLoS Negl Trop Dis* 2020, **14**(11):e0008843.
- 697

698 **Figure Titles and Legends**

699 **Figure 1.** Euler plot of proportion of severe dengue cases with each warning sign, Sentinel Enhanced
700 Dengue Surveillance System, Puerto Rico, 2012–2024.

701 **Figure 2.** Pearson’s correlation of predictions between machine learning models, Sentinel Enhanced
702 Dengue Surveillance System, Puerto Rico, 2012–2024. Pearson correlation coefficients measure the linear
703 agreement between the predictions of different machine learning models. Higher values indicate similar
704 prediction patterns across models, suggesting that models are identifying similar cases as severe dengue.
705 Darker colors represent higher correlations.

706 **Figure 3.** Forest plot of AUC values for Decision Trees (DT), K-Nearest Neighbors (KNN), Naïve Bayes,
707 Support Vector Machines (SVM), Artificial Neural Networks (ANN), Adaptive Boosting (AdaBoost),
708 Categorical Boosting (CatBoost), Light Gradient Boosting Machine (LightGBM), eXtreme Gradient
709 Boosting (XGBoost), and ensemble models for a 40-variable feature set and subsets excluding CBCs,
710 IgG, and serotype results, Sentinel Enhanced Dengue Surveillance System, Puerto Rico, 2012–2024.
711 DeLong method was used to obtain the 95% confidence intervals for the AUC-ROC. CBC = complete
712 blood count, IgG = immunoglobulin G, AUC-ROC = area under the receiver operating characteristic
713 curve.

714 **Figure 4.** SHapley Additive exPlanations (SHAP) values for the 40 Features in CatBoost, Sentinel
715 Enhanced Dengue Surveillance System, Puerto Rico, 2012–2024. SHAP values measure each feature’s
716 contribution to the prediction of severe dengue in the CatBoost model. Positive SHAP values indicate a
717 higher likelihood of severe dengue, while negative values suggest a lower likelihood (or protective
718 effect). Each dot represents a single case, with its horizontal position showing the SHAP value, reflecting
719 the strength and direction of the feature’s impact. The color of the dots indicates the actual feature value
720 for each case. For most features, values are binary (0 or 1), representing presence or absence (e.g., rash or
721 no rash). For age group, the scale ranges from 0 to 7, with 0 indicating the youngest age group (<1 year)
722 and 7 indicating the oldest age group (≥ 50 years). An example interpretation: if ‘persistent vomiting’ has
723 a positive SHAP value and the dot is green (value = 1), it indicates that the presence of persistent

724 vomiting strongly increases the likelihood of severe dengue for that case. The mean SHAP values shown
725 on the right represent the average absolute impact of each feature across all cases, indicating the overall
726 importance of that feature in the model’s predictions.

727 **Figure 5.** Iterative improvement in area under the curve (AUC) with additional variables in CatBoost
728 model for severe dengue prediction, Sentinel Enhanced Dengue Surveillance System, Puerto Rico, 2012–
729 2024. This figure shows the change in AUC as top-performing variables are sequentially added to the
730 CatBoost model. Starting with the highest-impact feature, “Days post onset,” each subsequent model
731 includes one additional variable in the order of their mean SHAP values. The combinations of variables
732 and their AUC, along with 95% confidence intervals, are shown to demonstrate the predictive gain with
733 each added variable.

734

Table 1. Demographic and clinical characteristics of participants with laboratory-confirmed dengue (RT-PCR and IgM ELISA) by severity, Sentinel Enhanced Dengue Surveillance System, Puerto Rico, 2012–2024.

	Total N = 1,708 n (column %)	Severe dengue N = 415 n (column %)	Not severe N = 1,293 n (column %)	p-value
Days post onset				<0.001
0	128 (7.5)	18 (4.3)	110 (8.5)	
1-3	810 (47.4)	133 (32.0)	677 (52.4)	
4-6	695 (40.7)	243 (58.6)	452 (35.0)	
7+	75 (4.4)	21 (5.1)	54 (4.2)	
Age group				<0.001
<1	32 (1.9)	5 (1.2)	27 (2.1)	
1-4	109 (6.4)	11 (2.7)	98 (7.6)	
5-9	238 (13.9)	42 (10.1)	196 (15.2)	
10-19	684 (40.0)	214 (51.6)	470 (36.3)	
20-29	225 (13.2)	38 (9.2)	187 (14.5)	
30-39	118 (6.9)	27 (6.5)	91 (7.0)	
40-49	93 (5.4)	20 (4.8)	73 (5.6)	
50+	209 (12.2)	58 (14.0)	151 (11.7)	
Female sex	802 (47.0)	185 (44.6)	617 (47.7)	0.290
DENV immune status				<0.001
Post-primary	577 (33.8)	180 (43.4)	397 (30.7)	
Primary	153 (9.0)	30 (7.2)	123 (9.5)	
Not tested	978 (57.3)	205 (49.4)	773 (59.8)	
DENV Serotype				<0.001
1	905 (53.0)	226 (54.5)	679 (52.5)	
2	102 (6.0)	15 (3.6)	87 (6.7)	
3	149 (8.7)	18 (4.3)	131 (10.1)	
4	50 (2.9)	19 (4.6)	31 (2.4)	
Unknown	502 (29.4)	137 (33.0)	365 (28.2)	
Comorbidities				
Chronic pulmonary disease or asthma	347 (20.3)	76 (18.3)	271 (21.0)	0.273
Cancer	27 (1.6)	6 (1.4)	21 (1.6)	0.978
Chronic kidney disease	12 (0.7)	5 (1.2)	7 (0.5)	0.285
Coronary heart disease	52 (3.0)	16 (3.9)	36 (2.8)	0.347
Diabetes	108 (6.3)	29 (7.0)	79 (6.1)	0.601
High cholesterol	84 (4.9)	24 (5.8)	60 (4.6)	0.420
Hypertension	156 (9.1)	41 (9.9)	115 (8.9)	0.611
Arthritis	23 (1.3)	5 (1.2)	18 (1.4)	0.965
Thyroid disease	74 (4.3)	16 (3.9)	58 (4.5)	0.682
Obesity	214 (12.5)	43 (10.4)	171 (13.2)	0.148
Gastritis	28 (1.6)	9 (2.2)	19 (1.5)	0.451
Warning signs				
Persistent vomiting	420 (24.6)	156 (37.6)	264 (20.4)	<0.001
Abdominal pain	995 (58.3)	328 (79.0)	667 (51.6)	<0.001
Restlessness	797 (46.7)	268 (64.6)	529 (40.9)	<0.001
Mucosal bleeding	274 (16.0)	95 (22.9)	179 (13.8)	<0.001
Hemoconcentration	129 (7.6)	86 (20.7)	43 (3.3)	<0.001
Hepatomegaly	47 (2.8)	19 (4.6)	28 (2.2)	0.015
Other clinical signs/symptoms				
Fever	1695 (99.2)	412 (99.3)	1283 (99.2)	1.000

Conjunctivitis	147 (8.6)	26 (6.3)	121 (9.4)	0.064
Chills	1362 (79.7)	369 (88.9)	993 (76.8)	<0.001
Nausea	1175 (68.8)	333 (80.2)	842 (65.1)	<0.001
No appetite	1330 (77.9)	363 (87.5)	967 (74.8)	<0.001
Rash	971 (56.9)	286 (68.9)	685 (53.0)	<0.001
Yellow skin	69 (4.0)	36 (8.7)	33 (2.6)	<0.001
Itchy skin	627 (36.7)	190 (45.8)	437 (33.8)	<0.001
Bruise	103 (6.0)	35 (8.4)	68 (5.3)	0.025
Headache	1432 (83.8)	372 (89.6)	1060 (82.0)	<0.001
Eye pain	1026 (60.1)	288 (69.4)	738 (57.1)	<0.001
Myalgia	1276 (74.7)	362 (87.2)	914 (70.7)	<0.001
Arthralgia	1062 (62.2)	301 (72.5)	761 (58.9)	<0.001
Back pain	912 (53.4)	261 (62.9)	651 (50.3)	<0.001
Calf pain	627 (36.7)	180 (43.4)	447 (34.6)	<0.001
Arthritis	251 (14.7)	84 (20.2)	167 (12.9)	<0.001
Nasal discharge	545 (31.9)	149 (35.9)	396 (30.6)	0.052
Sore throat	615 (36.0)	168 (40.5)	447 (34.6)	0.034
Cough	726 (42.5)	189 (45.5)	537 (41.5)	0.167
Diarrhea	736 (43.1)	228 (54.9)	508 (39.3)	<0.001
Seizure	30 (1.8)	30 (7.2)	0 (0.0)	<0.001
Pale skin	783 (45.8)	266 (64.1)	517 (40.0)	<0.001
Blue lips	72 (4.2)	35 (8.4)	37 (2.9)	<0.001
Laboratory				
Leukopenia	1012 (59.3)	320 (77.1)	692 (53.5)	<0.001

p-values were calculated using either the chi-square test or Fisher's exact test, as appropriate, based on the sample sizes in each category.

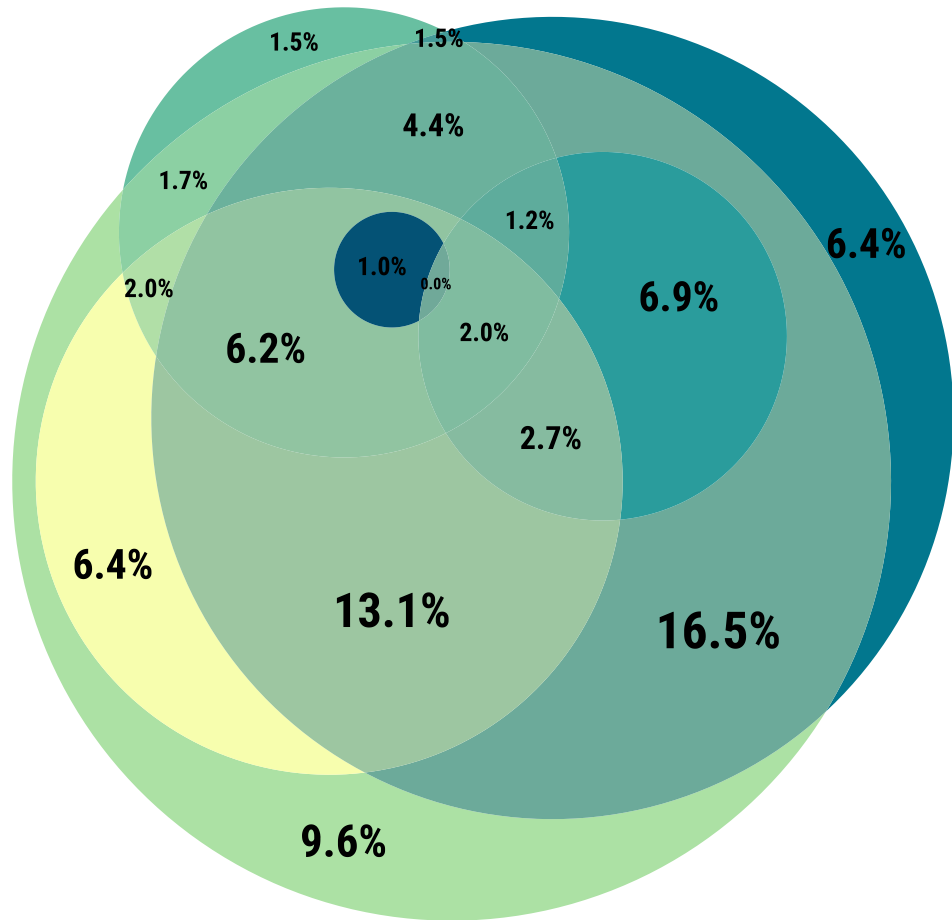
735

736

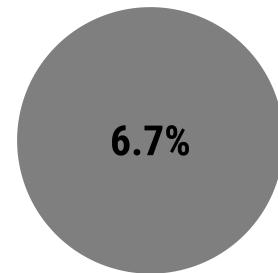
Table 2. Performance of warning signs for predicting severe dengue (n=1708), Sentinel Enhanced Dengue Surveillance System, Puerto Rico, 2012–2024.

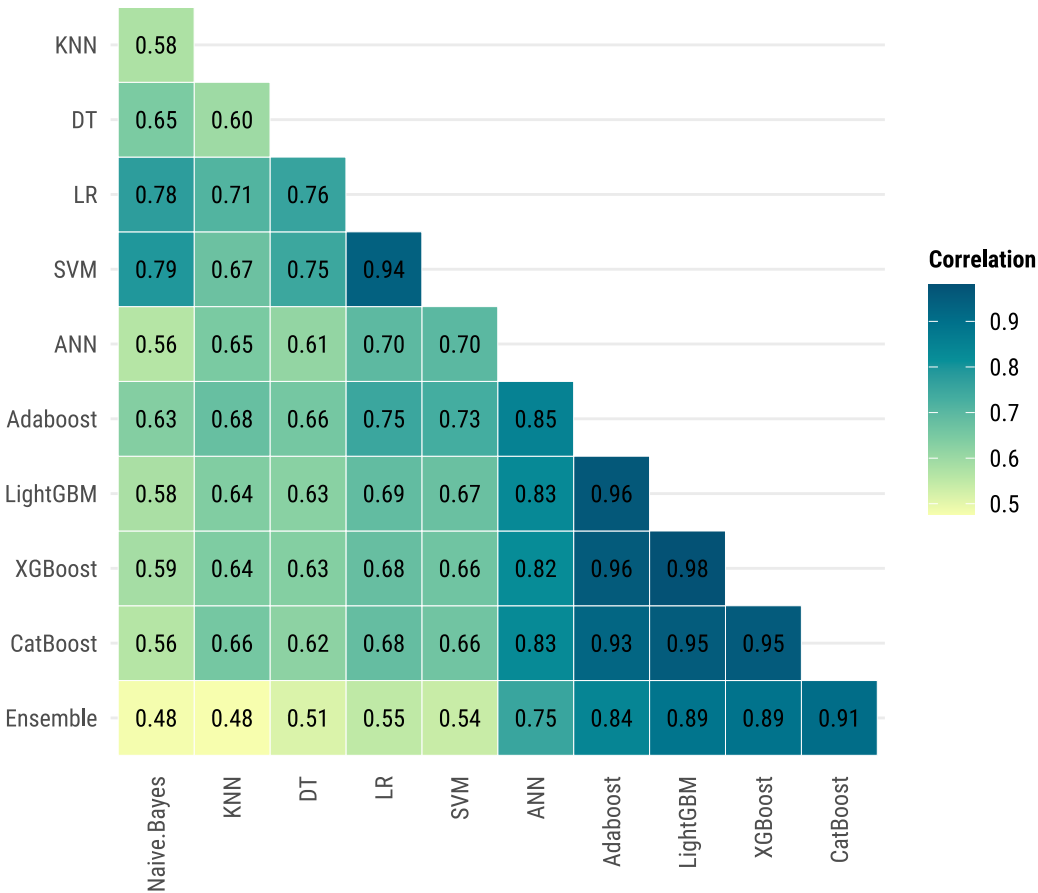
Warning sign	True Positive n (%)	True Negative n (%)	False Positive n (%)	False Negative n (%)	Sensitivity % (95% CI)	Specificity % (95% CI)	Positive Predictive Value % (95% CI)	Negative Predictive Value % (95% CI)	AUC-ROC % (95% CI)
Persistent vomiting	156 (9.1)	1029 (60.2)	264 (15.5)	259 (15.2)	37.6 (32.9, 42.4)	79.6 (77.3, 81.7)	37.1 (32.5, 42.0)	79.9 (77.6, 82.0)	58.5 (56.0, 61.1)
Abdominal pain	328 (19.2)	626 (36.7)	667 (39.1)	87 (5.1)	79.0 (74.8, 82.9)	48.4 (45.7, 51.2)	33.0 (30.0, 36.0)	87.8 (85.2, 90.1)	60.4 (58.5, 62.3)
Restlessness	268 (15.7)	764 (44.7)	529 (31.0)	147 (8.6)	64.6 (59.8, 69.2)	59.1 (56.4, 61.8)	33.6 (30.3, 37.0)	83.9 (81.3, 86.2)	58.7 (56.7, 60.8)
Mucosal bleeding	95 (5.6)	1114 (65.2)	179 (10.5)	320 (18.7)	22.9 (18.9, 27.2)	86.2 (84.2, 88.0)	34.7 (29.0, 40.6)	77.7 (75.4, 79.8)	56.2 (53.2, 59.2)
Hemoconcentration	86 (5.0)	1250 (73.2)	43 (2.5)	329 (19.3)	20.7 (16.9, 24.9)	96.7 (95.5, 97.6)	66.7 (57.8, 74.7)	79.2 (77.1, 81.1)	72.9 (68.7, 77.1)
Hepatomegaly	19 (1.1)	1265 (74.1)	28 (1.6)	396 (23.2)	4.6 (2.8, 7.1)	97.8 (96.9, 98.6)	40.4 (26.4, 55.7)	76.2 (74.0, 78.2)	58.3 (51.1, 65.5)
Any warning sign	388 (22.7)	347 (20.3)	946 (55.4)	27 (1.6)	93.5 (90.7, 95.7)	26.8 (24.4, 29.3)	29.1 (26.7, 31.6)	92.8 (89.7, 95.2)	60.9 (59.1, 62.7)
Only one warning sign	78 (9.0)	347 (40.1)	413 (47.7)	27 (3.1)	74.3 (64.8, 82.3)	45.7 (42.1, 49.3)	15.9 (12.8, 19.4)	92.8 (89.7, 95.2)	54.3 (52.2, 56.4)
Only two warning signs	126 (14.9)	347 (41.0)	347 (41.0)	27 (3.2)	82.4 (75.4, 88.0)	50.0 (46.2, 53.8)	26.6 (22.7, 30.9)	92.8 (89.7, 95.2)	59.7 (57.3, 62.1)
Three or more warning signs	184 (24.7)	347 (46.6)	186 (25.0)	27 (3.6)	87.2 (81.9, 91.4)	65.1 (60.9, 69.2)	49.7 (44.5, 54.9)	92.8 (89.7, 95.2)	71.3 (68.4, 74.1)

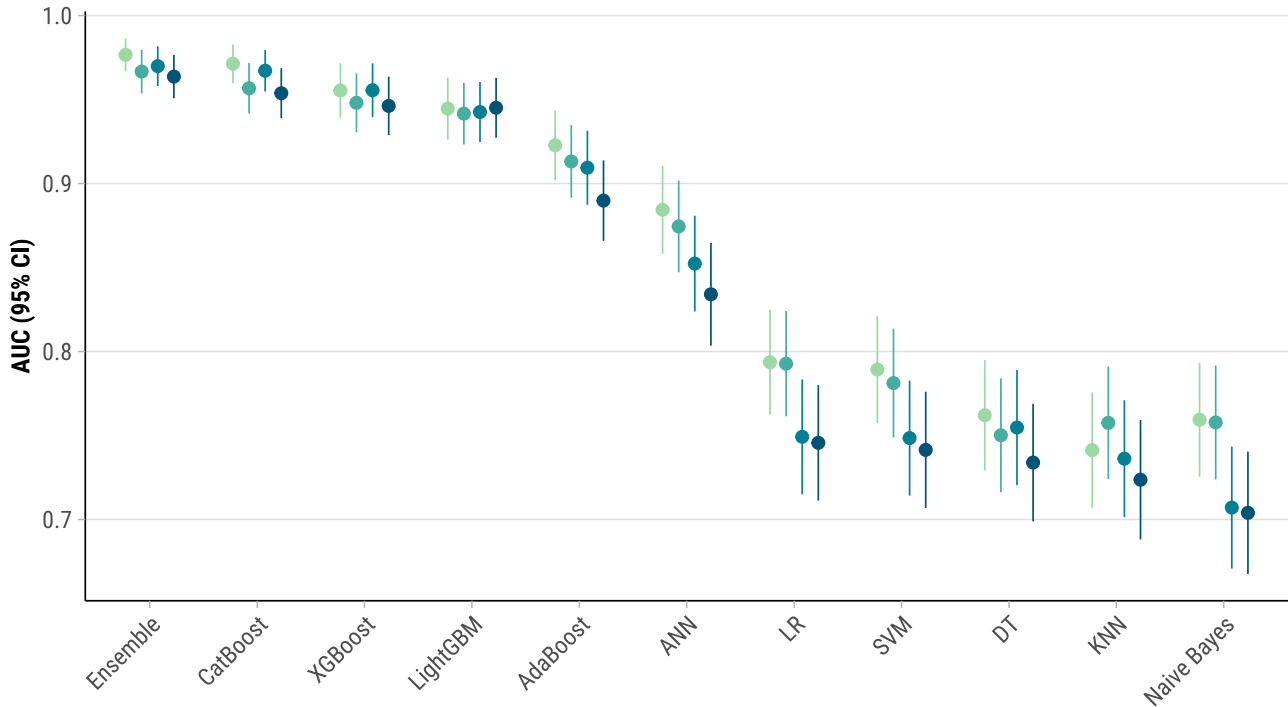
AUC-ROC: area under receiver operating characteristic curve.



- Persistent vomiting
- Abdominal pain
- Mucosal bleeding
- Hemoconcentration
- Restlessness
- Hepatomegaly
- No warning signs







● 40-variable feature set
 ● Excluding IgG test and serotype
 ● Excluding CBCs
 ● Excluding CBCs, IgG test, and serotype

