MEDICINE SOUTHEAST ASIA

SOPHISTICA
WHERE INNOVATION IN MEDICINE BEGINS

Article

Machine learning classifiers using echocardiographic parameters to predict shock occurrence in severe dengue: A secondary analysis of a prospective cohort

Novie Homenta Rampengan[1]*, Nizam Albar[2], Derren Rampengan[3]

[1] Department of Child Health, Faculty of Medicine, Universitas Sam Ratulangi, Manado, Indonesia
[2] Department of Computer Engineering, Faculty of Engineering, Universitas Serambi Mekkah, Banda Aceh, Indonesia
[3] Faculty of Medicine, Universitas Sam Ratulangi, Manado, Indonesia
* Correspondence: Novie Homenta Rampengan (email: novierampengan@unsrat.ac.id)

Abstract

Background: Shock is the most critical complication of dengue, yet early risk stratification remains challenging. This study performed a secondary analysis of a prospective cohort to evaluate machine learning (ML) models for predicting shock and to identify the most influential clinical and haemodynamic predictors.

Results: Eighty-six patients were included, of whom 67 (77.9%) developed shock. In basic features only Model, performance was limited, indicated by AUC ≤0.68. Addition of demographic and haemodynamic features to the model showed the best balance of discrimination and calibration, with Random Forest achieving an AUC of 0.80 and accuracy of 0.80. Combining all features yielded slightly higher accuracy (up to 0.82 with Random Forest) but no clear calibration advantage over Model 2. Feature importance analyses ranked haemodynamic parameters, such as stroke volume, stroke volume index, cardiac index, and aortic VTI, as the most influential predictors.

Methods: Data from patients admitted with dengue shock syndrome at the Hospital for Tropical Diseases, Ho Chi Minh City, were reanalyzed. Demographic, vital sign, haemodynamic, and echocardiographic variables were included. Shock occurrence served as the binary outcome. Seven ML classifiers (Decision Tree, Random Forest, XGBoost, LightGBM, AdaBoost, Gradient Boosting, Naive Bayes) were trained using stratified 10-fold cross-validation. Performance was assessed by accuracy, AUC, sensitivity, specificity, F1 score, and RMSE. Feature importance was analyzed to explore mechanistic relevance.

Conclusion: ML classifiers showed moderate ability to predict shock in dengue, with Random Forest, LightGBM, and XGBoost performing best. A parsimonious model using demographic and haemodynamic features achieved calibration similar to more complex models and was supported by feature importance findings.

Keywords: Dengue shock syndrome, Random Forest, LightGBM, haemodynamic parameters, echocardiography

Academic Editor: Maulana A. Empitu

Received: 15 October 2025
Revised: 1 November 2025
Accepted: 20 December 2025
Published: date

Citation: To be added by editorial staff during production.

1. Introduction

Dengue is a major vector-borne viral disease with a high burden in tropical and subtropical regions, where its age-adjusted incidence rate reach 752.04 per 100,000 population in 2021 [1]. Severe forms of dengue, characterized by plasma leakage, bleeding, and organ impairment, are associated with significant morbidity and mortality [2,3]. Among the critical complications, circulatory shock represents a leading cause of death in severe dengue, particularly when not recognized and managed promptly [3,4]. Early identification of patients at risk of shock remains a central challenge in clinical practice.

Traditional risk stratification relies primarily on demographic and basic clinical variables such as age, sex, and vital signs [4,5]. However, these parameters often lack the sensitivity and specificity needed to predict shock progression. Echocardiography has been increasingly recognized as a valuable tool for evaluating haemodynamic and myocardial function in dengue patients, providing insight into stroke volume, cardiac index, and ventricular performance [6,7]. Prior works have demonstrated that echocardiographic abnormalities are common in dengue and may contribute to the risk of shock and poor outcomes [6,7]. Nevertheless, the integration of these multidimensional data into predictive models has been limited.

Recent advances in machine learning (ML) enable the construction of models that incorporate complex clinical and echocardiographic features to improve risk prediction [8-10]. By leveraging such approaches, it is possible to capture nonlinear interactions among variables and generate clinically useful risk estimates. The present study aimed to develop and compare ML-based predictive models for shock occurrence in patients with severe dengue. This work represents a secondary analysis of a prospective observational cohort conducted at the Hospital for Tropical Diseases (HTD), Ho Chi Minh City, Vietnam [7]. Four feature sets were defined, ranging from demographic and basic clinical data to advanced echocardiographic haemodynamic and structural/functional parameters. The objective was to evaluate whether inclusion of haemodynamic and echocardiographic indices improves prediction of shock beyond basic clinical features.

2. Results

2.1. Characteristics

A total of 86 patients were included in the analysis, of whom 67 (77.9%) developed shock and 19 (22.1%) did not. The mean age was lower in the shock group compared with the non-shock group (12.22 ± 6.37 vs. 15.26 ± 11.96 years), although this difference was not statistically significant ($p = 0.143$). The distribution of sex was similar between groups (male: 58.2% vs. 47.4%; $p = 0.563$). Vital signs, including pulse, systolic and diastolic blood pressure, mean arterial pressure, temperature, and respiratory rate, did not differ significantly between groups. Several haemodynamic parameters showed significant differences. Patients with shock had lower stroke volume (26.45 ± 6.94 vs. 31.88 ± 8.61 mL, $p = 0.005$), stroke volume index (22.05 ± 3.67 vs. 26.01 ± 3.66, $p < 0.001$), and cardiac index (2.10 ± 0.39 vs. 2.40 ± 0.45, $p = 0.005$) compared with those without shock. Among echocardiographic measures, pulmonary ejection time was shorter in the shock group (264.19 ± 37.13 vs. 284.32 ± 33.87 ms, $p = 0.037$), and AorticVTIAve.1 was significantly lower (12.00 ± 1.88 vs. 13.81 ± 1.83, $p < 0.001$). Other structural and functional indices, including ejection fraction, MPI values, and chamber dimensions, were comparable between groups. The characteristics and echocardiography profiles of the dengue patients are summarized and presented in **Table 1**.

Table 1. Characteristics and echocardiography profiles of patients stratified by shock occurrence

Variable	mean ± SD	*p*-value[a]

	No shock (n = 19)	**Shock (n = 67)**	
Age, years	15.26 ± 11.96	12.22 ± 6.37	0.143
Pulse, /min	92.63 ± 14.15	95.99 ± 14.43	0.372
Systolic blood pressure (mmHg)	99.21 ± 8.54	101.79 ± 9.80	0.301
Diastolic blood pressure (mmHg)	65.26 ± 9.05	69.82 ± 12.66	0.147
Mean arterial pressure (mmHg)	142.72 ± 14.02	148.34 ± 15.67	0.162
Temperature (°C)	37.26 ± 0.45	37.27 ± 1.26	0.985
Respiratory rate (/min)	23.84 ± 4.86	24.93 ± 4.91	0.397
Stroke volume (mL)	31.88 ± 8.61	26.45 ± 6.94	0.005
Stroke volume index	26.01 ± 3.66	22.05 ± 3.67	<0.001
Cardiac index	2.40 ± 0.45	2.10 ± 0.39	0.005
Body surface area (m^2)	1.22 ± 0.30	1.20 ± 0.26	0.742
Left ventricular diastolic diameter	3.80 ± 0.58	3.63 ± 0.51	0.209
Left ventricular systolic diameter	2.68 ± 0.43	2.57 ± 0.44	0.321
Ejection fraction (%)	64.78 ± 4.87	64.45 ± 6.54	0.841
Left myocardial performance index	0.38 ± 0.15	0.38 ± 0.14	0.957
Right myocardial performance index	0.18 ± 0.08	0.21 ± 0.11	0.385
E/A ratio	0.61 ± 0.14	0.63 ± 0.17	0.507
Pulmonary ejection time	284.32 ± 33.87	264.19 ± 37.13	0.037
Mitral A-wave duration	344.78 ± 53.50	330.51 ± 45.29	0.248
Tricuspid A-wave duration	337.32 ± 53.52	317.50 ± 43.62	0.1
Aortic ejection time	250.86 ± 36.09	240.03 ± 31.59	0.205
LVOT	1.70 ± 0.21	1.67 ± 0.19	0.542
LVOT at site 2	2.92 ± 0.69	2.81 ± 0.62	0.505
Aortic velocity–time integral 1	13.81 ± 1.83	12.00 ± 1.88	<0.001
Aortic velocity–time integral 2	13.00 ± 2.36	12.03 ± 1.96	0.074
Inferior vena cava diameter (min.)	0.77 ± 0.30	0.66 ± 0.20	0.064
Inferior vena cava diameter (max.)	1.18 ± 0.29	1.10 ± 0.24	0.259
Inferior vena cava collapsibility index	0.36 ± 0.09	0.41 ± 0.09	0.052
Sex, n (%)			
Male	9 (47.4)	39 (58.2)	0.563[b]

Female	10 (52.6)	28 (41.8)	
Recurrent shock, n (%)			
No	19 (100.0)	43 (64.2)	NA
Yes	0 (0.0)	24 (35.8)	

*Otherwise stated, p-value is estimated from t-test; bChi-square NA, not applicable; MAP, mean arterial pressure; SV, stroke volume; SVI, stroke volume index; CI, cardiac index; BSA, body surface area; E/A ratio, ratio of early to late mitral inflow velocity; LVOT, left ventricular outflow tract diameter;

2.2. Machine learning predictive performance

Across all models, sensitivity was generally higher than specificity, reflecting stronger ability to identify patients with shock compared to those without (**Table 2**). Model 1 (basic features) showed modest predictive performance, with AUC values ranging from 0.54 (Decision Tree) to 0.68 (Gradient Boosting). Random Forest and LightGBM achieved the highest sensitivity (0.94 ± 0.15), although specificity remained low across classifiers (≤0.35). Model 2 (basic and haemodynamic features) improved overall discrimination. Random Forest achieved the best balance, with accuracy 0.80 ± 0.11 and AUC 0.80 ± 0.16, accompanied by high sensitivity (0.93 ± 0.09). XGBoost and LightGBM also performed well, with AUC values >0.70. Model 3 (basic and structural or functional features) performed less consistently, with lower AUC values overall. Most classifiers achieved moderate accuracy (0.66–0.73) but poor specificity, suggesting limited added predictive value from structural/functional parameters alone. Model 4 (all Combined) demonstrated the highest predictive ability. Random Forest again yielded the strongest performance (accuracy 0.82 ± 0.14, AUC 0.75 ± 0.24, sensitivity 0.95 ± 0.09). LightGBM also achieved robust results (accuracy 0.80 ± 0.12, AUC 0.74 ± 0.21, specificity 0.50 ± 0.32). Inclusion of haemodynamic parameters (Model 2) and the full combined feature set (Model 4) improved model discrimination compared with basic or structural/functional features alone. Tree-based ensemble methods, particularly Random Forest and LightGBM, consistently outperformed other classifiers (**Table 2**).

Table 2. Predictive performance of machine learning classifiers for shock occurrence across feature sets

Model and feature set	**Accuracy**	**AUC**	**Sensitivity**	**Specificity**	**F1**	**RMSE**
Model 1						
Decision Tree	0.70 ± 0.15	0.54 ± 0.19	0.83 ± 0.15	0.25 ± 0.33	0.81 ± 0.11	0.51
Random Forest	0.75 ± 0.13	0.64 ± 0.21	0.94 ± 0.15	0.15 ± 0.32	0.85 ± 0.10	0.424
XGBoost	0.71 ± 0.19	0.68 ± 0.29	0.87 ± 0.20	0.15 ± 0.23	0.81 ± 0.16	0.44
LightGBM	0.76 ± 0.16	0.60 ± 0.36	0.94 ± 0.15	0.20 ± 0.33	0.85 ± 0.11	0.43
AdaBoost	0.68 ± 0.15	0.60 ± 0.25	0.79 ± 0.16	0.30 ± 0.40	0.79 ± 0.12	0.49
Gradient Boosting	0.72 ± 0.17	0.68 ± 0.26	0.84 ± 0.15	0.30 ± 0.33	0.82 ± 0.12	0.47
Naive Bayes	0.61 ± 0.20	0.57 ± 0.27	0.69 ± 0.20	0.35 ± 0.32	0.72 ± 0.16	0.50
Model 2						
Decision Tree	0.70 ± 0.06	0.57 ± 0.13	0.79 ± 0.09	0.35 ± 0.32	0.80 ± 0.04	0.55

Random Forest	0.80 ± 0.11	0.80 ± 0.16	0.93 ± 0.09	0.35 ± 0.32	0.88 ± 0.07	0.36
XGBoost	0.78 ± 0.12	0.72 ± 0.14	0.92 ± 0.10	0.30 ± 0.33	0.87 ± 0.08	0.40
LightGBM	0.77 ± 0.12	0.77 ± 0.14	0.89 ± 0.10	0.35 ± 0.32	0.86 ± 0.08	0.38
AdaBoost	0.81 ± 0.11	0.55 ± 0.28	0.93 ± 0.10	0.40 ± 0.30	0.88 ± 0.07	0.46
Gradient Boosting	0.75 ± 0.11	0.66 ± 0.23	0.87 ± 0.10	0.35 ± 0.32	0.84 ± 0.07	0.45
Naive Bayes	0.71 ± 0.15	0.68 ± 0.23	0.79 ± 0.13	0.45 ± 0.27	0.80 ± 0.11	0.46
Model 3						
Decision Tree	0.68 ± 0.19	0.54 ± 0.24	0.78 ± 0.20	0.30 ± 0.40	0.78 ± 0.15	0.54
Random Forest	0.73 ± 0.15	0.58 ± 0.29	0.92 ± 0.15	0.10 ± 0.30	0.84 ± 0.11	0.43
XGBoost	0.68 ± 0.12	0.49 ± 0.23	0.85 ± 0.16	0.10 ± 0.20	0.80 ± 0.09	0.49
LightGBM	0.67 ± 0.11	0.48 ± 0.28	0.83 ± 0.15	0.10 ± 0.20	0.79 ± 0.09	0.48
AdaBoost	0.73 ± 0.14	0.49 ± 0.21	0.89 ± 0.16	0.15 ± 0.23	0.83 ± 0.11	0.48
Gradient Boosting	0.66 ± 0.18	0.66 ± 0.27	0.80 ± 0.21	0.20 ± 0.33	0.77 ± 0.15	0.51
Naive Bayes	0.59 ± 0.24	0.59 ± 0.28	0.65 ± 0.28	0.40 ± 0.37	0.68 ± 0.22	0.52
Model 4						
Decision Tree	0.69 ± 0.09	0.55 ± 0.15	0.79 ± 0.11	0.30 ± 0.33	0.79 ± 0.06	0.55
Random Forest	0.82 ± 0.14	0.75 ± 0.24	0.95 ± 0.09	0.40 ± 0.37	0.89 ± 0.09	0.37
XGBoost	0.75 ± 0.16	0.72 ± 0.18	0.86 ± 0.15	0.35 ± 0.32	0.84 ± 0.11	0.40
LightGBM	0.80 ± 0.12	0.74 ± 0.21	0.89 ± 0.13	0.50 ± 0.32	0.87 ± 0.09	0.37
AdaBoost	0.77 ± 0.14	0.57 ± 0.21	0.89 ± 0.10	0.35 ± 0.32	0.86 ± 0.09	0.46
Gradient Boosting	0.73 ± 0.12	0.72 ± 0.18	0.82 ± 0.12	0.40 ± 0.30	0.82 ± 0.08	0.48
Naive Bayes	0.70 ± 0.14	0.65 ± 0.24	0.76 ± 0.12	0.50 ± 0.32	0.79 ± 0.10	0.48

Receiver operating characteristic (ROC) curves for Model 2 (Basic + Haemodynamic) demonstrated varying discriminatory performance across classifiers (**Figure 1**). The Decision Tree model showed modest separation from the reference line, reflecting limited discriminative ability. Ensemble methods, particularly Random Forest, XGBoost, and LightGBM, achieved stronger ROC curves, with Random Forest demonstrating the most favorable profile, consistent with its higher AUC (0.80 ± 0.16). Gradient Boosting and AdaBoost also provided reasonable discrimination, though with lower stability. Naive Bayes achieved moderate performance, with ROC curves suggesting weaker separation compared to tree-based ensembles.

Calibration plots for the same classifiers are shown in **Figure 2**. The Decision Tree model demonstrated poor alignment with the ideal calibration line, with marked overestimation at lower predicted probabilities. Random Forest, XGBoost, and LightGBM exhibited improved calibration, with predicted risks more closely following the diagonal line, although some degree of underestimation was noted at intermediate probabilities. AdaBoost produced well-aligned calibration in the mid-range but showed deviations at the extremes. Gradient Boosting showed substantial miscalibration, particularly at lower probability ranges. Naive Bayes demonstrated moderate alignment, though with some underestimation at higher predicted risks. Overall, ensemble tree-based models (Random Forest, XGBoost, LightGBM) displayed the most favorable

calibration performance, whereas single learners and Gradient Boosting showed weaker agreement between predicted and observed outcomes.

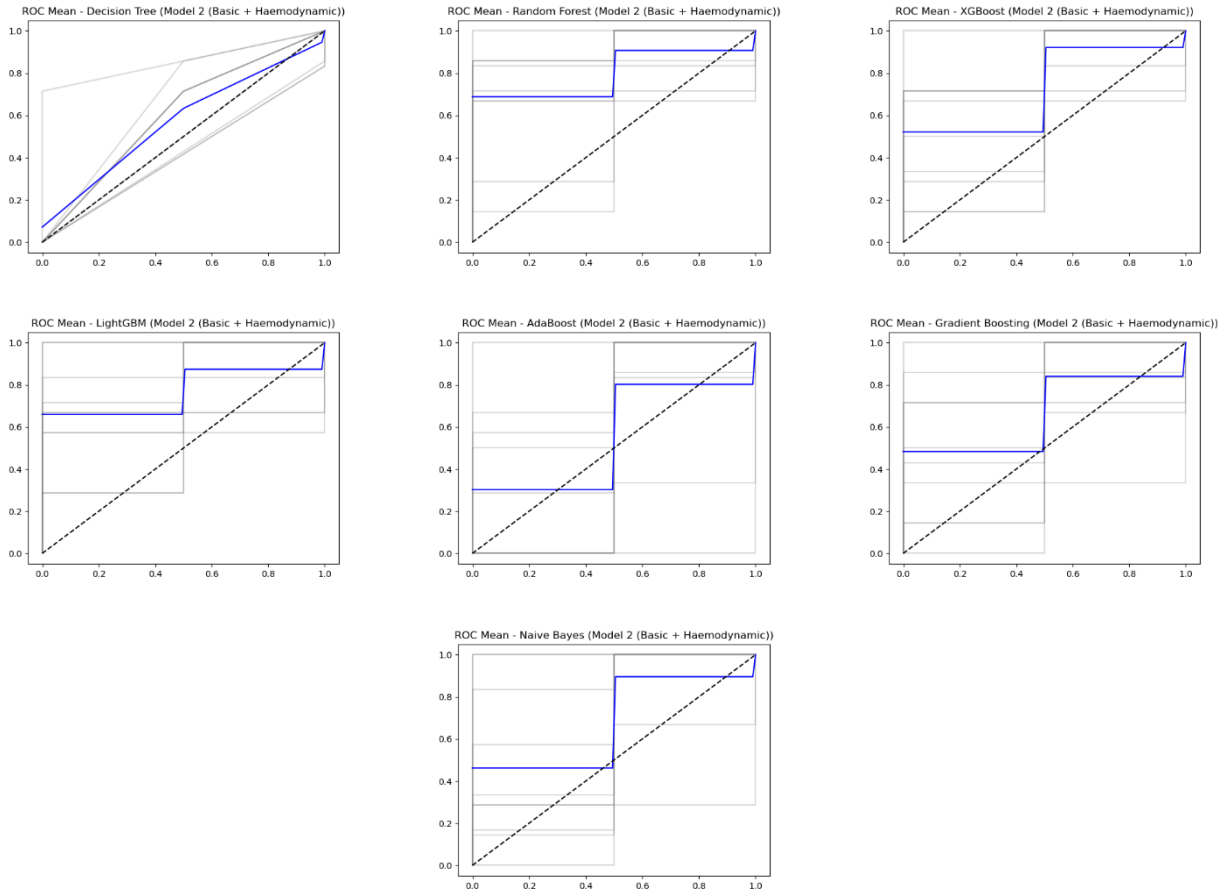
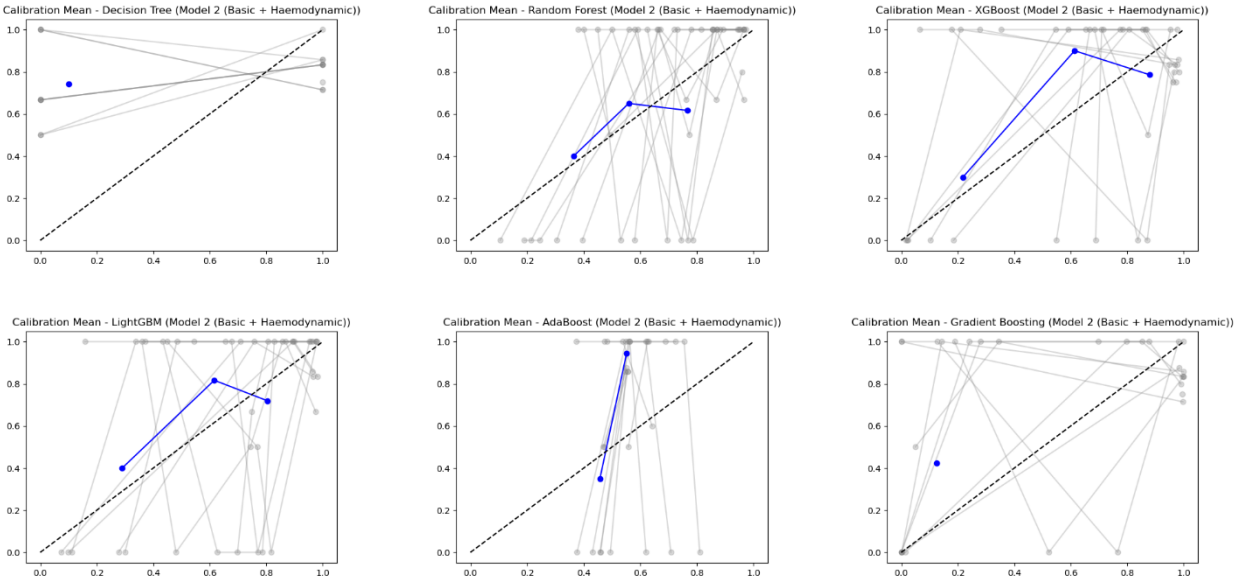


Figure 1. Receiver operating characteristic (ROC) curves for Model 2 (basic and haemodynamic features) across seven classifiers. Mean ROC curves are shown in blue, with the dashed line representing the reference line (AUC = 0.5).



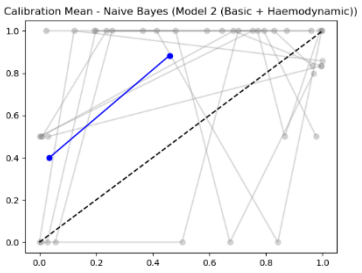
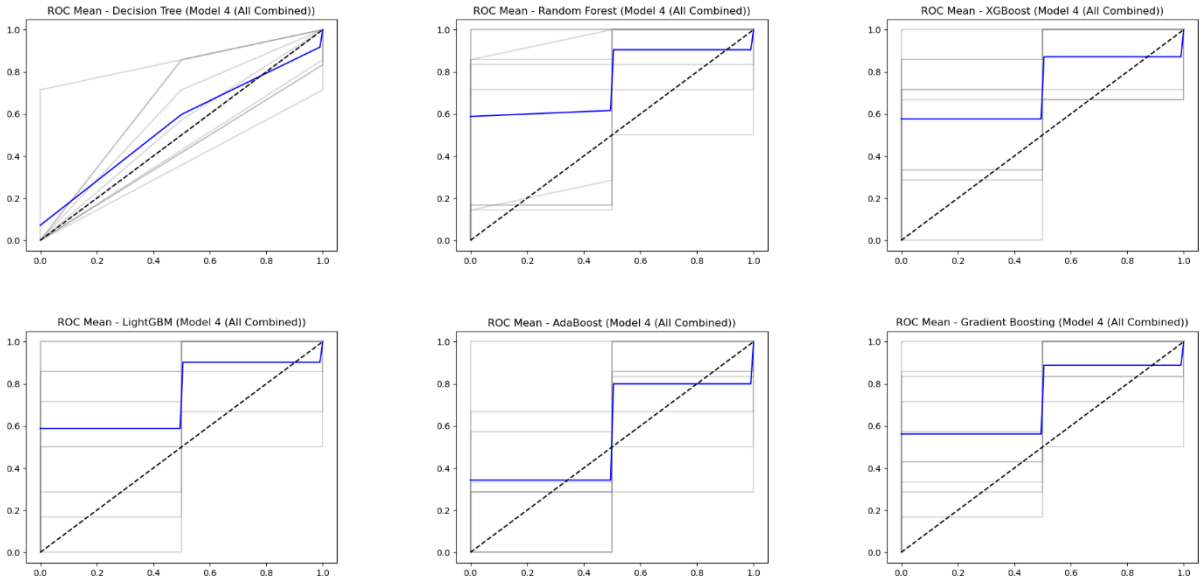


Figure 2. Calibration plots for Model 2 (basic and haemodynamic features) across seven classifiers. The dashed diagonal line represents perfect calibration. Blue lines indicate the mean calibration performance, with grey lines showing individual folds.

ROC curves for Model 4 are shown in Figure 3. Compared with basic or haemodynamic feature sets, combining all predictors resulted in improved discrimination across most classifiers. The Decision Tree model continued to show limited separation from the reference line. In contrast, ensemble methods such as Random Forest, XGBoost, and LightGBM achieved strong ROC profiles, with Random Forest showing the most favorable curve, consistent with its highest sensitivity (0.95 ± 0.09) and accuracy (0.82 ± 0.14). AdaBoost and Gradient Boosting also demonstrated good discrimination, while Naive Bayes achieved moderate performance. Overall, the inclusion of all feature categories enhanced model discrimination, with Random Forest and LightGBM yielding the most robust results.

Calibration plots for Model 4 are presented in Figure 4. The Decision Tree model remained poorly calibrated, with overestimation of predicted risk. Random Forest, XGBoost, and LightGBM displayed the most favorable calibration, with predicted risks tracking closer to the diagonal reference line, although minor underestimation occurred at intermediate probabilities. AdaBoost showed reasonable mid-range calibration but deviated at the extremes. Gradient Boosting and Naive Bayes demonstrated weaker calibration, with consistent misalignment between predicted and observed risks. Taken together, ensemble tree-based methods not only provided the strongest discrimination but also achieved the most reliable calibration. When compared with Model 2, Model 4 offered slightly improved discrimination. Nonetheless, calibration plots between the two models were broadly similar between the two models.



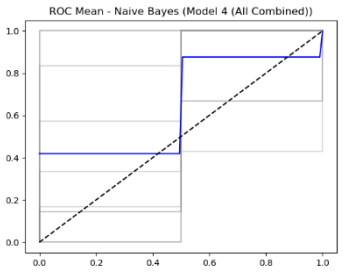


Figure 3. Receiver operating characteristic (ROC) curves for Model 4 (all features combined) across seven classifiers. Mean ROC curves are shown in blue, with the dashed line representing the reference line (AUC = 0.5).

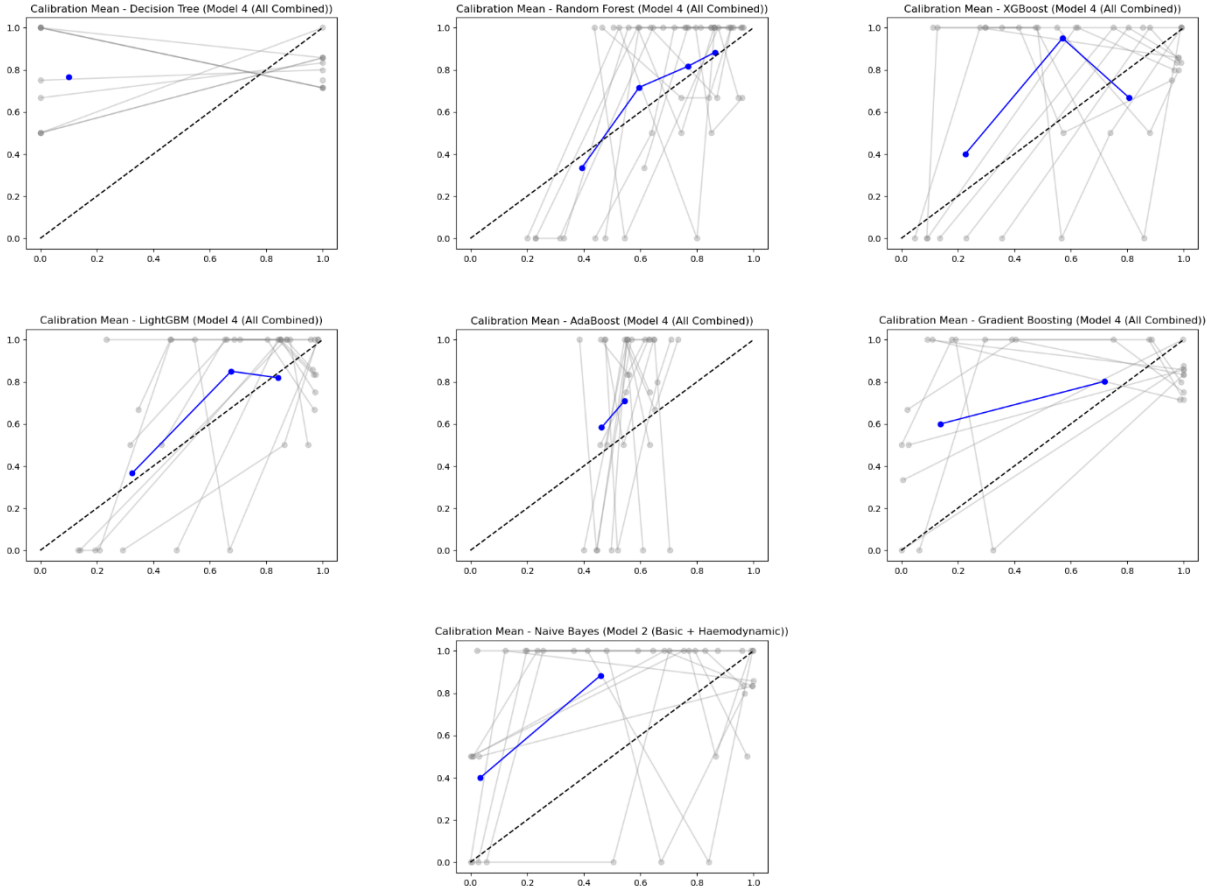


Figure 4. Calibration plots for Model 4 (all features combined) across seven classifiers. The dashed diagonal line represents perfect calibration. Blue lines indicate mean calibration performance, with grey lines showing individual folds.

2.3. Feature Importance and SHAP

Feature importance rankings for Model 2 and Model 4 are summarized in **Table 3**. Across all three classifiers, haemodynamic parameters such as AorticVTI, stroke volume, stroke volume index, and cardiac index consistently emerged among the most influential predictors (**Figures 5—7**). Age also contributed meaningfully in both models, particularly in Random Forest and XGBoost. When additional echocardiographic parameters were incorporated in Model 4, features such as LVOT at site 2, left myocardial performance index (LMPI), and pulmonary ejection time appeared in the top ranks for LightGBM and Random Forest. Nevertheless, the haemodynamic indices that dominated Model 2 remained important in Model 4, indicating stability of these predictors regardless of model complexity. These findings suggest that a compact

feature set centered on haemodynamic variables is sufficient to capture most of the predictive signal, reinforcing the practicality of Model 2 compared with the more complex Model 4.

Table 3. Top five features contributing to shock prediction in Model 2 and Model across Random Forest, LightGBM, and XGBoost classifiers.

Top-5 features in Model 2	Importance	Top-5 features in Model 4	Importance
Random Forest		Random Forest	
Aortic velocity–time integral	0.139	Aortic velocity–time integral	0.100
Stroke volume	0.125	Stroke volume	0.099
Stroke volume index	0.098	Stroke volume index	0.077
Cardiac index	0.083	Cardiac index	0.065
Age	0.062	LVOT at site 2	0.045
LightGBM		LightGBM	
LVOT	49.5	Stroke volume index	41.5
Stroke volume	46.8	Aortic velocity–time integral	36.8
Cardiac index	43.9	LMPI	35.9
Aortic velocity–time integral	31.5	Cadiac index	31.2
IVCmax	30.9	PulmonaryET	26.4
XGBoost		XGBoost	
Stroke volume	0.129	Body surface area	0.124
Body surface area	0.102	Stroke volume	0.119
Stroke volume index	0.089	Age	0.109
Aortic velocity–time integral	0.075	LVOT	0.073
Age	0.071	Stroke volume index	0.053

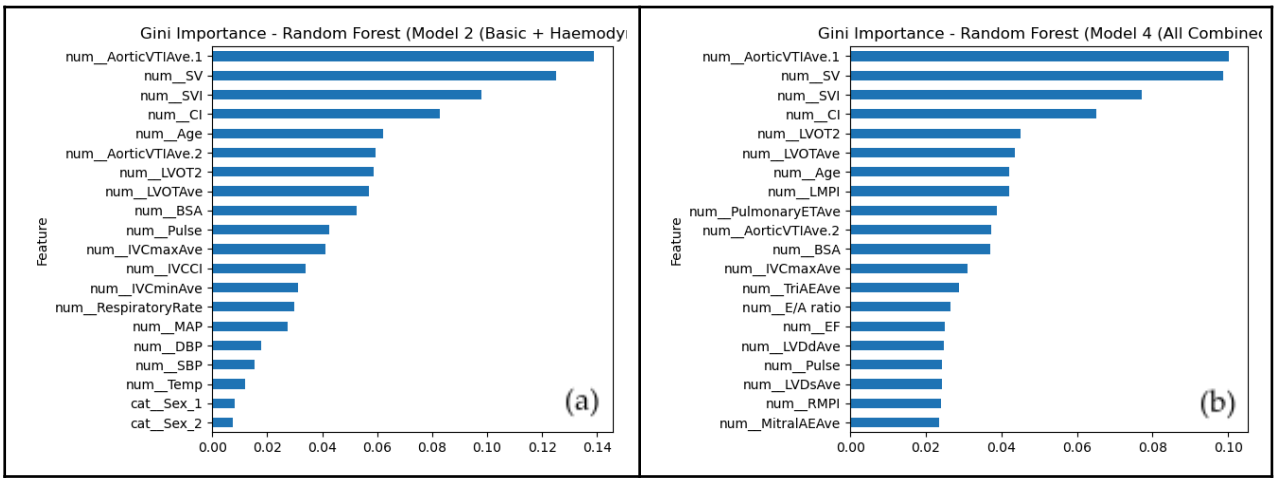


Figure 5. Gini importance rankings comparing Model 2 (a) and Model 4 (b) based on Random Forest classifier. Features are ordered from top to bottom by decreasing importance, with higher values indicating a greater influence on model predictions.

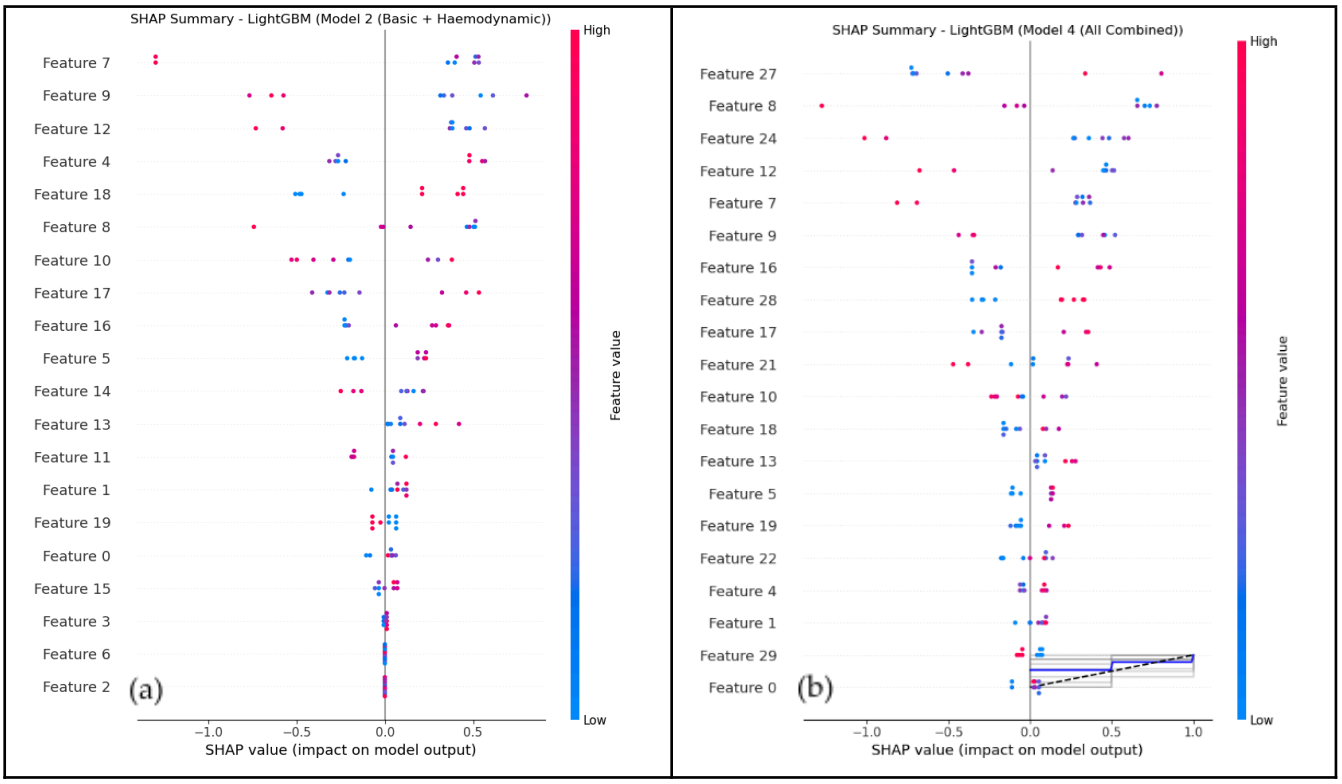


Figure 6. SHAP summary plots comparing Model 2 (a) versus Model 4 (b) based on LightGBM classifier. Each point represents an individual observation, with the horizontal position indicating the SHAP value. Feature values are color-coded from low (blue) to high (red). Features are ranked by their overall importance, defined as the mean absolute SHAP value across all observations.

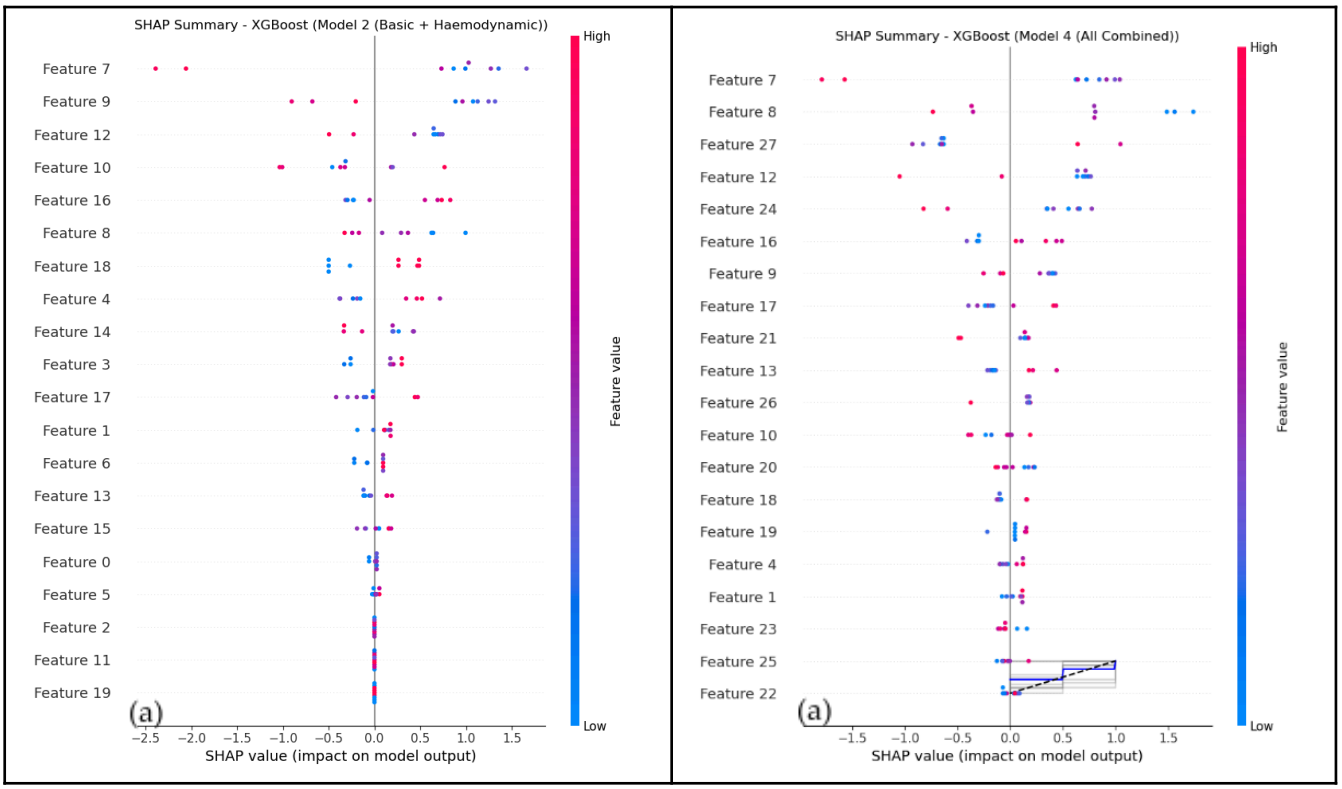


Figure 7. SHAP summary plots comparing Model 2 (a) versus Model 4 (b) based on XGBoost classifier. Each point represents an individual observation, with the horizontal position indicating the SHAP value. Feature values are color-coded from low (blue) to high (red). Features are ranked by their overall importance, defined as the mean absolute SHAP value across all observations.

3. Discussion

This secondary analysis evaluated the performance of ML algorithms in predicting shock among children with dengue shock syndrome using demographic, haemodynamic, and echocardiographic features. Ensemble-based classifiers, particularly Random Forest, LightGBM, and XGBoost, consistently provided the strongest discrimination and calibration. Importantly, a reduced feature set combining basic demographic and haemodynamic variables (Model 2) performed comparably to the more complex all-feature model (Model 4). The accurate risk stratification in the present study, therefore, can be achieved without reliance on extensive echocardiographic inputs. These suggest the feasibility of the generated parsimonious models for clinical deployment, as similarly reported previously [11-13].

Feature importance analyses provided mechanistic insights into the determinants of shock prediction. Parameters directly reflecting circulatory function, such as stroke volume, stroke volume index, cardiac index, and aortic velocity–time integral (AorticVTI), emerged as the most influential predictors across models. This observation is consistent with the established pathophysiology of dengue shock syndrome, in which increased vascular permeability leads to plasma leakage, reduced venous return, diminished preload, and subsequent decline in effective cardiac output [14, 15]. Stroke volume and cardiac index directly quantify forward flow, while stroke volume index accounts for body size–adjusted circulatory capacity, making it particularly informative in pediatric populations [4, 16-18]. AorticVTI, which represents the distance blood travels through the left ventricular outflow tract during systole, has been shown to correlate strongly with stroke volume and cardiac output and thus serves as a robust surrogate marker of systemic perfusion [14, 16].

Additional echocardiographic parameters, including left ventricular outflow tract measurements, myocardial performance indices (LMPI), and pulmonary ejection time, emerged among the top-ranked features in Model 4. These indices provide complementary insight into chamber geometry, systolic–diastolic coupling, global ventricular efficiency, and right ventricular function, all of which may be affected during severe dengue infection [17, 19]. Plasma leakage and systemic inflammation in dengue shock syndrome can impair preload and myocardial performance, leading to altered ventricular filling, prolonged ejection times, and transient myocardial dysfunction that may not be detected by blood pressure alone [6, 7]. The prominence of these parameters therefore supports a contributory role of myocardial involvement and altered ventricular timing in the haemodynamic phenotype of severe dengue, consistent with prior echocardiographic observations in both pediatric dengue and distributive shock states [6].

The comparable calibration observed between Model 2 and Model 4, in the present study, underscores the value of parsimonious models that rely on fewer, readily obtainable predictors. While echocardiographic parameters provide important mechanistic insight, their acquisition requires specialized equipment and operator expertise that may not be consistently available, particularly in low-resource or high-burden dengue settings [20, 21]. In contrast, Model 2 relies on fundamental haemodynamic measures that are more feasible for routine bedside use, enhancing its practicality for early risk stratification. This balance between physiological relevance and operational feasibility strengthens the translational potential of the model. These findings are consistent with prior evidence showing that machine learning approaches outperform traditional regression in capturing nonlinear interactions in critical illness, while also highlighting that simpler models are more likely to be implemented in real-world care pathways [8, 11, 22].

The strengths of this study include the use of a prospective, well-characterized pediatric cohort and rigorous internal validation using multiple performance and interpretability metrics. However, several limitations warrant consideration. The analysis was restricted to patients admitted to intensive care, which may limit generalizability to milder dengue cases or earlier stages of disease. In addition, the use of single-center data may constrain external validity. Future studies should focus on external validation in larger, multi-center cohorts and evaluate model performance in non–intensive care settings to determine broader applicability across the clinical spectrum of dengue infection.

4. Materials and Methods

4.1. Study Design

This study is a secondary analysis of a previously published prospective observational cohort conducted at the Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam [15]. The original study enrolled individuals aged above 3 years admitted to paediatric or adult intensive care units (ICUs) with a clinical diagnosis of dengue with warning signs or severe dengue, within 12 hours of ICU admission. Patients were reviewed daily until discharge or up to 5 days, with standardized documentation of clinical parameters, intravenous fluid therapy, and interventions. Portable echocardiograms were performed at enrolment and then daily until discharge, with additional follow-up 10–14 days later. Detailed methodology, including dengue diagnostics, laboratory protocols, and echocardiography acquisition, has been described previously [15]. For the present analysis, echocardiographic parameters measured at ICU admission or within the first 24 hours were used.

4.2 Ethical Considerations

Ethical approvals for the original study were obtained from the Oxford Tropical Research Ethics Committee and the Ethics Review Committee at HTD, with written informed consent secured from participants or their legal guardians [15]. The current analysis used anonymized, previously collected data and therefore did not require new ethical approval.

4.3 Data Preprocessing

The dataset included demographic information, vital signs, haemodynamic parameters, and structural/functional echocardiographic indices. The binary outcome was shock occurrence (shock vs. no shock). Mean arterial pressure (MAP) was estimated from systolic and diastolic blood pressure values rather than directly measured. Echocardiographic parameters were acquired at admission or within the first 24 hours and averaged across repeated measurements during the examination session, as described in the original study [7]. Categorical variables (e.g., Sex) were encoded as categorical data. Missing values were imputed using median values for numerical features and most frequent values for categorical features. Categorical predictors were one-hot encoded

4.4 Feature Sets

Four predictive models were constructed to assess the incremental contribution of different clinical and echocardiographic feature domains. Model 1 comprised basic demographic characteristics and vital signs, including age, sex, pulse rate, systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure (MAP), body temperature, and respiratory rate. Model 2 expanded Model 1 by incorporating hemodynamic and volumetric parameters derived from echocardiography, namely stroke volume (SV), stroke volume index (SVI), cardiac index (CI), left ventricular outflow tract average diameter (LVOT Ave), left ventricular outflow tract area ($LVOT^2$), average aortic velocity–time integral (Aortic VTI Ave.1 and Aortic VTI Ave.2), body surface area (BSA), average minimum inferior vena cava diameter (IVC min Ave), average maximum inferior vena cava diameter (IVC max Ave), and inferior vena cava collapsibility index (IVCCI). Model 3 extended Model 1 by adding structural and functional cardiac parameters, including average left ventricular diastolic diameter (LVDd Ave), average left ventricular systolic diameter (LVDs Ave), ejection fraction (EF), left myocardial performance index (LMPI), right myocardial performance index (RMPI), early-to-late ventricular filling ratio (E/A ratio), average pulmonary ejection time (Pulmonary ET Ave), average mitral A-wave acceleration (Mitral AEA Ave), average tricuspid A-wave acceleration (Tri AEA Ave), and average aortic ejection time (Aortic ET Ave). Model 4 represented the most comprehensive approach and included all variables from Models 1 through 3.

4.5 Model Training and Validation

Model performance was evaluated using 10-fold stratified cross-validation to ensure that the proportion of outcome classes was preserved across all folds. In each cross-validation iteration, models were trained on nine folds and evaluated on the remaining held-out fold, allowing robust estimation of generalization performance while minimizing sampling bias. A range of supervised machine-learning classifiers was implemented, including Decision Tree, Random Forest, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Adaptive Boosting (AdaBoost), Gradient Boosting, and Naive Bayes. For tree-based ensemble methods, key hyperparameters were predefined based on prior optimization and empirical performance, with the number of estimators set to 200, the learning rate ranging from 0.05 to 0.1, and the maximum tree depth fixed at 4 to balance model complexity and

overfitting risk. All analyses were conducted with fixed random seeds (seed = 123) to ensure reproducibility of model training and evaluation.

4.6 Statistical Parameters

Model performance was assessed using multiple complementary metrics, including accuracy, area under the receiver operating characteristic curve (AUC), sensitivity, specificity, F1 score, and root mean squared error (RMSE). To obtain stable and generalizable estimates, receiver operating characteristic (ROC) curves and calibration curves were averaged across all cross-validation folds. For interpretability, feature importance measures were extracted from tree-based classifiers to quantify the relative contribution of each predictor to model predictions. In addition, SHapley Additive exPlanations (SHAP) values were computed for selected models to provide a model-agnostic and consistent explanation of individual feature effects and their overall impact on prediction outcomes

5. Conclusions

Machine learning models showed significant discriminative ability in predicting shock in pediatric dengue, with Random Forest, LightGBM, and XGBoost performing best. Haemodynamic parameters, such as stroke volume, stroke volume index, cardiac index, and aortic VTI, were the strongest predictors. The model that only required demographic and haemodynamic features, achieved calibration comparable to the more complex model. These findings highlight the demographic and haemodynamic features-based model is clinically feasible for risk stratification in severe dengue. Nonetheless, external, multicenter validation is warranted in the future study before clinical application.

6. Patents

Acknowledgments: Authors appreciate the inter-institutional collaboration during the research and the making of this report.

Conflicts of Interest: The authors declare that they have no known conflicts of interest in relation to the publication of this work.

References

1. Zhang S-X, Yang G-B, Zhang R-J, *et al*. Global, regional, and national burden of dengue, 1990–2021: Findings from the global burden of disease study 2021. Decoding Infection and Transmission 2024; 2:100021. doi: 10.1016/j.dcit.2024.100021.
2. Yang X, Quam MBM, Zhang T, *et al*. Global burden for dengue and the evolving pattern in the past 30 years. Journal of Travel Medicine 2021; 28(8): taab146. doi: 10.1093/jtm/taab146.
3. Armenda S, Rusmawatiningtyas D, Makrufardi F, *et al*. Factors associated with clinical outcomes of pediatric dengue shock syndrome admitted to pediatric intensive care unit: A retrospective cohort study. Annals of Medicine and Surgery 2021; 66:102472. doi: 10.1016/j.amsu.2021.102472.
4. Nguyen TT, Le NT-H, Nguyen NM, *et al*. Clinical features and management of children with dengue-associated obstructive shock syndrome: A case report. Medicine 2022; 101(43):e31322. doi: 10.1097/MD.0000000000031322.
5. Singh RK, Tiwari A, Satone PD, *et al*. Updates in the management of dengue shock syndrome: a comprehensive review. Cureus 2023; 15(10):e46713. doi: 10.7759/cureus.46713.
6. Yacoub S, Trung TH, Lam PK, *et al*. Cardio-haemodynamic assessment and venous lactate in severe dengue: Relationship with recurrent shock and respiratory distress. PLOS Neglected Tropical Diseases 2017; 11(7):e0005740. doi: 10.1371/journal.pntd.0005740.
7. Thanachartwet V, Wattanathum A, Sahassananda D, *et al*. Dynamic Measurement of Hemodynamic Parameters and Cardiac Preload in Adults with Dengue: A Prospective Observational Study. PLOS ONE 2016; 11(5):e0156135. doi: 10.1371/journal.pone.0156135.
8. Vo LT, Vu T, Pham TN, *et al*. Machine learning-based models for prediction of in-hospital mortality in patients with dengue shock syndrome. World J Methodol 2025; 15(3):101837. doi: 10.5662/wjm.v15.i3.101837.
9. Zulkifli B, Fakri F, Odigie J, *et al*. Chemometric-empowered spectroscopic techniques in pharmaceutical fields: A bibliometric analysis and updated review. Narra X 2023; 1(1). doi: 10.52225/narrax.v1i1.80.
10. Idroes GM, Noviandy TR, Idroes GM, *et al*. Prognostication of differentiated thyroid cancer recurrence: An explainable machine learning approach. Narra X 2024; 2(3):e183-e183. doi: 10.52225/narrax.v2i3.183.
11. Nguyen RN, Lam HT, Phan HV, *et al*. Machine Learning Nomogram for Predicting Dengue Shock Syndrome in Pediatric Patients With Dengue Fever in Vietnam. Cureus 2025; 17(4):e81819. doi: 10.7759/cureus.81819..
12. Pollonini L, Padhye NS, Re R, *et al*. Pulse transit time measured by photoplethysmography improves the accuracy of heart rate as a surrogate measure of cardiac output, stroke volume and oxygen uptake in response to graded exercise. Physiol Meas 2015; 36(5):911-924. doi: 10.1088/0967-3334/36/5/911.
13. Thach TQ, Eisa HG, Hmeda AB, *et al*. Predictive markers for the early prognosis of dengue severity: A systematic review and meta-analysis. PLOS Neglected Tropical Diseases 2021; 15(10):e0009808. doi: 10.1371/journal.pntd.0009808.
14. Chanh HQ, Trieu HT, Tran Kim H, *et al*. Kinetics of cardiovascular and inflammatory biomarkers in paediatric dengue shock syndrome. Oxf Open Immunol 2024; 5(1):iqae005. doi: 10.1093/oxfimm/iqae005.
15. Teo A, Chua CLL, Chia PY, *et al*. Insights into potential causes of vascular hyperpermeability in dengue. PLoS Pathogens 2021; 17(12):e1010065. doi: 10.1371/journal.ppat.1010065.
16. Madanayake PMW, Jayawardena AEU, Wijekoon SL, *et al*. Fluid requirement in adult dengue haemorrhagic fever patients during the critical phase of the illness: an observational study. BMC Infectious Diseases 2021; 21(1):286. doi: 10.1186/s12879-021-05971-6.
17. Trieu HT, Khanh LP, Ming DKY, *et al*. The compensatory reserve index predicts recurrent shock in patients with severe dengue. BMC medicine 2022; 20(1):109. doi: 10.1186/s12916-022-02311-6.
18. Xu B, Thein T-L, Tay ZM, *et al*. The impact of blood pressure indicators and shock indices on hazard of complicated dengue in hospitalized adults with dengue infection. Journal of Clinical Virology 2025; 180:105855. doi: 10.1016/j.jcv.2025.105855.
19. Nachman D, Eisenkraft A, Rahamim E, *et al*. Assessing Cardiac Flow Measurements Using a Noninvasive Photoplethysmography-Based Device Compared to Invasive Pulmonary Artery Catheter. JACC Adv 2025; 4(9):102093. doi: 10.1016/j.jacadv.2025.102093.
20. Saxena A. Rheumatic heart disease screening by “point-of-care” echocardiography: an acceptable alternative in resource limited settings? Translational Pediatrics 2015; 4(3):210-213. 10.3978/j.issn.2224-4336.2015.06.01.
21. Qanitha A, Qalby N, Amir M, *et al*. Clinical cardiology in South East Asia: Indonesian lessons from the present towards improvement. Global Heart 2022; 17(1):66. doi: 10.5334/gh.1133.
22. Chaw JK, Chaw SH, Quah CH, *et al*. A predictive analytics model using machine learning algorithms to estimate the risk of shock development among dengue patients. Healthcare Analytics 2024; 5:100290. doi: 10.1016/j.health.2023.100290.