

A Novel Random Forest Algorithm-Based Prediction Model for in-Hospital Mortality Risk in Congolese Children with Severe Malaria

Ken Bisabu Kelu

`kelu.bisabu.ken@umontreal.ca`

University of Kinshasa

Vicky Mayifuila Niangi

University of Kinshasa

Gerrye Mubungu Lumbono

University of Kinshasa

Joseph Bodi Mabiala

University of Kinshasa

Celestin Nsibu Ndosimao

University of Kinshasa

Article

Keywords: Random Forest model, Severe malaria, in-hospital mortality, Risk prediction, Multiple organ dysfunction, Pediatric critical care

Posted Date: April 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9487430/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A Novel Random Forest Algorithm-Based Prediction Model for in-Hospital Mortality Risk in Congolese Children with Severe Malaria

Ken Bisabu Kelu, MD, M.Sc, PhD^{1,2*}, Vicky Mayifuila Niangi, MD^{1,2}, Gerrye Mubungu Lumbono, MD, PhD^{2,3}, Joseph Bodi Mabilia, MD, PhD^{1,2}, Celestin Nsibu Ndosimao, MD, PhD^{1,2*#}

¹ Pediatric Emergency and Intensive Care Unit, Department of Pediatrics, University Hospital of Kinshasa, Faculty of Medicine, University of Kinshasa, Kinshasa, Democratic Republic of the Congo

² Biomedical Sciences Department, Faculty of Medicine, University of Kinshasa, Kinshasa, Democratic Republic of the Congo

³ Reference Center for Rare and Undiagnosed Diseases, Genetics Center, Faculty of Medicine, University of Kinshasa, Kinshasa, Democratic Republic of the Congo

This author supervised the study

**Correspondence to: Ken Bisabu Kelu and Celestin Nsibu Ndosimao*

Pediatric Emergency and Intensive Care Unit, Department of Pediatrics, University Hospital of Kinshasa, University of Kinshasa, Kinshasa XIPO.BOX 123, Democratic Republic of Congo

Email : kelu.bisabu.ken@umontreal.ca and celensibu@gmail.com

Abstract

This study aimed to develop a machine learning-based model to predict in-hospital mortality among children with severe malaria (SM). Early identification of high-risk patients is critical to improving clinical outcomes.

A total of 52 clinical and laboratory variables were collected from children aged 2 to 9 years diagnosed with SM. Participants were randomly divided into training and testing datasets, maintaining a 1:5 death-to-survival ratio. Variable selection was performed sequentially using univariate analysis, the BORUTA algorithm, and LASSO regression to identify key predictors for the Random Forest (RF) model.

The observed in-hospital mortality rate was 17% (Table 1). The most important predictors of mortality included pH, respiratory rate, standard base excess (SBE), strong ion gap (SIG), albumin- and phosphate-corrected anion gap (AGCAP), glucose, potassium, creatinine, and the presence of cerebral malaria (Figure 2c-d). The final RF model demonstrated excellent predictive performance, with an area under the curve of 0.95 (Figure 4a-b-c) and an accuracy of 0.97 (95% CI: 0.83-0.99).

These findings suggest that acidosis, respiratory failure, hypoglycemia, hyperkalemia, renal impairment, and neurological dysfunction, are potential contributors to SM-associated mortality risk. Early recognition and strict management of SM-induced multi-organ dysfunction may significantly reduce pediatric SM-associated in-hospital mortality.

Keywords: Random Forest model, Severe malaria, in-hospital mortality, Risk prediction, Multiple organ dysfunction, Pediatric critical care.

Introduction

Severe malaria (SM), a potentially fatal tropical disease caused by Plasmodium parasites, remains a major global health challenge. Diagnosis is based on clinical presentation, supported by laboratory and radiologic findings, in accordance with World Health Organization (WHO) criteria. Children and pregnant women in malaria-endemic regions are at particularly high risk for severe illness and death(1). Among children, those aged 2 to 9 years are especially vulnerable, likely due to age-related differences in immune system development and response(2).

Microvascular disorders leading to multi-organ dysfunction (MOD), along with associated metabolic and electrolyte imbalances, are common in SM and represent key risk factors for disease severity and mortality. However, these complications may be underrecognized in pediatric patient(3). Given the substantial healthcare burden and societal impact of SM, clinical management would benefit from enhanced prevention strategies and improved risk prediction tools. Although several prognostic scoring models for pediatric SM have been developed, many were designed for use in resource-limited settings and often fail to incorporate critical prognostic factors, such as MOD syndrome and metabolic or electrolyte disturbances, that are strongly associated with increased mortality risk(4-7). Meaningful improvement in the prediction of pediatric SM severity through prognostic models will only be achieved if scientific advances are accompanied by critical evaluation of existing models, highlighting both their strengths and limitations, and by the development of more inclusive approaches. These future models should integrate clinical variables, laboratory findings, and medical imaging results that reflect SM-induced MOD.

Traditional logistic regression, while widely used in previous predictive models, is limited by its inability to effectively capture nonlinear relationships within data(3). In contrast, we chose a machine learning approach, specifically, the Random Forest (RF) algorithm, which can identify the most effective pathways to a predefined outcome without such constraints. This approach has been shown to outperform logistic regression in terms of predictive accuracy in various clinical cohorts(8, 9). RF-based prediction models provide clinicians with more precise risk assessments,

enabling tailored diagnosis and treatment strategies to better anticipate patient outcomes(10, 11). Accordingly, we developed a predictive model for in-hospital death risk in children presenting with SM using a RF machine learning algorithm by identifying key factors associated with SM severity and mortality.

Results

Study population characteristics

A total of 100 consecutive children diagnosed with SM were enrolled. Table 1 provides an overview of the study population's characteristics. 61% (61/100) of included children were males. The median age at admission was 4 (IQR 3-4) years and the median length of hospitalization was estimated at 3 (IQR 2-3) days. The in-hospital SM-associated mortality rate was 17% (17/100), higher in males and patients with a history of coma, convulsions, repeated febrile episodes, respiratory failure, and acidoketosis respiration ($P<0.001$ for all, Table 1). Both *Plasmodium falciparum*-induced SM and other *Plasmodium* species-induced SM were the responsible for the deaths ($P<0.001$ for all, Table 1). Increased SM-associated death risks were observed in children with CM, and SRM, as well as in those with renal, hemoglobinological, and hepatic dysfunctions ($P<0.001$ for all, Table 1). Additionally, hypercapnia, hypoxemia, high AG metabolic acidosis, decreased plasma bicarbonate levels, increased plasma AG, AGCAP, AGCAPL and SIG, hyperlactatemia, negative SBE and SBD CAP values, hypoglycemia, and hyperkalemia, were all significantly associated with SM-associated death risk ($P<0.001$ for all, Table 1).

Feature Selection

Twenty-nine features (of 52) have a significant association with the SM-associated death risk ($P\leq 0.1$ for all, Figure 2c-d). The BORUTA algorithm based feature selection ranked 16 of these 29 variables as being important (Figure 2b and Supplementary Material). Finally, all 16 variables identified as important by the BORUTA algorithm were included as inputs in the LASSO feature selection analysis. In addition, variables deemed unimportant by BORUTA were also considered and introduced gradually using a forward stepwise approach. Of these, nine variables showed a non-zero importance score and were retained as candidate predictors for the final RF model. The LASSO coefficient profiles of the selected features, along with the optimal penalization parameter (λ), are displayed in Figure 2a and the supplementary material.

Random Forest Algorithm-Based Prediction Model on the Training Sample

A model for predicting SM-associated death risk was constructed using the candidate predictors identified through LASSO, including pH, respiratory rate, SBE, SIG, AGCAP, Glucose, Potassium, Creatinine and CM. A 70:30 random split of the data sample (n = 100 patients) into training and test data resulted in a consistent death-to-alive ratio of 1/5 (0.2) in both subsets. The SMOTE algorithm was applied to address class imbalance in the training data, resulting in an approximately balanced death-to-survival ratio of 0.94 (~1:1) (Figure 1).

The final RF model was trained using 800 trees (ntree = 800), with 4 features randomly selected at each iteration (mtry=4), and a maximum of 24 terminal nodes per tree (maxnodes = 24). The model achieved an accuracy of 0.94 and a Kappa statistic of 0.78. Figure 3a-b show the distributions of SM-associated death risk predictors, grouped by patient outcome (deceased vs. survivors).

Validation of Severe Malaria-Associated Death Risk Prediction Model on the Test Sample

Our RF classifier yielded the highest AUC score 0.95 (Figures 4a-b-c). The final model demonstrated strong discriminatory power for predicting in-hospital SM-associated death risk in SM children, achieving an accuracy of 0.97 [95% CI: 0.83–0.99] on the testing subset. This performance was associated with a low OOB error rate of 3.33%, a Kappa statistic of 0.87, and a statistically significant *p*-value of 0.03. Furthermore, the model achieved an F1_score of 0.98, a precision of 0.96 and a recall of 0.99. Based on the variable important analysis of the final RF model, pH, respiratory rate and SBE were identified as the most important predictors. A decreasing trend in both pH and SBE, and respiratory failure was associated with the highest risk of SM-associated death (Figure 2c-d).

Individualized Prediction Interpretation on the Test Sample

The generated LIME plot and heatmap display individual predictions for each patient (Figure 5a), highlighting the main features contributing to the RF model's predictions. The length of the color bars represents the magnitude of each feature's contribution (weight), while the color indicates

the direction (red for negative, blue for positive) of the estimated effect. In Figure 5a, SM cases 1 and 3 (left) were predicted to have a high SM-associated death risk, with predicted probabilities of 0.55 (case 1) and 0.82 (case 3), and explanation fits of 0.52 for case 1 and 0.58 for case 3. In contrast, SM cases 2 and 4 (right) were classified as having no SM-associated death risk, each with a predicted probability of 1.00 and explanation fits of 0.56 (case 2) and 0.54 (case 4), respectively.

The constructed predictive nomogram provides individualized estimates of in-hospital SM-associated death risk, based on a validated logistic regression model derived from the RF outputs (Figure 5b). The Hosmer–Lemeshow goodness-of-fit test indicated that the logistic regression model derived from the original RF model fit the observed data well (p -value = 1). The calibration plot demonstrates good agreement between the predicted probabilities and the observed proportions of SM-associated death risk. (Figure 4d). Nomogram analysis indicates that a child with SM presenting with respiratory failure, metabolic acidosis, hypoglycemia, elevated potassium and creatinine levels, and CM is classified by the RF model as having a high risk of SM-associated death. This profile yields a total score of approximately 325, corresponding to a predicted probability greater than 0.9 (~100%). In contrast, a child without dyspnea, metabolic or electrolyte disturbances, or CM has a very low predicted risk, with a total score of around 110, corresponding to a probability of less than 0.1 (<10%).

Discussion

In this study, we present a novel RF algorithm-based prediction model for estimating in-hospital death risk among children with SM. We identified key predictors of SM severity, ranked in descending order of importance: pH, respiratory rate, SBE, SIG, AGCAP, glucose, potassium, creatinine, and CM. These variables were strongly associated with increased likelihood of in-hospital mortality.

Our study offers several strengths. First, to the best of our knowledge, this is the first application of a machine learning algorithm to predict in-hospital pediatric deaths due to SM. Second, we developed a robust risk prediction model incorporating both rigorous internal validation techniques and clinically relevant variables reflective of SM-associated MOD, including neurological, respiratory, renal, metabolic, and electrolyte disturbances.

Despite the relatively small sample size, the RF model demonstrated excellent performance, achieving an AUC of 0.95 and an accuracy of 0.97, with a lower bound of 0.83, highlighting its strong discriminatory power and the clinical utility of its predictor variables. These findings support the potential of such models to aid in personalized treatment planning for children presenting with SM.

Our results align with previous research and suggest future directions, including the development of SM-associated death risk prediction models using artificial intelligence methods and the inclusion of variables that reflect SM-induced MOD(3, 12, 13).

Our findings demonstrate that children with SM often die from metabolic and electrolyte disturbances, such as metabolic acidosis, hypoglycemia, and hyperkalemia, which result from SM-induced MOD. Previous studies have similarly reported elevated mortality rates in children diagnosed with SM-induced MOD(14, 15). These findings support the notion that multiple organs may be simultaneously affected due to microvascular obstruction caused by sequestration of parasitized red blood cells. Consistent with our observations, post-mortem histopathological

studies in children with SM have confirmed the presence of sequestration-induced microvascular obstruction in several organs, including the brain, lungs, and kidneys(16-21).

CM, recognized as one of the most severe and feared complications of malaria and a major predictor of SM-associated mortality, was present in 76.5% (13 out of 17) of the children who died (Table 1). The primary pathophysiological mechanism appears to be microvascular obstruction caused by sequestration of parasitized red blood cells in the cerebral capillaries. This process leads to a cascade of complications, including cerebral edema, elevated intracranial pressure, brain hypoxia with increased oxygen demand, coma, and ultimately, brain dysfunction and death(22, 23).

SRM, such as abnormal, deep, and labored breathing, as well as hypercapnia and hypoxemia caused by ventilation/perfusion (V/Q) mismatch, were documented in 58.8% (10 out of 17) of the children who died (Table 1). The underlying pathophysiological mechanisms include metabolic acidosis, acute respiratory distress syndrome (ARDS), and pulmonary edema, conditions known to contribute to severe respiratory complications in SM, which affect approximately 40% of children(24, 25). These disorders account for more than half of the deaths in affected patients, highlighting their strong predictive value for SM-associated mortality(19, 25, 26).

Our data indicate that children with SM who died exhibited signs of impaired renal function, as evidenced by elevated blood urea nitrogen (median: 68 [IQR 59-78.4] mg/dL, Table 1) and creatinine levels (median: 1.2 [IQR 0.9-1.4] mg/dL, Table 1). Renal dysfunction, often progressing to acute kidney injury, is a well-established risk factor for SM-associated mortality. This condition is likely driven by multiple contributing factors, including parasite sequestration in renal vasculature, hemodynamic instability, endothelial activation and damage, systemic inflammation, and hemolysis(27, 28). Our findings are consistent with previous studies that highlight the detrimental impact of impaired renal function on outcomes in children with SM(29, 30).

Hypoglycemia was observed in 94% of children with SM who died, with a mean blood glucose level of 47.65 ± 14.54 mg/dL (Table 1). These findings align with previous studies reporting a high incidence of hypoglycemia in SM and its significant contribution to SM-associated mortality. The underlying physiological mechanisms include excessive glucose consumption by malaria parasites within red blood cells, impaired gluconeogenesis, anemia, and quinine-induced hyperinsulinemia(31-34). Both hypoglycemia and sequestration-induced microvascular obstruction of parasitized red blood cells in cerebral capillaries contribute to neurological dysfunction, thereby increasing the risk of death in children with SM.

According to our data, all children who died from SM exhibited high anion AG metabolic acidosis. The mean corrected AG (AGCAP) was 14.94 ± 2.92 mmol/L (Table 1), and all deceased patients (100%, 17/17) had acidosis, with a median pH of 7.18 [IQR: 7.05-7.23] (Table 1). These patients also showed markedly low plasma bicarbonate levels (median: 8.8 [IQR: 6.7-9.6] mmol/L), elevated SIG values (median: 16.36 [IQR: 15.98-17.5] mmol/L), negative SBE (median: -1.59 [IQR: -1.80 to -1.28] mmol/L), and negative SBDCAP values (mean: -1.34 ± 0.38 mmol/L) (Table 1). Additionally, hyperkalemia was present, with a median potassium level of 4.8 [IQR: 4.7-5.1] mmol/L (Table 1). These findings reflect the accumulation of unmeasured strong anions, other than lactate, as well as anions from weak acids, indicating a complex acidosis profile. The presence of hyperbilirubinemia and impaired renal function in these children suggests a role for both hepatic and renal dysfunction in the pathogenesis of acidosis. Moreover, hyperkalemia likely resulted from the combined effects of metabolic acidosis and renal impairment. Our findings are consistent with previous studies reporting that metabolic acidosis occurs in approximately 50% of children with SM and is a major predictor of SM-associated mortality, typically arising from MOD(35-38).

Our model integrates both clinical criteria and laboratory parameters that are consistent with SM-induced MOD. With ongoing improvements in diagnostic infrastructure, resource-limited settings are increasingly gaining access to essential laboratory tests and patient monitoring. This RF-based prediction model is specifically designed for SM and aims to provide rapid, accurate estimation of

in-hospital death risk. The selected predictors demonstrate strong prognostic value for SM-associated mortality and should be considered by clinicians when managing children presenting with SM.

Limitations

A major limitation of this study was the small sample size. Further research is needed to validate our model in a larger, independent population of children with SM before it can be recommended for widespread clinical use.

Conclusions

Meaningful improvement in the prediction of pediatric SM severity through prognostic models will only be achieved if scientific advances are accompanied by critical evaluation of existing models, highlighting both their strengths and limitations, and by the development of more inclusive approaches. These future models should integrate clinical variables, laboratory findings, and medical imaging results that reflect SM-induced MOD.

In this retrospective study of prospectively collected data, breathing difficulty, metabolic acidosis, hypoglycemia, elevated plasma potassium and creatinine levels, and CM were strongly associated with in-hospital mortality among children with SM during hospitalization. Our RF model may assist clinicians in identifying the most critically ill patients within the first 24 hours of admission and support pediatric intensive care unit managers in anticipating the resources needed when caring for children with SM. However, external validation in larger, independent cohorts is necessary before this model can be implemented for routine bedside use.

Methods

Ethics Statement

The study was conducted in accordance with the principles of the Declaration of Helsinki and internationally recognized scientific standards(39, 40). Ethical approval was obtained from the local Ethics Committee of the Centre de Formation et d'Appui Sanitaire (CEFA)-Monkole (reference number: 004/CEFA-MONKOLE/CEL/2016), covering both the initial and subsequent phases of the study. This approval extended to all participating healthcare facilities due to their established partnership. Written informed consent was obtained from the legal guardians of all enrolled children prior to data collection. All collected data were fully anonymized before analysis to ensure confidentiality and privacy. The study was registered on ClinicalTrials.gov under the identifier NCT07355426.

Study Design, Population and Endpoint

This retrospective longitudinal study was based on existing data from 100 children aged 2 to 9 years who were admitted with SM between January 30 and July 31, 2017. The study was conducted at two healthcare facilities in Kinshasa, Democratic Republic of the Congo : the Monkole Mother and Child Hospital and the Kimbondo Pediatric Center, both selected through purposive sampling.

These hospitals are located in peri-urban areas of Kinshasa and serve large populations residing in regions with intense malaria transmission. In these settings, the burden of malaria remains particularly high, especially among children aged 2 to 9 years, who are at increased risk of infection and severe disease.

All children aged 2 to 9 years admitted to the emergency department with a diagnosis of SM, defined according to the WHO criteria (1, 41), during the study period, were consecutively enrolled. Children with confirmed non-malarial infections, concomitant comorbidities, or any condition likely to produce similar clinical or biological dysfunctions, were excluded to minimize potential confounding.

The primary outcome was in-hospital mortality, defined as any death occurring during the same hospitalization for SM, regardless of the duration between admission and death.

Data Collection

For each eligible patient, demographic, clinical, and laboratory data were collected prospectively during hospitalization, from admission until discharge or death. Blood samples, obtained through both arterial and venous puncture, were used to measure arterial blood gases, electrolytes, albumin, bilirubin, urea, and creatinine. Malaria diagnosis was confirmed using rapid diagnostic tests and light microscopy for *Plasmodium* species.

Parameters related to metabolic disturbances were calculated, including the anion gap (AG), albumin- and phosphate-corrected anion gap (AGCAP), standard base excess (SBE), albumin- and phosphate-corrected standard base deficit (SBD CAP), and strong ion gap (SIG). These were derived using the modified Van Slyke equation, with adjustments for plasma albumin and phosphate concentrations based on correction factors proposed by Wooten(42, 43), and according to the Stewart approach as modified by Figge(44, 45).

Patients were further categorized into key clinical syndromes based on WHO-defined criteria, integrating clinical and laboratory findings(1, 41, 46) (Supplementary Material). All children received standard treatment for SM in accordance with WHO guidelines, including intravenous artesunate, antipyretic therapy, and targeted management of organ dysfunction as well as metabolic and electrolyte imbalances(41).

Selection of Candidate Variables for Predictive Modeling

The predictive model was developed using potential SM associated-severity and mortality risk factors, including demographic, clinical, and laboratory variables, collected at admission or during hospitalization. Categorical variables were encoded as integers ranging from 0 to the maximum number of categories, as detailed in Table 1. Continuous variables were standardized by centering and scaling to ensure comparability across features.

The most important candidate predictors were selected through a sequential approach involving univariable logistic regression, the BORUTA algorithm, and Least Absolute Shrinkage and Selection Operator (LASSO) feature selection(47, 48). First, variables with a univariable logistic regression P -value <0.1 were identified and included in the BORUTA algorithm, which ranked features from most to least important. Variables deemed significant by BORUTA were then subjected to LASSO regression, and those with non-zero coefficients were retained as candidate predictors for the final RF model.

Random Forest Modeling

We first randomly split the data sample into training and testing data according to a division of 70:30, ensuring that the ratio of deceased and survivors is the same in all subsets. The Synthetic Minority Over-sampling Technique (SMOTE) was used to deal with unbalanced data and boost the the training data size(49).

A basic RF model using default parameters, including a grid search with 10-fold cross-validation, was first defined on the training set. The previous model was tested and evaluated with values of m try ranging from 1-10, and different $maxnodes$ values to identify the best values for both parameters that best reflected the RF model's accuracy. Finally, the selected optimal values for m try and $maxnode$ were used to tune the number of trees. The definitive RF model that was the most accurate was trained by using the best values for all these parameters.

Model Testing and Performance Evaluation

The confusion matrix was computed on the testing data using predictions from the final RF model trained on the training data. The RF model performance metrics, including $F1_score$, Accuracy, Precision, Recall, Kappa statistic, P -value and the Area Under the Curve (AUC) for each individual predictor and the overall RF model, were calculated to estimate the discriminatory power of the model to identify the SM-associated death risk in children presenting with SM.

The importance of each predictor was assessed by permuting the out-of-bag (OOB) data for that predictor while keeping all others unchanged, and measuring the resulting increase in prediction error. Predictors were then ranked based on the mean decrease in accuracy and Gini coefficient. The relative importance of each predictor in the RF model was assessed using global variable importance measures. Predictors with high importance values were considered to have a significant impact on the model's outcome.

Finally, the Local Interpretable Model-agnostic Explanations (LIME) approach was used to provide interpretable, patient-level explanations of how each predictor in the final RF model contributed to individual SM-associated death risk predictions(50). We lastly fitted a logistic regression analysis with the combined predicted probabilities from the original RF model, and the outcome was used to construct a predictive nomogram for SM-associated death risk(51). Model calibration was assessed by the Hosmer-Lomeshow goodness-of-fit test and by plotting the predicted versus actual deciles of predicted risk.

The study was conducted in accordance with the principles outlined in the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) recommendations(52), and the data flow diagram was created with [Biorender.com](https://biorender.com), as shown in Figure 1.

Statistical Analysis

Categorical variables were summarized as counts (N) and percentages (%), and comparisons were performed using the chi-square test or Fisher's exact test, as appropriate. The distribution of continuous variables was assessed using the Shapiro-Wilk test. Accordingly, normally distributed variables were expressed as mean $\pm$ standard deviation (SD), while non-normally distributed variables were presented as median with interquartile range (IQR). Group comparisons were conducted using Student's t-test or Welch's t-test for normally distributed data, and the Mann-Whitney U test (Wilcoxon rank-sum test) for non-normally distributed data, as appropriate.

A two-sided p-value ≤ 0.05 was considered statistically significant. All statistical analyses were performed using RStudio software (version 2025.05.1+513; released June 5, 2025; Public Benefit Corporation, Université de Montréal).

Acknowledgements

The authors are very grateful to the MONKOLE hospital management. Its involvement, support and insights have been crucial for this study. We would also like to thank all clinical and laboratory staff from Monkole Mother and Child Hospital Center, and Kimbondo Pediatric Center for their collaboration during their routine clinical care and surveillance activities.

Author contributions

KBK, VMN, JBM and CNN contributed to the study concept and design. VMN was responsible for data collection. KBK performed statistical analysis, data analysis and interpretation. KBK and CNN drafted the initial version of the manuscript. CNN and JBM supervised the study. All authors critically revised the manuscript for important intellectual content and approved the final version of the manuscript.

Funding

The authors have not received a specific grant for this research from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests

The authors have no conflicts of interest to disclose.

Data availability

The code and datasets used and/or analyzed in the current study are not publicly available but can be obtained from the corresponding author upon reasonable request by email.

Ethics approval and Consent to participate

The study was conducted in accordance with the principles of the Declaration of Helsinki and internationally recognized scientific standards(39, 40). Ethical approval was obtained from the local Ethics Committee of the Centre de Formation et d'Appui Sanitaire (CEFA)-Monkole (reference number: 004/CEFA-MONKOLE/CEL/2016), covering both the initial and subsequent phases of the study. This approval extended to all participating healthcare facilities due to their

established partnership. Written informed consent was obtained from the legal guardians of all enrolled children prior to data collection. All collected data were fully anonymized before analysis to ensure confidentiality and privacy. The study was registered on ClinicalTrials.gov under the identifier NCT07355426.

Consent for Publication

This manuscript has been reviewed and approved by all authors, who have agreed to its submission for publication.

Supplementary Information

Supplementary Material.

References

1. Daily JP, Parikh S. Malaria. *N Engl J Med*. 2025;392(13):1320-33.
2. Ranjha R, Singh K, Baharia RK, Mohan M, Anvikar AR, Bharti PK. Age-specific malaria vulnerability and transmission reservoir among children. *Glob Pediatr*. 2023;6:None.
3. Njim T, Tanyitiku BS. Prognostic models for the clinical management of malaria and its complications: a systematic review. *BMJ Open*. 2019;9(11):e030793.
4. Helbok R, Kendjo E, Issifou S, Lackner P, Newton CR, Kombila M, et al. The Lambarene Organ Dysfunction Score (LODS) is a simple clinical predictor of fatal malaria in African children. *J Infect Dis*. 2009;200(12):1834-41.
5. Kumar N, Thomas N, Singhal D, Puliye JM, Sreenivas V. Triage score for severity of illness. *Indian Pediatr*. 2003;40(3):204-10.
6. Gupta MA, Chakrabarty A, Halstead R, Sahni M, Rangasami J, Puliye A, et al. Validation of "Signs of Inflammation in Children that Kill" (SICK) score for immediate non-invasive assessment of severity of illness. *Ital J Pediatr*. 2010;36:35.
7. Berkley JA, Ross A, Mwangi I, Osier FH, Mohammed M, Shebbe M, et al. Prognostic indicators of early and late death in children admitted to district hospital in Kenya: cohort study. *BMJ*. 2003;326(7385):361.
8. Johnson AE, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD. Machine Learning and Decision Support in Critical Care. *Proc IEEE Inst Electr Electron Eng*. 2016;104(2):444-66.
9. Zampieri FG, Costa EL, Iwashyna TJ, Carvalho CRR, Damiani LP, Taniguchi LU, et al. Heterogeneous effects of alveolar recruitment in acute respiratory distress syndrome: a machine learning reanalysis of the Alveolar Recruitment for Acute Respiratory Distress Syndrome Trial. *Br J Anaesth*. 2019;123(1):88-95.
10. Li J, Tian Y, Zhu Y, Zhou T, Li J, Ding K, et al. A multicenter random forest model for effective prognosis prediction in collaborative clinical research network. *Artif Intell Med*. 2020;103:101814.
11. Pirneskoski J, Tamminen J, Kallonen A, Nurmi J, Kuisma M, Olkkola KT, et al. Random forest machine learning method outperforms prehospital National Early Warning Score for predicting one-day mortality: A retrospective study. *Resusc Plus*. 2020;4:100046.
12. Sypniewska P, Duda JF, Locatelli I, Althaus CR, Althaus F, Genton B. Clinical and laboratory predictors of death in African children with features of severe malaria: a systematic review and meta-analysis. *BMC Med*. 2017;15(1):147.
13. Agnandji ST, Recker M, Mordmuller B, Glockner S, Adegnik AA, Lell B, et al. Prostration and the prognosis of death in African children with severe malaria. *Int J Infect Dis*. 2023;134:240-7.
14. Namazzi R, Opoka R, Datta D, Bangirana P, Batte A, Berrens Z, et al. Acute Kidney Injury Interacts With Coma, Acidosis, and Impaired Perfusion to Significantly Increase Risk of Death in Children With Severe Malaria. *Clin Infect Dis*. 2022;75(9):1511-9.
15. Cunningham AJ, Walther M, Riley EM. Piecing together the puzzle of severe malaria. *Sci Transl Med*. 2013;5(211):211ps18.
16. Milner DA, Jr. Malaria Pathogenesis. *Cold Spring Harb Perspect Med*. 2018;8(1).
17. Newton CR, Hien TT, White N. Cerebral malaria. *J Neurol Neurosurg Psychiatry*. 2000;69(4):433-41.

18. Taylor TE, Fu WJ, Carr RA, Whitten RO, Mueller JS, Fosiko NG, et al. Differentiating the pathologies of cerebral malaria by postmortem parasite counts. *Nat Med*. 2004;10(2):143-5.
19. Marsh K, Forster D, Waruiru C, Mwangi I, Winstanley M, Marsh V, et al. Indicators of life-threatening malaria in African children. *N Engl J Med*. 1995;332(21):1399-404.
20. Milner DA, Jr., Lee JJ, Frantzreb C, Whitten RO, Kamiza S, Carr RA, et al. Quantitative Assessment of Multiorgan Sequestration of Parasites in Fatal Pediatric Cerebral Malaria. *J Infect Dis*. 2015;212(8):1317-21.
21. Wynkoop HJ, Bevan A, Galeano X, Raees M, Haque MR, Taylor T, et al. Multiple Organ Dysfunction Syndrome in Malawian Children with Cerebral Malaria. *Am J Trop Med Hyg*. 2024;111(6):1223-9.
22. Seydel KB, Kampondeni SD, Valim C, Potchen MJ, Milner DA, Muwalo FW, et al. Brain swelling and death in children with cerebral malaria. *N Engl J Med*. 2015;372(12):1126-37.
23. Birbeck GL, Molyneux ME, Kaplan PW, Seydel KB, Chimalizeni YF, Kawaza K, et al. Blantyre Malaria Project Epilepsy Study (BMPEs) of neurological outcomes in retinopathy-positive paediatric cerebral malaria survivors: a prospective cohort study. *Lancet Neurol*. 2010;9(12):1173-81.
24. Van den Steen PE, Deroost K, Deckers J, Van Herck E, Struyf S, Opdenakker G. Pathogenesis of malaria-associated acute respiratory distress syndrome. *Trends Parasitol*. 2013;29(7):346-58.
25. Taylor WRJ, Hanson J, Turner GDH, White NJ, Dondorp AM. Respiratory manifestations of malaria. *Chest*. 2012;142(2):492-505.
26. Mitran C, Opoka RO, Conroy AL, Namasopo S, Kain KC, Hawkes MT. Pediatric Malaria with Respiratory Distress: Prognostic Significance of Point-of-Care Lactate. *Microorganisms*. 2023;11(4).
27. Silva GBDJ, Pinto JR, Barros EJJ, Farias GMN, Daher EF. Kidney involvement in malaria: an update. *Rev Inst Med Trop Sao Paulo*. 2017;59:e53.
28. Katsoulis O, Georgiadou A, Cunningham AJ. Immunopathology of Acute Kidney Injury in Severe Malaria. *Front Immunol*. 2021;12:651739.
29. von Seidlein L, Olaosebikan R, Hendriksen IC, Lee SJ, Adedoyin OT, Agbenyega T, et al. Predicting the clinical outcome of severe falciparum malaria in african children: findings from a large randomized trial. *Clin Infect Dis*. 2012;54(8):1080-90.
30. Conroy AL, Opoka RO, Bangirana P, Idro R, Ssenkusu JM, Datta D, et al. Acute kidney injury is associated with impaired cognition and chronic kidney disease in a prospective cohort of children with severe malaria. *BMC Med*. 2019;17(1):98.
31. Krishna S, Waller DW, ter Kuile F, Kwiatkowski D, Crawley J, Craddock CF, et al. Lactic acidosis and hypoglycaemia in children with severe malaria: pathophysiological and prognostic significance. *Trans R Soc Trop Med Hyg*. 1994;88(1):67-73.
32. Taylor TE, Molyneux ME, Wirima JJ, Fletcher KA, Morris K. Blood glucose levels in Malawian children before and during the administration of intravenous quinine for severe falciparum malaria. *N Engl J Med*. 1988;319(16):1040-7.
33. Osier FH, Berkley JA, Ross A, Sanderson F, Mohammed S, Newton CR. Abnormal blood glucose concentrations on admission to a rural Kenyan district hospital: prevalence and outcome. *Arch Dis Child*. 2003;88(7):621-5.
34. Ogetii GN, Akech S, Jemutai J, Boga M, Kivaya E, Fegan G, et al. Hypoglycaemia in severe malaria, clinical associations and relationship to quinine dosage. *BMC Infect Dis*. 2010;10:334.

35. Leopold SJ, Ghose A, Allman EL, Kingston HWF, Hossain A, Dutta AK, et al. Identifying the Components of Acidosis in Patients With Severe Plasmodium falciparum Malaria Using Metabolomics. *J Infect Dis.* 2019;219(11):1766-76.
36. Dondorp AM, Chau TT, Phu NH, Mai NT, Loc PP, Chuong LV, et al. Unidentified acids of strong prognostic significance in severe malaria. *Crit Care Med.* 2004;32(8):1683-8.
37. Newton CR, Valim C, Krishna S, Wypij D, Olola C, Agbenyega T, et al. The prognostic value of measures of acid/base balance in pediatric falciparum malaria, compared with other clinical and laboratory parameters. *Clin Infect Dis.* 2005;41(7):948-57.
38. Mzumara G, Leopold S, Marsh K, Dondorp A, Ohuma EO, Mukaka M. Identifying prognostic factors of severe metabolic acidosis and uraemia in African children with severe falciparum malaria: a secondary analysis of a randomized trial. *Malar J.* 2021;20(1):282.
39. World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310(20):2191-4.
40. International Ethical Guidelines for Health-related Research Involving Humans: Prepared by the Council for International Organizations of Medical Sciences (CIOMS) in collaboration with the World Health Organization (WHO). 4th ed. Geneva 2016.
41. WHO guidelines for malaria. 2024 November 30.
42. Siggaard-Andersen O. Acid-Base Balance *ENCYCLOPEDIA OF RESPIRATORY MEDICINE/Acid-Base Balance* 2005. p. 173-208.
43. Wooten EW. Calculation of physiological acid-base parameters in multicompartiment systems with application to human blood. *J Appl Physiol* (1985). 2003;95(6):2333-44.
44. Stewart PA. Modern quantitative acid-base chemistry. *Can J Physiol Pharmacol.* 1983;61(12):1444-61.
45. Figge J, Mydosh T, Fencel V. Serum proteins and acid-base equilibria: a follow-up. *J Lab Clin Med.* 1992;120(5):713-9.
46. Postels DG, Katangwe-Chirwa T. Cerebral malaria. *Semin Pediatr Neurol.* 2025;54:101206.
47. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological).* 2018;58(1):267-88.
48. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform.* 2019;20(2):492-503.
49. Wang S, Dai Y, Shen J, Xuan J. Research on expansion and classification of imbalanced data based on SMOTE algorithm. *Sci Rep.* 2021;11(1):24039.
50. Hassan SU, Abdulkadir SJ, Zahid MSM, Al-Selwi SM. Local interpretable model-agnostic explanation approach for medical imaging analysis: A systematic literature review. *Comput Biol Med.* 2025;185:109569.
51. Zhang Y, Zhang Z, Wei L, Wei S. Construction and validation of nomograms combined with novel machine learning algorithms to predict early death of patients with metastatic colorectal cancer. *Front Public Health.* 2022;10:1008137.
52. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. *Br J Surg.* 2015;102(3):148-58.

FIGURES

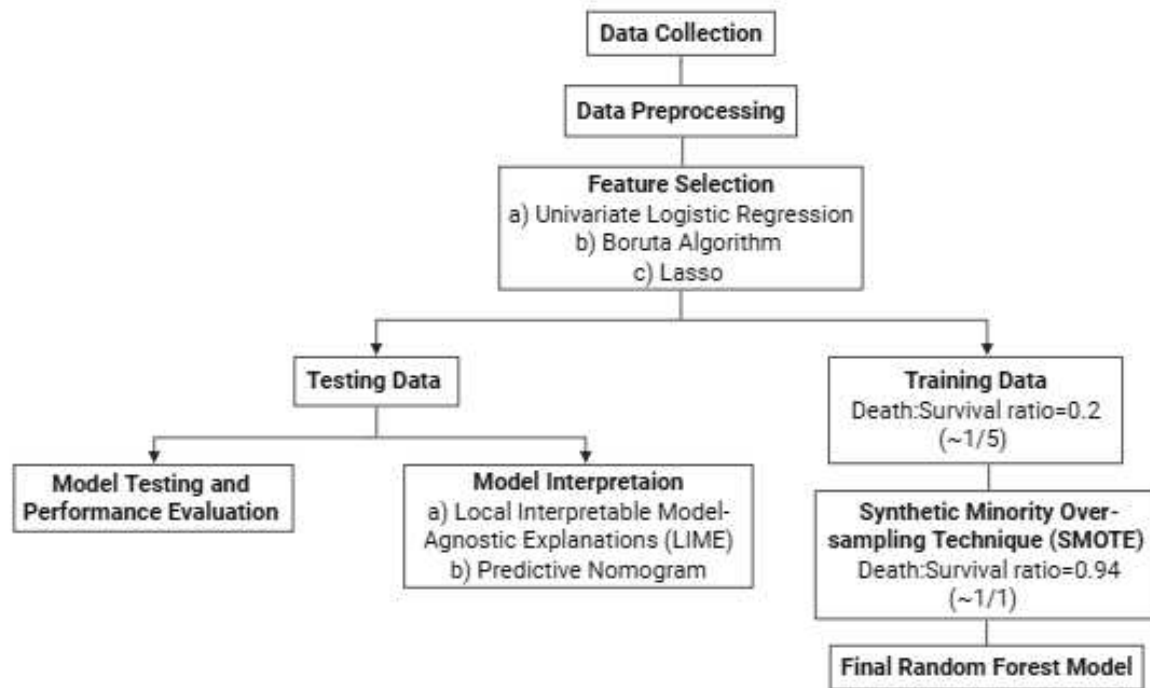


Figure 1. Random Forest Modeling Process for Pediatric Mortality Risk Stratification

Created with [Biorender.com](https://biorender.com).

Abbreviations: LASSO, Least absolute shrinkage and selection operator.

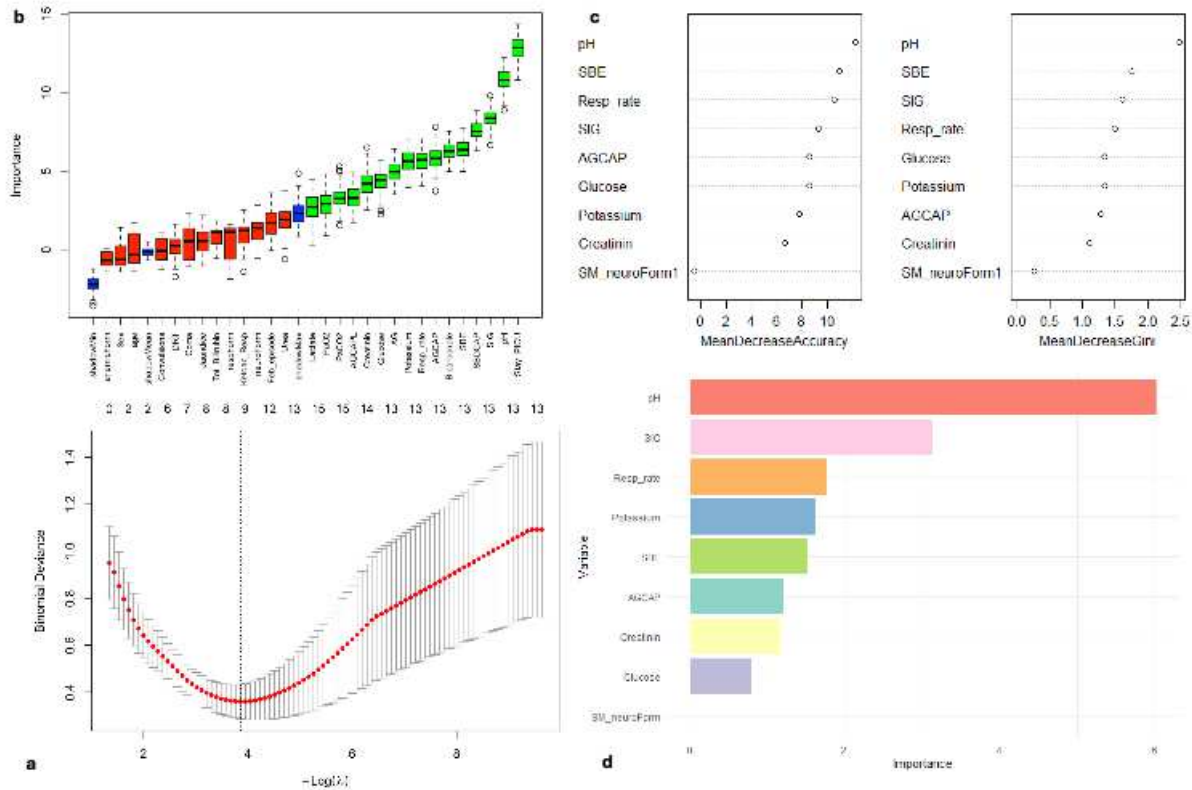


Figure 2. Feature Importance Ranking.

(a) LASSO coefficient profiles of the selected features, along with the optimal penalization parameter, lambda (λ). (b) Boruta algorithm-based feature selection. Boxplots in blue, red and green, are shadowMin, rejected as unimportant variables and confirmed as important variables, respectively. (c) Candidate predictors for the Random Forest model ranked by mean decrease in accuracy (left) and mean decrease in Gini coefficient (right). (d) Ranking of Predictor Variables Based on Random Forest Model for SM-associated Death Risk.

Abbreviations: AG, Anion Gap; AGCAP, albumin- and phosphate-corrected anion gap; AGCAPL, albumin-, phosphate- and lactate-corrected anion gap; Creatinin, Creatinine; Feb_episode, Febrile episode; Ketoac_Res, Ketoacidosis respiration; neuroForm, Cerebral malaria; PICU, Pediatric Intensive Care Unit; RespForm, Severe Respiratory Malaria; RDT, Rapid Diagnostic Test; Resp_rate, Breathing rate; Tot_Bilirubin, Total bilirubin; SBE, Standard Base Excess; SBD CAP, albumin- and phosphate-corrected standard base deficit; SIG, Strong Ion Gap.

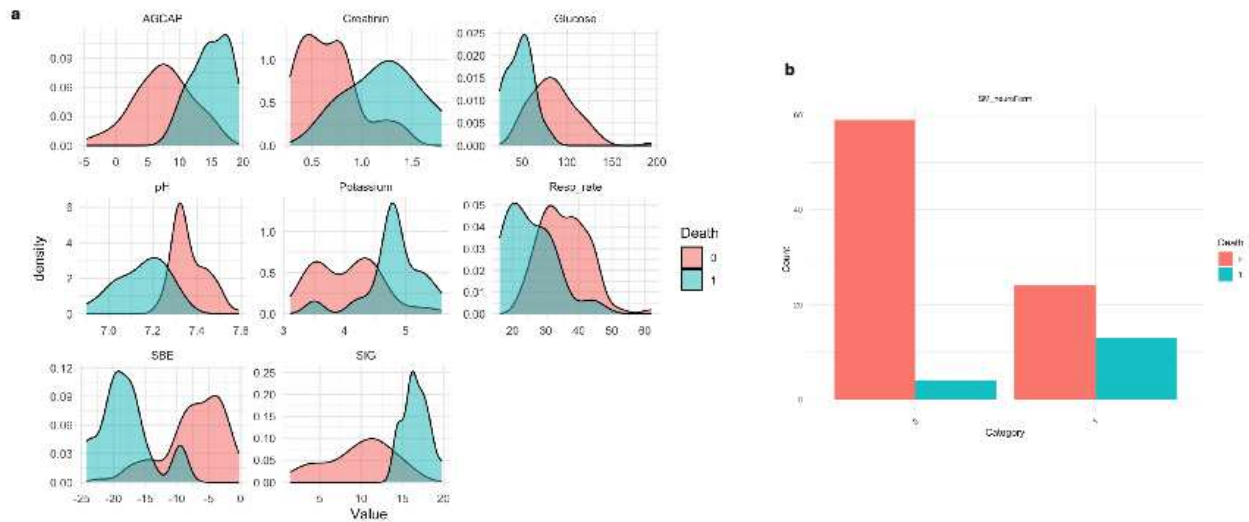


Figure 3. Severe Malaria-Associated Death Risk Predictors, grouped by patient outcome (deceased vs. survivors). (a) Density distribution of continuous predictors. (b) Frequency distribution of the categorical predictor, cerebral malaria.

Abbreviations: AGCAP, albumin- and phosphate-corrected anion gap; Creatinin, Creatinine; pH, Potential of Hydrogen; Resp_rate, Breathing rate; SBE, Standard Base Excess; SIG, Strong Ion Gap; SM_neuroform, Cerebral Malaria.

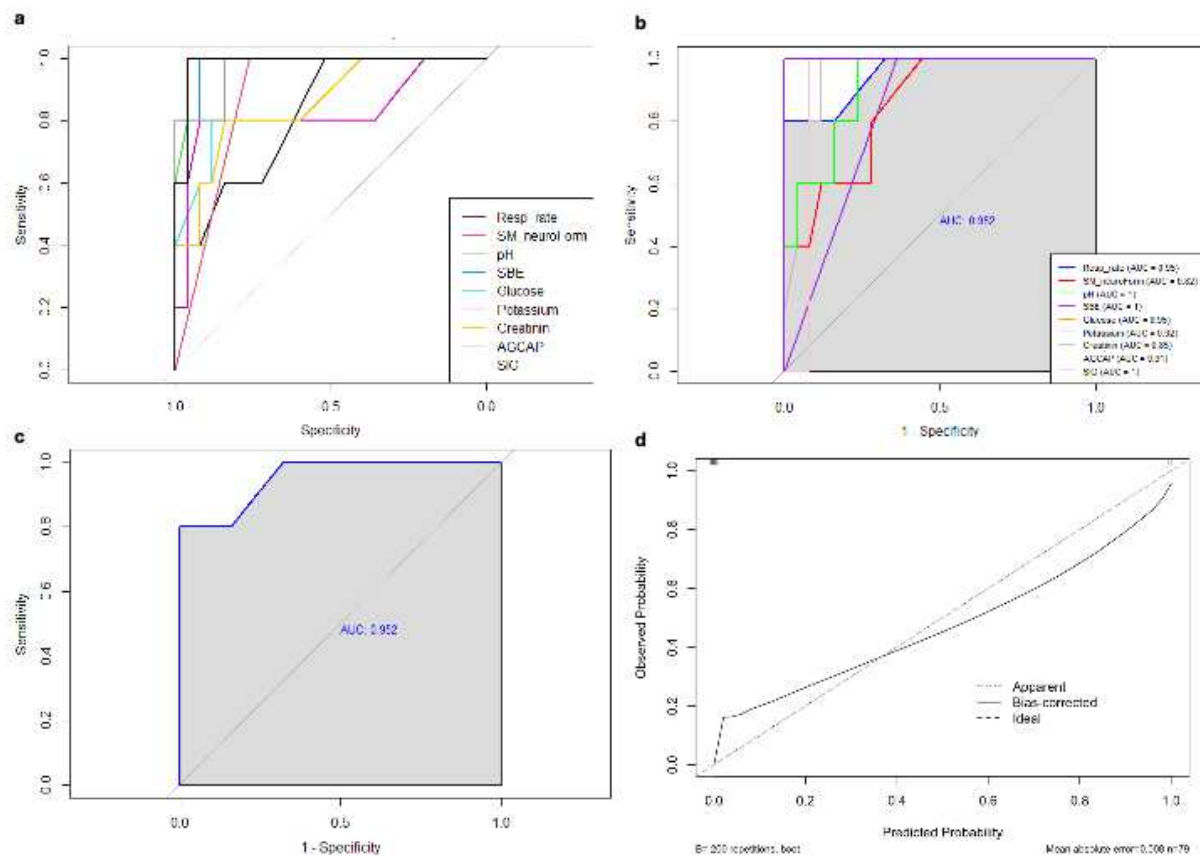


Figure 4. ROC Curves and Calibration Plot of the Random Forest Model.

(a, b) Individualized ROC Curves Illustrating Discriminatory Power of Each Predictor. (c) Overall ROC curve. (d) Calibration Plot Showing Agreement Between Predicted and Observed SM-Associated Death Risk.

Abbreviations: AGCAP, albumin- and phosphate-corrected anion gap; Creatinin, Creatinine; pH, Potential of Hydrogen; Resp_rate, Breathing rate; ROC, Receiver Operating Characteristic; SBE, Standard Base Excess; SIG, Strong Ion Gap; SM_neuroform, Cerebral Malaria.

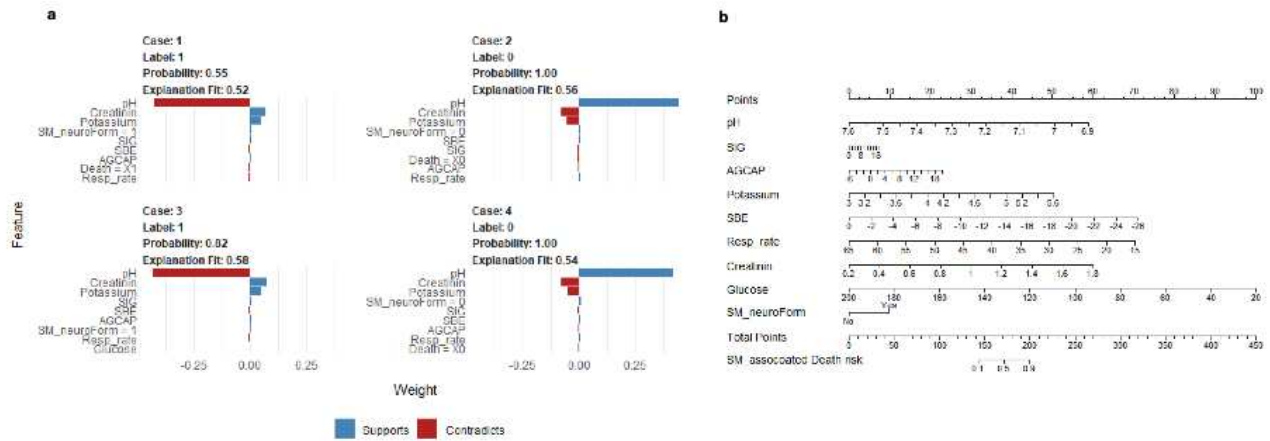


Figure 5. Individualized Risk Interpretation for SM-Associated Death. (a) LIME Results Illustrating Feature Contributions to Individual Death Risk Predictions. This plot illustrates the main contributing features behind the RF model’s predictions for the first four test cases. For each case (observation), all nine predictor variables are displayed. The length of each color bar indicates the magnitude of that feature’s contribution (weight) to the predicted death risk, while the color denotes the direction of the effect [positive (blue) or negative (red) influence on the risk]. (b) Predictive Nomogram Predicting the likelihood of SM-Associated Death. To obtain the nomogram-predicted probability, patient values on each axis must be located and a vertical line must be drawn upward to the point axis (Points scale) to determine how many points are attributed for each predictor value. Points for all predictors must be summed and accessed on the point line (Total Points scale) to find SM-associated death probability (SM_Death risk). SIG, AGCAP, SBE and Potassium are expressed in mmol/L; Creatinine and Glucose in mg/dL, and Breathing rate in cycles per minute.

Abbreviations: AGCAP, albumin- and phosphate-corrected anion gap; Creatinin, Creatinine; LIME, Local Interpretable Model-Agnostic Explanations; pH, Potential of hydrogen; Resp_rate, Breathing rate; SBE, Standard Base Excess; SIG, Strong Ion Gap; SM_neuroform, Cerebral Malaria.

Table 1. Study Population Characteristics

Characteristics	Overall (N=100)	Survivors (n=83)	Deceased (n=17)	P-value
Demographic Characteristics				
Age, years	4 [3-4] [†]	4 [3-4] [†]	4 [3-4] [†]	0.442
Male sex, No (%)	61 (61)	50 (60)	11 (65)	0.943
Clinical Examination				
Febrile episode	4 [3-4] [†]	3 [3-4] [†]	4 [4-5] [†]	0.0002
Temperature, °C	38.25 [37.78-39] [†]	38.20 [37.8-39] [†]	38.30 [37.7-39.5] [†]	0.577
Coma, No (%)	40 (40)	28 (33.73)	12 (70.6)	0.011
Convulsions, No (%)	37 (37)	24 (28.91)	13 (76.47)	0.0006
Pallor, No (%)	85 (85)	69 (83.13)	16 (94.12)	0.433
Jaundice, No (%)	22 (22)	15 (18.07)	7 (41.18)	0.076
Heart rate, per minute	121.28 (20.31)*	121.08 (16.94)*	122.24 (32.88)*	0.889
Resp_rate, per minute	32 [28-40] [†]	36 [30.5-40] [†]	24 [20-31] [†]	2.28e ⁻⁰⁵
Ketoacidosis Resp, No (%)	31 (31)	18 (21.69)	13 (76.47)	3.16e ⁻⁰⁵
Hepatomegaly, No (%)	39 (39)	35 (42.17)	4 (23.53)	0.245
Splenomegaly, No (%)	68 (68)	57 (68.67)	11 (64.71)	0.973
Length of stay in the PICU	3 [2-3] [†]	3 [2-4] [†]	1 [1-1] [†]	<2.2e ⁻¹⁶
Final Diagnoses				
CM, No (%)	37 (37)	24 (28.92)	13 (76.47)	0.0006
SMA, No (%)	50 (50)	46 (55.42)	4 (23.53)	0.033
SRM, No (%)	29 (29)	19 (22.89)	10 (58.82)	0.007
Medications				
Artesunate, No (%)	90 (90)	75 (90.36)	15 (88.24)	1
Quinine, No (%)	10 (10)	8 (9.64)	2 (11.76)	1
Laboratory Assessments				
P _a CO ₂ , mm Hg	27.74 (5.97)*	29.18 (5.26)*	20.68 (3.87)*	1.4e ⁻⁰⁸
P _a O ₂ , mm Hg	78 [64.6-102.78] [†]	79.2 [69.90-104.50] [†]	58.7 [48-65] [†]	6.67e ⁻⁰⁵
pH	7.32 [7.28-7.4] [†]	7.34 [7.32-7.42] [†]	7.18 [7.05-7.23] [†]	4.89e ⁻⁰⁷
SBE ^a , mmol/L	0.19 [-0.79 - 0.83] [†]	0.43 [-0.05 - 0.86] [†]	-1.59 [-1.8 to -1.28] [†]	8.36e ⁻¹⁰
Bicarbonate, mmol/L	15.75 [10.8-19.15] [†]	17 [12.5-20.25] [†]	8.8 [6.7-9.6] [†]	<2.2e ⁻¹⁶
L-Lactate, mmol/L	5.27 (2.19)*	4.78 (2.04)*	7.68 (0.97)*	6.25e ⁻¹²
Glucose, mg/dL	78.82 (28.57)*	85.2 (26.48)*	47.65 (14.54)*	3.13e ⁻¹⁰
Potassium, mmol/L	4.2 [3.6-4.6] [†]	4.1 [3.5-4.4] [†]	4.8 [4.7-5.1] [†]	2.74e ⁻⁰⁶
Sodium, mmol/L	136 [135-138] [†]	136 [135-138] [†]	136 [135-137] [†]	0.429
Calcium, mmol/L	1.18 [1.1-1.28] [†]	1.17 [1.1-1.27] [†]	1.21 [1.04-1.3] [†]	0.922
Magnesium, mmol/L	0.98 [0.89-1.12] [†]	0.98 [0.89-1.12] [†]	0.99 [0.89-1.12] [†]	0.766
Chlore, mmol/L	100 [98.38-102] [†]	100 [98.45-102] [†]	98.6 [98-101] [†]	0.098
Phosphore, mmol/L	1.51 [1.29-1.86] [†]	1.52 [1.28-1.86] [†]	1.47 [1.32-1.87] [†]	0.629
Blood urea nitrogen, mg/dL	53.5 [40.5-65.3] [†]	45.5 [35.8-62.5] [†]	68 [59-78.4] [†]	0.0005
Creatinine, mg/dL	0.7 [0.5-0.9] [†]	0.6 [0.4-0.8] [†]	1.2 [0.9-1.4] [†]	8.68e ⁻⁰⁶
Albumin, g/L	35.36 [34-37.2] [†]	35.4 [34-37.3] [†]	34.7 [34.4-36.5] [†]	0.609
Phosphate, mg/dL	9.99 [8.51-12.21] [†]	10.09 [8.55-12.35] [†]	9.61 [8.41-11.77] [†]	0.907
Hematocrit, %	16 [13-22.25] [†]	15 [13-22] [†]	21 [16-23] [†]	0.628
Total Bilirubin, mg/dL	2.8 [1.78-3.93] [†]	2.4 [1.51-3.6] [†]	4.4 [3-5.6] [†]	0.002
Parasitemia ^b , parasites/μL	54.43 (2.15) ^b	52.35 (2.22) ^b	65.87 (1.75) ^b	0.629

Table 1 (Continued)

RDT_PfHRP ₂ , No (%)	72 (72)	64 (77.11)	8 (47.06)	0.027
RDT_PAN, No (%)	28 (28)	19 (22.89)	9 (52.94)	0.03
AG, mmol/L	20.96 (4.99)*	19.67 (4.37)*	27.25 (2.27)*	1.83e ⁻¹³
AGCAP, mmol/L	8.65 (5.24)*	7.36 (4.65)*	14.94 (2.92)*	2.9e ⁻¹⁰
AGCAPL, mmol/L	3.38 (4.1)*	2.58 (4.65)*	7.26 (2.55)*	4.75e ⁻⁰⁷
SID _a , mmol/L	36.74 (2.73)*	37.05 (2.76)*	35.22 (2.02)*	0.003
SID _e , mmol/L	25.84 (5.39)*	27.31 (4.63)*	18.66 (1.97)*	<2.2e ⁻¹⁶
SIG, mmol/L	11.55 [8.06-14.48] [†]	10.22 [7.13-12.46] [†]	16.36 [15.98-17.5] [†]	<2.2e ⁻¹⁶
SBDCAP ^a , mmol/L	-7e ⁻⁵ (1)*	0.27 (0.85)*	-1.34 (0.38)*	<2.2e ⁻¹⁶

Continuous variables were summarized according to their distribution: normally distributed variables are presented as mean (standard deviation) and indicated by an asterisk (*), whereas non-normally distributed variables are reported as median [interquartile range] and indicated by a dagger (†). Continuous variables marked with the superscript letter *a*, were centered and scaled using Z-score standardization. Parasitemia, denoted by the superscript letter *b*, is expressed as the geometric mean and geometric standard deviation, along with the 95% geometric confidence interval. Categorical variables are presented as counts (n) with corresponding percentages (%).

Abbreviations: AG, Anion Gap; AGCAP, albumin- and phosphate-corrected anion gap; AGCAPL, albumin-, phosphate- and lactate-corrected anion gap; CM, Cerebral Malaria; Ketoacidosis_Resp, ketoacidosis respiration; PaO₂, Partial pressure of oxygen in arterial blood; PaCO₂, Partial pressure of carbon dioxide in arterial blood; PAN, Pan-plasmodium; PfHRP₂, Plasmodium falciparum histidine-rich protein 2; pH, Potential of hydrogen; PICU, Pediatric Intensive Care Unit; RDT, Rapid Diagnostic Test; Resp_rate, respiratory rate; SBE, Standard Base Excess; SBDCAP, albumin- and phosphate-corrected standard base deficit; SID_a, apparent Strong Ion Difference; SID_e, efferent Strong Ion Difference; SIG, strong Ion Gap; SMA, Severe Malarial Anemia; SRM, Severe Respiratory Malaria.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialRFModelFinalSciRep.pdf](#)