Supplementary Table 1

Table S1: Rank of variable importance by reduction in residual sum of squares (RSS) using random forest regression.

Bacterial Etiology

Variable Name RSS Reduction

Age 73.9

Blood in stool 33.9 Breastfed 29.1

HAZ 25.9 MUAC 24.3

Resp. Rate 20.6 Season 19.9

Wealth Index 19.7

Temperature 18.9

Vomiting 18.0

Supplementary Table 2

Any Bacteria

Variable Name

Beta Estimate (95%CI)

P-value

|  |  |  |
| --- | --- | --- |
| Intercept | -1.9330 (-5.3748 - 1.5089) | 0.2710 |
| Age (mo.) | 0.0567 (0.0455 - 0.0679) | <0.0001 |
| Blood in stool | 2.0062 (1.7533 - 2.2591) | <0.0001 |
| Breastfed |  |  |
| None | Reference |  |
| Partially | -0.7296 (-0.9592 - -0.4999) | <0.0001 |
| Exclusively | -0.6245 (-1.0411 - -0.2078) | 0.0033 |
| HAZ | -0.1036 (-0.1768 - -0.0305) | 0.0055 |
| MUAC | -0.0353 (-0.1029 - 0.0322) | 0.3050 |
| Resp. Rate (per min.) | 0.0131 (0.0028 - 0.0234)  Reference | 0.0124 |
| Season |
| Dry/Cold |
| Rainy/Cold | 0.9156 (0.6344 - 1.1969) | <0.0001 |
| Dry/Hot | 0.9725 (0.7595 - 1.1855) | <0.0001 |
| Rainy/Hot | 0.8389 (0.6164 - 1.0614) | <0.0001 |
| Wealth Index | -0.0211 (-0.1053 - 0.0631) | 0.6230 |
| Temperature (◦C) | 0.0299 (-0.0621 - 0.1218) | 0.5241 |
| Vomiting | -0.6766 (-0.846 - -0.5071) | <0.0001 |

Table S2: The estimate, 95% confidence interval, and p-value from a logistic regression model.

Supplementary Table 3

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Africa | |  |  | Asia |  |
| Country  Variable | The Gambia | Mali | Mozambique | Kenya | India | Bangladesh | Pakistan |
| 1 | Age | Age | Age | Age | Age | Age | Age |
| 2 | Blood in stool | MUAC | Breastfed | Breastfed | MUAC | Blood in stool | MUAC |
| 3 | HAZ | Breastfed | HAZ | HAZ | HAZ | Eyes | HAZ |
| 4 | Breastfed | HAZ | MUAC | MUAC | Season | Vomiting | Breastfed |
| 5 | MUAC | Resp. Rate | Season | Resp. Rate | Resp. Rate | Season | Resp. Rate |
| 6 | Temp. | Wealth Index | Resp. Rate | Temp. | Blood in stool | MUAC | Wealth Index |
| 7 | Resp. Rate | Temp. | Temp. | Wealth Index | Wealth Index | Rectal Straining | Temp. |
| 8 | Season | Ppl. of House | Wealth Index | Ppl. of House | # Share Fac. | Temp. | Ppl. of House |
| 9 | Wealth Index | Vomiting | Vomiting | # Share Fac. | Temp. | HAZ | Blood in stool |
| 10 | Ppl. of House | Season | Blood in stool | Water Source | Ppl. of House | Wealth Index | Days of Episode |
| Country AUCs | **0.827** (0.818-0.836)  0.840 (0.832-0.849) | **0.741** (0.731-0.752)  0.756 (0.746-0.767) | **0.802** (0.791-0.813)  0.831 (0.821-0.841) | **0.763** (0.752-0.773)  0.773 (0.763-0.784) | **0.775** (0.767-0.784)  0.807 (0.799-0.815) | **0.892** (0.885-0.898)  0.921 (0.915-0.927) | **0.760** (0.750-0.770)  0.765 (0.755-0.775) |
| Continent AUCs |  | **0.779 (0.774-0.785)**  0.797 (0.792-0.802) | |  |  | **0.826 (0.821-0.830)**  0.850 (0.845-0.854) |  |
| Overall AUC |  |  | | **0.806 (0.803-0.809)**  0.829 (0.825-0.832) |  |  |  |
| Continent External Validation | **0.746** (0.696-0.797)  0.767 (0.719-0.815) | **0.711** (0.657-0.765)  0.790 (0.736-0.842) | **0.773** (0.722-0.823)  0.831 (0.787-0.876) | **0.723** (0.672-0.773)  0.715 (0.664-0.766) | **0.772** (0.735-0.810)  0.811 (0.777-0.846) | **0.897** (0.868-0.926)  0.930 (0.906-0.953) | **0.765** (0.719-0.811)  0.782 (0.738-0.826) |

Table S3: The table contains both site-specific variable importance ordering and a cross-validated average overall AUC, AUC by country, and AUC by continent and confidence intervals from a 5 (bold) and 10 (ital.) variable logistic regression model for predicting a viral etiology with variables based on the overall variable importance. Lastly, it shows the AUC and a 95% confidence interval resulting from testing the logistic regression with variables based on the overall variable importance on each site individually following its training on the other countries in the same continent.

9