Supplementary Material

Smartphone conjunctiva photography for malaria risk stratification in asymptomatic school age children

Shaun G. Hong[1], Sang Mok Park[1], Semin Kwon[1], Haripriya Sakthivel[1], Sreeram Nagappa[1], Jung Woo Leem[1], Steven R. Steinhubl[1,2], Pascal Ngiruwonsanga[3], Jean-Louis Ndikumana Mangara[4], Célestin Twizere[5], and Young L. Kim[1,2,6,7*]

[1]Weldon School of Biomedical Engineering, Purdue University, West Lafayette, Indiana, USA
[2]Regenstrief Center for Healthcare Engineering, Purdue University, West Lafayette, Indiana, USA
[3]Gakoma District Hospital, Gisagara District, Southern Province, Rwanda
[4]Malaria, Neglected Tropical Diseases and Other Parasitic Diseases Division, Rwanda Biomedical Center, Kigali, Rwanda
[5]Center of Excellence in Biomedical Engineering and eHealth, University of Rwanda, Kigali, Rwanda
[6]Purdue Institute for Cancer Research, Purdue University, West Lafayette, Indiana, USA
[7]Purdue Quantum Science and Engineering Institute, Purdue University, West Lafayette, Indiana, USA
***youngkim@purdue.edu**

Supplementary Table 1. Multiple logistic regression analysis of associations among malaria RDT, selected radiomic features, and demographic information

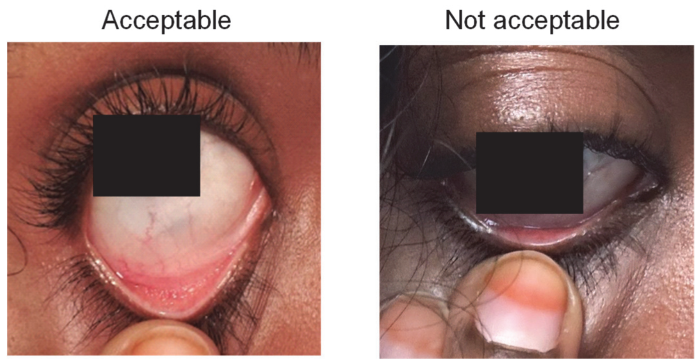
	RDT results	
	Odds ratio (p-value)	95% confidence interval
Age	1.07 (0.000)	1.05, 1.08
Sex	1.01 (0.798)	0.94, 1.09
Body temperature	0.96 (0.075)	0.915, 1.00
Blood Hgb level	0.90 (0.000)	0.88, 0.93

Adjusted for the ten selected radiomic features.

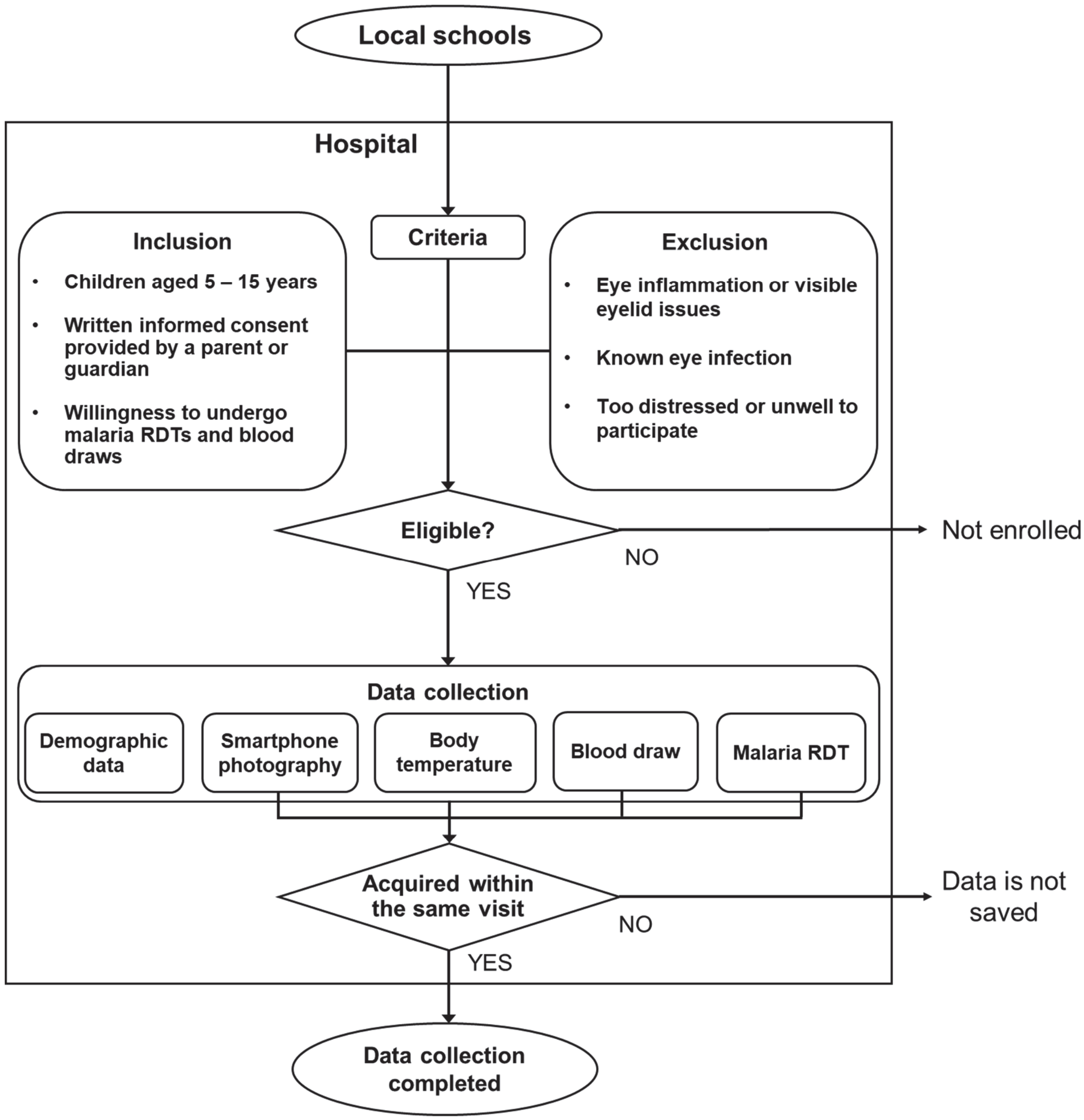
Supplementary Table 2. Mathematical definitions and key descriptions of selected radiomics features

Radiomic feature	Description	Mathematical definition
G, first order, root mean squared (#1)	The root mean squared of the green-channel intensity values, indicating the overall magnitude of intensity values.	$\text{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} G_i^2}$, where G_i represents the green (G)-channel intensity of pixel i and N is the total number of pixels.
G, first order, mean (#2)	The average green-channel intensity of all the pixels in the region of interest (ROI).	$\text{Mean} = \frac{1}{N}\sum_{i=1}^{N} G_i$, where G_i represents the green (G)-channel intensity of pixel i.
G, first order, median (#3)	The central tendency of the green-channel intensity values, less affected by extreme values.	The median is the middle value of the sorted green (G)-channel intensity values such that 50% of the values are below and 50% are above.
G, first order, minimum (#4)	The smallest green-channel intensity value within the ROI.	$\text{Minimum} = \min(G)$, where G represents the green (G)-channel intensity.
G, first order, 10 percentile (#5)	The lower end of the green-channel intensity distribution, capturing the lower-intensity values.	The value below which 10% of the green (G)-channel intensity values is calculated.
G, first order, 90 percentile (#6)	The upper range of the green-channel, indicating the threshold below which most values lie.	The value below which 90% of the green (G)-channel intensity values is calculated.
R, first order, maximum (#7)	The highest red-channel intensity value in the region of interest, indicating the peak of the distribution.	$\text{Maximum} = \max(R)$, where R represents the red (R)-channel intensity values.
G, first order, variance (#8)	The spread or dispersion of the green-channel intensity values from the mean.	$\text{Variance} = \frac{1}{N}\sum_{i=1}^{N}(G_i - \text{Mean})^2$, where G_i represents the green (G)-channel intensity of pixel i and Mean is the mean green-channel intensity.
G, first order, energy (#9)	The sum of squared green-channel intensity values, reflecting the magnitude of voxel values in the ROI. Higher energy indicates greater uniformity of pixel intensity values.	$\text{Energy} = \sum_{i=1}^{N} G_i^2$, G_i represents the green (G)-channel intensity of pixel i.
G, first order, mean absolute deviation (#10)	The mean of the absolute differences between each pixel's intensity and the overall mean intensity of the ROI, assessing heterogeneity.	$\text{Mean absolute deviation} = \frac{1}{N}\sum_{i=1}^{N} \lvert G_i - \mu \rvert$, where G_i represents the green (G)-channel intensity of pixel i and μ represents the mean intensity of the region of interest.

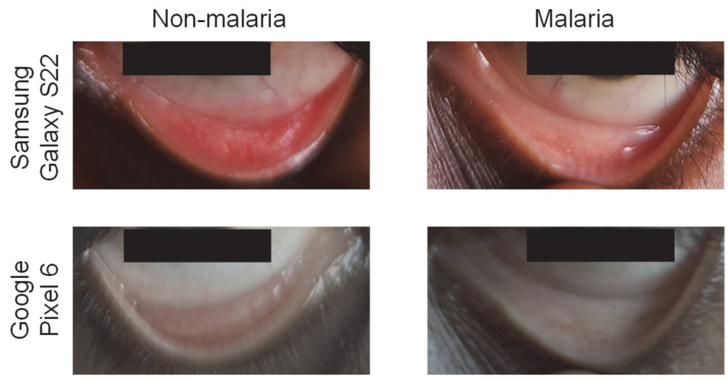
All radiomic features are computed within the inner eyelid area, delineated using the Mask R-CNN segmentation model.



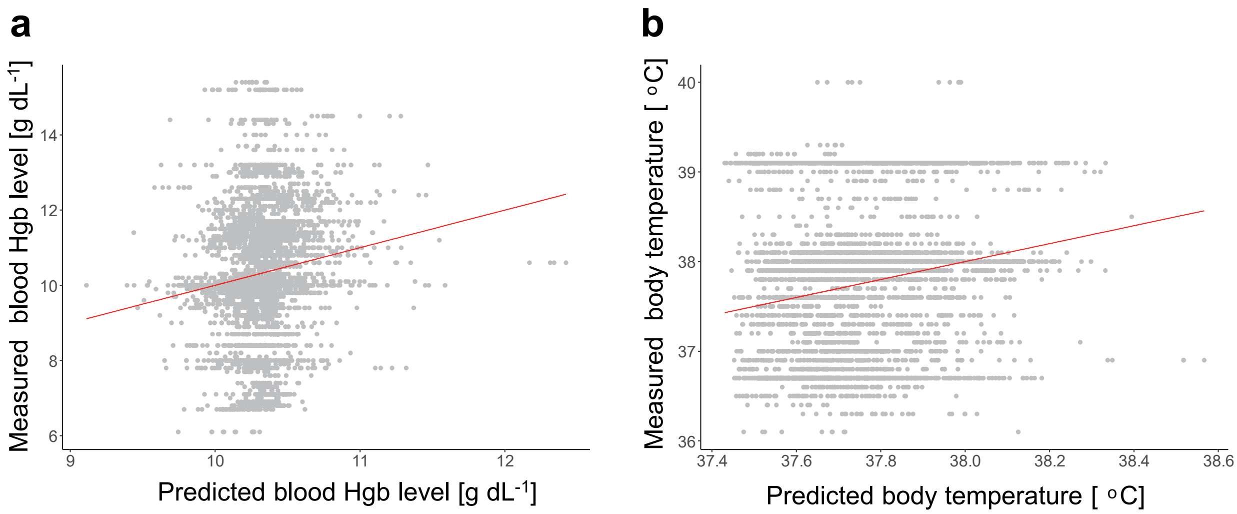
Supplementary Figure 1. Photo acquisition protocol. The user-friendly protocol guides study personnel in capturing high-quality photos of the inner eyelid. It includes instructions along with visual examples (left: acceptable example, right: unacceptable example) to ensure consistent and reliable photo acquisition. The individual in the photos is one of the authors and grants permission for the use in publication.



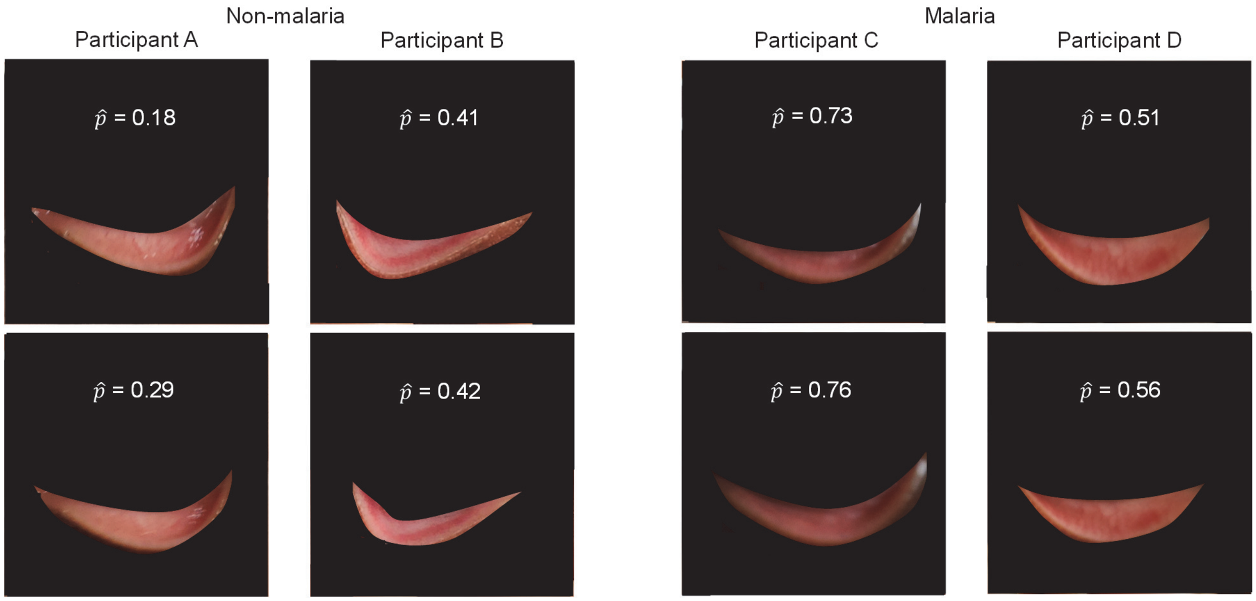
Supplementary Figure 2. Flowchart of recruitment, enrollment, and data collection. This single-center, cross-sectional observational study builds upon Rwanda's ongoing community-level malaria diagnosis and treatment initiatives, conducted in collaboration with the Rwanda Biomedical Centre (implementation arm of the Ministry of Health).



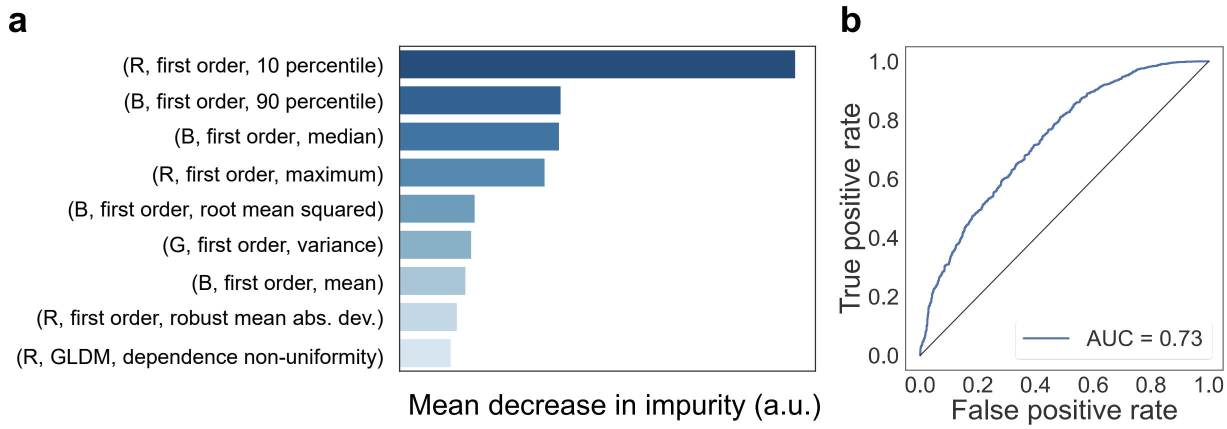
Supplementary Figure 3. Representative photos of the inner eyelid for malaria and non-malaria cases, acquired using Samsung Galaxy S22 and Google Pixel 6. Differences in onboard camera specifications and features between the two smartphone models lead to variations in color tones, contrast, and signal-to-noise ratios. Notably, the reported malaria classification model, based on the selected radiomic features, is unaffected by the smartphone models used in this study (**Figure 3e**).



Supplementary Figure 4. Predictions of blood Hgb levels and body temperature using multiple linear regression models based on the selected radiomic features (**Table 2**). (**a**) Scatter plot comparing the predicted versus actual blood Hgb levels. The high prevalence of anemia in the region may partially contribute to the observed overall association. (**b**) Scatter plot comparing the predicted versus actual body temperature. The low correlation coefficients between the selected radiomic features and both blood Hgb levels and body temperature suggest limited specificity of these features for predicting blood Hgb levels and body temperature.



Supplementary Figure 5. Representative examples of the predicted probability $\hat{p}$ from the malaria classification model for non-malaria and malaria cases. For non-malaria cases, output results from two representative participants are selected to illustrate relatively good and poor prediction performance. Similarly, for malaria cases, output results from two representative participants are selected to demonstrate relatively good and poor prediction performance. All photos acquired using Samsung Galaxy S22 are selected. The observed outcomes are 0 for a negative RDT and 1 for a positive RDT. Despite photo-to-photo variations within the same participants, the malaria classification model is relatively consistent when the predicted outcome is classified as positive if $\hat{p} > 0.5$.



Supplementary Figure 6. Radiomic analyses for detecting anemia. Using laboratory blood Hgb levels, separate radiomic analyses are performed to detect anemia following the WHO guidelines. (**a**) Feature importance of the top ten significant radiomic features for anemia, excluding shape-related features. The bar graph presents the mean decrease in impurity for each feature. (**b**) ROC curve of the anemia prediction model. The distinct set of radiomic features identified for anemia suggests that the features selected for malaria may exhibit some degree of specificity to malaria, rather than to systemic inflammation.

Supplementary Data 1
The clinical and radiomic dataset. The original photos are not publicly accessible because of the sensitive nature of patient information, which could compromise the participants' privacy and consent.

Supplementary Code 1
The custom codes and examples supporting the findings of this study.