



#5378

Malaria RDT interpretation accuracy of health workers compared to artificial intelligence (AI) and Panel Read in Kano State, Nigeria

Frade, Sasha¹, Cooper, Shawna¹, Smedinghoff, Sam¹, Hattery, David¹, Ruan, Yongshao¹, Isabelli, Paul¹, Ravi, Nirmal², McLaughlin, Megan³, Metz, Lynn³, Finette, Barry^{3;4}

¹ Audere, Seattle, WA, USA; ² EHA Clinics, Kano, Nigeria, ³ THINKMD, PBC, Burlington, VT, USA; ⁴ Department of Pediatrics, University of Vermont Larner College of Medicine, Burlington, VT, USA

Introduction

- Rapid and accurate testing, diagnosis, and treatment are crucial elements of public health malaria strategies found to decrease malaria prevalence and mortality rates.
- Proper treatment can prevent the further spread of infection and potentially combat the development of drug-resistant malaria.
- mRDTs, the test of choice of malaria diagnostic testing, require no special equipment, advanced medical training, or specific experience to administer. They are small and easily transportable, do not require samples to be stored, and are used with minimal training at the point of care; even in remote settings.
- Although mRDTs are known for their cost effectiveness, speed of result, and ease of use – health worker misadministration and misinterpretation errors remain a concern, including misinterpretation of faint lines.

Study Goal

Investigate whether the use of a mobile application powered by artificial intelligence (AI) could improve the accuracy of mRDT interpretations amongst Health Workers (HWs) in Kano State, Nigeria.



Figure 1: Kano State, Nigeria

Methods

- Forty-four (44) HWs were included in the study - 8 REACH community HW, 6 EHA Clinic (EHAC) HW, and 30 Kano state government community HW.
- THINKMD's clinical risk assessment platform (THINKMD platform) was used by HWs to assess patients presenting with malaria symptoms and recommend testing of suspected cases using an SD Bioline P.f mRDT.
- Android Intent protocols were used to direct HWs to the HealthPulse AI app to capture a photo of the mRDT and bidirectionally pass information between the THINKMD platform app and the HealthPulse app.

Conclusion

- The **AI accuracy surpassed HWs, especially on faint positives**, and was **resilient** to a range of mRDT and photo capture conditions that are challenging even for highly trained HWs.
- AI interpretation is **consistent**, whereas HWs can display variability.
- **AI-powered mRDT image capture can improve HW RDT result interpretation accuracy.**

HWs + AI = improved RDT result interpretation accuracy

A mobile app such as the THINKMD clinical decision support platform coupled with AI-powered mRDT image capture and interpretation can elevate human evaluation of mRDT results, even in low resource settings.

User App Flow

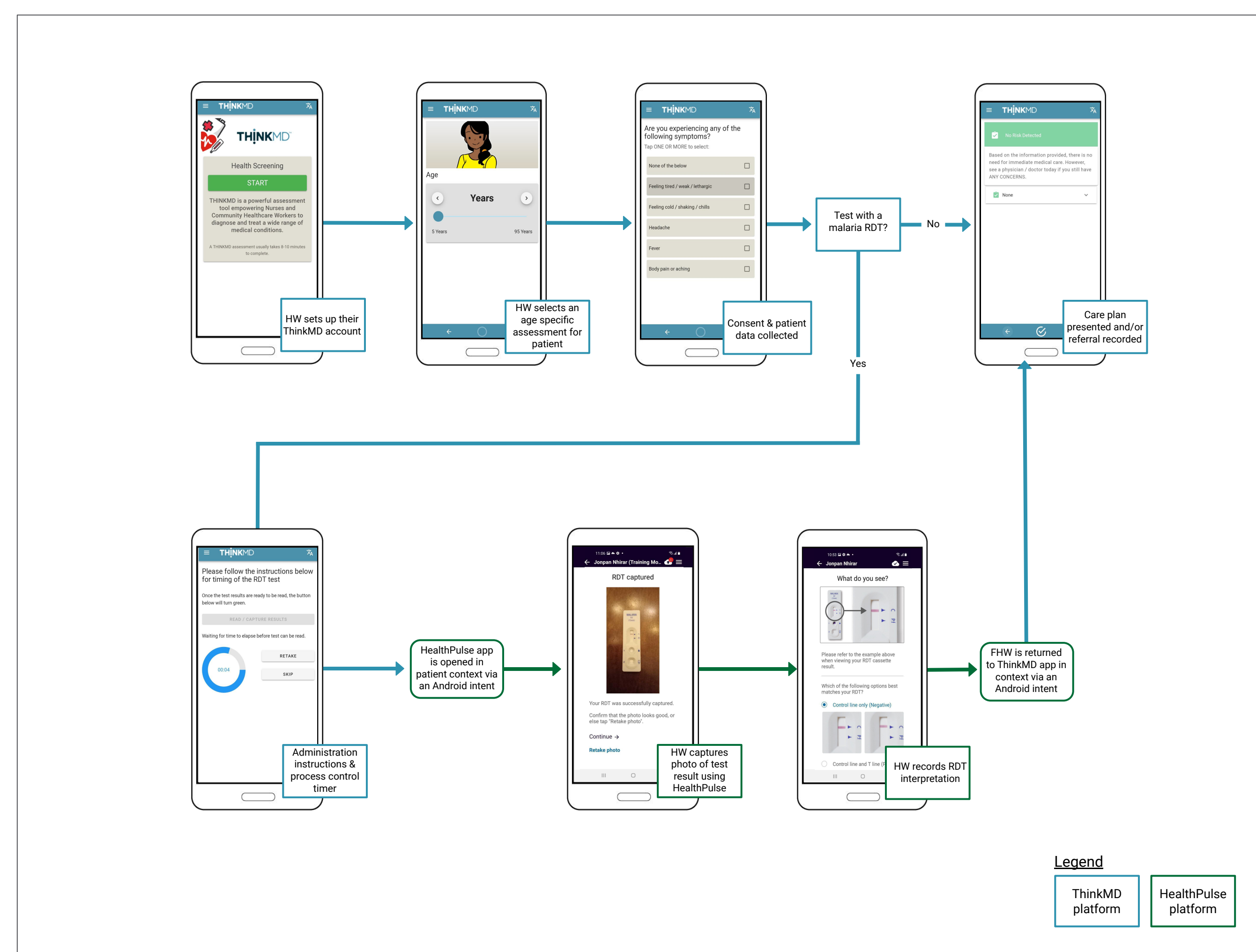


Figure 2: ThinkMD platform app and HealthPulse photo capture flow

Analysis Methods

A set of comparisons were performed to assess the accuracy of the HWs and AI using a Panel Read interpretation as the reference standard.

Result Interpretations Performed

HW Interpretation	Panel Read Interpretation	HealthPulse AI Interpretation
After taking a photo of the mRDT, HWs indicated in the HealthPulse app whether they saw a control line only (negative mRDT result), a control and P.f line (positive mRDT result), or no control line (invalid mRDT result).	A standard reference interpretation for each mRDT image was performed by an external, independent panel of three readers (Panel Readers). Where the Panel Readers agreed with the HW's interpretation of the mRDT result, the HW's interpretation was deemed correct.	To investigate the applicability of an AI-powered interpretation, HealthPulse AI performed an interpretation using only the mRDT photo.*
HWs evaluate the physical mRDT in person.	The panel readers evaluate a photo of the mRDT.	The AI evaluates a photo of the mRDT.

*The multi-stage flow confirms the image contains the correct mRDT (SD Bioline P.f for this study), confirms the image meets the needed quality bar, locates the mRDT result window in the image, identifies the presence and absence of control and test lines, and classifies the result as positive, negative, invalid, or uninterpretable.

Results

Quality of Images Captured

99.3% of images captured were interpretable by AI. Filters could be further refined to request retakes for uninterpretable images.

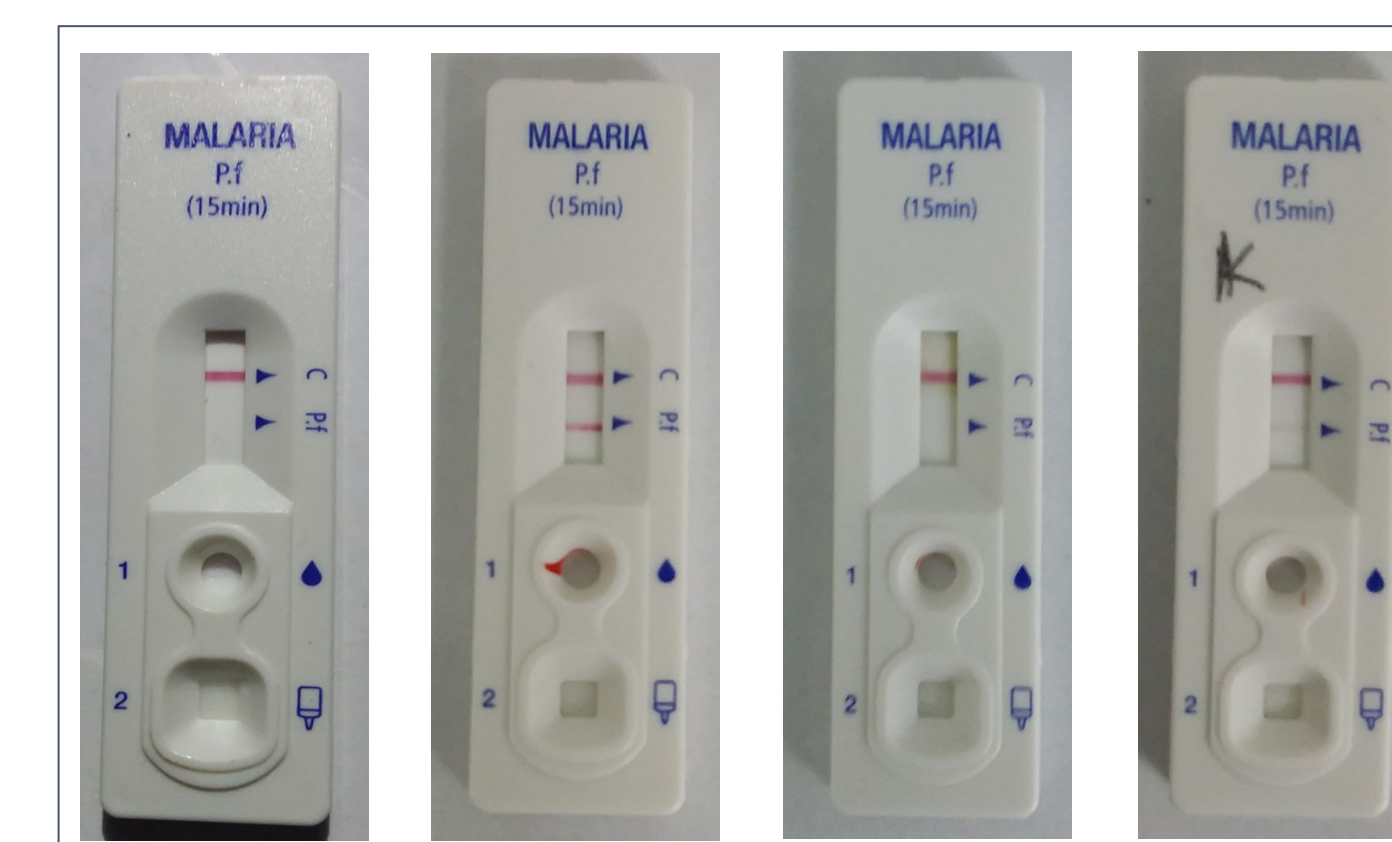


Figure 3: Good quality images

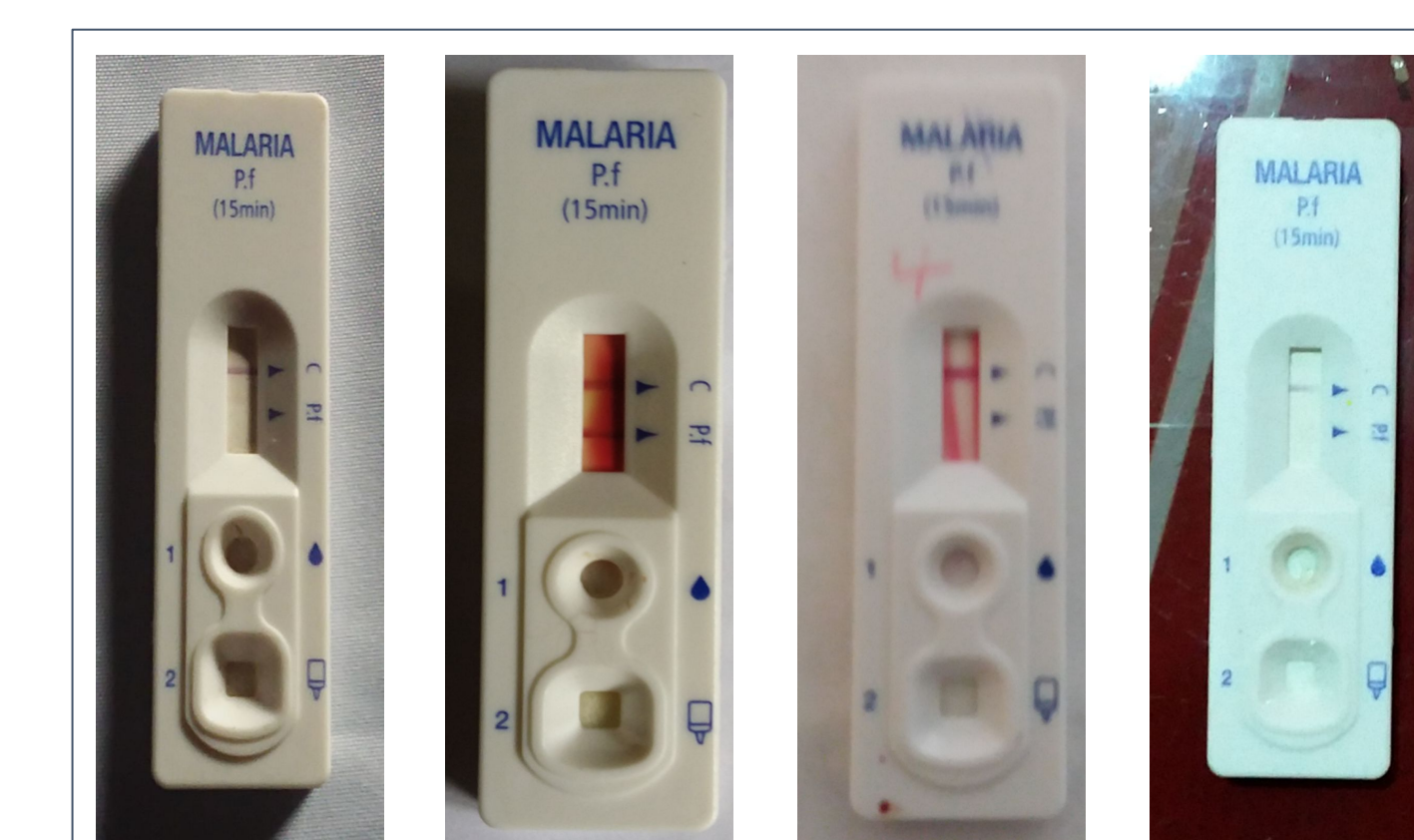


Figure 4: Sufficient quality for use with AI

mRDT Interpretation Accuracy Comparisons

Comparison of AI and Panel Read Interpretation

Panel Read	AI				Total
	Invalid % (n)	Negative % (n)	Positive % (n)	Uninterpretable % (n)	
Uninterpretable	28.57 (2)	1.52 (18)	0.77 (10)	11.1 (1)	30
Invalid	57.14 (4)	0.00 (0)	0.08 (1)	0.00 (0)	5
Negative	14.29 (1)	97.12 (1 148)	2.77 (36)	2.22 (2)	1 185
Positive	0.00 (0)	1.35 (16)	96.38 (1 251)	6.66 (6)	1 267
Total	100 (7)	100 (1 182)	100 (1 298)	100 (9)*	2 496

*Images where the AI ran but did not output an interpretation, as either no RDT could be located in the photo or the result window could not be located on the RDT.

Strong Panel Read-AI agreement indicates potential to use AI as the reference comparator.

Comparison of HW Interpretation and Panel Read Interpretation

Panel Read	HW Interpretation			Total
	Invalid % (n)	Negative % (n)	Positive % (n)	
Uninterpretable	0.00 (0)	1.30 (16)	1.19 (15)	31
Invalid	28.57 (2)	0.08 (1)	0.16 (2)	5
Negative	57.14 (4)	94.31 (1 160)	1.83 (23)	1 187
Positive	14.29 (1)	4.31 (53)	96.82 (1 219)	1 273
Total	100 (7)	1 230	1 259	2 496

Panel Read results indicates there was room for improvement in HW interpretations.

AI Accuracy

	AI	HWs
Weighted F1* "accuracy" compared to to panel read reference standard	96.4	95.3
Faint Positive Accuracy	90.2%	76.1%

HealthPulse AI had strong accuracy, performing similarly to HWs on easier to read mRDTs and better than HWs on faint positives.

*Weighted F1 is an accuracy metric typically used to evaluate AI model performance.

Number of Images	2,496
% Removed by Image Quality Assurance	0.7% (2,479 remain)
Interpretation Distribution After IQA	47.7% Negative 51.1% Positive 0.2% Invalid 1.0% Uninterpretable
# Faint Positives	163

Next steps

- Pursue opportunities to increase the efficiency and accuracy of malaria control programs through HealthPulse AI and the ThinkMD platform.

