Transforming Rapid Diagnostic Tests into Trusted Diagnostic Tools in LMIC using AI

Krishnam Gupta, Yongshao Ruan, Ahmed Ibrahim, Rouella Mendonca,
Shawna Cooper, Sarah Morris, and David Hattery
Audere
Seattle, USA
{krishnam, yongshao, ahmed, rouella, scooper, sarah, hattery}@auderenow.org

***Abstract*—In low and middle-income countries (LMICs), Rapid Diagnostic Tests (RDTs) are often the only way to diagnose diseases such as malaria, HIV, and COVID efficiently and cost effectively, especially in rural settings. However, basic RDTs are often misinterpreted, reducing their reliability for medical treatment or official case counts. AI-based mobile solutions are difficult to implement in LMICs due to limited resources available on commonly used phones and unstable Internet connectivity. HealthPulse AI algorithms aim to address these issues by providing a lightweight, yet highly accurate library of Computer Vision (CV) models for the detection and interpretation of common RDTs for conditions such as malaria, HIV, and COVID. The complete system can function end-to-end offline on phones with as little as 1 GB of total device memory. In addition to detecting the RDT type and interpreting the results, the system can flag image quality issues such as bad lighting or blurriness. If required, it can ask the user for a photo retake in real-time, reducing the need for re-testing. The system provides accurate and consistent result interpretation for surveillance or decision support use cases, helping health systems better understand current disease prevalence which may help mitigate the next pandemic. The AI algorithm pipeline uses deep learning to analyze RDT images, with multiple computer vision models working together to confirm the presence of the expected RDT, flag adverse image conditions, and provide accurate and consistent results. HealthPulse AI prioritizes privacy, accountability, and accessibility while aiming to revolutionize care delivery in LMICs by transforming low-cost RDTs into trusted diagnostic tools using computer vision and AI.**

***Index Terms*—RDT, rapid diagnostic test, rapid test, AI, artificial intelligence, AI algorithms, ML, machine learning, CV, computer vision**

I. Introduction

Basic pregnancy tests are a type of inexpensive Rapid Diagnostic Test (RDT), similar to at-home COVID tests which have recently become common. However, they can be difficult to use and interpret, leading people to prefer more expensive digital versions [1]. In low and middle income countries (LMIC), RDTs are crucial for diagnosing diseases like malaria, HIV, and COVID-19. However, studies show that basic RDTs are often misinterpreted [2], hindering their effectiveness. Using artificial intelligence (AI), simple smartphones can provide the same interpretation capabilities as more expensive tests with easier to read results, turning basic RDTs into trusted diagnostic tools. This could be a game-changer for LMICs,

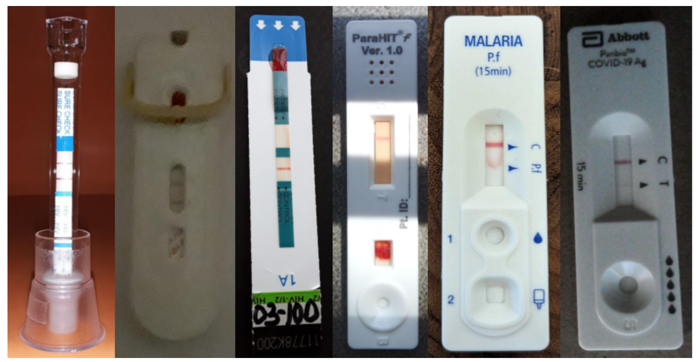

Fig. 1. Different types of RDTs for HIV, Malaria & Covid. We show some examples with adversarial conditions seen in the real world that make RDT interpretation more challenging. From left to right : *Sure Check HIV 1/2 Assay* with glare on test area, *Mylan HIV Self Test* low resolution image, *Abbott Determine - HIV 1/2* rotated image, *ParaHIT-F Malaria RDT* with a shadow on Test line region, *SD Bioline Malaria Ag Test* with staining on result window, *Panbio COVID-19 Ag Rapid Test* with a faint test line.

where 1 billion malaria and HIV RDTs are procured globally [3].

Many economic and technological constraints impact access in low and middle-income countries including limited network coverage, slow internet speeds, access to power for charging mobile devices, and access to modern mobile hardware. When developing mobile solutions for these regions, these accessibility constraints must be considered.

Computer vision models are trained using deep learning to analyze RDT images, providing revolutionary functionality for care delivery by providing a diagnostic solution that can:

- Run with intermittent or no Internet connectivity.
- Run with operating system as old as Android 6.
- Run on phones with total device memory as low as 1 GB.
- Flag adverse photo capture conditions.
- Provide accurate results for commonly used RDTs, even for faint line identification.
- Work across RDTs of different types, shapes and sizes.
- Run on images captured without use of any stands, controlled backgrounds, or lighting.

II. Related Work

The current body of research on AI interpretation of RDT results [4] [5] primarily focuses on closed room settings with controlled environments, which simplifies the problem. RDTScan [6] is a non-AI system that detects and interprets RDTs with a smartphone. However, current analyses performed in controlled lab settings are unlikely to exactly mimic conditions in the field.

While these studies are beneficial, they fail to address the many challenges that arise when photographing RDTs in real-world scenarios. These challenges include, but are not limited to: different lighting conditions, adversarial backgrounds, rotated RDTs, blurred or skewed images, faint test lines, blood or staining in the result window, the presence of dirt on the RDTs, and RDTs obstructed by barcodes or QR codes. These adversarial conditions often co-occur, making the problem more difficult.

III. Methodology

The HealthPulse AI training pipeline consumes the images and associated labels, producing multiple CV models that work together:

- An object detector that locates the RDT and its sub-parts within the image, and identifies the RDT type.
- A second object detector that examines the RDT result window (found by the prior object detector), and locates the test and control line regions.
- A classifier that examines each line region of the result window (found by the second object detector) and outputs line presence probability.
- An Image Quality Assurance (IQA) pipeline that flags adverse image conditions like blur, low lighting, and overexposure.

Google's MediaPipe (GMP) framework [7] is used to route images through this sequence of models and return results. We use a modular pipeline with multiple smaller models (versus one large model) to reduce memory requirements and improve compatibility for low-end smartphones. This approach allows us to break down the problem into three smaller sub-problems, enabling focused and flexible training of models that can be generalized for a wide range of RDT types.

A. ML Models Training Details

Customized versions of YOLOX Nano [8] are used for both object detectors. The YOLOX models are pretrained on the COCO dataset [9]. A smaller, customized version of MixNet [10] is used as a classifier for the final stage. The models can be converted to Tensorflow Lite (tflite) format to be run directly on the edge using GMP. The largest tflite model used is only 1 MB in size. Our combined pipeline has a runtime speed of 2500 ms for a 1 GB phone, which reduces to 500 ms for a high memory 8 GB phone.

B. Image Quality Assurance Pipeline

IQA involves utilizing algorithms to measure the quality of captured images based on a set of quality metrics. The IQA module focuses on extracting significant features like edges and textures from the photo. These features are then aggregated and compared to our internal predefined quality standards based on image interpretability by AI. If any quality issues are detected, the user will be prompted to retake the photo. This critical step in the pipeline plays an important role in improving overall system accuracy, ensuring that only images with sufficient quality to be accurately interpreted proceed to subsequent stages of the pipeline.

IV. Data and Results

A. Building Privacy, Accountability & Accessibility in AI

To ensure privacy, AI algorithms are initially trained with lab-generated data and fine-tuned with sets of field images. Field data is encrypted in transit and at rest and is stored in secure, HIPAA and GDPR-compliant environments with strict access control policies. Shared field data is stripped of PII before being used for training or validation purposes. We build safety and accountability into AI by ensuring that there is a human in the loop when diagnostics are used for prescribing medications. This is in-line with USAID's guide to machine learning [11]. To ensure accessibility, we build AI algorithms to operate on commodity smartphones available in LMIC. Further, we monitor AI performance as it is used in different locales, tuning as needed.

B. Dataset Details

To date, we've acquired or created over 150,000 labeled images, covering RDTs for COVID, HIV, and malaria. Image labels/metadata come from two sources: Photo capture in a lab that activates RDTs and records metadata/test results from the physical test, and from human labellers who receive only images (either lab-created or field). We outsource RDT test interpretation of images to a trained panel of reviewers for consistency and to remove bias in AI performance evaluation.

C. Results

The RDTs pictured in Figure 1 differ in material type, shape, size as well as location of test lines. Weighted F1 scores for field image sets with the number of images in the set is displayed in Table I below. In the last column, we report the impact of Image Quality Assurance on the weighted F1 score. The weighted F1 score is calculated by taking the mean of all per-class F1 scores while considering each class's support. We also report Health Workers Score, where relevant, in the last column.

We observe across all three diseases, that using IQA has a positive impact on model performance. The IQA pipeline filters out images which the model might fail on, thus helping the overall performance. Another advantage of IQA is that it has a high degree of overlap with human interpretability of the captured images. So the IQA filter can provide real-time training on how to get better photos. For each dataset in the table, the classifier accuracy with IQA exceeds human accuracy reading the test result in the field.

TABLE I
WEIGHTED F1 SCORE AND IMPACT OF IQA ON MODEL PERFORMANCE

Condition dataset	**Dataset size**	**Without IQA**	**With IQA**	**Health Workers Score**
COVID-19	731	0.993	0.993	N/A
HIV	888	0.981	0.988	N/A
Malaria Set-1	2479	0.963	0.964	0.953
Malaria Set-2	4205	0.938	0.949	0.929

V. CONCLUSION & FUTURE WORK

HealthPulse AI can substantially improve access to timely quality care across a variety of private, public, community, and self-testing use cases. For instance, using only an RDT image, it can power virtual care models by providing valuable decision support and quality control to clinicians, especially in cases where lines may be faint and hard to detect with the human eye. In the private sector, it can automate and scale incentive programs so auditors only need to review automated alerts based on test anomalies; procedures which presently require human reviews of each incoming image and transaction. In community care programs, HealthPulse AI can be used as a training tool for health workers who are learning how to correctly administer and interpret tests. In the public sector, it can strengthen surveillance systems by augmenting real-time disease tracking, through verification of incoming facility, community, and self-testing results to augment current disease responses and pandemic preparedness responses for any health condition identified via an RDT.

HealthPulse AI algorithms for malaria, HIV, and COVID tests are used in programs, pilots, and studies across Kenya, Nigeria, South Africa, Côte d'Ivoire, Uganda, Rwanda, and Benin. In the future, we plan to move towards a general purpose IQA using deep learning and make use of synthetic imagery for model training, especially for rare and difficult to obtain training examples.

VI. ACKNOWLEDGMENT

Financial support for Audere's work was provided by the Bill and Melinda Gates foundation, grant numbers INV-007492 and INV-054895.

REFERENCES

[1] Pike, Joanna, Sonya Godbert, and Sarah Johnson. "Comparison of volunteers' experience of using, and accuracy of reading, different types of home pregnancy test formats." Expert Opinion on Medical Diagnostics 7.5 (2013): 435-441.

[2] Kabaghe, Alinune N., et al. "Health workers' compliance to rapid diagnostic tests (RDTs) to guide malaria treatment: a systematic review and meta-analysis." Malaria Journal 15.1 (2016): 1-11.

[3] Habiyambere, V., et al. "Forecasting the global demand for HIV monitoring and diagnostic tests: A 2016-2021 analysis." PLoS One 13.9 (2018): e0201341.

[4] Bermejo-Peláez, David, et al. "A Smartphone-Based Platform Assisted by Artificial Intelligence for Reading and Reporting Rapid Diagnostic Tests: Evaluation Study in SARS-CoV-2 Lateral Flow Immunoassays." JMIR public health and surveillance 8.12 (2022)

[5] Mendels, David-A., et al. "Using artificial intelligence to improve COVID-19 rapid diagnostic test result interpretation." Proceedings of the National Academy of Sciences 118.12 (2021): e2019893118.

[6] Park, Chunjong, et al. "The design and evaluation of a mobile system for rapid diagnostic test interpretation." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5.1 (2021): 1-26.

[7] Lugaresi, Camillo, et al. "Mediapipe: A framework for building perception pipelines." arXiv preprint arXiv:1906.08172 (2019).

[8] Ge, Zheng, et al. "Yolox: Exceeding yolo series in 2021." arXiv preprint arXiv:2107.08430 (2021).

[9] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014.

[10] Chen, Tianqi, et al. "Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems." arXiv preprint arXiv:1512.01274 (2015).

[11] Managing Machine Learning Projects in International Development: A Practical Guide. https://www.usaid.gov/digital-development/managing-machine-learning-projects, Accessed Apr. 2023.