

The Design and Evaluation of a Mobile System for Rapid Diagnostic Test Interpretation

CHUNJONG PARK, HUNG NGO, and LIBBY ROSE LAVITT, University of Washington, USA
VINCENT KARURI, SHIVEN BHATT, and PETER LUBELL-DOUGHTIE, Ona, Kenya
ANURAJ H. SHANKAR, University of Oxford, United Kingdom
LEONARD NDWIGA, VICTOR OSOTI, JULIANA K. WAMBUA, PHILIP BEJON, and LYNETTE ISABELLA OCHOLA-OYIER, KEMRI-Wellcome Trust Research Programme, Kenya
MONIQUE CHILVER and NIGEL STOCKS, University of Adelaide, Australia
VICTORIA LYON, BARRY R. LUTZ, and MATTHEW THOMPSON, University of Washington, USA
ALEX MARIAKAKIS, University of Toronto, Canada
SHWETAK PATEL, University of Washington, USA

Rapid diagnostic tests (RDTs) provide point-of-care medical screening without the need for expensive laboratory equipment. RDTs are theoretically straightforward to use, yet their analog colorimetric output leaves room for diagnostic uncertainty and error. Furthermore, RDT results within a community are kept isolated unless they are aggregated by healthcare workers, limiting the potential that RDTs can have in supporting public health efforts. In light of these issues, we present a system called RDTScan for detecting and interpreting lateral flow RDTs with a smartphone. RDTScan provides real-time guidance for clear RDT image capture and automatic interpretation for accurate diagnostic decisions. RDTScan is structured to be quickly configurable to new RDT designs by requiring only a template image and some metadata about how the RDT is supposed to be read, making it easier to extend than a data-driven approach. Through a controlled lab study, we demonstrate that RDTScan's limit-of-detection can match, and even exceed, the performance of expert readers who are interpreting the physical RDTs themselves. We then present two field evaluations of smartphone apps built on the RDTScan system: (1) at-home influenza testing in Australia and (2) malaria testing by community healthcare workers in Kenya. RDTScan achieved 97.5% and 96.3% accuracy compared to RDT interpretation by experts in the Australia Flu Study and the Kenya Malaria Study, respectively.

CCS Concepts: • Applied computing → Consumer health; Health informatics; • Human-centered computing → Field studies; Smartphones.

Additional Key Words and Phrases: rapid diagnostic tests (RDTs), malaria, influenza, mobile health, image processing, image quality control

Authors' addresses: Chunjong Park, cjparkuw@cs.washington.edu; Hung Ngo, hvn297@cs.washington.edu; Libby Rose Lavitt, lrlavitt@cs.washington.edu, University of Washington, USA; Vincent Karuri, vkaruri@ona.io; Shiven Bhatt, sbhatt@ona.io; Peter Lubell-Doughtie, peter@ona.io, Ona, One Padmore Place, George Padmore Lane, Nairobi, Kenya; Anuraj H. Shankar, anuraj.shankar@ndm.ox.ac.uk, Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, New Richards Building, Old Road Campus, Roosevelt Drive, Oxford, United Kingdom, OX3 7LG; Leonard Ndwigga, info@kemri.org; Victor Osoti, info@kemri.org; Juliana K. Wambua, jwambua@kemri-wellcome.org; Philip Bejon, pbejon@kemri-wellcome.org; Lynette Isabella Ochola-Oyier, liochola@kemri-wellcome.org, KEMRI-Wellcome Trust Research Programme, P.O. Box 54840 00200 Off Mbagathi Road, Nairobi, Kenya; Monique Chilver, monique.chilver@adelaide.edu.au; Nigel Stocks, nigel.stocks@adelaide.edu.au, University of Adelaide, Helen Mayo North Building, Frome Road, Adelaide, SA, 5005, Australia; Victoria Lyon, vlyon@uw.edu; Barry R. Lutz, blutz@uw.edu; Matthew Thompson, mjt@uw.edu, University of Washington, USA; Alex Mariakakis, atm15@cs.washington.edu, University of Toronto, 40 St. George Street, Toronto, ON, M5S 2E4, Canada; Shwetak Patel, shwetak@cs.washington.edu, University of Washington, 185 Stevens Way NE, Seattle, WA, 98195, USA.



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

© 2021 Copyright held by the owner/author(s).

2474-9567/2021/3-ART29

<https://doi.org/10.1145/3448106>

ACM Reference Format:

Chunjong Park, Hung Ngo, Libby Rose Lavitt, Vincent Karuri, Shiven Bhatt, Peter Lubell-Doughtie, Anuraj H. Shankar, Leonard Ndwiga, Victor Osoti, Juliana K. Wambua, Philip Bejon, Lynette Isabella Ochola-Oyier, Monique Chilver, Nigel Stocks, Victoria Lyon, Barry R. Lutz, Matthew Thompson, Alex Mariakakis, and Shwetak Patel. 2021. The Design and Evaluation of a Mobile System for Rapid Diagnostic Test Interpretation. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 29 (March 2021), 26 pages. <https://doi.org/10.1145/3448106>

1 INTRODUCTION

Over the past couple of decades, rapid diagnostic test (RDTs) have emerged as a potential solution to the pressing need for accessible medical testing. RDTs utilize biochemistry to transduce the load of a biological sample (e.g., blood, urine, nasal swab) to an analog colorimetric output. As such, RDTs enable point-of-care diagnostics without the need for expensive equipment. Pregnancy tests are one of the most well-known types of RDTs, but RDTs exist for many other health purposes, such as malaria [2, 33, 38], influenza [23, 48], and HIV [27, 56]. Recently, RDTs are being developed to support convenient COVID-19 testing during the pandemic^{1,2}. RDTs are viable to produce at scale, making them an inexpensive (~\$1 USD each) alternative to laboratory tests and ideal for point-of-care medical screening [32, 53]. RDTs are often associated with community healthcare settings in low- and middle-income countries where resources and access to sophisticated testing facilities are limited, but RDTs are used worldwide in clinics and homes as well [5, 44].

As RDTs become more commonplace, one concern is that people may misinterpret the visual results that appear on their tests—overlooking faint lines, thinking they are seeing lines that are not actually present, or misunderstanding the lines' meanings [20]. A system that interprets RDTs on the user's behalf would limit such errors, leading to improved test accuracy and higher utility among end-users. The ability to digitally document RDT results could also facilitate community-wide reporting, contact tracing, and surveillance networks during disease outbreaks. The ideal system would satisfy the following requirements:

- **High interpretation accuracy:** The ideal system would have comparable diagnostic accuracy to an expert directly reading the RDTs themselves. In addition, the ideal system would be consistent across settings (e.g., smartphone model, ambient environment). Such a system would improve the effective diagnostic accuracy of RDTs in the field, particularly amongst novice users, while removing the potential for subjective, biased, or rushed decision-making.
- **Smartphone-only:** RDTs are an attractive option for community use because all of the equipment needed to run an RDT comes within an inexpensive kit. Introducing hardware or smartphone accessories, as past researchers have proposed [11, 12, 22, 37, 42], hinders deployability. Roughly 45% of the global population owns a smartphone as of April 2020 [3]. Specifically in sub-Saharan Africa, the Global System for Mobile Communications (GSMA) estimates that the fraction of people who have a SIM connection will grow from 77% in 2019 to 86% in 2025, and the fraction of those connections coming from smartphones will rise to 65% in 2025 [19]. Therefore, requiring a smartphone does not necessarily introduce significant burden.
- **Configurability to new RDT designs:** Like many other products, lateral flow RDTs are produced by manufacturers with no overarching design standard beyond having results displayed as a set of lines. Existing smartphone-based RDT readers are catered to specific RDT brands^{3,4}, and a machine learning approach would require a dataset with hundreds of images in diverse settings for model training to accommodate new RDT designs. The ideal system would be quickly configurable so that new RDTs can be used in response to epidemics.

¹<https://cellexcovid.com/>

²<https://en.wondfo.com.cn/product/wondfo-sars-cov-2-antibody-test-lateral-flow-method-2/>

³<https://www.novarumdx.com/>

⁴<http://www.albagaia.com/hydrosense-app>

With these requirements in mind, we present RDTScan, an open-source system that supports automatic RDT detection and interpretation on smartphones. RDTScan provides real-time guidance to users so that they capture a high-quality photograph of their completed RDT; RDTScan then analyzes that photograph to infer the RDT's result. RDTScan builds off of our previous work [45] by improving the robustness of our feature-based template matching approach to RDT detection, extending our system to accommodate new RDT designs, adding new quality assurance checks to ensure accurate interpretation across different RDT form factors, and rigorously testing our system through multiple studies. We first evaluated our image processing pipeline in a controlled laboratory study to demonstrate RDTScan's ability to automatically interpret RDTs. After we established that our pipeline could either match or exceed human interpretation capabilities, we engaged with community health surveillance programs to deploy RDTScan in two contrasting scenarios: (1) at-home influenza testing in Australia and (2) malaria testing by CHWs in Kenya. Participants in both studies had success capturing a high-quality photograph of their completed RDT, reaching an overall success rate of 83.3% and 91.9% in the Australia Influenza Study and Kenya Malaria Study, respectively. RDTScan showed comparable or slightly better interpretation performance to that of experts. In the Australia Influenza Study, RDTScan achieved 97.5% (85.7% sensitivity, 98.7% specificity) compared to the experts and 83.6% accuracy (33.9% sensitivity, 98.4% specificity) compared to gold-standard clinical measurements. In the Kenya Malaria Study, RDTScan achieved 96.3% accuracy (95.5% sensitivity, 98.7% specificity) and 85.5% accuracy (92.9% sensitivity, 65.0% specificity) compared to the experts and gold-standard clinical measurements, respectively. In summary, our research delivers the following contributions:

- (1) An open-source, smartphone-based RDT interpretation system⁵ that achieves comparable accuracy to experts who are experienced in RDT administration without the need for additional hardware,
- (2) An in-lab validation study showing RDTScan's interpretation accuracy across RDTs, analyte concentrations, smartphone devices, and lighting conditions, and
- (3) Two field evaluations that demonstrate the efficacy of RDTScan in vastly different settings.

2 RELATED WORK

RDTs leverage an assortment of techniques to detect medical conditions [32, 53]. Covering these techniques is out of the scope of this work, so we instead focus on how RDTs have been incorporated into healthcare workflows around the world. We then describe past approaches to automatic RDT interpretation, after which we discuss object interpretation for broader object categories.

2.1 Current Practices with RDTs

RDTs are typically used in settings outside of hospital environments, namely community and primary care settings. RDTs are popular in these areas because of their low cost, ease of use by non-lab technicians, and portability. RDTs have been used both for illnesses that are endemic in certain low- and middle-income countries (e.g., malaria in sub-Saharan Africa [2, 33, 38], leishmaniasis in India [50]) as well as illnesses that are clinical priorities in high-income countries (e.g., group A streptococcus [9], influenza [54, 58], HIV [27, 56]). Some RDTs are already seeing use for at-home testing [5, 44], and we anticipate this trend will grow as more RDTs gain regulatory clearance.

Although RDTs are becoming increasingly user-friendly, subjective interpretation of their output leaves room for diagnostic uncertainty and error. This topic has mostly been explored in the context of malaria RDTs. Harvey et al. [20] observed CHWs as they administered malaria RDTs and noted that only 54% were able to correctly interpret the test results; the most common mistakes were failures to identify faint positive lines or invalid results. Harvey et al. found that multi-day training programs improved interpretation accuracy to 93%, but such programs are impractical in many cases since they can take CHWs away from other responsibilities. Although

⁵<https://github.com/cjpark87/rdt-scan>

there have been attempts to improve procedural adherence and mitigate confusion by standardizing terminology, labeling, and instructions across RDT manufacturers, these efforts have typically been slow to implement and have been focused around RDTs for the same condition [25]. Our work seeks to shift interpretation burden from novice users to an automated analysis platform while being flexible enough to accommodate new designs with significantly less overhead than what would be required for a data-driven model.

2.2 Automated RDT Capture and Interpretation

Image interpretation is most successful when a clear image has been taken of the target object. Therefore, many researchers have proposed standalone devices and smartphone adapters to control the imaging environment (i.e., ambient lighting, shadows, camera position) for automatic RDT interpretation [42]. An example of a standalone device for this purpose is Fio's battery-operated Deki Reader⁶, which includes an internal chamber with controlled lighting. Herrera et al. [22] compared the RDT interpretation accuracy of the Deki Reader against visual inspection of the RDTs by experts and saw 99% concordance between the two. Mudanyali et al. [37] translated many of the Deki Reader's features to a smartphone-based system, utilizing an LED array and other optical components for imaging. Targeting a less expensive solution, Dell et al. [11, 12] and Ozkan et al. [43] independently proposed the use of plastic stands for consistent positioning between the smartphone's camera and the RDT. Because hardware and accessories enforce constraints on the imaging environment, the accompanying software can be highly tuned and efficient; however, incorporating hardware imposes additional financial costs to end-users and thus reduces the ease-of-access and ubiquity that RDTs engender in the first place. In contrast, RDTScan transfers the burden of image quality control from hardware to software while still being mindful of on-device computational limits.

There are commercial products for automatic RDT interpretation, as well. For example, Ellume⁷ produces a custom RDT cartridge with embedded sensors that can read a custom immunoassay made with fluorescent nanoparticles. Apps like Novarum's DX Mobile Reader³ and Albagia's Hydrosense app⁴, on the other hand, use computer vision to analyze lateral flow RDTs; unfortunately, there is no documentation about their algorithms or performance because they are proprietary apps. Regardless, these products are catered to specific RDT brands, so the underlying software can rely on design-specific features to interpret those RDTs. Our approach is unique in that we aim to accommodate new RDT designs with a single template image and metadata, which is significantly less overhead than the large image datasets that would be needed to support a machine learning approach. Our prior work [45] demonstrates the first step we took towards supporting RDT interpretation: a smartphone app that uses real-time image processing to ensure high-quality image capture. We have since improved upon RDTScan in a few ways, including a better method for RDT detection and additional quality assurance methods for blood and glare detection, fiducial tracking, and color-aware line interpretation. We also rigorously evaluate our automatic result interpretation algorithm across RDTs, analyte concentrations, smartphone devices, and lighting conditions through both an in-lab evaluation and two case studies.

2.3 Guided Media Capture

Whether a person is taking a picture of an RDT or another object, post-processing can only do so much to improve the quality of a poorly captured image. In light of this issue, researchers have explored ways of introducing real-time guidance for media capture in various domains. NudgeCam [6] leverages the smartphone's inertial sensors to track the camera's stability and orientation. NudgeCam also assesses the video content itself, using real-time image processing to check the overall brightness of the scene and to detect faces. EasySnap [26, 55] provides text-to-speech audio cues to help blind and low-vision photographers take well-framed pictures. For pictures involving people, EasySnap uses face detection to continuously monitor the size and position of the

⁶<http://fio.com/>

⁷<https://www.ellumehealth.com/>

subject's face in the photo. For pictures involving objects, the photographer can walk up to a target object so that EasySnap can register its visual features; once the object is registered, the photographer can move away from the object and EasySnap ensures that the object remains in view.

Guided image capture has also been created for specific object categories. One notable example is bank check recognition for online banking^{8,9}. To the best of our knowledge, these apps rely on optical character recognition to localize the consistent features of the bank check (e.g., routing number, check ID) and use those regions to infer the remaining contents [24]. Because countries have unique bank check formats, companies often develop one system per country [17, 18], which limits generalizability. Chen et al. [8] created an app called SmartDCap to help people scan paper documents with a smartphone. SmartDCap continuously checks the framing and sharpness of the document, consolidating these metrics into a score that is shown to the end-user along with audio feedback.

RDTScan builds upon these developments with a specific focus on RDTs. RDTScan performs the RDT-specific task of result interpretation while being configurable to a variety of RDT designs. RDTScan also provides real-time, human-readable feedback so that users can capture high-quality photographs of their RDTs and maximize the likelihood of correct interpretation.

3 RDTSCAN DESIGN

In this section, we first introduce the standards and terminology for lateral flow RDTs—the specific subgroup of RDTs that this work addresses. We then use this vocabulary to describe the RDTScan system. Throughout this section, we refer to the HLS (hue-saturation-lightness) color space, which is an alternate image representation to the standard RGB (red-green-blue). For our purposes, the HLS color space is defined as follows: $H \in [0, 179]$, $L \in [0, 255]$, $S \in [0, 255]$.

3.1 Standards and Terminology

In a typical lateral flow test, a liquid biological sample migrates across a strip via capillary flow. The sample first passes over a conjugate pad that holds the particles needed to create the colorimetric output (e.g., colloidal gold, latex) and then over zones that immobilize a target antigen, antibody, or protein. As more of the target is immobilized, visible colored lines appear to the end-user.

Lateral flow RDTs come in a variety of form factors, the two most common being *cassettes* and *dipsticks*. RDT cassettes are activated by putting a liquid biological sample in the sample well and then adding a buffer solution to the buffer well. RDT dipsticks are activated by mixing a biological sample with a buffer solution in a test tube and then dipping the strip into the tube. In this work, we focus our attention on the two RDTs shown in Figure 1: (1) AccessBio's CareStart Malaria Pf/Pv test¹⁰, an RDT cassette that analyzes whole blood; and (2) Quidel's QuickVue Influenza A+B test¹¹, an RDT dipstick that analyzes a nasal swab specimen.

An RDT's results appears in the *result window*—the thin rectangular region where the immunoassay itself is exposed to the user. Whenever an RDT has been activated properly, its *control line* will be visible to the user. Any other lines in its result window, called *test lines*, indicate the presence of a target analyte. The intensity of the test lines are a function of the biological sample's analyte concentration; the higher the load, the more intense the test lines will appear. The control line, on the other hand, is typically intense as long as the user administered the RDT properly.

Lateral flow RDTs can vary by more than just their form factor. RDTs can have one or many test lines depending on how many strains of the same pathogen they can detect. Going from left to right in Figure 1, the CareStart RDT can show a control line and then two lines indicating the presence of different malaria parasites (*Plasmodium*

⁸https://play.google.com/store/apps/details?id=com.wf.wellsfargomobile&hl=en_US

⁹https://play.google.com/store/apps/details?id=com.infonow.bofa&hl=en_US

¹⁰http://www.accessbio.net/eng/products/products01_02.asp

¹¹<https://www.quidel.com/immunoassays/rapid-influenza-tests/quickvue-influenza-test>

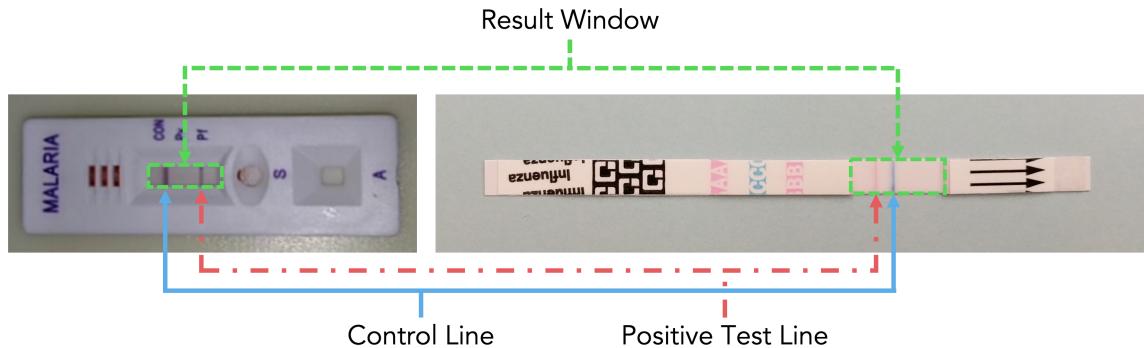


Fig. 1. The two RDTs evaluated in this work: (left) the CareStart Malaria Pf/Pv RDT cassette, showing a valid positive case of *Plasmodium falciparum*, and (right) the QuickVue Influenza A+B RDT dipstick, showing a valid positive case of influenza A.

Table 1. The metadata required to accommodate a new RDT design with RDTScan.

Data Field	Description
Template image	A non-skewed, tightly cropped photo of the RDT
Result window corners	The (x, y) pixel coordinates denoting the top-left and bottom-right corners of the general region where the results will appear
Line positions	The (x, y) pixel coordinate denoting the center of the various lines in the RDT
Line meanings	The meaning of the top line (e.g., “control”, “influenza A”)
Viewfinder scale	The relative height of the viewfinder compared to the screen’s height
Fiducial locations (optional)	The (x, y) pixel coordinates denoting the top-left and bottom-right corners of variable dark-colored markings that have a fixed location
Line color (optional)	The range of color hue values (H channel) expected for each of the RDT’s lines
Line intensity (optional)	The minimum expected difference in brightness (L channel) between the immunoassay and the lines; increasing the threshold requires a more intense line to appear

falciparum and *Plasmodium vivax*). The QuickVue RDT, on the other hand, can show a control line between two test lines for different types of influenza (A and B). The color, position, and order of these lines depend on the manufacturer’s specifications.

The diversity of RDT designs, combined with a lack of standardized markings, makes it difficult for a single algorithm to interpret multiple RDT designs. A data-driven algorithm would work if hundreds or thousands of annotated images are available for each new RDT design. However, it would take significant effort and resources to activate and photograph all of those RDTs, creating a barrier for quickly adapting such an algorithm to new RDT designs. These challenges led to the development of RDTScan, which utilizes feature-based template matching to accommodate new RDT designs.

3.2 Required Metadata for New RDTs

To leverage feature-based template matching, RDTScan requires a reference template image and the metadata listed in Table 1. The template image is a photograph of an unused, unmodified RDT with as few imperfections

as possible (i.e., no shadows or perspective skew). One way to create such an image is by taking a photograph of the target RDT on an uncluttered background with a free document scanning app like Microsoft’s Office Lens¹². Once the reference template image has been created, the image is loaded in an image-editing software program (e.g., Microsoft Paint, Photoshop) to identify the rough locations of the result window and lines.

RDTScan also accepts three optional pieces of metadata to improve performance for a particular RDT. To aid with detection, RDTScan can take advantage of *fiducials*—dark markers that are consistent in location, but not appearance, across RDTs from the same manufacturer (e.g., bar codes). To aid with interpretation, RDTScan can leverage the expected color hues of the different lines. Finally, RDTScan has a default value for the minimum expected difference in brightness (L channel) between the lines and the background immunoassay for a positive test result, but that parameter can be adjusted to prioritize sensitivity or specificity. The following subsections describe how RDTScan uses the aforementioned metadata.

3.3 Detection and Quality Checking

The clearer the image of an RDT, the easier it is for an algorithm to automatically interpret the test results. Because there is an upper limit to how much an image can be post-processed, RDTScan is designed to facilitate high-quality image capture. In our initial work [45], we presented an early instantiation of RDTScan catered towards a single RDT. Participants in that first deployment noted a few issues that hindered their ability to capture a clean photograph of their RDTs, such as instances of glare, shadows, and blood appearing in the result window. Furthermore, we ran into many challenges as we tried to apply that version of RDTScan to different RDT designs like Quidel’s QuickVue RDT. These challenges motivated improvements to our RDT detection approach and quality assurance checks. We summarize RDTScan’s complete functionality below.

3.3.1 Camera Configuration. One way to enforce consistency during image capture is through intelligent hardware configuration. Ambient lighting can affect the appearance of objects, so RDTScan activates the smartphone’s flash by default as a dominant illuminator around the RDT. RDTScan also utilizes the operating system’s auto-focus, auto-exposure, and auto-white-balance functions to adjust the camera’s properties with respect to the target RDT. By default, these functions are designed for global optimization—adjusting the camera’s properties with respect to its entire field-of-view. RDTs with a narrow aspect ratio only take up a small part of the overall image frame, so global optimization would cater to the RDT’s background in those cases. Instead, RDTScan enforces local optimization at the center of the camera’s field-of-view where the RDT is expected to be.

3.3.2 RDT Detection. RDTScan employs feature-based template matching to minimize the bootstrapping effort required to locate an RDT design within an image. RDTScan first extracts unique visual features from both the template image and the camera frame and then identifies matches between feature keypoints via brute-force. RDTScan then uses a least-squares procedure to calculate a 3×3 homography matrix that maps the corners of the template to an irregular quadrilateral that surrounds the RDT in the camera frame. In our prior work [45], we used the BRISK [29] feature extractor because of its efficiency on low-end devices at a slight cost to accuracy. However, getting an accurate homography matrix is particularly critical for dipstick RDTs since their uneven aspect ratio leaves little room for error. In the updated version of RDTScan, we prioritize accuracy over latency by utilizing the SIFT [30] feature extractor. SIFT is designed to be scale-invariant, which provides two benefits: (1) it allows RDTScan to provide more accurate guidance if the size of the RDT is significantly different between the camera frame and the template image, and (2) it allows RDTScan to downsample incoming camera frames by $\times 1/2$ to expedite computation without significantly impacting accuracy. We further reduce computation time by only extracting SIFT features within the middle of the screen where the viewfinder is located (see details on the user interface in Section 3.5). These two optimizations reduce the processing time per frame by a factor of $\times 1/16$.

¹²<https://www.microsoft.com/en-us/p/office-lens/9wzdnrcfj3t8>

compared to using SIFT to extract feature keypoints from the entire frame. Computational complexity during image capture is an important consideration because RDTScan analyzes video frames in real-time to provide end-users with feedback; as computations become longer, the delay between updated feedback updates increases.

3.3.3 Quality Assurance – Exposure. Images that are dim due to underexposure or washed out due to overexposure make it difficult to see lines in the RDT’s result window. To check the camera frame’s brightness, RDTScan computes a histogram of the frame’s L channel. Frames are considered underexposed if the histogram’s maximum value is less than 125, and frames are considered overexposed if at least 20% of the pixels are greater than 255.

3.3.4 Quality Assurance – Sharpness. Because the underlying lateral flow immunoassay enacts changes in line intensity within the result window, image sharpness is the principle factor that determines whether a line will be visible. To check that there is sufficient sharpness, RDTScan computes edge intensity variance of an image I according to the Laplacian operator [46]:

$$\text{var}(\text{Laplacian}(I)) = \text{var}\left(\frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}\right) \forall (x, y) \in I \quad (1)$$

High variance indicates the presence of both smooth regions and intense edges, which is typical of most focused images. However, the ideal amount of edge variance depends on the target RDT design since some are more visually complex than others. RDTScan computes the edge variance of the template image before the app is loaded to serve as a baseline. Because end-users are expected to position their RDT at the center of the frame within a viewfinder, RDTScan computes the edge variance in that region for incoming frames. A frame passes this check if its edge variance is at least 80% that of the template image.

3.3.5 Quality Assurance – Position, Size, and Orientation. Feature-based template matching can extract any affine transformation between a template and a target image. However, imposing some standards regarding the size, position, and orientation of the RDT in a given camera frame is important for consistent review by both humans and software. To check these properties, the irregular quadrilateral that defines the border of the detected RDT in the camera frame is converted to a rotated rectangle by computing the average length of opposing sides and placing them around the center of the original shape. RDTScan checks the position, size, and orientation of the resulting rectangle as follows:

- **Position:** The distance between the center of the rotated rectangle and the center of the camera frame should be within 10% of the image height.
- **Size:** The area of the rotated rectangle should be close to the ideal size of the RDT in the camera frame as dictated by the viewfinder scale parameter in the metadata. This check has a tolerance of 10% of the camera frame’s height.
- **Orientation:** The angle between the rotated rectangle’s longer axis and the vertical axis of the camera frame should be between -10° and $+10^\circ$.

3.3.6 Quality Assurance – Glare. Some lateral flow tests have a clear, glossy membrane to protect the underlying immunoassay from environmental damage. Unfortunately, this membrane can create glare on the result window depending on the relative positions and orientations of the camera, flash, RDT, and ambient lighting during image capture. Just as how RDTScan checks the overall illumination on the entire RDT, it also checks the brightness histogram specifically within the result window for any clipping. As before, the result window is considered underexposed if the histogram’s maximum value is less than 125 and overexposed if at least 20% of the pixels are greater than 255.

3.3.7 Additional Fiducial Detection. Some RDTs, particularly dipsticks, do not have identical markings across tests due to how they are manufactured. Quidel’s QuickVue RDTs, for example, are cut from a single sheet at an

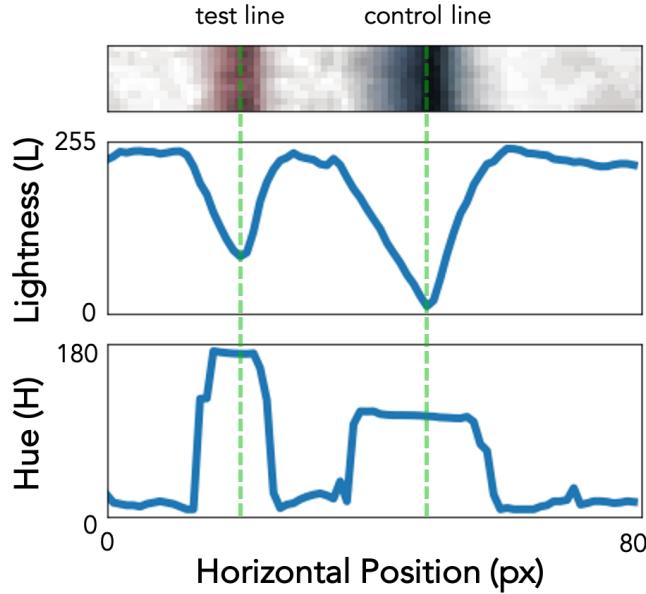


Fig. 2. An illustration of RDTScan’s interpretation algorithm, which leverages both the lightness (L) and hue (H) channels to identify lines within the result window.

interval that is different than the underlying pattern. This process creates strips with the same general design but with offset features, posing a challenge for many computer vision-based approaches. To accommodate such designs, RDTScan can leverage fiducials in the target RDT design. RDTScan searches for these fiducials using k-means clustering within a 5D spatial-chromatic space according to RGB color and (x,y) -position. We assume $k=5$ for the number of clusters since most RDT designs have few unique colors. The position of the darkest cluster is then compared against the fiducial locations specified in the metadata. Incoming camera frames pass this check if the distance between the x -coordinates of the detected and pre-specified fiducial locations is within 50% of the fiducial’s width and the corresponding y -coordinates are within 50% of the fiducial’s height.

3.4 Interpretation

RDTScan analyzes the result windows of captured RDTs for the presence of test and control lines. Our initial version of RDTScan [45] relied strictly upon the L channel to identify lines, making it susceptible to false positives from blood, shadows, or other irregularities in the result window. Below, we describe RDTScan’s interpretation algorithm and the improvements that have been made to make it robust to such issues.

3.4.1 Result Window Crop. To isolate the result window from the rest of the image, RDTScan crops the camera frame around the detected RDT’s bounding box and rectifies the remainder so it has the same dimensions as the template image. The resulting image is cropped a second time according to the coordinates specified in the metadata to highlight the RDT’s result window. In many RDT designs, the result window is sunken into the cassette or dipstick, creating shadows along the edges; therefore, a thin border (10% of the result window’s height) is cropped from the edges parallel to the control line to ignore those potential false positives.

3.4.2 Result Window Enhancement. RDTScan applies image post-processing to emphasize faint lines. Techniques like brightness histogram equalization typically work best on images that have a variety of dark and bright pixels. RDTs tend to be mostly white, so applying equalization globally would darken most of the pixels and make faint lines less distinctive. To avoid this issue, RDTScan uses contrast-limited adaptive histogram equalization (CLAHE) [47] to enhance the image's contrast on a tile-by-tile basis.

3.4.3 Presence of Blood. One common issue we saw in our prior work with blood-based RDT cassettes was that CHWs sometimes overused or misplaced blood samples on the RDT to the point where blood would seep into the result window and lines could not be discerned. End-users can easily recognize a bloody RDT without any assistance, but a means of automatically identifying these cases can help rule out algorithmic false positives or expedite manual review by back-end supervisors. To detect blood within the result window, RDTScan applies the following filter to isolate red pixels: $H \in [0, 10] \cup [160, 179]$, $L \in [100, 255]$, $S \in [100, 255]$. An RDT is deemed to have too much blood if over 25% of the result window's area is red.

3.4.4 RDT Result Interpretation. After cropping and enhancing the result window, RDTScan computes the average intensity value (L channel) for each row parallel to the expected orientation of the lines themselves. Those intensity values are plotted against their position along the result window to represent how the intensity varies along the underlying immunoassay (Figure 2). Lines are defined as local minima that are at least 60 units deep and cover at least 5% of the result window's main axis. Because the metadata includes the location of the different lines, RDTScan uses that information to both confirm line validity and to assign semantic meaning to them. In some cases, visual noise can manifest at the line locations in a way that creates a trough in the intensity plot. If the metadata includes the expected range of color hue values (H channel) for that line, RDTScan extracts the average hue of the row, and lines with an average hue outside of the expected range are rejected. Otherwise, RDTScan only checks the intensity values for result interpretation.

3.5 User Interface

Passing all of the quality assurance checks requires having vigilant end-users who are able to position their camera in the optimal manner during image capture. As shown in Figure 3, RDTScan provides an overlay that lies on top of a camera preview while end-users move their smartphone. This overlay includes a viewfinder with the same aspect ratio as the target RDT design and is scaled according to the metadata provided by the developer. RDTScan also displays a list of the quality checks that the most recently processed frame failed. By showing all of the checks simultaneously, end-users can identify regularly occurring or concurrent issues. At the same time, end-users may want to know what steps need to be taken to pass those checks and increase their chance of a successful capture. RDTScan provides human-readable instructions that describe how end-users should adjust their smartphone to capture the best image possible. Only one instruction is provided at a time, and the instructions are prioritized to minimize user effort. The feedback order is as follows:

- (1) **Image brightness and shadows:** Since end-users may need to change their environment to improve the ambient lighting, these instructions are prioritized first. If the image is overexposed and the flash is on, end-users are instructed to turn off the flash; likewise, end-users are instructed to turn on the flash if it is off and the image is underexposed. If the image is overexposed and the flash is already off, end-users are instructed to decrease the ambient lighting; likewise, end-users are instructed to increase ambient lighting if the flash is already on and the image is underexposed.
- (2) **Blur:** Image blur is often indicative of a person moving their camera, so it does not make sense to check the other characteristics of the camera frame until the camera is back in focus. If blur is detected, RDTScan simply notifies end-users that they should hold the phone still until the camera can refocus itself.

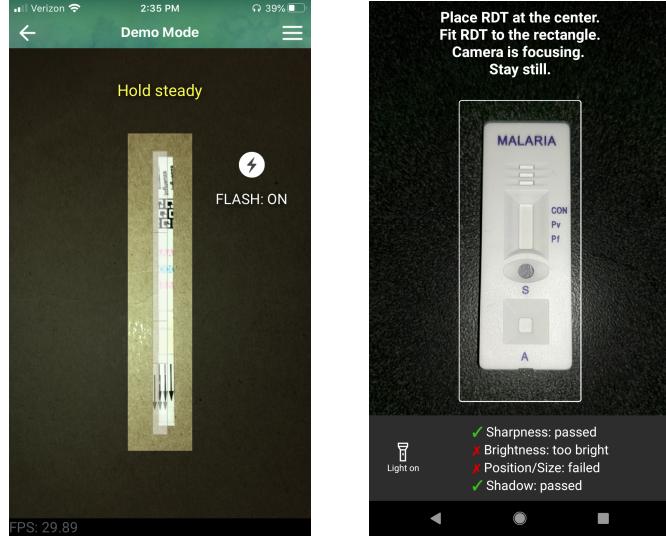


Fig. 3. Screenshots of the two apps created by our collaborators using RDTScan: (left) flu@home for the Australia Influenza Study and (right) RDT Open Reader for the Kenya Malaria Study.

- (3) **RDT size:** If the camera is at an improper distance from the RDT, RDTScan may not have enough information to detect an accurate bounding box around the RDT to correct its other properties. RDTScan instructs end-users to move their camera towards or away from the RDT depending on whether the RDT appears too small or too big relative to the viewfinder.
- (4) **RDT position and orientation:** Once the camera is at the proper height, translation and rotation suggestions are more likely to be accurate since all of the RDT's corners will be visible and well-spaced. RDTScan provides guidance for camera movements that are parallel to the RDT (i.e., translation and rotation) until the RDT is aligned with the viewfinder.
- (5) **Glare:** At this point, the position of the camera relative to the RDT should be nearly optimal. However, glare can be created by the flash or ambient lighting due to the angle of the camera relative to those light sources. If glare is detected, RDTScan instructs end-users to slightly tilt their smartphone at an angle until the glare disappears.

Once the current camera frame passes all of the quality checks, RDTScan automatically stores the image and advances so that the user does not have explicitly capture the photograph.

4 IN-LAB EVALUATION

Before we released RDTScan, we tested our interpretation algorithm in a controlled lab environment. This evaluation examined our algorithm's limit-of-detection: the lowest analyte concentration level at which a positive result can still be recovered from the RDT. We compared the limit-of-detection across different RDTs, analyte concentrations, smartphones, and lighting conditions according to three interpretation methods: inspection of the RDT directly, inspection of the RDT image, and automatic interpretation.

Table 2. A comparison of the four different interpretation methods for the (left) QuickVue and (right) QuickVue RDTs. Each box indicates the variance across lighting conditions. The horizontal lines indicate the limit-of-detection for the different interpretation methods.

QUICKVUE					CARESTART				
Dilution	DIRECT_READ	IMAGE_READ	ENHANCED_IMAGE_READ	ALGO_READ	Dilution	DIRECT_READ	IMAGE_READ	ENHANCED_IMAGE_READ	ALGO_READ
1:1	+	16/16	16/16	16/16	1:1	+	16/16	16/16	16/16
1:10	+	16/16	16/16	16/16	1:5	+	16/16	16/16	16/16
1:100	+	8/16	12/16	16/16	1:10	+	14/16	16/16	16/16
1:200	+	4/16	12/16	16/16	1:20	+	6/16	16/16	16/16
1:400	+	4/16	4/16	16/16	1:40	+	4/16	10/16	16/16
1:600	+	0/16	4/16	16/16	1:60	+	0/16	6/16	16/16
1:800	+	0/16	0/16	9/16	1:80	-	0/16	0/16	8/16
1:1000	-	0/16	0/16	0/16	1:100	-	0/16	0/16	0/16
0:1 (control)	-	0/16	0/16	0/16	1:200–1:1000	-	0/80	0/80	0/40
					0:1 (control)	-	0/16	0/16	0/16

4.1 Procedure

Three researchers with experience in administering RDTs created two dilution series using analyte proxies for the RDTs that would eventually be used in our case studies: malaria *Pv/Pf* protein mixed with deionized water for the Carestart RDTs and extracted influenza A protein for the QuickVue RDTs. The sample-to-buffer ratios for the CareStart series spanned 1:1–1:200, whereas the QuickVue series spanned 1:1–1:1000. These ranges were empirically selected so that a spectrum of line intensities would be created. To account for possible variables in the manufacturing of the immunoassays, the researchers used each analyte concentration to activate two RDTs. The researchers also activated two RDTs without any analyte to serve as negative controls.

The researchers first interpreted the physical tests under white fluorescent ambient lighting (DIRECT_READ). The researchers then took photos of the RDTs using an app built with RDTScan on two smartphones: a Google Pixel 1 and a Samsung Galaxy S8. The researchers took photos with both smartphones under four lighting sources: fluorescent light, warm white LED (2700 K), cool white LED (5000 K), and ambient light (distant fluorescent + sunlight). The automatic interpretation decisions returned by RDTScan were saved as the algorithm readings (ALGO_READ). The captured images were saved and then pushed through RDTScan’s detection and interpretation pipeline up until contrast enhancement to generate an enhanced image of the result window.

Later that day, the same three researchers interpreted the RDT results in the unmodified photographs (IMAGE_READ) so that it could be understood how much information was lost during the image capture process alone. After that, the researchers interpreted the enhanced images that were produced by RDTScan (ENHANCED_IMAGE_READ). For each interpretation method, the RDTs were shuffled and the researchers made their decisions independently to avoid any biases. At the end of this process, the researcher’s decisions were aggregated using majority consensus to form a single decision per RDT and per interpretation method.

To summarize, the four interpretation methods in order of increasing digitization and automation are DIRECT_READ, IMAGE_READ, ENHANCED_IMAGE_READ, and ALGO_READ. For each RDT brand, there were 2 DIRECT_READs, 16 IMAGE_READs (2 trials \times 2 smartphones \times 4 lighting conditions), 16 ENHANCED_IMAGE_READs, and 16 ALGO_READs.

4.2 Results

Inter-rater reliability was moderate amongst the three researchers according to Fleiss’ kappa ($\kappa = 0.69$) [28]. Table 2 compares the four different interpretation methods. Note that for all interpretation methods, the same decisions were reached for two RDTs at the same concentration. We consider DIRECT_READ to be the baseline for our analysis since the reported limits-of-detection for both tests are inapplicable to the analyte proxies we used. We compare interpretation methods using McNemar’s test [35] when there is a sufficient number of samples to do so (e.g., at least 25 samples with different results from the two methods); otherwise, we use an exact binomial test.

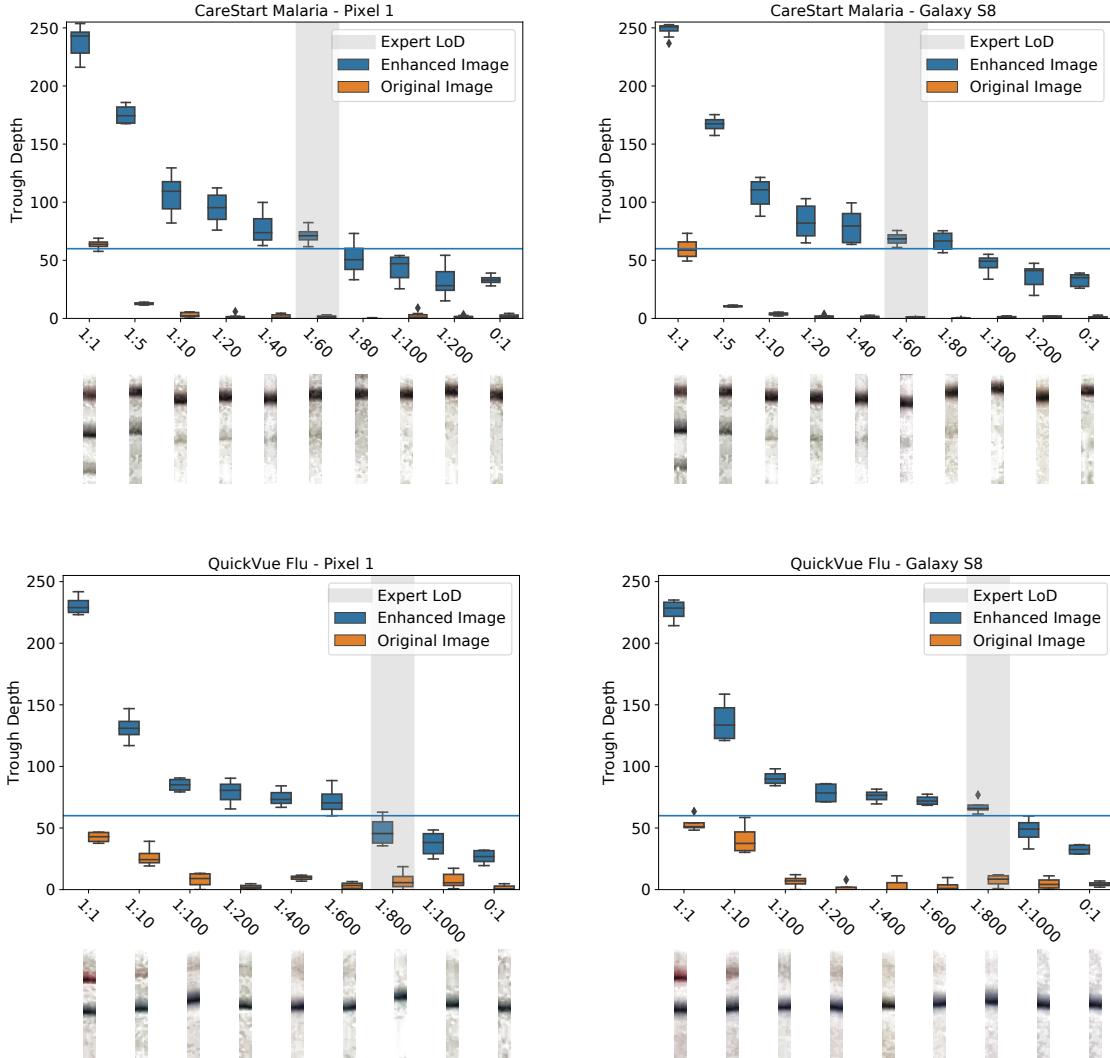


Fig. 4. The magnitude of the intensity troughs that were measured by RDTScan across the different concentrations. The results are split according to RDT brand (rows) and smartphone models (columns). The variance within each bar aggregates the results across different lighting conditions. Note that the horizontal axes are not to scale given that the respective tests have different sensitivities. Images on the x-axis are examples of contrast enhanced result lines used for interpretation.

For the QuickVue RDT, IMAGE_READ performed worse than DIRECT_READ ($\chi^2(1) = 54.018, p < .001$), highlighting the fact that camera sensors can diminish small details during digitization. ENHANCED_IMAGE_READ was also worse than DIRECT_READ ($\chi^2(1) = 46.021, p < .001$) but outperformed IMAGE_READ ($Bin(24, 0.5) = 0.000, p < .001$). Contrast enhancement emphasized subtle details, which lowered the limit-of-detection and increased sensitivity. ALGO_READ performed comparably to DIRECT_READ ($Bin(8, 0.5) = 0.000, n.s.$), showing that the combination of

contrast enhancement and an automated peak-finding algorithm were able to overcome the information that was lost when the RDT was converted into an image.

The different interpretation methods were more similar to one another for the CareStart RDT. As before, ENHANCED_IMAGE_READ detected more positive cases than IMAGE_READ ($\text{Bin}(24, 0.5) = 1.000, p < 0.01$). ALGO_READ was able to detect lower concentrations than DIRECT_READ, although not to a statistically significant degree ($\text{Bin}(8, 0.5) = 0.000, n.s.$). One might argue that the ALGO_READ positive results at the 1:80 dilution could have been RDTScan misinterpreting image noise as a line that was not actually present. If that was the case, however, RDTScan would have also produced positive results at even lower dilutions. Therefore, it is more likely that RDTScan was able to detect lines that researchers missed on the CareStart RDTs.

Figure 4 shows a box plot of the intensity differences at the RDTs' test line locations. As a reminder, RDTScan detects lines that have a trough depth greater than 60 units. The variance of the boxes captures the different trough depths that were measured across lighting conditions. As expected, the trough depth was large when the analyte concentration was high, and the depth decreased at lower concentration. Contrast enhancement not only significantly increased the trough depths to make faint lines visible, but also increased the resolution between different concentrations. In fact, contrast enhancement increased the trough depth of the negative controls since it intensified visual noise on the surface of the immunoassay. To avoid potentially classifying these cases as diagnostic positives, RDTScan's interpretation algorithm checks trough width (and potentially hue) to distinguish real lines from noise. There were no statistically significant differences in ALGO_READ decisions across the different lighting conditions or RDTs with the same concentration, as illustrated by the low variances in Figure 4. This result demonstrates the robustness of RDTScan across environments without the need for accessories or hardware.

5 FIELD EVALUATIONS

Although controlled lab testing was instrumental for characterizing RDTScan's interpretation limits, it only represents the upper limits of RDTScan's performance in the field. RDTScan automates much of the quality control process during image capture, but the fact that researchers with RDT image capture experience handled the smartphones during that study would be assumed to boost the quality of each image. Furthermore, analyte concentrations in a biological sample vary depending on the duration of the infection, the severity of the infection, the site from which the sample was taken, etc. Therefore, deploying RDTScan was the best way to assess its performance in real-world settings.

In this section, we present two field evaluations: at-home influenza testing in Australia and in-clinic malaria testing in Kenya. For each deployment, we describe how our collaborators modified RDTScan for their needs, the demographics and context of the participants, the study procedures, and the performance of RDTScan itself.

5.1 Australia Influenza Study

5.1.1 Motivation & Context. Influenza causes seasonal epidemics and occasional pandemics that impact the entire world. As with many other infectious diseases, it is believed that early detection can reduce transmission and improve the outcomes of patients who are at higher risk of complications [7]. At-home influenza testing could reduce the need for patients to travel to a healthcare facility for diagnosis confirmation, enhance the efficiency of healthcare delivery during busy influenza seasons, and even reduce the risk of transmission during a pandemic. In most countries, there are currently no diagnostic tests for influenza that have regulatory approval for use by untrained individuals outside of clinical settings. We partnered with a group of clinicians and smartphone app developers who were interested in assessing the potential efficacy of Quidel's QuickVue RDT for at-home influenza testing in Australia, where there are an estimated 310,000 general practitioner consultations and 18,000 hospital admissions related to influenza each year [40].

5.1.2 Study App Integration. The existing study protocol included a smartphone app called fluhome (Figure 3, left) for an optional home-testing component to a broader study on the spread of influenza in Australia. Our collaborators wanted their app to help participants take a clean photo of their QuickVue RDT so that the test's results could be compared to a clinical gold-standard with minimal reporting errors. There was great interest in the possibility of using an app for automatic interpretation to eventually remove that element of human error; however, there were ethical concerns about showing those results to end-users without regulatory approval. Therefore, their app ran RDTScan's automatic interpretation algorithm in the background and uploaded the resulting decision and image to a back-end server without showing the results to the end-users.

To mitigate participant frustration and ensure that RDTScan did not significantly impact the rest of the study protocol, we added a 30-second timeout for RDTScan within the fluhome app. The app automatically advanced if RDTScan did not successfully detect the participant's RDT during that time, regardless of whether the issue was due to human error or RDTScan's inability to detect the RDT. During pilot testing, we found that RDTScan occasionally had issues detecting the QuickVue RDT due to its thin and inconsistent design. Therefore, we utilized the optional fiducial checking step to ensure proper detection.

5.1.3 Deployment Site and Participants. Study participants were recruited from clinics participating in the Australian Sentinel Practices Research Network (ASPREN)¹³. Clinics within ASPREN share de-identified information on influenza-like illness and other infectious disease conditions as part of national surveillance efforts in Australia each year. Adult patients who visited a clinic participating in ASPREN while exhibiting influenza-like symptoms were invited by their general practitioner (GP) to participate in the study. Details about the study can be found in the protocol published by Lyon et al. [31]. In total, 299 participants enrolled in study (196 female, 103 male) with an average age of 41 ± 13.9 years.

5.1.4 Procedure. The study was conducted from July 2019 to February 2020; although the typical influenza season in Australia covers these same months, the peak of that particular season happened in July [21]. GPs obtained nasal swabs from participants during clinic visits, which were sent to a reference laboratory for reverse transcription quantitative polymerase chain reaction (RT-qPCR), serving as the gold-standard indicator of influenza status (PCR_RESULT). Participants who opted into the home-testing protocol were then sent home with a custom kit that included the materials needed to operate Quidel's QuickVue RDT and a link to download fluhome on their smart-device (Android or iOS smartphone or tablet). Participants were asked by their GP to complete the RDT using the instructions provided in the app. After completing the RDT, participants used the app to take a photograph of the dipstick. In the app, participants completed a survey to indicate if they saw a blue control line on their dipstick, as well as any red or pink test lines on their QuickVue RDT (DIRECT_READ). Usability was measured according to capture success rate within the 30-second timeout and the average capture time when participants had a successful capture. The captured images were analyzed in the background using RDTScan (ALGO_READ), and they were also later reviewed by an independent researcher who underwent a half-day training session on RDT interpretation to serve as a bronze-standard (EXPERT_IMAGE_READ).

5.1.5 Results: Usability. Table 3 summarizes the usability metrics across different platforms and age groups. Overall, participants were successful 83.3% of the time in capturing an image of their RDT using RDTScan within the 30-second timeout. When participants had a successful capture, they typically did so in under 10 seconds. Capture success rate was significantly higher amongst iOS users than Android users ($\chi^2(1) = 7.598, p < .01$), and capture time was significantly lower ($t(245) = -3.161, p < .01$) for iOS users. Although there may be confounding factors due to the typical demographics of Android and iOS users, we hypothesize that this can be partly attributed to the processing power of the operating systems themselves. SIFT operations are roughly 4× slower on Android than on iOS [10], and Android throttles the CPU more aggressively than iOS; together, these differences cause

¹³<http://www.aspren.com.au/>

Table 3. A summary of RDTScan's usability during the Australia Influenza Study. Capture success rate quantifies how often participants were able to use RDTScan to get a high-quality photograph of their RDT within the 30-second timeout, while capture time measures the average time it for a successful capture.

Mobile OS	N	Capture Success Rate	Capture Time (s)
iOS	175	88.0%	8.73
Android	124	75.8%	11.52
Age	N	Capture Success Rate	Capture Time (s)
≤ 29	71	90.1%	9.06
30–39	65	96.9%	9.71
40–49	53	81.1%	9.56
50–59	52	72.6%	10.62
≥ 60	34	58.8%	11.00
Total	299*	83.3%	9.77

*A total of 275 participants reported their age, while the rest chose to not disclose that information.

a lower effective frame rate on Android and thus fewer cycles for real-time feedback. Furthermore, there is a broader diversity of Android device specifications, some of which are far inferior to the typical iPhone for the sake of lower prices. There were also noteworthy differences in capture performance between age groups. The capture success rate amongst participants under the age of 40 was greater than 90%, whereas the older participants had far more difficulties. We suspect that younger participants had more success using RDTScan because of their greater familiarity with smartphones and camera-based apps.

5.1.6 Results: Accuracy. A total of 249 images passed all of RDTScan's quality checks and were subsequently interpreted. Five of those images had inconclusive results, meaning that no control line appeared within the result window. For the remaining 244 images, we compared ALGO_READ against EXPERT_IMAGE_READ and PCR_RESULT. Table 4 shows the confusion matrices for both comparisons. When compared against the bronze-standard EXPERT_IMAGE_READ, RDTScan achieved 97.5% accuracy (sensitivity = 85.7%, specificity = 98.7%). When compared against the gold-standard PCR_RESULT, RDTScan achieved 83.6% accuracy (sensitivity = 33.9%, specificity = 98.4%). In evaluating how the other RDT interpretation methods compared to PCR_RESULT, DIRECT_READ had 76.6% accuracy (sensitivity = 35.2%, specificity = 91.2%), while EXPERT_IMAGE_READ had 82.6% accuracy (sensitivity = 32.9%, specificity = 100%). RDTScan demonstrated better accuracy than the novice end-users ($\chi^2(1) = 28.195, p < .001$) and comparable accuracy to the expert ($Bin(14, 0.5) = 6.000, n.s.$) when those readings were compared to PCR_RESULT. RDTScan and the other RDT interpretation methods had poor sensitivity when compared to the gold-standard. PCR_RESULT was generated from nasal swabs collected by trained GPs, whereas the other three readings were generated from self-obtained nasal swabs that were collected by the study participants a few days later. Therefore, the low sensitivity of the QuickVue RDT can be attributed to poor swabbing technique, lower viral load at the time of testing, or improper RDT administration at home. Even when all of the instructions have been properly followed, researchers have witnessed low sensitivity with the QuickVue RDT. For instance, Agoritsas et al. [1] found that QuickVue RDTs had a sensitivity of 78% in a study where nasal swabs were collected from pediatric patients by research nurses. With our collaborator's study design, we are unable to separate out swabbing and test sensitivity issues.

Table 4. Confusion matrices showing the performance of RDTScan against (left) expert interpretation of RDT images and (right) RT-qPCR for the Australia Influenza Study. As points of comparison, DIRECT_READ and EXPERT_IMAGE_READ achieved 76.6% and 82.6% accuracy when compared to PCR_RESULT, respectively.

ALGO_READ	EXPERT_IMAGE_READ		PPV = 85.7%	NPV = 98.7%	ALGO_READ	PCR_RESULT		PPV = 86.4%	NPV = 83.3%
	Pos (+)	Neg (-)				Pos (+)	Neg (-)		
Pos (+)	18	3			Neg (-)	19	3		
Neg (-)	3	220				37	185		
SNS = 85.7%		SPC = 98.7%		ACC = 97.5%		SNS = 33.9%		SPC = 98.4%	

SNS = sensitivity, SPC = specificity, ACC = accuracy, PPV = positive predictive value, NPV = negative predictive value

5.2 Kenya Malaria Study

5.2.1 Motivation & Context. The World Health Organization estimates that there were 228 million cases of malaria in 2018 [57]. Malaria is endemic to many regions that have limited resources [16], so populations within those regions rely on community health workers (CHWs) to serve their healthcare needs. CHWs often travel from patient-to-patient, making RDTs a convenient point-of-care solution for malaria diagnosis. CHWs can have competing demands on their time, leading to occasional mistakes and rushed decisions [49]. We partnered with a Kenya-based software company who wanted to support CHWs in the field during their malaria testing duties with the CareStart RDT.

5.2.2 Study App Integration. The protocol for this deployment included a smartphone app called RDT Open Reader, which is designed to provide CHWs with instructions as they administer CareStart RDTs. Similar to the Australia Influenza Study, RDTScan was used to document RDT completion and the automatic interpretation results were hidden from the CHWs. We applied the same 30-second timeout to ensure that RDTScan did not significantly impact CHWs' typical workflow. Pilot testing revealed that our detection algorithm was sufficient for detecting the CareStart RDT without incorporating any of the optional checks (e.g., fiducials, line hues).

5.2.3 Deployment Site and Participants. The study was conducted at a health clinic in Kilifi, Kenya from December 2019 to March 2020. Six CHWs (5 male, 1 female) were responsible for administering RDTs and using the RDT Open Reader app (Figure 3, right). All of the CHWs had at least ten years of job experience, which included administering and interpreting RDTs. One CHW was in their 40s, while the rest were in their 30s. The CHWs shared four smartphones that were provided by the clinic: an Oppo A3s, an Infinix Hot 7, a Samsung Galaxy A10s, and a TECNO Pouvoir 3 Air.

5.2.4 Procedure. Participating CHWs were trained to use the RDT Open Reader app during two half-day training sessions in a classroom setting, a half-day session in the field, and then a half-day refresher course a few days later. The CHWs were instructed to use the app whenever they provided services to patients presenting with malaria-like symptoms. After administering a CareStart RDT, the CHWs used the app to capture a photograph of it. The CHWs were not consistently supervised to avoid disrupting their duties. The CHWs administered and inspected RDTs in various location in clinic due to the mixed use of facilities, which means that the CHWs interacted with RDTScan in various lighting conditions and background settings. The CHWs were asked to record their interpretation of the physical RDT (DIRECT_READ), which was later compared to the results of RDTScan (ALGO_READ) and whole blood samples that were later processed through RT-qPCR (PCR_RESULT). Similar to the

Table 5. A summary of RDTScan's usability during the Kenya Malaria Study. Capture success rate quantifies how often participants were able to use RDTScan to get a high-quality photograph of their RDT within the 30-second timeout, while capture time measures the average time it took for a successful capture.

Smartphone Models	N	Capture Success Rate	Capture Time (s)
Oppo A3s	120	99.2%	8.54
Infinix Hot 7	38	63.2%	10.99
Samsung Galaxy A10s	113	90.3%	6.95
TECNO Pouvoir 3 Air	62	96.8%	9.12
Total	332	90.4%	8.38

Australia Flu Study, usability was measured according to capture success rate within the 30-second timeout and the average capture time for a successful capture.

5.2.5 Results: Usability. Table 5 summarizes capture performance across different smartphone models. Since CHWs shared smartphones within the clinic and they were often under considerable time pressure to perform their duties, we were unable to separate usability measures between individuals. Overall, the six CHWs were successful 90.4% of the time in capturing an image of their RDT using RDTScan within the 30-second timeout. When the CHWs had a successful capture, they typically did so in around 8.3 seconds. The CHWs had the least success with the Infinix Hot 7 smartphone ($\chi^2(1) = 47.131, p < .001$), which can be attributed to its inferior processing power relative to the other devices (1.3 GHz vs. 1.8–2.0 GHz CPU). In fact, when the smartphone models are sorted according to their processing power, that ordering aligns with the ranking of average capture time. As mentioned before, lower processing power results in a slower effective frame rate, which leads to fewer opportunities for user feedback. The Infinix Hot 7 also does not fully support Android's Camera2 API, but rather legacy support for camera hardware control. This limited support could have affected RDTScan's ability to adjust the exposure and focus while the CHWs captured images using the Infinix Hot 7 phone.

5.2.6 Results: Accuracy. A total of 300 images passed all of RDTScan's quality checks and were subsequently interpreted. Due to mismatches between RDTs that were successfully captured using RDTScan and samples that could be processed by RT-qPCR, 228 images were used for comparison between ALGO_READ and PCR_RESULT. Table 6 shows the confusion matrices for both comparisons. When compared to DIRECT_READ, RDTScan achieved 96.3% accuracy (sensitivity = 95.5%, specificity = 98.7%). When compared against PCR_RESULT, RDTScan achieved 85.5% accuracy (sensitivity = 92.9%, specificity = 65.0%). As a point of comparison, DIRECT_READ compared against PCR_RESULT had 84.6% accuracy (sensitivity = 93.5%, specificity = 60.0%). RDTScan demonstrated comparable accuracy to the CHWs ($Bin(4, 0.5) = 1.000, n.s.$). There were 3 false positive cases when the CHWs saw lines that did not exist according to PCR_RESULT. RDTScan was able to correctly interpret these cases as true negatives, showing that RDTScan may be particularly helpful in situations when clinical personnel are asked to administer many RDTs under time pressure. Similar to the QuickVue RDT, the CareStart RDT had significantly lower specificity relative to RT-qPCR. Unlike in the Australia Influenza Study, however, the CHWs had far more experience administering RDTs than the first-time end-users. Instead, the discrepancy may be due to the fact that RT-qPCR is a molecular-based technique that undergoes stages of processing to detect the presence of the pathogen genome, while the RDT mainly works to detect immune reactive elements like antigens. Most of the false positives between ALGO_READ and PCR_RESULT had clearly visible test lines, so the CareStart RDT itself may have had low specificity. Similar findings have been reported in previous work [39, 51], which found that CareStart RDTs have a specificity around 80% in the field.

Table 6. Confusion matrices showing the performance of RDTScan against (left) expert interpretation of RDT images and (right) RT-qPCR for the Kenya Malaria Study. As a point of comparison, DIRECT_READ achieved 84.6% accuracy when compared to PCR_RESULT.

ALGO_READ	DIRECT_READ		PCR_RESULT	
	Pos (+)	Neg (-)	Pos (+)	Neg (-)
Pos (+)	212	1	PPV = 99.5%	
Neg (-)	10	77	NPV = 88.5%	
	SNS = 95.5%	SPC = 98.7%	ACC = 96.3%	
				SNS = 92.9% SPC = 65.0% ACC = 85.5%

SNS = sensitivity, SPC = specificity, ACC = accuracy, PPV = positive predictive value, NPV = negative predictive value

5.3 Comparison across the Two Case Studies

The two case studies were conducted in vastly different situations, cultures, and environments. In this section, we highlight some of the insights that can be gleaned from comparing and contrasting the studies, and we also talk about some of the protocol decisions that may guide future deployments of RDTScan.

5.3.1 RDT Form Factor. The robustness of our RDT detection approach depends on the appearance of consistent visual keypoints around the result window. Template images with an even aspect ratio (i.e., as wide as they are tall) are less susceptible to localization errors since they are more likely to have keypoints surrounding the result window. As such, the initial version of RDTScan was able to support the CareStart RDT because of the unique lettering and markings throughout its design. The thin and inconsistent design of the QuickVue RDT, on the other hand, posed greater challenges. We overcame these issues by introducing the notion of fiducial detection in RDTScan (Section 3.3.7), but the lower capture success rate we observed in the Australia Flu Study may still be partly attributable to the QuickVue RDT’s design. The contrasting form factors also highlighted the utility of different functions supported by RDTScan. Blood can seep into the result window of the CareStart RDT if it is not administered properly, and glare can appear on the result window because of the glossy film that protects its result window. In the Kenya Malaria Study, RDTScan correctly rejected 15 images because of blood and 6 images because of glare, and none of the remaining images had enough blood or glare to obscure results. Meanwhile, the use of color information to validate control and test lines played a critical role in avoiding false positives since RDT detection was less robust for the QuickVue RDTs.

5.3.2 Usability-Accuracy Tradeoff. The guiding principle of RDTScan is that the higher the quality of the RDT image being analyzed, the more accurate RDTScan can be with automatic interpretation results. In other words, loose quality checks lead to better usability and lower accuracy, while tight quality checks lead to worse usability and higher accuracy. We arrived at the thresholds described in this paper through our prior work [45] and conversations with our collaborators about their slight preference for high accuracy over usability. Both studies used the same settings when possible to avoid further confounds in evaluating RDTScan. In a real-world deployment, however, the prioritization of usability versus accuracy depends on the context in which RDTScan is being used. For example, developers may want to prioritize accuracy when RDTScan is being used as the sole mechanism for diagnosis. On the other hand, developers may want to prioritize usability if RDTScan is being used as a screening tool to identify individuals in a diverse population who need a more expensive clinical test. Space precluded us from presenting a complete grid search of the numerous thresholds and quality checks within RDTScan, and their impact on the user experience, so we leave that to future work. However, researchers and

developers may want to explore adaptive settings for RDTScan’s quality checks so that they become less strict over time. The 30-second timeout implemented in both study apps is an extreme case of an adaptive setting, removing all quality checks at the end of the timer for the sake of capturing a picture.

5.3.3 Experience. In our prior work [45], we found that CHWs became more adept at using RDTScan over time. Participants in the Australia Flu Study only used RDTScan once, while the CHWs in the Kenya Malaria Study used RDTScan multiple times; the practice that CHWs’ gained from repeatedly using RDTScan may explain the superior usability results in the latter study. Beyond people’s familiarity with RDTScan, familiarity with smartphones in general also could have driven the differences we saw in usability metrics. Studies have shown that the elderly face many obstacles when learning new features on their smartphones [4, 34], and our results from the Australia Flu Study reveal that younger participants had better success using RDTScan than older participants.

5.4 Comparison to Previous Version of RDTScan

The deployment from our prior work (Mali Malaria Study) was similar to the Kenya Malaria Study. Although the studies involved different RDT brands, both were conducted in sub-Saharan Africa with CHWs as study participants. Furthermore, the CHWs in both studies used low-end smartphones to capture cassette-based RDTs. In light of these similarities, RDTScan’s improved RDT detection algorithm led to significantly better usability statistics in the more recent study. The capture success rate rose from 67% to 90.4%, and the average capture time went down from 14.3 seconds to 8.38 seconds. The interpretation accuracy metrics were much more similar across the two studies. The original version of RDTScan had 98.7% accuracy relative to expert readings in the Mali Malaria Study. Meanwhile, the updated version of RDTScan had 96.3% accuracy. The slight reduction in accuracy could be attributed to the fact that the baseline for the Mali Malaria Study came from the consensus of two lab technicians, while the baseline for the Kenya Malaria Study came from one of six CHWs. For a fairer comparison, we re-ran the original RDTScan interpretation algorithm on the images from the Kenya Malaria Study. We found that the original algorithm would have accepted 21 images that were rejected by the updated version of RDTScan—6 because of glare and 15 because of blood. By letting those images be interpreted, the original version of RDTScan would have been 6.3% less accurate. We also compared the two versions of RDTScan’s interpretation algorithm on the images from the Australia Flu Study. The updated version of RDTScan uses color information to validate interpretation results, which led to a 7.4% accuracy improvement from the previous version. This performance boost came from improving RDTScan’s ability to localize the QuickVue RDT in 18 cases.

6 DISCUSSION

RDTs have already made significant impact on healthcare diagnostics because of their low cost and relevant application to various medical conditions and diseases. As RDTs become approved for use by broader populations, their roles will likely expand even further. We believe that RDTScan will significantly contribute to the expanding role of RDTs, but there is more research to be done to this end. Below, we describe the limitations of our approach and potential areas of future exploration.

6.1 Expanding RDTScan to New RDT Designs

We chose to use feature-based template matching for RDT detection because it requires a single template image and some simple metadata for each new RDT design. Although this requirement has significantly less overhead than the large datasets needed to support a machine learning model, our approach is not optimal in all cases. In order to accurately localize an RDT, RDTScan relies on the assumption that unique visual keypoints occur naturally in RDT designs (e.g., logos, labels, and arrows). To assess the validity of this assumption, we conducted an informal study to determine if RDTScan could reliably detect RDTs from other manufacturers. The images of

the RDTs we examined for this informal study can be found in Appendix A. We applied RDTScan’s detection algorithm to RDTs from two lists curated by major health agencies—one by the World Health Organization for malaria RDTs [41] and one by the Centers for Disease Control for influenza RDTs [14]. Out of the 12 RDTs in those lists, 10 were cassettes and 2 were dipsticks. RDTScan was able to detect 8 of the 10 cassettes, all without the need for fiducial checking; both failure cases were blank white cassettes without any lettering or logos. Thresholding an image based on the cassette’s color would be the simplest way of handling such RDT designs, but that would also require end-users to capture their RDT on a distinct background. RDTScan was able to detect both of the dipsticks provided that fiducial checking was enabled. We also examined 3 more pregnancy test dipsticks and found that RDTScan was able to support all of them with fiducial checking. As an alternative to fiducial checking, we considered leveraging multiple templates to cover the range of possible markings; however, that approach would have incurred significantly more computation per frame and thus would have affected the user experience.

A data-driven recognition model may eventually become the ideal approach to RDT detection if RDT designs become more diverse and deep learning on smartphones becomes more computationally efficient. Training such a model will require a large training dataset of RDT images, so we hope that RDTScan can accelerate the creation of such datasets in the future.

6.2 Supporting Developers

RDTScan⁵ is publicly available as an open-source codebase, and we have witnessed other global health organizations utilize and extend RDTScan to suit their needs since its release. Most notably, we have seen organizations leverage RDTScan for COVID-19 RDTs that are being developed worldwide. Because such RDTs are being actively developed and evaluated, one organization is using RDTScan to support the automatic interpretation of 10 different brands of COVID-19 RDTs. RDTScan’s ability to rapidly accommodate new RDT designs without the need for extensive data collection has been critical to this end.

Beyond the fact that incorporating RDTScan into a smartphone app requires mobile programming experience, using RDTScan also requires familiarity with tools that would enable a person to extract metadata for their target RDT. These tools include a mobile scanner app (e.g., Microsoft Lens) and a photo editing program (e.g., Microsoft Paint), both of which can be downloaded for free. Using these tools does not require programming experience since they have GUIs, and we include detailed instructions on this process for future developers on the RDTScan website. We were able to coach project managers in our case studies to generate this metadata for themselves within a phone call. Nevertheless, we recognize that a smoother process for expanding RDTScan to new RDT designs would make RDTScan even easier for developers to use. We envision a software tool that could help developers more easily extract metadata from their template RDT image, whether through a domain-specific user interface or automatic image analysis. The automatic image analysis approach would share many of the same characteristics as RDTScan’s interpretation algorithm, but the tool would be able to take advantage of assumptions about the template to make the analysis easier, such as the fact that all of the lines will be activated.

6.3 Standards for RDT Design

Because the concept of RDTs predates that of computer vision, RDTs are not typically designed with image processing in mind. After demonstrating that an unmodified smartphone can be used to improve the effective accuracy of their product in the field, we have begun to collaborate with RDT manufacturers to create a standard RDT design that is optimized for image processing. This prototypical RDT design includes a flat, wide cassette that minimizes shadows and provides sufficient tolerance for RDT localization. To eliminate the need of an RDT template image, our RDT prototype includes consistent fiducials at each corner of the cassette to simplify the detection of the RDT’s edges, similar to what is done with the detection of QR codes [13]. In fact, we envision each RDT can be printed with a QR code that includes the same metadata needed to initialize RDTScan, such as

the location and meaning of the different lines along the result window. A thick colored border around the edge of the RDT is included for further correcting any errors in result window localization. Finally, our prototype also includes color references so that the effects of ambient lighting can be removed through color-calibration. RDTScan would still have useful functionality for an RDT with all of these features, such as its result interpretation algorithm and many of its quality assurance checks.

6.4 RDT Procedure Controls

RDTScan only checks the end result of the RDT, but that result is the culmination of an entire procedure that often goes unchecked. RDTs are prone to many procedural errors: the use of expired tests, sampling and handling errors of the biological sample, and improper fluid usage, to name a few [15, 20]. Some health organizations have used periodic supervision [36] and training programs [20] to educate their workforce on proper RDT administration, but such measures are only possible for repeated RDT users like CHWs. Instead, we believe that image processing and computer vision can be used as a proxy supervisor of the RDT procedure. For example,

- Optical character recognition can be used to confirm that the expiration date has not passed,
- Color blob detection can be used to check that enough blood has been drawn into a capillary tube, and
- Optical flow can be used to ensure that the user has swirled their nasal swab in a tube for enough time.

Implementing procedural controls will incur additional power consumption on the smartphone, so adding these checks will require a balance of utility and simplicity.

6.5 RDTScan’s Long-Term Efficacy

We demonstrated that RDTScan provides accurate image capture and interpretation performance when used by people with varying levels of technological and clinical expertise. However, many solutions that rely on mobile apps to address human-centered issues (e.g., health, finance, education) have failed due to the complexity of workflows and relationships, particularly in low- and middle-income settings [52]. Longitudinal deployments in which people are shown the automatic results generated by RDTScan would be needed to assess how RDTScan could be integrated into people’s workflows and the level of trust that people are willing to place in an automated system like RDTScan. As a direction of future work, we plan on deploying RDTScan within community health clinics and using both naturalistic observations and interviews to capture people’s experience with the system.

7 CONCLUSION

RDTs are becoming an increasingly popular option for point-of-care medical testing. To complement this trend in healthcare, we have created a mobile system that enables RDT detection and interpretation with a smartphone. We demonstrated that template-based feature matching is a feasible approach to RDT interpretation, and we argue that our approach is easier to scale to new RDT designs than a data-driven one. We also showed that RDTScan provides interpretation accuracy levels that are comparable to experts reading the RDTs themselves, opening the door to new workflows for at-home and in-clinic scenarios, particularly in low-resource settings. RDTScan is an open-source library, so we encourage public health organizations to utilize our system for their RDT-related needs. We anticipate that RDTScan can reveal novel insights and support new systems in public health, and we hope that this work inspires engineers and researchers to create software libraries that support the interpretation of other analog devices.

ACKNOWLEDGMENTS

Funding for this study was provided by the Bill & Melinda Gates Foundation (Grant OPP1184006, OPP1203115). The Australia flu@home study is a sub-study within the Seattle Flu Study, which is funded by Gates Ventures; the funder was not involved in the design of the study and does not have any ownership over the management

and conduct of the study, the data, or the rights to publish. ASPREN is supported by the Australian Government Department of Health. This manuscript is published with the permission of the Director of KEMRI. The opinions expressed in this paper are those of the authors, and do not necessarily represent the views of the funding entities or the RDT manufacturers.

For the Australia Influenza Study, we would like to express our gratitude to the numerous general practitioners, nurse practitioners, and practice staff who participate in ASPREN. We would also like to thank Daniel Blakeley and Lana Allen for their administrative assistance, as well as John Tamerius and Andy Pal of Quidel for their guidance on RDT administration and activation. Finally, we would like to thank the digital health nonprofit Audere for leading the app development effort for the broader Australia Flu Study protocol and for their data curation efforts. For the Kenya Malaria Study, we would like to express our gratitude to the Pingilikani Community and the Pingilikani Dispensary Administration, and the CHWs for facilitating and executing the study protocol. Specifically, we would like to thank Samuel Shuma, Daudi Shume, Jemimah Pole, Oscar Deche, Oscar Chidui, and Kennedy Genya for enrolling participants and collecting data. We would also like to thank Edward Otieno for managing app updates for the team and Margaret Chifwete for her help with laboratory testing. Lastly, we would like to thank Arunan Skandarajah from the Bill & Melinda Gates Foundation for his oversight and feedback throughout this work.

REFERENCES

- [1] Konstantinos Agoritsas, Kathy Mack, Bema K Bonsu, Debbie Goodman, Douglas Salamon, and Mario J Marcon. 2006. Evaluation of the Quidel QuickVue test for detection of influenza A and B viruses in the pediatric emergency medicine setting by use of three specimen collection methods. *Journal of clinical microbiology* 44, 7 (2006), 2638–2641.
- [2] Chukwuemeka CA Azikiwe, C C Ifezulike, Iyeopu M Siminialayi, Louis U Amazu, J C Enye, and O E Nwakwunite. 2012. A comparative laboratory diagnosis of malaria: Microscopy versus rapid diagnostic test kits. *Asian Pacific Journal of Tropical Biomedicine* 2, 4 (apr 2012), 307–310. [https://doi.org/10.1016/S2221-1691\(12\)60029-X](https://doi.org/10.1016/S2221-1691(12)60029-X)
- [3] BankMyCell. 2020. How Many Smartphones Are In The World? , 26 pages. <https://www.bankmycell.com/blog/how-many-phones-are-in-the-world>
- [4] Anabela Berenguer, Jorge Goncalves, Simo Hosio, Denzil Ferreira, Theodoros Anagnostopoulos, and Vassilis Kostakos. 2017. Are Smartphones Ubiquitous?: An in-depth survey of smartphone adoption by seniors. *IEEE Consumer Electronics Magazine* 6, 1 (jan 2017), 104–110. <https://doi.org/10.1109/MCE.2016.2614524>
- [5] L. Bissonnette and M. G. Bergeron. 2010. Diagnosing infections—current and anticipated technologies for point-of-care diagnostics and home-based testing. , 1044–1053 pages. <https://doi.org/10.1111/j.1469-0691.2010.03282.x>
- [6] Scott Carter, John Adcock, John Doherty, and Stacy Branham. 2010. NudgeCam: Toward targeted, higher quality media capture. In *MM'10 - Proceedings of the ACM Multimedia 2010 International Conference*. 615–618. <https://doi.org/10.1145/1873951.1874034>
- [7] Centers for Disease Control and Prevention. 2016. Influenza Signs and Symptoms and the Role of Laboratory Diagnostics. <https://www.cdc.gov/flu/professionals/diagnosis/labrolesprocedures.htm>
- [8] Francine Chen, Scott Carter, Laurent Denoue, and Jayant Kumar. 2013. SmartDCap: Semi-automatic capture of higher quality document images from a smartphone. In *International Conference on Intelligent User Interfaces, Proceedings IUI*. 287–296. <https://doi.org/10.1145/2449396.2449433>
- [9] F Chiadmi, J Schlatter, B Mounkassa, P Ovetchkine, and N Vermerie. 2004. Fast diagnostic tests in the management of group A beta-hemolytic streptococcal pharyngitis. *Annales de biologie clinique* 62, 5 (sep 2004), 573–577.
- [10] Claudiu Cobârzan, Marco A. Hudelist, Klaus Schoeffmann, and Manfred Jürgen Primus. 2015. Mobile Image Analysis: Android vs. iOS. In *MultiMedia Modeling*, Xiangjian He, Suhuai Luo, Dacheng Tao, Changsheng Xu, Jie Yang, and Muhammad Abul Hasan (Eds.). Springer International Publishing, Cham, 99–110.
- [11] Nicola Dell and Gaetano Borriello. 2013. Mobile tools for point-of-care diagnostics in the developing world. In *Proc. DEV '13*. ACM Press, New York, New York, USA, 1–10. <https://doi.org/10.1145/2442882.2442894>
- [12] Nicola Dell, Ian Francis, Haynes Sheppard, Raiva Simbi, and Gaetano Borriello. 2014. Field evaluation of a camera-based mobile health system in low-resource settings. In *Proc. MobileHCI '14*. 33–42. <https://doi.org/10.1145/2628363.2628366>
- [13] Denso ADC. 2011. *QR Code Essentials*. Technical Report. 1–12 pages. <http://www.nacs.org/LinkClick.aspx?fileticket=D1FpVAvvJuo=&tabid=1426&mid=480>
- [14] Centers for Disease Control and Prevention. [n.d.]. <https://www.cdc.gov/flu/professionals/diagnosis/table-ridt.html>

- [15] Philippe Gillet, Marcella Mori, Jef Van Den Ende, and Jan Jacobs. 2010. Buffer substitution in malaria rapid diagnostic tests causes false-positive results. *Malaria Journal* 9, 1 (jul 2010), 215. <https://doi.org/10.1186/1475-2875-9-215>
- [16] Douglas Gollin and Christian Zimmermann. 2007. Malaria: Disease Impacts and Long-Run Income Differences. *Department of Economics Working Paper Series* 2997 (2007), 33. <https://doi.org/10.1126/science.1061573>
- [17] Nikolai Gorski, Valery Anisimov, Emmanuel Augustin, Olivier Baret, and Sergey Maximov. 2001. Industrial bank check processing: The A2iA CheckReader™. *International Journal on Document Analysis and Recognition* 3, 4 (2001), 196–206. <https://doi.org/10.1007/PL00013561>
- [18] N. Gorski, V. Anisimov, E. Augustin, O. Baret, D. Price, and J. C. Simon. 1999. A2iA Check Reader: A family of bank check recognition systems. In *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*. IEEE Computer Society, 527–530. <https://doi.org/10.1109/ICDAR.1999.791840>
- [19] GSMA. 2020. *The Mobile Economy Sub-Saharan Africa*. Technical Report. <https://www.gsma.com/mobileeconomy/sub-saharan-africa/>
- [20] Steven A Harvey, Larissa Jennings, Masela Chinyama, Fred Masaninga, Kurt Mulholland, and David R Bell. 2008. Improving community health worker use of malaria rapid diagnostic tests in Zambia: package instructions, job aid and job aid-plus-training. *Malaria Journal* 7, 1 (dec 2008), 160. <https://doi.org/10.1186/1475-2875-7-160>
- [21] Nigel Hawkes. 2019. Flu: Australia sees early start to season. *BMJ (Clinical research ed.)* 366 (jul 2019), l4603. <https://doi.org/10.1136/bmj.l4603>
- [22] Sócrates Herrera, Andrés F Vallejo, Juan P Quintero, Myriam Arévalo-Herrera, Marcela Cancino, and Santiago Ferro. 2014. Field evaluation of an automated RDT reader and data management device for Plasmodium falciparum/Plasmodium vivax malaria in endemic areas of Colombia. *Malaria Journal* 13, 1 (mar 2014), 87. <https://doi.org/10.1186/1475-2875-13-87>
- [23] Shichu Huang, Koji Abe, Steven Bennett, Tinny Liang, Paula D. Ladd, Lindsay Yokobe, Caitlin E. Anderson, Kamal Shah, Josh Bishop, Mike Purfield, Peter C. Kauffman, Sai Paul, AnneMarie E. Welch, Bonnie Strelitz, Kristin Follmer, Kelsey Pullar, Luis Sanchez-Erebia, Emily Gerth-Guyette, Gonzalo Domingo, Eileen Klein, Janet A. Englund, Elain Fu, and Paul Yager. 2017. Disposable Autonomous Device for Swab-to-Result Diagnosis of Influenza. *Analytical Chemistry* 89, 11 (jun 2017), 5776–5783. <https://doi.org/10.1021/acs.analchem.6b04801>
- [24] L. Huette, P. Barbosa-Pereira, O. Bougeois, J. V. Moreau, B. Plessis, P. Courtellemont, and Y. LeCourtier. 1997. Multi-Bank Check Recognition System: Consideration on the Numeral Amount Recognition Module. 133–156. https://doi.org/10.1142/9789812797681_0006
- [25] Jan Jacobs, Barbara Barbé, Philippe Gillet, Michael Aidoo, Elisa Serra-Casas, Jan Van Erps, Joelle Daviaud, Sandra Incardona, Jane Cunningham, and Theodoor Visser. 2014. Harmonization of malaria rapid diagnostic tests: Best practices in labelling including instructions for use. *Malaria Journal* 13, 1 (dec 2014), 505. <https://doi.org/10.1186/1475-2875-13-505>
- [26] Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P. Bigham. 2011. Supporting blind photography. In *ASSETS'11: Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*. 203–210. <https://doi.org/10.1145/2049536.2049573>
- [27] Eleni Kakalou, Vasileios Papastamopoulos, Panagiotis Ioannidis, Kostas Papanikolaou, Ourania Georgiou, and Athanasios Skoutelis. 2014. Early HIV diagnosis through use of rapid diagnosis test (RDT) in the community and direct link to HIV care: a pilot project for vulnerable populations in Athens, Greece. *Journal of the International AIDS Society* 17 (nov 2014), 19619. <https://doi.org/10.7448/ias.17.4.19619>
- [28] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159–174. <http://www.jstor.org/stable/2529310>
- [29] Stefan Leutenegger, Margarita Chli, and Roland Y Siegwart. 2011. BRISK: Binary robust invariant scalable keypoints. In *2011 International conference on computer vision*. Ieee, 2548–2555.
- [30] David G. Lowe. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 2 (nov 2004), 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94> arXiv:cs/0112017
- [31] Victoria Lyon, Monica Zigman Suchsland, Monique Chilver, Nigel Stocks, Barry Lutz, Philip Su, Shawna Cooper, Chunjong Park, Libby Rose Lavitt, Alex Mariakakis, et al. 2020. Diagnostic accuracy of an app-guided, self-administered test for influenza among individuals presenting to general practice with influenza-like illness: study protocol. *BMJ open* 10, 11 (2020), e036298.
- [32] David Mabey, Rosanna W. Peeling, Andrew Ustianowski, and Mark D. Perkins. 2004. Diagnostics for the developing world. , 231–240 pages. <https://doi.org/10.1038/nrmicro841>
- [33] Christine Manyando, Eric M Njunju, Justin Chileshe, Seter Sizya, and Clive Shiff. 2014. Rapid diagnostic tests for malaria and health workers' adherence to test results at health facilities in Zambia. *Malaria Journal* 13, 1 (may 2014), 166. <https://doi.org/10.1186/1475-2875-13-166>
- [34] Ronald E. McGaughey, Steven M. Zeltmann, and Mark E. McMurtrey. 2013. Motivations and obstacles to smartphone use by the elderly: Developing a research framework. *International Journal of Electronic Finance* 7, 3-4 (2013), 177–195. <https://doi.org/10.1504/IJEF.2013.058601>
- [35] Quinn McNemar. 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12, 2 (jun 1947), 153–157. <https://doi.org/10.1007/BF02295996>
- [36] Pablo J Montoya, Sheila A Lukehart, Paula E Brentlinger, Ana J Blanco, Florencia Floriano, Josefa Sairosse, and Stephen Gloyd. 2006. Comparison of the diagnostic accuracy of a rapid immunochromatographic test and the rapid plasma reagins test for antenatal syphilis screening in Mozambique. *Bulletin of the World Health Organization* 84, 2 (2006), 97–104. <https://doi.org/10.2471/BLT.04.018663>

- [37] Onur Mudanyali, Stoyan Dimitrov, Uzair Sikora, Swati Padmanabhan, Isa Navruz, and Aydogan Ozcan. 2012. Integrated rapid-diagnostic-test reader platform on a cellphone. *Lab on a Chip* 12, 15 (aug 2012), 2678–2686. <https://doi.org/10.1039/c2lc40235a>
- [38] David Mukanga, James K. Tibenderana, Juliet Kiguli, George W. Pariyo, Peter Waiswa, Francis Bajunirwe, Brian Mutamba, Helen Counihan, Godfrey Ojiambo, and Karin Kallander. 2010. Community acceptability of use of rapid diagnostic tests for malaria by community health workers in Uganda. *Malaria Journal* 9, 1 (2010). <https://doi.org/10.1186/1475-2875-9-203>
- [39] Marshal M. Mweu, Juliana Wambua, Fixtan Njuga, Philip Bejon, and Daniel Mwanga. 2019. Bayesian evaluation of the performance of three diagnostic tests for Plasmodium falciparum infection in a low-transmission setting in Kilifi County, Kenya. *Wellcome Open Research* 4 (oct 2019), 67. <https://doi.org/10.12688/wellcomeopenres.15204.3>
- [40] Anthony T Newall, Paul A Scuffham, and Brent Hodgkinson. 2007. *Economic Report into the Cost of Influenza to the Australian Health System Report to the Influenza Specialist Group Executive summary*. Technical Report March. <http://www.isg.org.au/assets/assets/isg-cost-influenza-report-30-2007.pdf>
- [41] World Health Organization. [n.d.]. <https://www.who.int/malaria/areas/diagnosis/rapid-diagnostic-tests/rdt-rd7-final-list-of-products.pdf>
- [42] Aydogan Ozcan. 2014. Mobile phones democratize and cultivate next-generation imaging, diagnostics and measurement tools. *Lab on a Chip* 14, 17 (2014), 3187–3194. <https://doi.org/10.1039/c4lc00010b>
- [43] Haydar Ozkan and Osman Semih Kayhan. 2016. A Novel Automatic Rapid Diagnostic Test Reader Platform. *Computational and Mathematical Methods in Medicine* 2016 (apr 2016), 1–10. <https://doi.org/10.1155/2016/7498217>
- [44] Nitika Pant Pai, Caroline Vadnais, Claudia Denkinger, Nora Engel, and Madhukar Pai. 2012. Point-of-Care Testing for Infectious Diseases: Diversity, Complexity, and Barriers in Low- And Middle-Income Countries. *PLoS Medicine* 9, 9 (sep 2012). <https://doi.org/10.1371/journal.pmed.1001306>
- [45] Chunjong Park, Alex Mariakakis, Jane Yang, Diego Lassala, Yasamba Djiguiba, Youssouf Keita, Hawa Diarra, Beatrice Wasunna, Fatou Fall, Marème Soda Gaye, et al. 2020. Supporting Smartphone-Based Image Capture of Rapid Diagnostic Tests in Low-Resource Settings. In *Proceedings of the 2020 International Conference on Information and Communication Technologies and Development*. 1–11.
- [46] J L Pech-Pacheco, G Crist, J Chamorro-Mart Nez, and & J Fern Andez-Valdivia. [n.d.]. *Diatom autofocusing in brightfield microscopy: a comparative study*. Technical Report. www.iv.optica.csic.es/ssp/autofocus.html
- [47] Stephen M Pizer, E. Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. 1987. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing* 39, 3 (1987), 355–368. [https://doi.org/10.1016/S0734-189X\(87\)80186-X](https://doi.org/10.1016/S0734-189X(87)80186-X)
- [48] S. W. Ryu, J. H. Lee, J. Kim, M. A. Jang, J. H. Nam, M. S. Byoun, and C. S. Lim. 2016. Comparison of two new generation influenza rapid diagnostic tests with instrument-based digital readout systems for influenza virus detection. *British Journal of Biomedical Science* 73, 3 (sep 2016), 115–120. <https://doi.org/10.1080/09674845.2016.1189026>
- [49] Osama M.E. Seidahmed, Muneir M.N. Mohamedein, Afrah A. Elsir, Fayezy T. Ali, El Fatih M. Malik, and Eldirdieri S. Ahmed. 2008. End-user errors in applying two malaria rapid diagnostic tests in a remote area of Sudan. *Tropical Medicine & International Health* 13, 3 (feb 2008), 406–409. <https://doi.org/10.1111/j.1365-3156.2008.02015.x>
- [50] Shri Prakash Singh, Siddhivinayak Hirve, M. Mamun Huda, Megha Raj Banjara, Narendra Kumar, Dinesh Mondal, Shyam Sundar, Pradeep Das, Chitra Kumar Gurung, Suman Rijal, C. P. Thakur, Beena Varghese, and Axel Kroeger. 2011. Options for active case detection of visceral leishmaniasis in endemic districts of India, Nepal and Bangladesh, comparing yield, feasibility and costs. *PLoS Neglected Tropical Diseases* 5, 2 (2011). <https://doi.org/10.1371/journal.pntd.0000960>
- [51] Rene Ning Teh, Irene Ule Ngole Sumbele, Gillian Asoba Nkeudem, Derick Ndelle Meduke, Samuel Takang Ojong, and Helen Kuokuo Kimbi. 2019. Concurrence of CareStart™ Malaria HRP2 RDT with microscopy in population screening for Plasmodium falciparum infection in the Mount Cameroon area: predictors for RDT positivity. *Tropical medicine and health* 47, 1 (2019), 17.
- [52] Kentaro Toyama. 2015. *Geek heresy: Rescuing social change from the cult of technology*.
- [53] Piia Von Lode. 2005. Point-of-care immunotesting: Approaching the analytical performance of central laboratory methods. , 591–606 pages. <https://doi.org/10.1016/j.clinbiochem.2005.03.008>
- [54] Siriorn Watcharananan, Sasisopin Kiertiburanakul, and Wasun Chantratita. 2010. Rapid influenza diagnostic test during the outbreak of the novel influenza A/H1N1 2009 in Thailand: An Experience with Better Test Performance in Resource Limited Setting. *Journal of Infection* 60, 1 (jan 2010), 86–87. <https://doi.org/10.1016/j.jinf.2009.10.049>
- [55] Samuel White, Hanjie Ji, and Jeffrey P. Bigham. 2010. EasySnap: Real-time audio feedback for blind photography. In *UIST 2010 - 23rd ACM Symposium on User Interface Software and Technology, Adjunct Proceedings*. 409–410. <https://doi.org/10.1145/1866218.1866244>
- [56] Rapeeporn Wongkanya, Tippawan Pankam, Shauna Wolf, Supanit Pattanachaiwit, Jureeporn Jantarapakde, Supabhorn Pengnongyang, Prasopsuk Thapwong, Apichat Udomjirasirichot, Yutthana Churattanakraisri, Nanthika Prawepray, Apiluk Pakornsit, Thidadaow Sitthipau, Sarayut Petchaithong, Raruy Jitsakulchaidej, Somboon Nookhai, Cheewanan Lertpiriyasuwat, Sumet Ongwandee, Praphan Phanuphak, and Nittaya Phanuphak. 2018. HIV rapid diagnostic testing by lay providers in a key population-led health service programme in Thailand. *Journal of virus eradication* 4, 1 (jan 2018), 12–15. <http://www.ncbi.nlm.nih.gov/pubmed/29568547>

- [57] World Health Organization. 2019. *World Malaria Report*. Technical Report. 1–232 pages. <https://www.who.int/publications-detail/world-malaria-report-2019>
- [58] Hassan Zaraket and Reiko Saito. 2016. Japanese Surveillance Systems and Treatment for Influenza. *Current Treatment Options in Infectious Diseases* 8, 4 (dec 2016), 311–328. <https://doi.org/10.1007/s40506-016-0085-5>

A APPENDIX

A.1 Configurability Test with Other RDTs

To assess the extent to which RDTScan can be configured to new RDT designs, we examined whether RDTScan could reliably detect a range of RDT products not covered in our case studies. We generated the list of RDTs for this study by combining two lists curated by major health agencies—one by the World Health Organization for malaria RDTs [41] and one by the Centers for Disease Control for influenza RDTs [14]. After finding that the lists had very few dipstick RDTs (9 cassettes vs. 2 dipsticks), we added 3 additional dipsticks that are used for pregnancy testing. The results of this study are shown below. RDTScan was able to detect 8 of the 10 cassettes, all without the need for fiducial checking; both failure cases were blank white cassettes without any lettering or logos. RDTScan was able to detect all of the dipsticks provided that fiducial checking was enabled.

RDT Name	Form Factor	Compatibility with RDTScan
Binax Now Influenza A&B Card (Alere)	Card	Yes
QuickVue Influenza A+B (Quidel)	Dipstick	Yes (w/ fiducial detection)
LifeSign LLC Status Flu A&B (Princeton BioMeditech)	Cassette	No (blank cassette)
XPECT Flu A&B (Remel/Thermo Fisher)	Cassette	Yes
OSOM Ultra Plus Flu A&B Test (Sekisui Diagnostics)	Dipstick	Yes (w/ fiducial detection)
CareStart Flu A&B Plus (Access Bio)	Cassette	Yes
CareStart Malaria Pf/Pv (Access Bio)	Cassette	Yes
SD Bioline Malaria Ag Pf/Pv (Alere)	Cassette	Yes
PALUTOP +4 optima (All. Diag)	Cassette	Yes
One Step Malaria HRP2/pLDH (Wondfo)	Cassette	No (blank cassette)
AllTest Malaria P.f./Pan (AllTest)	Cassette	Yes
Asan Easy Test Malaria Pf/Pan Ag (ASAN Pharm.)	Cassette	Yes
Clearview hCG II Dipsticks (Clearview)	Dipstick	Yes (w/ fiducial detection)
Pregnancy Urine Dip-Strips (CLIAwaived, Inc)	Dipstick	Yes (w/ fiducial detection)
OSOM hCG DipStick Urine Test (Sekisui Diagnostics)	Dipstick	Yes (w/ fiducial detection)