Title: Harnessing Mechanistic Simulators for Rapid Diagnostic Test Capture and Deep Learning Classification

Authors: Elliott Rogers[1], Valérian Turbé[1], Dickman Gareta[2,5,6,7], Carina Herbst[2], Kobus Herbst[2,3], Maryam Shahmanesh[2,4], Rachel A. McKendry[1*]

Affiliations:
[1] London Centre for Nanotechnology and Division of Medicine, University College London, London, UK
[2]Africa Health Research Institute, Durban, South Africa
[3]DSI-SAMRC South Africa Population Research Infrastructure Network, Durban, South Africa
[4]UCL Institute of Global Health
[5] Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
[6] Graduate School for Health Sciences, University of Bern, Bern, Switzerland
[7] School of Laboratory Medicine and Medical Sciences, University of KwaZulu Natal, Durban, South Africa

**Corresponding author*

Abstract:
Rapid Diagnostic Tests (RDTs) are affordable tools for disease diagnosis and surveillance. There is growing evidence of the benefits of Machine Learning models to aid accurate decision making through classification of test results, improving sensitivity and specificity and reducing the risk of false positives/negatives. Yet to date these models have relied on large, clinically sensitive, costly, and time-intensive real-world image training libraries. Herein, we present SynSight – a classification pipeline for RDTs trained using mechanistically generated synthetic data. SynSight combines test region of interest segmentation using real-time object detection and result classification. We validate SynSight with HIV RDTs achieving 98.2% sensitivity and 99.2% specificity, and further with COVID-19 RDTs (two antigen, achieving 99% and 97% accuracy; one antibody, achieving 95% accuracy). This agile approach opens training opportunities for machine learning classifiers for RDTs with no available real-world image libraries, keeping pace with new RDT development and innovation.

Main

The humble lateral flow test giving results within minutes has become an indispensable rapid diagnostic test (RDT) for disease control valued for their ability to meet the REASSURED criteria[1]. The simplest tests, for example, COVID-19 RDTs, display a visual red coloured test line when a target antigen is present. More advanced multiplex RDTs simultaneously detect different targets in a single assay, for example, RDTs differentiating immune response by detecting Immunoglobulin G (IgG) and IgM antibodies.

RDT use and reliance in the management of acute and chronic diseases is ever increasing[1]. By late-2022, over 3 billion COVID-19 RDTs had been performed[2]. In 2024, more than 1.2 million Cholera RDTs will be shipped by GAVI in the largest-ever global deployment to support vaccine roll out[3]. The number of HIV RDTs used globally is projected to reach 206 million by 2027[4]. Moreover, there is increasing harmonisation of RDTs. For example, in June 2024, The Africa Centres for Disease Control and Prevention launched the Africa Collaborative Initiative to Advance Diagnostics to streamline regulatory approval for RDTs in six months, facilitating access for 55 member states (https://africacdc.org/programme/laboratory-systems-and-networks/africa-collaborative-initiative-to-advance-diagnostics/). Hence, RDTs play an increasingly pivotal role at point-of-care (POC) testing. However, their accuracy is risked by subjective human visual interpretation, and human error can result in misclassification[5-7].

Artificial intelligence (AI) and deep learning can aid diagnoses of many diseases, including for breast cancer[8] and tuberculosis[9]. Recently, research has applied machine learning (ML) to improve the classification of RDTs, including for HIV[7,10], COVID-19[10-14], influenza[15], and malaria[15,16]. Our previous study, *Turbé et al.*, trained convolutional neural networks (CNNs), termed ClassifyAI, on over 11,000 field acquired HIV RDTs images. ClassifyAI improved test specificity by 11% and sensitivity by 2.3% compared to visual interpretation by community health workers and nurses[7]. However, resource intensive real-world image library curation hinders the ability to quickly develop and deploy CNNs[11,15]. For ClassifyAI, 60 highly trained workers spent two years collecting images which were manually cropped to the region of interest (ROI), labelled by experts and then used for training[7]. More recently, few-shot learning approaches, such as AutoAdaptPOC, need 1/50th of the images for training but rely on a base model trained on hundreds of real-world images[20]. As an alternative, we explore a novel approach using synthetic images generated by mechanistic simulators.

Herein, we present SynSight, an agile, accurate, ML enabled RDT classification pipeline that overcomes the bottleneck of real-world image libraries (Fig. 1). SynSight features two models–segmentation and classification–trained on synthetic data (Fig. 1c). We evaluate SynSight's performance using the framework of "train on synthetic, test on real" (TSTR) compared to "train on real, test on real" (TRTR). A small TRTR - TSTR indicates high utility[17,18]. We demonstrate the versatility and performance of our SynSight pipeline (TSTR) on four RDTs: (i) one antibody HIV RDT (TRTR: *Turbé et al*[7], Fig. 2a-c), and (ii) two antigen and one antibody COVID-19 RDTs (TRTR: *Arumugam et al*[10], Fig. 2c).

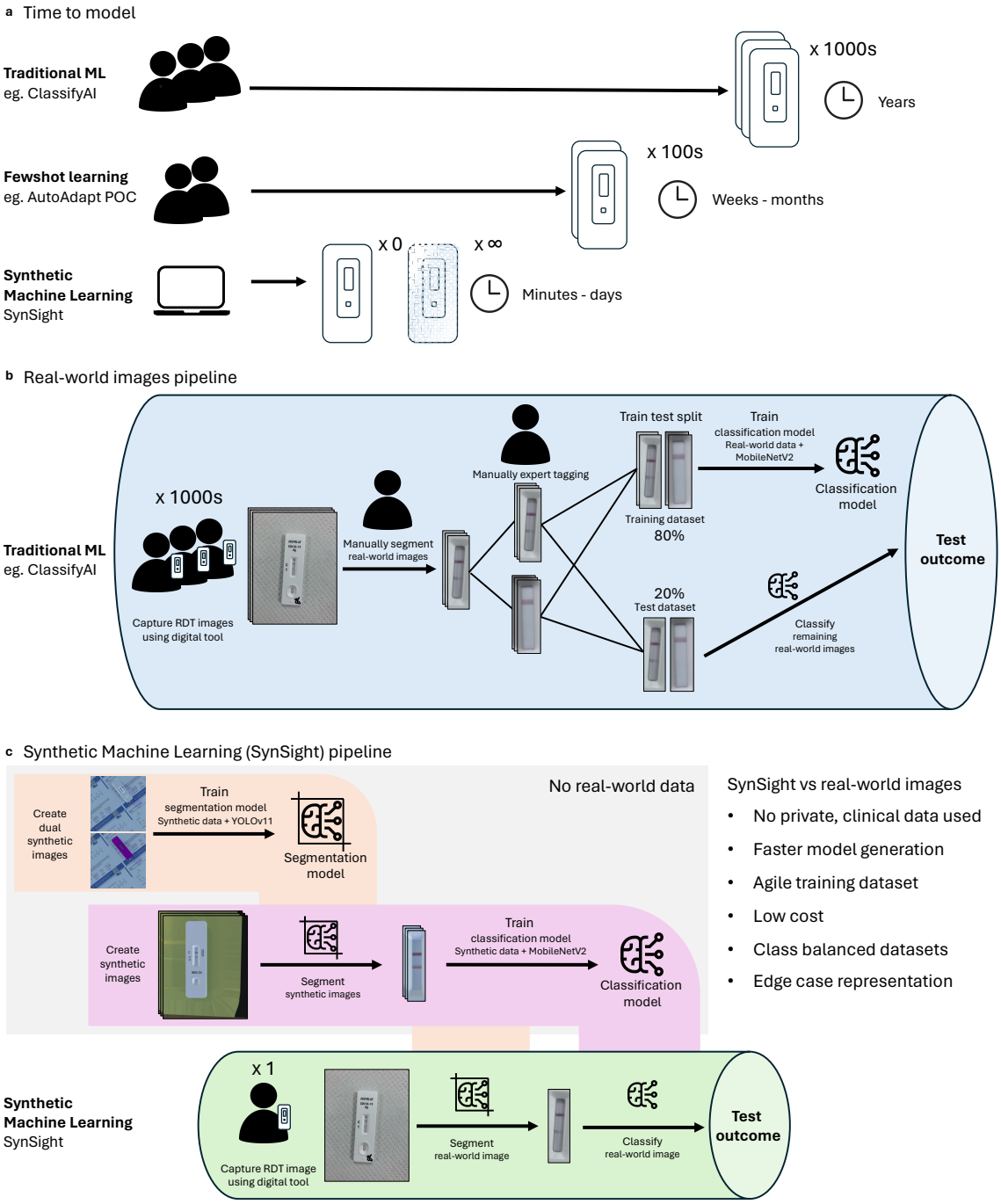


Fig. 1: Different approaches to training RDT classification models, including SynSight, our Machine Learning pipeline trained on synthetic data. a, Illustration of the significant time and human resources needed to capture real-world image libraries needed for current training approaches for Machine Learning models, such as ClassifyAI (*Turbé et al*[7]), and few shot learning approaches, such as AutoAdaptPOC (*Arumugam et al*[10]), compared to our synthetic approach, termed SynSight, using synthetic images and outputting a high performance model in minutes to days. **b**, Detailed current Convolutional Neural Network (CNN) training approach using real-world image libraries, (e.g., ClassifyAI). This involved the collection of thousands of real-world images by highly skilled individuals, manual segmentation of Region of Interests (ROI), manual expert tagging, training the model on 80% of the image library, and classifying the remaining 20%. **c**, Our SynSight approach, training a segmentation model (orange shading: a state of the art Real-Time Objection Detection model, YoloV11, and synthetic data) and classification model (pink shading:

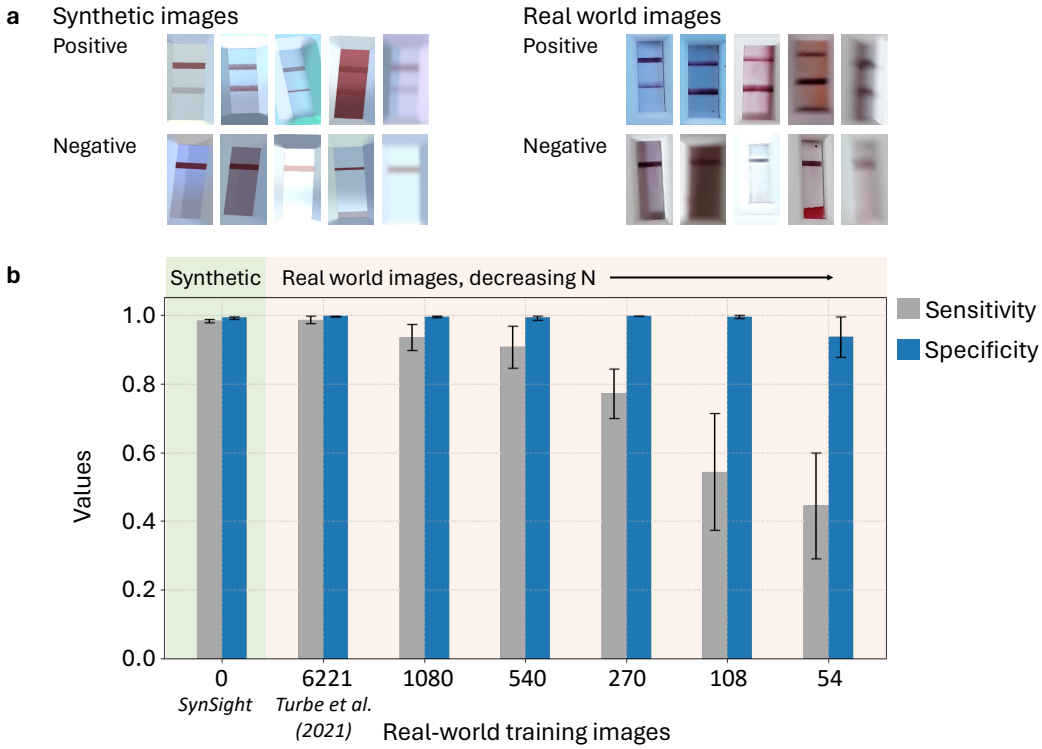
MobileNetV2 and synthetic data) on synthetic data. Each real-world image is automatically segmented then classified using models trained on synthetic data (green shading). SynSight outputs high performance Machine Learning segmentation and classification models using no private, clinical data, is faster, more agile, cheaper, and can produce customised training datasets and models.

To start, we generated large and diverse image libraries of synthetic RDT images focusing on challenging images for published ML models[6,11]. This highlights a core strength of synthetic image generation: increasing the variety of image conditions, frequency of underrepresented edge cases with poor ML classification performance (eg. blood in the ROI[11]), and rare images (eg. no control line[7]). Most ML training for RDT classification involves capturing a library of 100s to 1000s of real-world clinical images, expert tagging, manual or automated ROI cropping, training and validating on subsets, and finally classifying any remaining real-world images (Fig. 1a and b)[7,11]. By contrast, SynSight generated tagged synthetic data in minutes, removing the reliance on sensitive patient data (Fig. 1c). This approach could be applied to any RDT and image type.

To address challenges of sim-to-real generalisation, we use a mechanistic simulator, Unity, instead of deep generative models to generate images informed by our previous studies and a review of current literature[7,11,19]. All features – such as lighting (e.g. intensity) and camera parameters (e.g. angle) – were controlled through C# scripts (Fig. 2a). For example, overlaying particle distribution systems fixed to two dimensions can generate a synthetic library of the poorly classified weak positive result[6,11]. We can generate class-balanced synthetic datasets, whereas ClassifyAI's real-world HIV RDT image library has a 1:8 positive-to-negative class imbalance[7]. These images are then used to train segmentation and classification models.

Segmentation of the ROI of RDTs is a key step for ML classification[7,10,20]. To address time and labour intensive manual cropping, researchers have segmented ROIs by combining Mask R-CNN, an instance segmentation model, with the dimensions of RDTs and ROIs[10]; while others have used a state of the art real-time objection detection model, You Only Look Once (YOLO), for object detection of RDTs[20]. Here, we present the first demonstration of a YOLO model trained on synthetic data for the segmentation of RDT ROIs[21]. We leverage image generation to create dual RDT images with image processing automating ROI mapping (Fig. 1c, orange shading). The resulting model segments ROIs and requires no manually labelled real-world images for training. We have also developed and implemented an RDT segmentation model which could be used to guide image capture through Augmented Reality overlays (Supplementary Figure 1).

Next, we trained classification models on synthetic images (Fig. 1c, pink shading). For the HIV RDT, we trained a classification model on 8,000 synthetic images and achieved a TSTR sensitivity of 98.2% (± 0.8%) and specificity 99.2% (± 0.7%). This showed comparable accuracy to our model, ClassifyAI, trained on 6,221 real-world images (TRTR sensitivity: 98.7% ± 1.8%, specificity: 99.8% ± 0.2%)[7] (Fig. 2b-c). We also evaluated SynSight against ClassifyAI when trained on fewer real-world images (Fig. 2b)[7]. When trained on 270 real-world images the TRTR sensitivity of ClassifyAI dropped to 77.3% (± 11.6%), over 20 percentage points lower than SynSight (Fig. 2b, Supplementary Table 2). The advantage of our new synthetic approach is clear when large real-world image libraries are not available, are costly to capture, or do not accurately represent all image types.



c

			Train on synthetic, test on real (TSTR)	Train on real, test on real (TRTR)
			Synthetic (classification) 0 real-world images	*Turbé et al. (2021)* 6,221 real-world images
HIV	Antigen	Sensitivity	98.2% (± 0.8%)	98.7% (± 1.8%)
		Specificity	99.2% (± 0.7%)	99.8% (± 0.2%)
			Synthetic (pipeline) 0 real-world images	*Arumugam et al. (2023)* 403 real-world images
COVID	Antigen, Jinwofu	Accuracy	99.0%	100%
	Antigen, DeepBlue	Accuracy	97.3%	99%
	Antibody, Ecotest	Accuracy	95.1%	Used for training – not classified

Fig. 2. Trained on synthetic, tested on real (TSTR) performance of Machine Learning (ML) segmentation and classification models compared to published trained on real, tested on real (TRTR) models. a, Selection of synthetic images – showing varying test result, line intensity, background strip colour, blur, bleed, shadows – compared to real-world HIV RDTs (*Turbé et al*[7]). **b**, Plot showing the average sensitivity and specificity of HIV RDT classification using our classification model trained on synthetic data (TSTR, n=8,000; green shading) compared to a model trained on thousands of real-world images (TRTR, n=6,221)[7] and decreasing number of real-world images (TRTR, n=1080 to n=54; orange shading). The threshold for the synthetic model was calculated using 18% of the real-world dataset, as per our previous publication[7]. The TRTR model was trained using 10 fold cross-validation, with sensitivity and specificity obtained using a hold-out dataset (TRTR, n=693)[7]. **c**, High classification sensitivity, specificity, and accuracy of ML models trained on synthetic images (TSTR) compared to ML models trained on real-world images (TRTR) (HIV, 6,221 images[7];

COVID-19, 403 images[10]). The threshold for the COVID-19 RDTs was obtained using a synthetic dataset.

Next, we applied the full SynSight pipeline, implementing both the segmentation model and the classification model, to two antigen COVID-19 RDTs and one antibody COVID-19 RDT. 99.6% (921/925) of images were successfully segmented. Classification accuracy of the two antigen RDTs were 99% (Jinwofu) and 97% (DeepBlue), meeting the performance of the few-shot approach which used 403 training images[10] (Fig. 2c). As SynSight used synthetic images, we were also able to classify the antibody RDT images used for base-model training of the few-shot learning approach, achieving 95% accuracy (Fig. 2c).

The SynSight pipeline demonstrates a solution for quickly generating large image libraries to train high performance ML segmentation and classification models. The image generation and training was fast – under 30 minutes for the classification models on a laptop – compared to the weeks, months or years needed to curate a real-world image library[7,10,11]. An image processing pipeline could be prepared using synthetic data before a single real-world image of an RDT test is captured. This agility also lends itself to improving model performance over time by generating images of RDTs with reported classification issues, for example blood leaking into the ROI leading to false positives[11].

Synthetic data poses significant potential in imaging informatics for ML tasks in healtcare[18]. However, limitations of this approach may arise when image generation does not capture the variability of real-world images. This could lead to poor model performance, bias, and confabulation, a plausible but incorrect outcome based on patterns in the synthetic images. Precursor qualitative research is key for developing representative synthetic image libraries. For this study, this consisted of our experience from previous studies and evaluating existing research. For future synthetic image libraries for specific RDTs, available experts – individuals creating, implementing, and quality controlling RDTs – should be consulted to identify gaps in RDT classification performance. Further, when collecting real-world images for classification, clear guidance and image capture standard operating procedures (SOP) should be developed.

In conclusion, digital tools to improve the classification of the humble lateral flow are essential to support data capture, as well as improving accuracy by minimising false positives and false negatives. Developing accurate ML models quickly is crucial as novel RDTs are continuously developed for emerging and re-emerging diseases. Our SynSight pipeline segments and classifies RDTs with high accuracy. Our ML models trained on synthetic data demonstrate high utility, mirroring the performance of models trained on hundreds or thousands of real-world images. We validated this approach on four different RDTs, including antigen and multiplex antibody RDTs, for two different diseases. The high accuracy of SynSight is marked when the real-world training dataset is small. Synthetic data removes privacy concerns of training with sensitive, clinical data. These digital support tools have the potential to deliver an RDT expert review to the hands of any individual, and are increasingly important as novel, multiplex RDT use and development continues.

Methods

Ethics

The work presented are secondary analyses of existing real-world image libraries of HIV and COVID-19 RDTs. Ethical approval for the Africa Health Research Institute's (AHRI) HIV image library was collected as part of the demographic surveillance study, with ethics granted by the Biomedical Research Ethics Committee of the University of KwaZulu-Natal, South Africa (no. BE435/17)[7,22]. The ethics of the COVID-19 RDT image library are outlined in previous publications[10].

Real-world data

Images of HIV RDTs were collected by the AHRI, whereby fieldworkers conducted and imaged the ADVANCED QUALITY ONE STEP Anti-HIV (1&2) Test (InTec PRODUCTS, INC.)), as previously described[7]. The two antigen COVID-19 RDTs were the Jinwofu SARS-CoV-2 Antigen Rapid Test and the Anhui DeepBlue SARS-CoV-2 Antigen Test. The antibody COVID-19 RDT was the AssureTech EcoTest COVID-19 IgG/IgM Antibody Test kit. These COVID-19 RDTs were conducted and imaged by researchers at Columbia University using nasopharyngeal swabs from Mayo Clinic Hospital patients (Mayo Clinic IRB 20-010688)[10].

Synthetic data generation

Synthetic images of RDTs were mechanistically generated using 3D modelling software, Unity. This allowed control of all parameters of each image using C# scripts and features in Unity, and can be grouped into test specific and condition specific parameters. Example images can be found in the Supplementary Information.

Test specific parameters include test size, test result, control and test line sizes, control and test line opacities, control and test line appearances (solid and particle), control and test line positions, test strip colour, simulations of liquid progressing into the test window, appearance of randomised alphanumeric text above and to the right of the test window, appearance of outcome labelling text to the left hand side of the test window, and appearance of symbols below the test window. A thin or thick solid test line was used to replicate homogeneous control and test line results, with intensity set between 0.5 to 100% opacity. To replicate non-homogenous real-world test lines, multiple overlayed particle systems fixed to two dimensions of varying sizes, number and area were used. Condition specific parameters include lighting, camera yaw, position of the camera by translating randomly in three dimensions, camera layer culling masks, camera field of view, camera aspect ratio, and output image size.

Most parameters were governed by probabilistic mechanisms to create a wide variety of image properties. This probabilistic variation across multiple parameters results in a virtually infinite number of unique images. The fixed lighting array of the synthetic HIV image library can be used to illustrate this. Six lights were randomly selected from 12 possible lights in an array surrounding the RDT, generating 924 possible light combinations. Each light combination generates unique shadows on the ROI. Each light has a predefined likelihood of being white or a random colour. For simplicity, we can approximate the random colours through the Red, Green and Blue channels (0 to 1 assigned in Unity maps to 0 to 255 in RGB colour space in 10 discrete steps) giving 1000 possible random colours (10 x 10 x 10), and 1,001 unique colour combinations (including white). This results in over 10^{18} possible configurations across the lights (1001^6). Light intensity is randomly set between 0 and 0.4, giving 41 possible light intensities (assuming discrete 0.01 units), generating over 4 billion possible intensity combinations across six lights (41^6). Light conditions alone, including selection, intensity, and colour from the array of 12 lights can produce approximately 10^{30} unique lighting conditions. Further probabilistic values used to generate the synthetic RDT image training libraries can be found in provided C# scripts.

The image creation parameters were configured differently depending on the desired training dataset. 4000 positive and 4000 negative images were generated to train the HIV RDT image classification model. 1000 positive and 1000 negative images were generated to train the COVID-19 RDT image classification model.

Images for training the segmentation model were created in pairs. In one image, the ROI (test window) is assigned a clear block colour receiving no shadows (green shading), and the full cassette assigned a different clear block colour receiving no shadows (purple shading) (Figure 1). 4871 training images were used for training the segmentation model.

Image processing

Image processing consisted of several transformations aimed at standardising and improving data before training. Training images for the segmentation model were processed using various data augmentation techniques, including scaling images, mosaicking, random erasing, adjusting Hue, Saturation and Value (HSV), rotations, translations, and other data augmentation techniques. These were applied randomly through a predefined augmentation policy (RandAugment, https://docs.ultralytics.com/modes/train/).

Training images for the HIV classification model were subject to four image processing steps: resizing to 224 by 224 pixels preparing the image for the model input; blurring by applying a random intensity blur filter to soften image details; inverting by subtracting the value of each pixel from 255; and, finally, normalising by subtracting the mean pixel value and dividing by the standard deviation of all pixels in the image, closely following previous image processing to allow comparision[7]. Real-world images were subject to the same image processing excluding blurring. Training images for the COVID-19 classification model had slightly different processing steps to reduce computational demand. Images were first segmented to the ROI using a segmentation model, resized to 224 by 224, normalised by dividing each pixel value by 255, blurred and inverted.

Image labels

All synthetic images were labelled according to the virtual camera capturing the image from the mechanistic simulation environment. The C# script for generating image parameters, layer assignment and camera culling masks were checked before image generation to ensure correct labelling. Parameter extremes, such as highest light intensity and lowest test line intensity, were produced and visually checked to ensure the test outcome label was accurate.

All real-world RDT images were classified into two categories (positive/negative). The HIV RDT real-world image library was labelled by a panel of three RDT experts[7]. The COVID-19 RDT real-world image library was labelled as per previous publication[10].

Machine Learning: Training

For both classification models, a base model of MobileNetV2 was used with weights pre-trained on the ImageNet dataset, excluding the top layer. The model was prepared to receive images of 224 by 224 pixels with three channels corresponding to Red, Green, and Blue (RGB). After the base model, the output undergoes Global Average Pooling reducing the spatial dimensions to a single vector per image. This vector is processed through a fully connected layer consisting of 256 neurons with rectified linear unit (ReLU) activation. For the HIV classification model, the final dense layer is a single neuron with a sigmoid activation function outputting a probability value. This is the same as previously described[7] with two changes: the number of epochs was reduced from 50 to 10 in order to output high performance with faster training, image preprocessing was updated to include the aforementioned blurring and inversing steps. Models were trained using GPU for comparison to previous work and training speed. For the COVID-19 classification model, an additional dropout layer with a rate of 0.5 was added to prevent overfitting and allow the model to

generalise well to new data. Training was conducted on a laptop CPU demonstrating the technological accessibility of model development.

For the segmentation model, a nano YOLOv11 segmentation (yolo11n-seg.pt) base model was used, pre-trained on Ultralytics' library. A custom synthetic dataset defined in a YAML file, including synthetic images and annotations, was used for further training. The model was trained for 5 epochs on images resized to 640 by 640 pixels. Mixed precision training was used to balance memory usage and efficiency. All training scripts are either previously published[7] or available on request.

Machine Learning: evaluation and application

For the HIV RDT models, following training of the model on synthetic data, a best threshold was calculated by using 18% of the full dataset for validation, the same proportion of data for validation as the comparative ClassifyAI model trained on real-world data[7]. First, receiver operating characteristic curves (ROC) were constructed for the validation dataset, providing false positive rate (FPR) and true positive rate (TPR) across various thresholds. The best threshold was determined by selecting the threshold which maximised the sum of TPR and 1 – FPR, maximising true positives and minimising false positives. The ratio of classes in the validation set was matched to 1:8 positive to negative ratio in the full HIV RDT image library. For the models trained on increasing real-world data (Fig. 2b), a 1:8 positive to negative ratio was also maintained in training datasets. For the first training dataset (N = 54, positive n = 5), the 18% validation dataset had 1 positive image and 8 negative images. For each for the increasing real-world image datasets this ratio is maintained. The remaining real-world images are then classified based on the threshold, and sensitivity and specificity obtained. This process was repeated for 10 random validation subsets of each dataset, as previously described[7].

For COVID-19 RDTs, no real-world RDT images were used during training, validation, or best threshold calculation. Instead, a separate synthetic image library which the classification model was not trained on was used to calculate the best threshold. This best threshold was then applied to all three COVID-19 RDTs.

Acknowledgements
We acknowledge funding for this research: Medical Research Council for the PhD studentship (MR/W006774/1 to E.R); EPSRC Digital Health Hub for AMR (EP/X031276/1 to R.A.M); MS is an NIHR Research Professor (NIHR 301634); AHRI is supported by core funding from the Wellcome Trust (082384/Z/07/Z and 201433/Z/16/A to KH and MS). For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Data availability
All synthetic images can be requested from the authors and will be made available to download through the following link: https://doi.org/10.5522/04/28351934. Source data for Fig. 2b is available in Supplementary Table 1 and Supplementary Table 2. The real-world image library used for evaluating the performance of HIV RDT classification models is available through the AHRI repository, as previously published: https://data.ahri.org/index.php/catalog/923#metadata-data_access. Source data for Fig. 2c is available in Supplementary Table 2 and Supplementary Figure 5. The real-world image library used for evaluating the performance of the COVID-19 RDT segmentation and classification models is available through the following link, as previously published: https://figshare.com/projects/AutoAdapt/188457.

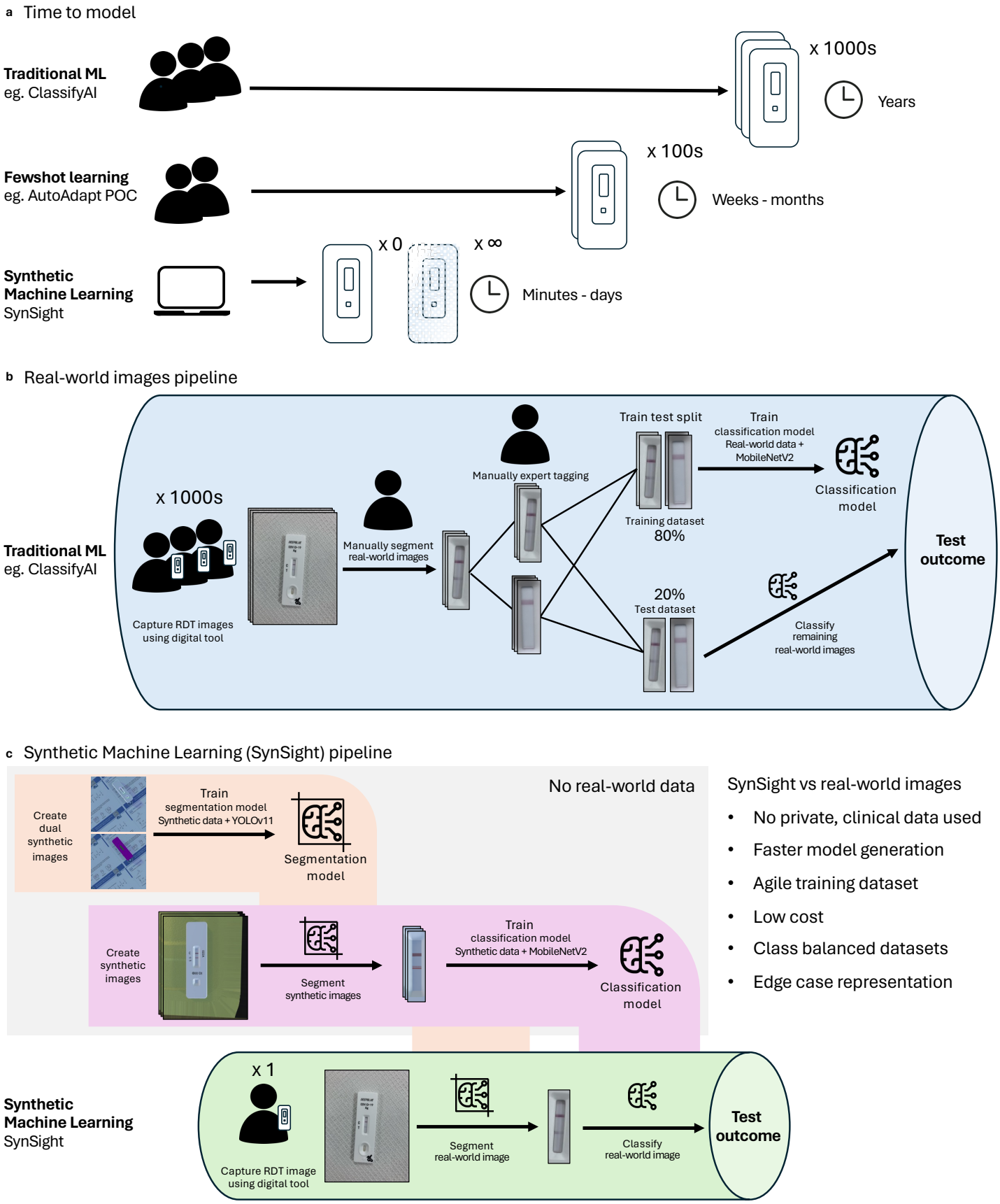
Code availability
Custom code used in training and analysing the machine learning models presented in this study can be requested from the authors and will be made available on GitHub at https://github.com/ElliottRogers/SynSight. The code was developed for Python 3.9. A detailed description of the code, including usage instructions and dependencies, is provided in several README files within the repository. This includes the core machine learning models used in this study (HIV classification, COVID segmentation, and COVID classification). In addition, previously published, free and easy to access code used to train the ClassifyAI model and the classification models trained on increasing real-world HIV data (Fig. 2b) is available here: https://xip.uclb.com/product/classify_ai.

The code was primarily tested on macOS 14.6.1. Other operating systems may be compatible but were not extensively tested. The following hardware was used for training: HIV Classification: Apple M2 (GPU - Metal), COVID Segmentation: Apple M2 (GPU - Metal), COVID Classification: CPU (Tested on Google Colab GPU).

References

1. Land, K.J., Boeras, D.I., Chen, X.-S., Ramsay, A.R. & Peeling, R.W. REASSURED diagnostics to inform disease control strategies, strengthen health systems and improve patient outcomes. *Nature microbiology* **4**, 46-54 (2019).
2. Organization, W.H. The ACT-accelerator: Two years of impact. *URL: https://www.who.int/publications/m/item/the-act-accelerator--two-years-of-impact* (2022).
3. WHO. Global deployment of rapid diagnostic tests to boost fight against cholera. (2024).
4. Consortium, E.I. HIV rapid diagnostic test market landscape. (2023).
5. Gaydos, C.A.*, et al.* Will patients "opt in" to perform their own rapid HIV test in the emergency department? *Annals of emergency medicine* **58**, S74-S78 (2011).
6. Mukadi, P.*, et al.* SMS photograph-based external quality assessment of reading and interpretation of malaria rapid diagnostic tests in the Democratic Republic of the Congo. *Malaria Journal* **14**, 1-12 (2015).
7. Turbé, V.*, et al.* Deep learning of HIV field-based rapid tests. *Nature medicine* **27**, 1165-1170 (2021).
8. Shen, L.*, et al.* Deep learning to improve breast cancer detection on screening mammography. *Scientific reports* **9**, 12495 (2019).
9. Breuninger, M.*, et al.* Diagnostic accuracy of computer-aided detection of pulmonary tuberculosis in chest radiographs: a validation study from sub-Saharan Africa. *PloS one* **9**, e106381 (2014).
10. Arumugam, S.*, et al.* Rapidly adaptable automated interpretation of point-of-care COVID-19 diagnostics. *Communications Medicine* **3**, 91 (2023).
11. Wong, N.C.*, et al.* Machine learning to support visual auditing of home-based lateral flow immunoassay self-test results for SARS-CoV-2 antibodies. *Communications medicine* **2**, 78 (2022).
12. Beggs, A.D.*, et al.* Machine learning for determining lateral flow device results for testing of SARS-CoV-2 infection in asymptomatic populations. *Cell Reports Medicine* **3**(2022).
13. Lee, S.*, et al.* Sample-to-answer platform for the clinical evaluation of COVID-19 using a deep learning-assisted smartphone-based assay. *Nature communications* **14**, 2361 (2023).
14. Mendels, D.-A.*, et al.* Using artificial intelligence to improve COVID-19 rapid diagnostic test result interpretation. *Proceedings of the National Academy of Sciences* **118**, e2019893118 (2021).
15. Park, C.*, et al.* The design and evaluation of a mobile system for rapid diagnostic test interpretation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **5**, 1-26 (2021).
16. Park, C.*, et al.* 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 (2020).
17. Jordon, J.*, et al.* Synthetic Data--what, why and how? *arXiv preprint arXiv:2205.03257* (2022).
18. van Breugel, B., Liu, T., Oglic, D. & van der Schaar, M. Synthetic data in biomedicine via generative artificial intelligence. *Nature Reviews Bioengineering*, 1-14 (2024).
19. Gibney, E. AI models fed AI-generated data quickly spew nonsense. *Nature* **632**, 18-19 (2024).
20. Lee, S.*, et al.* Rapid deep learning-assisted predictive diagnostics for point-of-care testing. *Nature Communications* **15**, 1695 (2024).
21. Glenn Jocher, A.C., and Jing Qiu. Ultralytics YOLO. Vol. 2023 (Ultralytics, https://github.com/ultralytics/ultralytics, 2023).

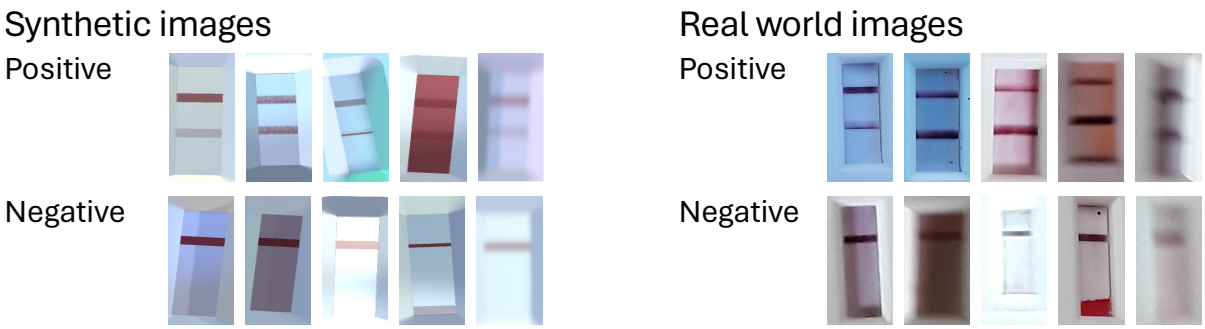
22. Gareta, D.*, et al.* Cohort profile update: Africa Centre Demographic Information System (ACDIS) and population-based HIV survey. *International journal of epidemiology* **50**, 33-34 (2021).



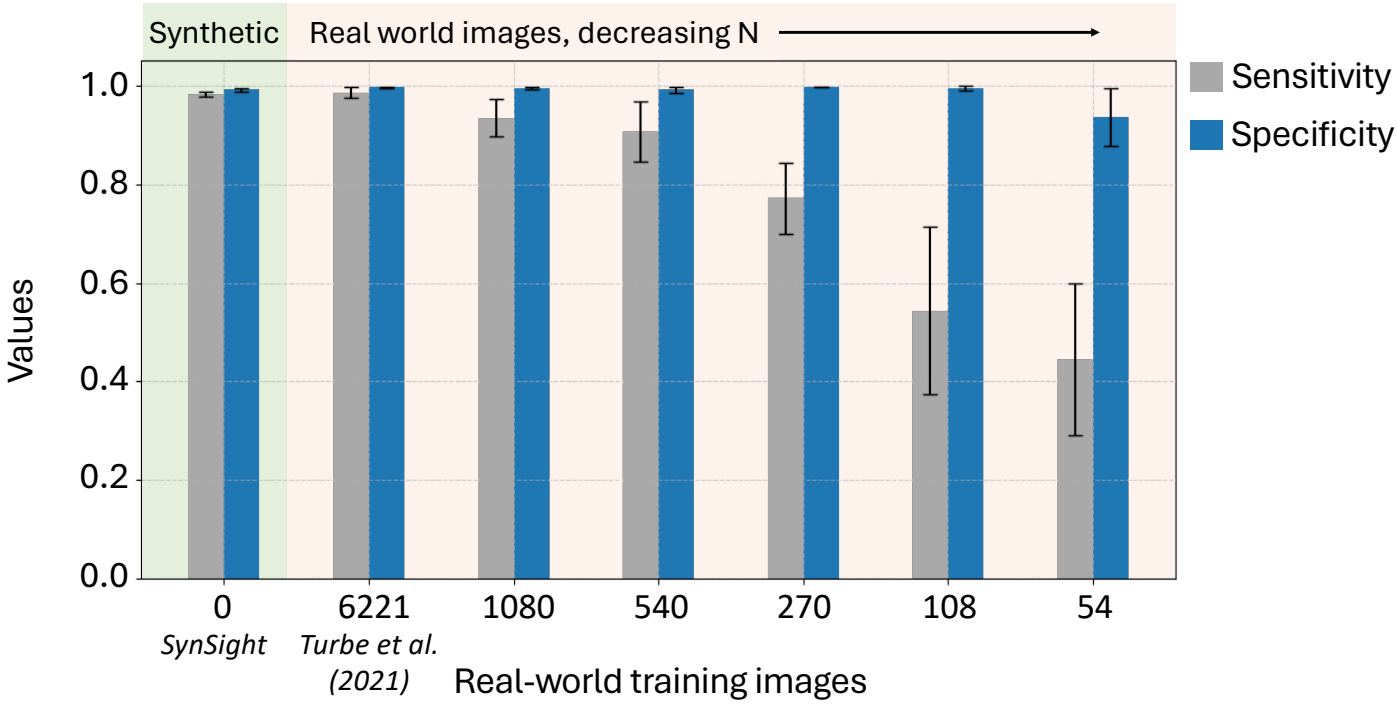
SynSight vs real-world images

- No private, clinical data used
- Faster model generation
- Agile training dataset
- Low cost
- Class balanced datasets
- Edge case representation

a



b



c

			Train on synthetic, test on real (TSTR)	Train on real, test on real (TRTR)
			Synthetic (classification) 0 real-world images	*Turbé et al. (2021)* 6,221 real-world images
HIV	Antigen	Sensitivity	98.2% (± 0.8%)	98.7% (± 1.8%)
		Specificity	99.2% (± 0.7%)	99.8% (± 0.2%)
			Synthetic (pipeline) 0 real-world images	*Arumugam et al. (2023)* 403 real-world images
COVID	Antigen, Jinwofu	Accuracy	99.0%	100%
	Antigen, DeepBlue	Accuracy	97.3%	99%
	Antibody, Ecotest	Accuracy	95.1%	Used for training – not classified