


Machine learning in point-of-care testing: innovations, challenges, and opportunities

Received: 17 December 2024

Accepted: 25 March 2025

Published online: 02 April 2025

 Check for updates

Gyeo-Re Han^{1,10}, Artem Goncharov^{1,10}, Merve Eryilmaz^{1,2}, Shun Ye^{2,3}, Barath Palanisamy^{2,3}, Rajesh Ghosh^{2,3}, Fabio Lisi⁴, Elliott Rogers⁵, David Guzman⁵, Defne Yigci⁶, Savas Tasoglu^{6,7,8}, Dino Di Carlo^{2,3}, Keisuke Goda⁴, Rachel A. McKendry⁵ & Aydogan Ozcan^{1,2,3,9} ✉

The landscape of diagnostic testing is undergoing a significant transformation, driven by the integration of artificial intelligence (AI) and machine learning (ML) into decentralized, rapid, and accessible sensor platforms for point-of-care testing (POCT). The COVID-19 pandemic has accelerated the shift from centralized laboratory testing but also catalyzed the development of next-generation POCT platforms that leverage ML to enhance the accuracy, sensitivity, and overall efficiency of point-of-care sensors. This Perspective explores how ML is being embedded into various POCT modalities, including lateral flow assays, vertical flow assays, nucleic acid amplification tests, and imaging-based sensors, illustrating their impact through different applications. We also discuss several challenges, such as regulatory hurdles, reliability, and privacy concerns, that must be overcome for the widespread adoption of ML-enhanced POCT in clinical settings and provide a comprehensive overview of the current state of ML-driven POCT technologies, highlighting their potential impact in the future of healthcare.

The landscape of diagnostic testing is undergoing a significant transformation, shifting from traditional centralized laboratory testing to more decentralized, rapid, and accessible methods through point-of-care testing (POCT)^{1,2}. Historically, centralized lab testing has played a crucial role in diagnosing and managing diseases by analyzing biological samples³. However, this often faces challenges related to lengthy turnaround times, high operational costs, and limited accessibility. The COVID-19 pandemic highlighted some of these limitations, as the surge in testing demand exceeded the capacity of centralized labs⁴. During the pandemic, at-home antigen tests⁴ and point-of-care nucleic acid testing⁵ became widespread for testing large populations effectively, demonstrating the feasibility and accuracy of POCT outside traditional lab environments. The widespread implementation of this

testing paradigm has revolutionized diagnostics, providing timely and accessible solutions essential for effective disease management and rapid medical response in diverse healthcare settings.

Current trends in POCT are guided by the updated REASSURED criteria—Real-time connectivity, Ease of specimen collection, Affordable, Sensitive, Specific, User-friendly, Rapid and Robust, Equipment-free, and Deliverable to end-users—which set the standard for modern POCT devices^{6–8}. Despite recent advancements, several significant challenges persist across various POCT modalities, including paper-based sensors such as lateral flow assays (LFAs) and vertical flow assays (VFAs), nucleic acid amplification tests (NAATs), and imaging-based sensor technologies^{4,9–13}. Achieving high analytical sensitivity and precision, detecting low-abundance biomarkers in biological samples,

¹Electrical & Computer Engineering Department, University of California, Los Angeles, CA, USA. ²Bioengineering Department, University of California, Los Angeles, CA, USA. ³California NanoSystems Institute (CNSI), University of California, Los Angeles, CA, USA. ⁴Department of Chemistry, The University of Tokyo, Tokyo, Japan. ⁵London Centre for Nanotechnology and Division of Medicine, University College London, London, UK. ⁶Department of Mechanical Engineering, Koç University, Istanbul, Türkiye. ⁷Koç University Translational Medicine Research Center (KUTTAM), Koç University, Istanbul, Türkiye. ⁸Physical Intelligence Department, Max Planck Institute for Intelligent Systems, Stuttgart, Germany. ⁹Department of Surgery, University of California, Los Angeles, CA, USA. ¹⁰These authors contributed equally: Gyeo-Re Han, Artem Goncharov. ✉e-mail: ozcan@ucla.edu

and high diagnostic accuracy comparable to conventional laboratory settings remains a primary challenge. Additionally, most POCT platforms have limited multiplexing capabilities and lack advanced analytical algorithms to interpret complex multivariable patterns, restricting their diagnostic applications for co-infections and multi-biomarker panel detection. Furthermore, the subjective interpretation of results by untrained users, such as determining whether a faint test line on a rapid test indicates a positive or negative result, poses significant hurdles in the scalability of POCT^{14,15}. The need for rapid, real-time data processing and error detection algorithms to ensure the reliability and accuracy of results further complicates the deployment of POCT in diverse clinical settings. Addressing these challenges is critical for the continued advancement and widespread adoption of POCT within the medical infrastructure.

Integrating artificial intelligence (AI) and machine learning (ML) into point-of-care sensors can potentially address many of the limitations currently hindering the advancement of POCT (Fig. 1)^{16–25}.

Combining advanced sensing modalities, assay platforms, and portable readers with ML algorithms enhances image/data analysis, signal processing, and quantitative interpretation^{16–25}. These algorithms can process complex datasets and accurately identify patterns or subtle changes in biomarker profiles despite the noisy nature of biological samples and POCT platforms' imperfections, potentially improving sensitivity and accuracy. ML/deep learning can also be used to optimize the properties of sensors in POCT platforms, enabling various applications, including the development of wearable sensors and non-invasive diagnostic tests^{26–29}. ML also enhances the multiplexing capabilities of point-of-care sensors through parallel analysis of multiple sensing channels using techniques such as e.g., neural networks and deep learning^{18,19}. For instance, deep learning has been successfully applied for computational optimization of multiplexed VFA designs, effectively identifying the optimal set of immunoreaction conditions, thereby enhancing the diagnostic performance and reducing the cost per test. Moreover, neural network-based analyte concentration

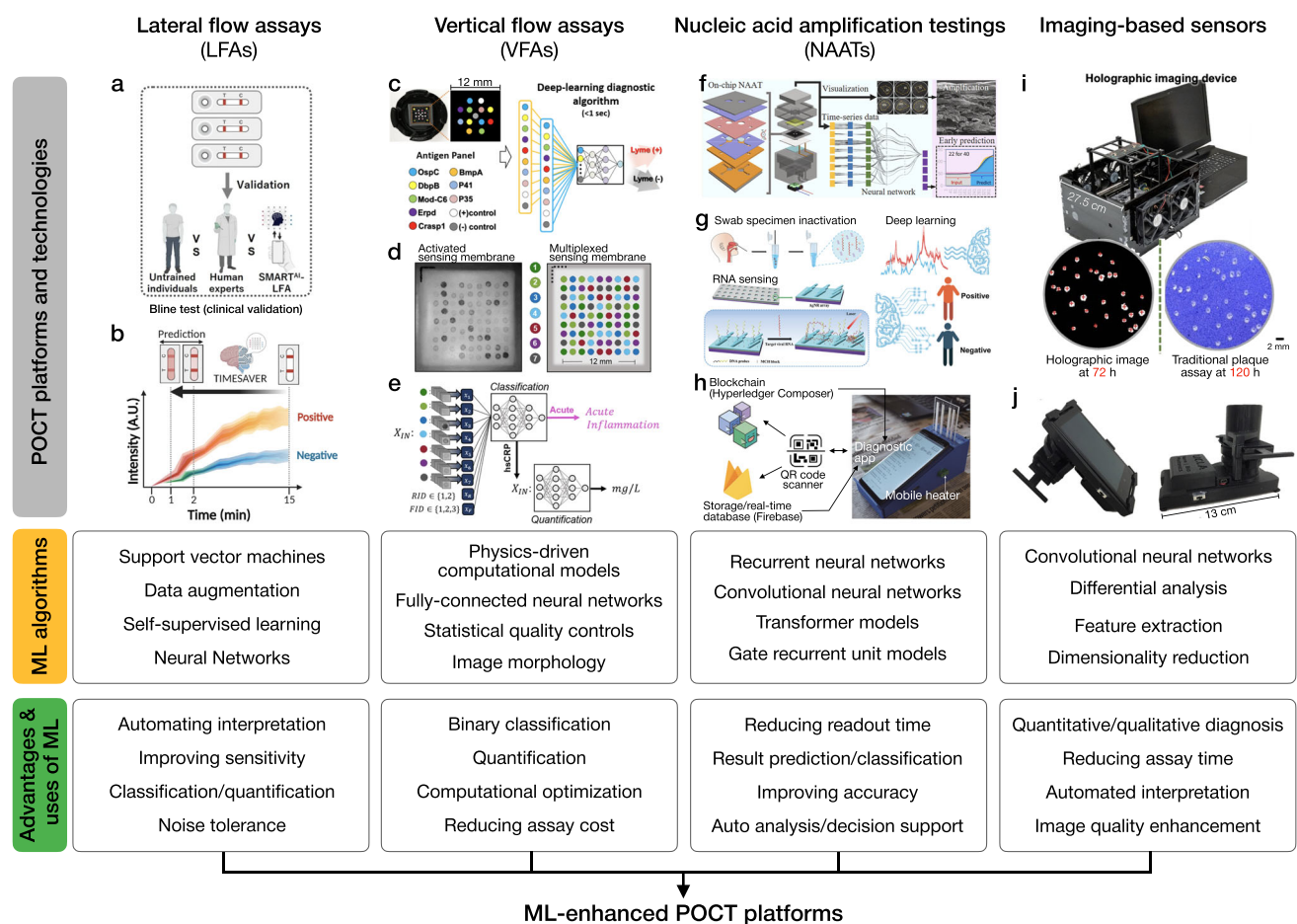


Fig. 1 | Overview of ML algorithms and advantages and uses of ML in various POCT platforms: LFAs, VFAs, NAATs, and imaging-based sensors. **a** Integration of ML in the SMART^{AL}-LFA platform enhanced the performance of LFAs by automating result interpretation, improving sensitivity, and enabling accurate predictions. **b** The TIMESAVER-LFA system demonstrated its high accuracy in distinguishing between positive and negative results while reducing assay time through the prediction algorithm. **c** xVFA platform used ML-driven diagnostic algorithms to detect Lyme disease using multiplexed antigen panels, leveraging deep learning to classify positive and negative test results. **d** The multiplexed sensing membrane inside the xVFA cassette features an algorithmically determined layout of immunoreaction spots with seven distinct spotting conditions that specifically interact with the target analyte and gold nanoparticles (AuNPs) conjugates. **e** ML algorithms applied to xVFA for binary classification and quantification of

target biomarker levels. **f** An on-chip NAAT platform with ML integration for real-time data visualization, time-series analysis, and early prediction of infectious diseases. **g** A Raman spectroscopy sensor-based NAAT platform incorporating ML for RNA sensing and deep learning to improve the accuracy of classifications for infectious disease diagnosis. **h** Smartphone-enabled DNA testing for malaria detection used deep learning to assist with local decision-making and blockchain technology to enhance data security. **i** A holographic imaging prototype for label-free live plaque assay to quantify virus. **j** A custom-made, 3D-printed smartphone-based bright-field microscope. **a, b, d, e, h, i** These are adapted with permission from refs. 16,17,19,22,23 by Springer Nature; **c, g** These are adapted with permission from ref. 18,21 by American Chemical Society; **f, j** These are adapted with permission from ref. 20,24 by Elsevier.

inference benefits from multiplexed sensing channels, significantly improving quantification accuracy and repeatability compared to standard multi-variable regression methods^{19,30}. Furthermore, ML can reduce assay time by automating data analysis and interpretation^{17,20,23}, facilitating quicker diagnostic decisions. Finally, convolutional neural networks (CNNs) have been widely applied to advance imaging-based POCT platforms owing to their ability to recognize/process patterns and extract task-specific useful features from image datasets. CNN-based POCT devices can offer shorter diagnostics times compared to standard laboratory equipment and provide automated analysis without compromising the sensitivity and accuracy of diagnostics³¹.

In this Perspective, we discuss some of the significant advancements and remaining challenges in the current landscape of POCT, with a specific focus on the applications of ML in sensing methods for LFAs, VFAs, NAATs, and imaging-based POCT technologies. Specifically, we explore how ML technologies can potentially address the limitations of these methods by improving analytical sensitivity, test accuracy, repeatability, sample-to-answer time, multiplexing capabilities, result interpretation and prediction, as well as scalability across different testing platforms. We also present some examples where ML has been successfully integrated into POCT devices. Beyond the technical aspects of ML-enhanced POCT, we provide a broader perspective by addressing critical regulatory, ethical, and implementation challenges unique to AI-driven diagnostics—an area that remains underexplored in existing literature. Additionally, we discuss ethical considerations such as data privacy, algorithmic transparency, and equitable deployment of AI in POCT settings, ensuring that this review is not only technically comprehensive but also provides actionable insights for real-world implementation. By providing a comprehensive overview of the current state, remaining challenges, and future opportunities of ML-enhanced POCT, this Perspective aims to highlight emerging opportunities for diagnostic innovations to improve healthcare. While ML has also been used in one-dimensional (1D) signal processing methods, such as electrochemical impedance or amperometric sensing, these have been extensively reviewed in recent literature^{32,33}. To avoid redundancy in discussing fundamental ML terminologies, concepts, and methodologies that have been comprehensively covered in prior works^{32–35}, here we provide a more focused discussion with an emphasis on in-depth case study analyses, where data processing complexity is inherently higher and requires advanced computational techniques.

The role of ML in advancing POCT: methods and impact

As POCT becomes more widely adopted in healthcare, its applications are becoming increasingly complex. These include testing panels comprising multiple biomarkers within a single cartridge, screening diverse populations across various age groups and races, and conducting tests over broad geographical regions and time periods. This widespread use of POCT generates large datasets with intricate patterns and sophisticated relationships between the output testing signals and underlying conditions at the level of an individual patient as well as the tested population. ML algorithms are particularly well-suited for these tasks owing to their ability to learn complex functional relationships in a data-driven manner³⁶. In addition, advanced sensor designs, especially for multiplexed sensing applications, benefit from computational co-optimization of the sensor hardware/design, where ML algorithms are used not only for diagnostics but also to enhance sensor design and performance¹⁹. For example, the use of metamaterials has improved sensor functionality by expanding the intrinsic sensitivity limits of materials^{26,37,38}. AI-based strategies have been employed in the design and manufacturing processes of metamaterials to overcome manual human design limitations and facilitate faster exploration of design and manufacturing parameters. Over the past decade, POCT has also advanced significantly through the

computational power of ML methods, enabling more accurate, higher-throughput diagnostics on inexpensive and widely accessible platforms. In this Perspective, we explore the role of AI and ML in some of the most prominent POCT technologies, including LFAs, VFAs, NAATs, and imaging-based methods.

Several different POCT-related use cases warrant the need for digital tools featuring ML to support the classification of the results. For example, self-administration and reading of POCT, as well as testing by less trained staff, are growing and require advancements in diagnostic accuracy³⁹. As LFA testing becomes increasingly decentralized, accurate interpretation of test results in different testing environments by different users is essential and can be facilitated by ML in POCT, which has been shown to reduce false positives and false negatives when used by individuals with less training^{40,41}.

In this context, ML can be categorized into three primary categories: supervised learning, unsupervised learning, and semi-supervised learning⁴². In supervised learning, the algorithm utilizes datasets with known true labels (i.e., ground truth) to learn the relationship between input patterns and the target outcome. During the learning (training) phase, these algorithms make predictions on the input data (i.e., training data) and utilize a predefined rule (loss function) to gradually improve the predicted values until they reach an acceptable performance metric. Supervised learning algorithms can be divided into classification and regression types, where the former classifies data between discrete categories while the latter predicts a continuous variable. In unsupervised learning methods, data do not contain labels, and the algorithms learn from the data structure to identify similar examples within the dataset (through, e.g., clustering). Finally, semi-supervised learning combines supervised and unsupervised learning methods, using ground truth labels from a smaller dataset to make predictions on larger unlabeled sets. In POCT and diagnostics, supervised learning approaches have been more frequently used due to the large amounts of labeled data available in this field³⁴. However, unsupervised algorithms have also been used for various applications, such as the diagnostics of sepsis⁴³ and cancer⁴⁴, successfully identifying relevant patient groups. In some other scenarios, semi-supervised models have also been applied to extract features from noisy image data, potentially reducing the cost of diagnostics⁴⁵. Due to their prevalence and competitive performance that has been demonstrated, the following sections of this Perspective will mainly focus on applications of supervised learning methods in POCT.

Various supervised ML approaches have been applied in POCT, including the k -nearest neighbor (kNN) method, support vector machines (SVMs), Naive Bayes method, random forest, fully-connected neural networks (FCNN), and convolutional neural networks (CNNs)³³. A typical pipeline for developing an ML-based method for the analysis of point-of-care sensors involves data preprocessing, splitting of the data (into training, validation, and blind testing datasets), model optimization, feature selection, and blind testing with new samples never seen before. Data preprocessing involves manipulating the dataset before it is input into the ML model, which consists of data denoising, augmentation, quality checks, normalization, and background subtraction. These preprocessing methods can dramatically improve ML model performance by lowering the impact of outlier samples and reducing variabilities present in raw signals. The pre-processed dataset is typically split into 60% training, 20% validation, and 20% blind testing sets, although the ratio of these can routinely change from application to application. After splitting the data, the ML model configuration and model hyperparameters are optimized based on the model performance on the validation set. The model with the optimal performance on the *validation set* is selected as the optimal model. Next, prior to training and blind testing of the final model, input features are also optimized by training the optimal model using different subsets of input features and comparing predictions on the

validation dataset. Finally, the optimized model (i.e., the fixed model with optimal configuration and input features) is tested on the blind testing dataset consisting of the samples not used during the training and validation stages.

Among different ML methods, neural networks and CNNs, in particular, are some of the most powerful ML techniques with an increasing prevalence in different fields of diagnostics, including POCT and imaging⁴⁶. This is because CNNs excel in feature extraction and pattern recognition tasks, enabling accurate and efficient analysis of image data. Neural networks consist of an input layer, hidden layers, and an output layer, and each layer contains units (nodes) associated with a particular set of weights (and biases) and a nonlinear activation function. The signal per unit is computed by applying the activation function to a weighted sum of the input values and potential bias terms from the previous layer. Consequently, the strength of the connection between successive units is determined by these weights. Although artificial neural networks are often regarded as a simplified representation of biological neural networks, wherein neurons are activated by the transmission of biochemical signals through synaptic connections, modern AI tools have advanced far beyond this simple analogy, presenting major differences in terms of their mathematical learning and inference principles, as well as computing hardware compared to biological systems; therefore such intuitive explanations should be taken with caution due to their limited coverage of modern deep learning-based information processors. During the training process, the digital weights are gradually updated to minimize the predefined loss function computed between the predicted and ground truth values. These weights are updated iteratively through a stochastic error backpropagation process^{47–49}, and the training process terminates either when the loss function meets certain criteria or when no statistically significant improvement of the loss is observed. Optimizing the model architecture and input features is critical to achieving optimal performance in a supervised training process. These optimizations are typically completed through a grid search on the validation dataset prior to the blind testing phase.

The large degrees of freedom in a deep neural network make them powerful, universal function approximators, capable of learning complex patterns between point-of-care sensor outputs and target diagnostics measurements and/or classifications despite the interference of noise factors from biological samples and imperfections of the individual sensors⁵⁰. For example, CNN-based models, since they are best suited for grid-structured data, such as image data, can efficiently extract spatial information from image data (e.g., shapes and edges) and perform pattern recognition, which makes them a powerful tool for imaging-based diagnostic tests and sensors⁵¹. In the following sections, we will explore specific applications/uses of ML and neural networks for various POCT platforms, including paper-based LFAs and VFAs, NAAT platforms, as well as different imaging-based point-of-care sensors. We will also discuss regulatory challenges, ethical considerations, and the integrity of ML, concluding with future perspectives on the broader adoption of AI and ML in healthcare systems.

Emerging innovations in POCT using ML LFAs

LFAs are widely used diagnostic tools for POCT due to their simplicity, rapid results, and cost-effectiveness. Typically, LFAs are developed to target a single biomarker and consist of two detection lines – a test line and a control line. In recent years, LFAs have become increasingly multiplexed, covering multiple test lines for various targets. For example, COVID-19 antibody LFAs have two test lines for IgG and IgM antibodies, while carbapenemases LFAs can have up to five test lines, targeting the most prevalent carbapenemases families. This feature increases the complexity of the visual interpretation of the outcome of the test. LFAs consist of a test strip that includes a sample pad, a conjugate pad, a nitrocellulose membrane with immobilized

antibodies, and an absorbent pad⁵². When a sample (e.g., whole blood, serum, plasma, urine, stool, saliva, or swab in liquid media) is applied to the sample pad, it moves through the strip by capillary action. If the biomarker of interest is present, it reacts with specific antibodies and binds to the capture regions, forming visible detection lines that indicate the presence or absence of the target analyte⁵³. This straightforward design makes LFAs suitable for various applications, including infectious disease detection^{54–56}, cardiac biomarker detection⁵⁷, pregnancy tests⁵⁸, cancer diagnostics⁵⁹, and environmental monitoring^{60,61}, among many others. From a commercial and engineering perspective, LFAs offer significant manufacturing advantages due to their simplified device architecture with low-cost materials and a long shelf life of approximately two years under ambient conditions, making them highly practical for commercial applications⁶². The latest progress in LFA research mainly focuses on improving the analytical performance by engineering physical components/parameters of LFAs, such as introducing sensitive conjugate labels/modalities^{63–69} and engineering advanced fluidic structures^{70–72}. Additionally, distance-based LFAs have emerged as a promising alternative to traditional LFAs by using changes in visual distance as a quantitative readout instead of relying on color intensity at a fixed test line^{73,74}. Various efforts to improve LFA performance through image/data-processing algorithms have also been introduced⁷⁵.

Integration of AI and ML in LFAs presents new methodologies to improve performance. This integration has been driven by the need to further improve sensitivity, specificity, result interpretation and quantification and to shorten the testing time. A notable trend in this integration is the use of smartphone or tablet-based digital devices to read test results^{75–77}. This not only enhances the usability and access to LFAs but also allows for the capture and collection of images of the test results in a digital format suitable for computational processing and interpretation without bias and reduces data loss⁴⁰. Specifically, ML technologies for LFAs focus on several key areas as outlined in Table 1: (i) automating the interpretation of test results to reduce human error and improve result consistency^{16,17,40,41,78–84}, (ii) improving sensitivity and accuracy through noise-tolerant analysis algorithms^{16,85–87}, and (iii) transforming traditionally qualitative tests into quantitative assays for more precise diagnostics^{16,78,81,83,85,86}.

Several studies have highlighted the critical role of ML in improving the performance of LFAs (Fig. 2); for example, a team at University College London (UCL) and the Africa Health Research Institute created an image library of over 11,000 real-world images of HIV LFAs acquired in KwaZulu South Africa (Fig. 2a)⁴⁰. The ML models accurately classified tests as positive or negative, significantly enhancing the specificity and sensitivity of decision support compared to visual interpretation by nurses and community health workers. This improvement was measured at 11% (from 89% to 100%) for specificity and 2.8% (from 95.6% to 97.8%) for sensitivity, reducing the risk of both false positives and negatives. Consequently, the system elevated the positive predictive value by 11.3% (from 88.7% to 100%) and the negative predictive value by 2.3% (from 95.7% to 98%), demonstrating its increased reliability in minimizing misclassification errors in POCT settings. Given that hundreds of millions of these tests are performed annually worldwide, this improvement could have major health and economic benefits to populations. Moreover, the UCL team also applied the models to COVID-19, and evaluated over 500,000 LFA images as part of the world's largest sero-surveillance study, REACT-2 (Real-time Assessment of Community Transmission-2) (Fig. 2b)⁴¹.

As another example, for automated image processing and result interpretation, Lee et al. introduced a deep learning-assisted smartphone-based LFA (SMART^{AI}-LFA) platform integrating a two-step CNN model (Fig. 2c)¹⁶. This model includes automated object detection using the You Only Look Once (YOLO)v3 network and test result classification with ResNet-18⁸⁸, providing accurate COVID-19 test results without the need for expert interpretation. YOLOv3 is a real-

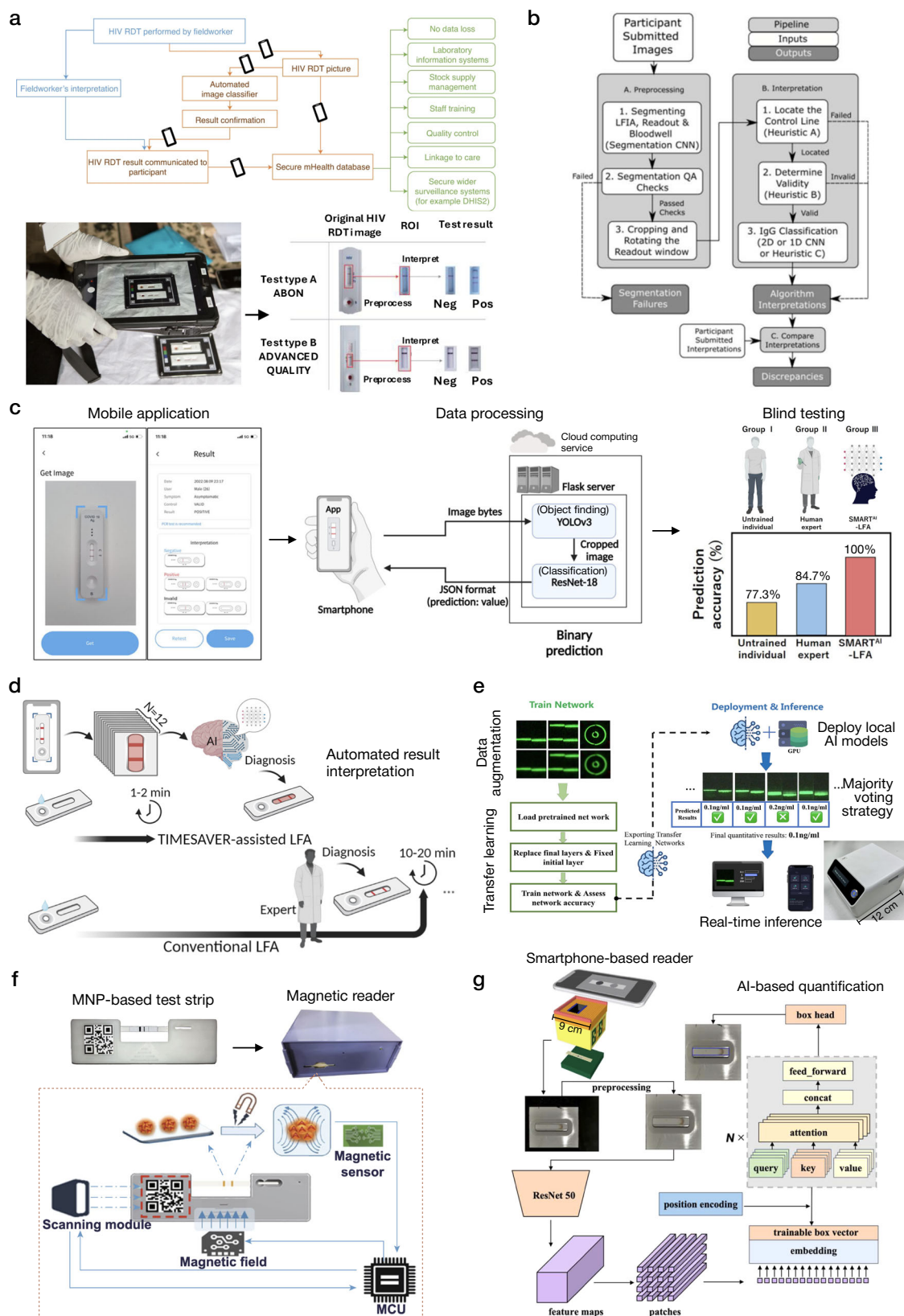
Table 1 | Applications of ML to LFAs

Ref. no.	Sensing modality	Analytes/targets	Specimens	AI/ML		Reader		Cost	Performance metrics	
				Method	Purpose	Portability	Size and weight		Diagnostic indicators	Analytical indicators
16	Colorimetric (AuNP)	COVID-19 Antigen	Nasopharyngeal (NP) and Oropharyngeal (OP) swabs, Saliva	CNNs (YOLOv3, ResNet-18), Data augmentation	Image processing (automated object detection), Result classification, Sensitivity improvement	Yes (Smartphone)	n/a ^a	n/a	Sensitivity: 98%, Specificity: 100%, Accuracy: 99% LoDs: 1.25 ng/mL (untrained user), 0.625 ng/mL (expert user), 0.156 (SMART ^{AI} -LFA)	>15 min
17	Colorimetric (AuNP)	COVID-19 antigen, Influenza A/B, cTnI, hCG	NP and OP swabs, Saliva, Buffer, Serum,	YOLO, CNN-LSTM, FC NN	Image processing, Result prediction, Assay time reduction	n/a (Custom CCD camera)	n/a	n/a	Sensitivity: 94.5% Specificity: 93.5% Accuracy: 94.2% for COVID-19 antigen assay (5 different LFA models), Sensitivity: 93.8%, Specificity: 100%, Accuracy: 95.8% for Influenza assay; Sensitivity: 96.9%, Specificity: 98.4%, Accuracy: 97.9% for TnI assay; Sensitivity: 97.5%, Specificity: 95%, Accuracy: 96.7% for hCG assay.	1–2 min
40	Colorimetric (AuNP)	HIV	Whole blood	SVM, CNNs (ResNet50, MobileNetV2, and MobileNetV3)	Image processing/classification, Result Interpretation	Yes (Tablet)	n/a	n/a	Sensitivity: 98.18 ± 0.79%, Specificity: 99.05 ± 0.15%	>10–40 min (varies by LFA product type)
41	Colorimetric (AuNP)	SARS-CoV-2 IgG	Whole blood	CNNs based on MobileNetV2, Visual transformer network	Image processing, Result classification	Yes (Smartphone)	n/a	n/a	Sensitivity: 90.1–97.1%, Specificity: 98.7–99.4%	>10–20 min
78	Colorimetric (AuNP)	COVID-19 neutralizing antibody	Serum	VIT, CNN (ResNET 50)	Image processing, Quantification, Analysis automation	Yes (Smartphone-embedded light box)	9 × 9 × 11 (cm)	n/a	n/a LoD: 160 ng/mL, Detection range: 625–10000 ng/mL	>20 min
79	Colorimetric (AuNP)	COVID-19 Ag, COVID-19 IgG/IgM	NP swabs, Whole blood	Mask R-CNN, Supervised contrastive learning, Self-supervised learning	Image processing (automated membrane extraction/interpretation), Rapid adaptation to new LFA kits, Classification	Yes (Smartphone)	n/a	n/a	Accuracy: 99–100% (COVID-19 antigen and antibody tests), 100% classification accuracy on both antigen and antibody tests in the drive-through study	>10 min
80	Colorimetric (AuNP)	SARS-CoV-2 Antibody and Antigen	Serum, Nasal swabs	CNN (MobileNet V2)	Automated result interpretation	Yes (Smartphone)	n/a	n/a	Sensitivity (IgG): 98%, Specificity (IgG): 100%, Sensitivity (IgM): 80%, Specificity (IgM): 89%, Sensitivity (Antigen): 100%, Specificity (Antigen): 100%	n/a

Table 1 (continued) | Applications of ML to LFAs

Ref. no.	Sensing modality	Analytes/targets	Specimens	AI/ML		Reader		Performance metrics		Time to result	
				Method	Purpose	Portability	Size and weight	Diagnostic indicators	Analytical indicators		
81	Colorimetric (AuNP)	Serum Amyloid A	n/a	SVM	Image processing, Quantification, Noise tolerance	Yes (Smartphone-embedded light box)	n/a	n/a	Accuracy: 94.23%	n/a	>5 min
82	Colorimetric (AuNP and enzymatic assay)	Drugs-of-abuse (amphetamine, ketamine, cocaine, methamphetamine, opiates, marijuana, and alcohol)	Saliva	Computer vision, Artificial neural networks	Image processing (automatic extraction of results), Histogram classification	Yes (Smartphone-embedded light box)	15 × 7 × 7 (cm), 0.3 kg	n/a	Total precision: 96 %	n/a: qualitative testing	>10 min
83	Surface-Enhanced Raman Scattering (SERS)	Escherichia coli O157	Milk, Beef	Extreme Gradient Boosting Regression	Quantitative analysis	No (Benchtop microscope)	113.3 × 47.3 × 81.7 (cm), 95 kg	n/a	n/a	LoD: 6.94 × 10 ¹ CFU/mL, Recovery: 86.41–128.25%	>5–10 min
84	Colorimetric (AuNP)	SARS-Cov-2 IgG/IgM	Serum, Whole blood	CNN	Automated result interpretation, Adaptation to different kits, users, readers	Yes (Smartphone)	272 x76 x 229 (cm)	n/a	Sensitivity: 99.5%, Specificity: 99.9%	n/a	<20 min
85	Fluorescence (UCNP)	Methamphetamine, Morphine	n/a	ResNet50, ResNet101, VGG16, VGG19, GoogleNet, MobileNet V2, AlexNet, DenseNet201	Image processing, Transfer learning, Quantification, Sensitivity improvement, Noise tolerance	Yes (Custom fluorescence reader paired with a smart-phone app)	10 × 12 × 7.4 (cm), 0.35 kg	n/a	Accuracy: up to 100% with transfer learning models	LoD: 1 ng/mL for MET and 0.1 ng/mL for MOP. Detection range: 1–20 ng/mL for MET and 0.1–100 ng/mL for MOP.	>0.33 min
86	Magnetic signal (MNP)	hCG, cTnI, CK-MB, Myoglobin	Serum	SVM, Custom waveform reconstruction	Signal classification, Improvements in sensitivity and specificity	No (Benchtop magnetic immunoassay reader)	n/a	n/a	n/a	LoD for hCG: 0.014 mIU/mL, Detection range: 1–1000 mIU/mL; Linear correlation with standard values of cTnI (R ² = 0.9902), CK-MB (R ² = 0.9870), and Myoglobin (R ² = 0.9866)	<18 min
87	Colorimetric (AuNP)	SARS-Cov-2 Ag	Serum, Whole blood	MagnifEye-powered algorithm	Automated result interpretation, Adaptation to different kits, users, readers	Yes (Smartphone)	n/a	n/a	Substudy 1: Accuracy: 99.78%, Sensitivity: 97.6%, Specificity: >99.9% Substudy 2: Accuracy: 98.6%, Sensitivity: 100%, Specificity: 99.28%	n/a	n/a

*n/a indicates not available in the associated reference.



time object detection model that efficiently identifies multiple objects in a single pass through the image, ensuring high speed and accuracy. This system, which incorporates a smartphone application for cloud-based processing, achieved 98% accuracy across clinical tests captured by different users and smartphones. In the blind testing phase, this SMART^{AI}-LFA outperformed both untrained users and human experts,

demonstrating the superior diagnostic accuracy of ML-enabled LFAs. The same authors demonstrated another approach, a time-efficient immunoassay with AI-based verification (TIMESAVER) (Fig. 2d)¹⁷. This system used a combination of neural network models to accurately identify detection regions in LFA. TIMESAVER reduced the assay time to 1–2 min compared to the traditional methods of 15 min, powered by

Fig. 2 | Application of ML to LFAs. **a** Infographics illustrating the benefits of HIV LFA processing in field settings (top), and HIV LFA capturing and analysis procedures used to enhance sensitivity and specificity for the detection of HIV in field settings (bottom). **b** LFA processing pipeline used for the automated interpretation of the REACT-2 dataset. **c** The SMART^{AI}-LFA platform used deep learning for automated image processing and result interpretation, utilizing this technology to detect the presence of the SARS-CoV-2 antigen. **d** The TIMESAVER-LFA platform employed an AI-based verification algorithm to reduce assay time to 1–2 min. **e** The fluorescent UCNP-based LFA used transfer learning to enhance sensitivity and

quantification capability for the detection of methamphetamine and morphine. **f** An ML approach was used in MNP-based LFA to improve analytical performance for the detection of hCG and multiple cardiac biomarkers. **g** The PDA nanoparticle-based LFA for the detection of COVID-19 neutralizing antibodies offered precise quantification of antibody concentrations through AI-based analysis. **a–d, f** These are adapted with permission from refs. 16,17,40,41,86, by Springer Nature; **e** This is adapted with permission from ref. 85 by Elsevier; **g** This is adapted with permission from ref. 78 by Elsevier.

a time-series deep learning algorithm. The system was applied to infectious diseases (e.g., COVID-19, influenza) and various biomarkers (e.g., cardiac troponin I [cTnI], human chorionic gonadotropin [hCG]), achieving high sensitivity and specificity on a spiked dataset (by spiking the target protein in running buffer). When blindly tested on clinical samples, the TIMESAVER system showed high accuracy in ~2 min of testing time, significantly faster than manual analysis by human experts, which took 15 min, showcasing the potential of ML to enhance paper-based point-of-care diagnostics.

To further improve the analytical performance (i.e., sensitivity and accuracy) and quantification capability of LFAs, ML algorithms were coupled with enhanced sensing modalities (i.e., fluorescence, magnetic nanoparticles, or optimized particles for enhanced absorption). For example, a study by Wang et al. utilized a transfer learning approach in combination with an LFA with fluorescent upconversion nanoparticles (UCNP) for the quantitative detection of methamphetamine and morphine (Fig. 2e)⁸⁵. Transfer learning is an ML technique in which a pre-trained AI model applies its acquired knowledge to a new problem, rather than learning from scratch, thereby enhancing efficiency and adaptability to new datasets. In biomedical sensing-related applications, transfer learning addresses challenges such as limited annotated data, balancing computational costs with large datasets, and improving generalization across various application scenarios. In a study by Wang et al., transfer learning improved the model's ability to detect low concentrations of analytes, reducing overfitting and improving generalization. It also ensured reliable analyte quantification in noisy environments. This approach simplified the detection process, making it feasible for use in portable POCT devices. A comparative analysis revealed that models trained without transfer learning exhibited drastically increased training times. These findings underscore the efficiency of transfer learning in biomedical POCT applications, ensuring reliable analyte quantification while optimizing computational resources.

As another important example, Yan et al. developed an approach to improve the analytical performance of magnetic nanoparticle (MNP)-labeled LFAs (Fig. 2f)⁸⁶. SVMs are classification algorithms that identify the optimal decision boundary to separate data points with maximum margin. Their effectiveness with small datasets and capacity to manage complex, high-dimensional, and noisy data makes them widely applicable in medical diagnostics and image recognition. By integrating an SVM-based classifier, the study improved the sensitivity and accuracy of detecting weak signals by effectively classifying complex patterns based on MNP signals. For example, at hCG concentrations of 0.25 mIU/mL, the accuracy of the SVM classifier was 100%, while that of visual reading was 7%. In addition, a novel waveform reconstruction method was introduced to accurately restore and interpret distorted waveforms of weak magnetic signals, thereby enhancing the assay's quantification capability. The application was validated by successfully quantifying hCG with a limit-of-detection (LoD) of 0.014 mIU/mL and a dynamic range of 1–1000 mIU/mL. This method has also been applied to analyze multiple test lines on a single test strip for several cardiac biomarkers – cTnI, creatine kinase isoenzyme MB (CK-MB), and myoglobin – demonstrating strong correlation with standard concentrations and showcasing the multiplexing potential in LFAs. For CK-MB and myoglobin, their quantification

cut-offs reached 0.5 ng/mL and 5 ng/mL in diluted serum, respectively, covering clinically relevant ranges when compared to negative samples. However, for cTnI, the quantification cut-off was limited to 0.5 ng/mL, falling short of the ~0.01 ng/mL sensitivity required for clinical use.

As another method for improved LFA performance, Tong et al. utilized polydopamine (PDA) nanoparticles to enhance the colorimetric signal using a vision transformer (ViT), which was trained to quantify the test results to detect COVID-19 neutralizing antibodies (Fig. 2g)⁷⁸. ViT processes images by dividing them into patches and leveraging a transformer model to capture spatial relationships throughout the entire image. Compared to CNNs, ViTs achieve high performance with less data and effectively comprehend long-range dependencies. This study employed the ViT framework to accurately predict the position of the test strip in the smartphone-captured image. The use of PDA nanoparticles improved the sensitivity of the assay, which was further advanced by using a ViT-based neural network that measured the intensity changes on the test strips captured by a smartphone-based reader. The ability of the neural network to accurately interpret these changes resulted in precise quantification of target antibody concentrations, overcoming the limitations of traditional visual assessment and making it an effective tool for evaluating vaccine efficacy in clinical and point-of-care settings.

VFAs

VFAs have recently emerged as a promising alternative testing platform to LFAs, offering shorter sample-to-answer time and improved multiplexing capabilities⁸⁹. In contrast to LFAs, which rely on lateral fluid flow, VFAs utilize a vertical liquid flow through stacked paper layers within a millimeter distance range, ensuring rapid sample/reagent migration and faster assay time. This unique fluidic design also enhances multiplexing capability by compartmentalizing sensing regions with a patterned reaction membrane, thereby minimizing cross-talk between detection zones. While VFAs require a relatively more complex fabrication and assembly process compared to LFAs and involve a slightly more intricate operation, they remain highly viable for POCT applications. Moreover, both LFAs and VFAs share the same materials and fundamental assay principles, such as antibody immobilization within the paper substrate and sample interaction with optical probe conjugates⁹⁰, making them complementary approaches. Typical VFAs utilize colorimetric⁹⁰ or surface-enhanced Raman scattering (SERS) detection modalities^{91–93}, covering various applications demonstrated so far^{89,90}. In this section, we will primarily focus on the applications where computational models and ML were used for the interpretation of the VFA signals, the improvement of VFA diagnostic/sensing accuracy and the optimization of the VFA design (see Table 2).

Conventional ML methods, including linear and polynomial regression models, have been widely applied for VFA analysis. Some of these applications include multiplexed detection of cancer biomarkers⁹⁴, quantification of metabolites in sweat⁹⁵, and detection of inflammatory markers in serum⁹⁶ (Fig. 3a). However, these previous approaches often fail to accurately capture the statistical features of the input signals, leading to erroneous predictions and high variability, especially for multiplexed analyses and real-life biological samples affected by various noise factors such as the sample matrix effect. In

Table 2 | Applications of ML to VFAs

Ref. no.	Sensing modality	Analytes/targets	Specimens	AI/ML	Reader		Cost	Performance metrics		Time to result
					Purpose	Portability		Diagnostic indicators	Analytical indicators	
19	Colorimetric (AuNPs)	CRP	Serum	FCNN, Linear regression	Overcoming hook effect, Improving CRP quantification accuracy and precision, Quantification of CRP in hs-CRP range, Detection of acute inflammation patients	Yes (Smart-phone-based)	\$2/test (Cartridge), \$270 (Reader)	Correlation between predicted and ground truth concentrations: $R^2 = 0.95$, CV: 11.2 %	LoD: 0.1 µg/mL, Detection range: 0.1–1000 µg/mL	<12 min
30	Fluorescent (CPN)	Myoglobin, CK-MB, h-FABP	Buffer, Serum	FCNN, Linear regression, Polynomial regression	Computational optimization of assay design, Quantification of analytes	Yes (Smart-phone-based)	\$3.3/test (Cartridge), \$410 (Reader)	Correlation between predicted and ground truth concentrations: $R^2 = 0.92$ (Myoglobin), $R^2 = 0.93$ (CK-MB), $R^2 = 0.95$ (FABP); CV: 12.4% (Myoglobin), 12.6% (CK-MB), 12.5% (FABP)	LoDs (buffer): 0.52 ng/mL for Myoglobin, 0.3 ng/mL for CK-MB, 0.49 ng/mL for FABP; Detection range (buffer, serum): ~0.5–50 ng/mL; Correlation with spiked values (buffer): $R^2 = 0.983$ for Myoglobin, $R^2 = 0.998$ for CK-MB, $R^2 = 0.999$ for FABP.	<15 min
94	Fluorescence (FITC)	Carcinoembryonic antigen (CEA), Alpha-fetoprotein (AFP), Cancer antigen 199 (CA199)	Buffer, Serum	Linear regression	Sensor calibration, Result Interpretation	Yes (Smartphone)	n/a	n/a	LoDs (buffer): 0.01 ng/mL for CEA, 0.02 ng/mL for AFP, and 0.04 U/mL for CA199; LoDs (serum): 0.03 ng/mL for CEA, 0.05 ng/mL for AFP, and 0.09 U/mL for CA199; Detection ranges (buffer and serum): 0.1–1000.0 ng/mL for CEA, 0.1–1000.0 ng/mL for AFP, and 0.1–1000.0 U/mL for CA199.	5 min
95	Colorimetric (TMB), Electrochemical (Prussian blue)	Glucose, Lactate, Uric acid, Magnesium ions, pH value	Sweat	Linear regression	Sensor calibration, Result Interpretation	Yes (Smart-phone-based)	15.8 × 7.2 × 9 (cm)	n/a	Correlation with spiked values (artificial sweat): $R^2 = 0.997$ for glucose, $R^2 = 0.997$ for lactate and magnesium ions, $R^2 = 0.995$ for uric acid, $R^2 = 0.992$ for magnesium ions.	15 min for glucose, uric acid, and pH; 10 min for lactate and magnesium ions.

Table 2 (continued) | Applications of ML to VFAs

Table 2 (continued) Applications of ML to VFAs											
Ref. no.	Sensing modality	Analytes/targets	Specimens	AI/ML		Reader		Cost	Performance metrics		Time to result
				Method	Purpose	Portability	Size and weight		Diagnostic indicators	Analytical indicators	
96	SERS	CRP, Interleukin-6 (IL-6), Serum amyloid A (SAA), Procalcitonin (PCT)	Buffer, Serum	Linear regression, Polynomial regression	Sensor calibration, Result Interpretation	No (Raman microscope)	-90 kg	> \$10,000 (Reader)	n/a	LoDs (buffer): 53.4 fg/mL for CRP, 4.72 fg/mL for IL-6, 48.3 fg/mL for SAA, and 7.53 fg/mL for PCT; Detection ranges (buffer): 0.01–1000 ng/mL for CRP and SAA; 0.001–1000 ng/mL for IL-6 and PCT; CV (serum): <10% in clinically relevant ranges	≤5 min
97	Colorimetric (ELISA, 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium)	Rabbit IgG	Buffer	CNNs: AlexNet, GoogleNet, ResNet34, MobileNet_V2	Image processing (automated detection of reaction areas), Classification between analyte concentrations	Yes (Smartphone)	n/a	n/a	Accuracy: 95.05% (MobileNet_V2), 96.32% (AlexNet), 96.88% (ResNet34), 97.14% (GoogleNet); AUC: 0.98 (MobileNet_V2), 0.99 (AlexNet), 0.98 (ResNet34), 0.99 (GoogleNet) for classification between analyte concentrations	LoDs for different illumination conditions: 550 pM (no light), 780 pM (fluorescent light), 840 pM (table lamp), and 418 pM (natural light); Correlation with spiked values: R ² = 0.945 (no light), R ² = 0.978 (fluorescent light), R ² = 0.989 (table lamp), and R ² = 0.913 (natural light)	~1 h
98	Colorimetric (HRP and potassium iodide)	Glucose	Buffer, Artificial urine	FCNN	Quantification, Classification of the analyte	No (Scanner)	48 × 28 × 11.7 (cm), 4.1 kg	\$350 (Reader)	The correlation between predicted and ground truth concentrations: Pearson's r (3D μPAD) = 0.974, Pearson's r (μTPAD) = 0.965; Classification between low, normal and high glucose: Accuracy (3D μPAD) = 91.2% Accuracy (μTPAD) = 94.4%	LoD (μTPAD and 3D μPAD): 0.5 mM	~10 min for 3D μPAD, ~1 h for μTPAD

Table 2 (continued) | Applications of ML to VFAs

Ref. no.	Sensing modality	Analytes/targets	Specimens	AI/ML	Reader		Cost	Performance metrics		Time to result
					Purpose	Portability		Diagnostic indicators	Analytical indicators	
99	Colorimetric (TMB)	Tuberculosis Rv-1656 antigen, SARS-CoV-2	Buffer, Urine, Saliva, Sweat	Computational physics-driven models	Result interpretation, Improvement of VFA scalability and adoption	Yes (Digital camera)	12.9 × 7.5 × 3.7 (cm), 0.35 kg	The deviation between model prediction and experimental results: <9% (when changing label concentration and washing volume), <20% (when altering sample volume), <20% (when changing assay formation procedures), 22 % (when switching from buffer to saliva sample matrix)	LoD (saliva): 0.1–1 nM	n/a
100	Colorimetric (AuNPs with signal amplification)	cTnI	Serum	FCNN, Power law fitting	Exclusion of outliers, Optimization of assay precision, Quantification of analyte	Yes (Raspberry Pi-based)	16 × 9.5 × 5.5 (cm), 0.32 kg	The correlation between predicted and ground truth concentrations: Pearson's $r = 0.965$, CV: 6.2 %	LoD: 0.2 pg/mL, Detection range: 1–10 ⁵ pg/mL, Correlation with spiked values (serum): $R^2 = 0.998$	<15 min
101	Chemiluminescence (Polymerized enzyme conjugated with AuNPs)	cTnI	Serum	FCNN, Power law fitting	Exclusion of outliers, Optimization of assay precision, Quantification of analyte	Yes (Raspberry Pi-based)	8 × 17.5 × 10.5 (cm), 0.65 kg	The correlation between predicted and ground truth concentrations: Pearson's $r = 0.979$, CV: 14.3 %	LoD: 0.16 pg/mL, Detection range: 1–10 ⁵ pg/mL, Correlation with spiked values (serum): $R^2 = 0.9929$	<25 min
18,102	Colorimetric (AuNP)	IgM/IgG antibodies for Lyme disease	Serum	FCNN	Computational optimization of assay design, Classification of positive and negative samples	Yes (Smart-phone-based)	16 × 8 × 4.5 (cm), 0.3 kg	Sensitivity: 90.5% (antigen) and 95.5% (peptide), Specificity: 87.0% (antigen) and 100% (peptide)	LoD (serum): 100 ng/mL, Detection range (serum): 0.1–5 µg/mL	15 min
104	Colorimetric (AuNPs)	IgM/IgG antibodies for COVID-19	Serum	FCNN, Random forests, Logistic regression	Computational optimization of assay design, Classification of immunity level, Longitudinal tracking of immunity levels	Yes (Smart-phone-based)	16 × 8 × 4.5 (cm), 0.3 kg	Accuracy: 89.5%	LoD (serum): 100 ng/mL, Dynamic range (serum): 0.1–5 µg/mL	<20 min
105	Electrochemical (FET-based)	Cholesterol	Serum	FCNN	Optimization of the assay procedure, Reduction of the test time, Quantification of analyte	No (Keithley 4200 A semi-conductor analyzer)	n/a	The correlation between predicted and ground truth concentrations: $R^2 > 0.976$, CV: -6.46%	LoD: 28.5 µg/dL, Detection range: 0–300 µg/dL	<2.5 min

^an/a indicates not available in the associated reference.

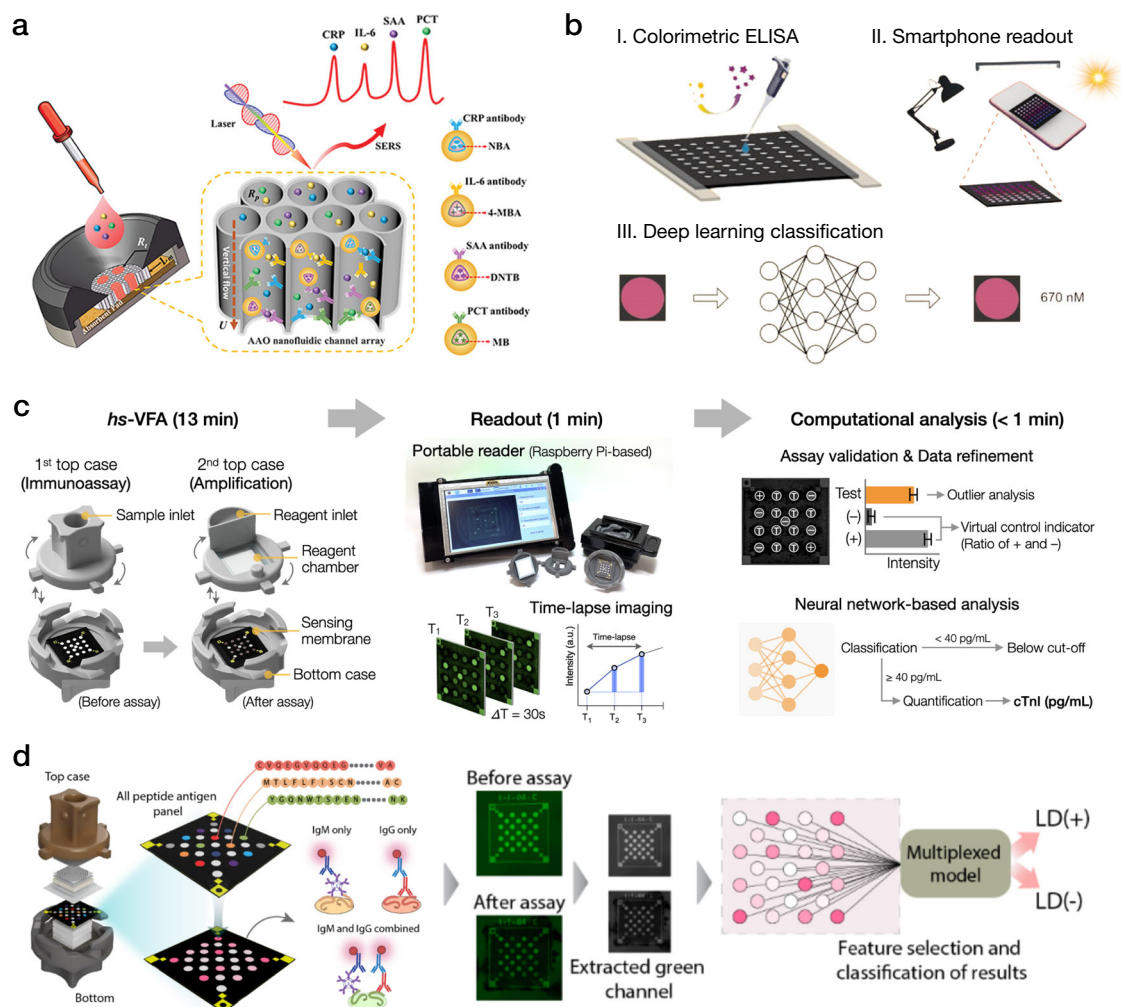


Fig. 3 | Applications of ML to VFAs. **a** SERS-based VFA processed by linear regression models for quantitative detection of inflammation markers. **b** Colorimetric ELISA (c-ELISA) processed by neural networks for accurate detection of Rabbit IgG under different illumination conditions. **c** Deep learning-enhanced xVFA for high-sensitivity detection of cTnI. **d** Peptide-based xVFA

processed by deep learning models for single-tier testing of Lyme disease. **a** This is adapted with permission from ref. 96 by Wiley; **b** This is adapted with permission from ref. 97 by Elsevier; **c** This is adapted with permission from ref. 100 by American Chemical Society; **d** This is adapted with permission from ref. 102 by Springer Nature.

contrast, advanced ML models like neural networks can provide better consistency due to their representation power and ability to learn complex relationships between noisy assay signals and ground truth information³⁶.

Neural network-based inference models (AlexNet, GoogleNet, ResNet34, MobileNetV2) were used to enhance the adaptability and robustness of paper-based colorimetric enzyme-linked immunosorbent assays (c-ELISA)⁹⁷. The activated assays were captured by a smartphone camera under three different lighting conditions and then processed by deep learning algorithms, achieving a detection accuracy of >97% in detecting Rabbit IgG in synthetic samples with GoogleNet (Fig. 3b). As another example, Lee et al. employed a three-dimensional (3D) μ PAD to detect glucose through a colorimetric assay processed by neural network models, achieving 91.2% in classifying glucose levels as low, normal, and high when tested on glucose-spiked buffer samples. The same study also proposed a microfluidic thread/paper-based analytical device (μ TPAD) to detect glucose levels in artificial saliva and showed a high correlation between the predicted and the ground truth glucose levels⁹⁸. These platforms applied ML methods to enhance the scalability and adaptability of VFA readout. However, the performances of these μ PAD platforms were primarily demonstrated on synthetic and spiked samples, necessitating further testing on biological specimens to confirm their viability for real-life applications.

Although the fundamental principles of various VFA platforms are similar, their design and operational characteristics may differ, leading to potential inconsistencies in the output signals across different assays. These variations stem from a sophisticated interaction between assay and readout parameters, making the standardization of these processes a challenging endeavor. To address this issue, Tay et al. proposed a comprehensive physics-driven framework to model key assay operation parameters, demonstrating reliable generalization to different VFA formats and sample matrices⁹⁹. The framework incorporated adaptable physical equations, covering typical paper-based assay procedures such as sample mixing, sample propagation through paper layers, analyte immobilization, and assay readout. The researchers validated their model on VFAs with varying immunoassay formation procedures, reporter label concentrations, and sample matrices (i.e., buffer, urine, saliva, and sweat), demonstrating good agreement with their experimental results. For example, the observed discrepancy between the predicted and experimental results was <10% when varying the label concentration or the washing volume, and <20% when altering the assay formation procedure. The overall high concordance of the model with experimental data and the minimal requirements for the training set size make the proposed workflow an appropriate tool for accelerated assay optimization and adoption.

In a different application, ML-driven VFAs were employed to achieve accurate cardiac biomarkers quantification in patient serum samples. For instance, Ballard et al. proposed a multiplexed VFA (xVFA) for the quantification of C-reactive protein (CRP), i.e., for high-sensitivity CRP (hs-CRP) testing, leveraging xVFA's multiplexing capabilities and ML to improve the sensor's dynamic range and quantification accuracy¹⁹. Researchers demonstrated accurate quantification of CRP in the high-sensitivity range (0–10 µg/mL), achieving an R^2 of 0.95 compared to a Food and Drug Administration (FDA)-approved analyzer. Additionally, they were able to successfully identify elevated CRP concentrations outside the high-sensitivity range, within the acute inflammation range (up to ~1000 µg/mL), mitigating the hook effect present in singleplexed testing platforms. Furthermore, Han et al. used the xVFA platform for high-sensitivity cTnI (hs-cTnI) testing (Fig. 3c)^{100,101}. cTnI is a gold-standard protein biomarker for diagnosing myocardial infarction (MI), and yet presents a challenge for POCT due to its extremely low clinical concentrations. To achieve high sensitivity, researchers integrated Au-ion amplification chemistry into the xVFA and demonstrated a high-sensitivity VFA (hs-VFA) with a colorimetric modality, which achieved trace-level cTnI detection by increasing the diameter of the optical absorption labels (AuNPs). During the amplification stage, the AuNP diameter increased from 15 nm to >200 nm, leading to a significantly higher signal. The hs-VFA showed a LoD of 0.2 pg/mL, which is >10 times lower than the clinical cutoff concentration in hs-cTnI testing. Fully connected neural network (FCNN) models were employed to exclude outlier sensors, enhance the precision of cTnI testing, and accurately quantify cTnI concentration in serum. The computationally optimized hs-VFA demonstrated a high correlation with an FDA-approved analyzer (Pearson's $r > 0.96$) and good repeatability with a <6.2% coefficient of variation (CV) between duplicate repeats of the tested samples, meeting clinical requirements for hs-cTnI testing.

The xVFA platform was also used for the multiplexed quantification of cardiac markers in serum. Researchers integrated fluorescence-based detection into the xVFA platform and applied this fluorescence-based xVFA (fxVFA) for the parallel quantification of three cardiac markers (i.e., myoglobin, CK-MB, and heart-type fatty acid-binding protein [h-FABP]) in serum samples³⁰. The assay contained a total of 17 immunoreaction spots/channels, including 6 testing spots with 2 spots per biomarker type and 11 control spots. The activated fxVFA was captured by a hand-held and cost-effective fluorescence reader, and the signals of the immunoreaction channels were processed by neural network models for the multiplexed quantification of the biomarkers. Three distinct FCNN models were developed, one model per cardiac marker, and in addition to inferring the biomarker concentration in the sample, the models were also used to optimize the subset of immunoreactions needed for each biomarker. The optimal spot configurations included spots specific to the target biomarker, as well as cross-reactive spots and control spots. The neural networks successfully learned from complex cross-reactive patterns between different immunoreactions, achieving superior performance compared to standard linear and polynomial regression models that were applied to the same patient samples. Quantification on the blind testing set, composed of 16 patient serum samples, showed a high correlation ($R^2 > 0.9$) with ground truth ELISA measurements for all three cardiac markers.

In addition, the xVFA platform processed by ML algorithms was applied for various serological testing applications, providing binary diagnostics of bacterial and viral diseases. For example, ML was employed to achieve accurate Lyme disease diagnostics using multiplexed serological testing on an xVFA platform^{18,102,103}. Conventional Lyme diagnostics involves a laborious two-tier testing procedure and has sensitivity limitations, particularly at the early stages of Lyme disease. To facilitate Lyme disease diagnostics without compromising the accuracy of two-tier testing, Joung et al. proposed an xVFA processed

by a smartphone-based reader and FCNN models for the serological testing of Lyme disease. This xVFA requires only 20 µL of patient serum and takes 15 min to operate¹⁸. The assay leveraged multiplexed detection of IgG/IgM antibodies across a panel of Lyme antigens, with ML utilized for two distinct purposes: (1) computational optimization of the assay design and (2) diagnostics of Lyme disease from patient serum samples. Computationally optimized assay achieved 90.5% sensitivity and 87.0% specificity on a blind testing set. Importantly, in addition to the improvements in diagnostics performance, computational assay optimizations also reduced the per-test cost, achieving 44% cost reduction when implementing the optimal subset of immunoreactions at the xVFA. Further improvement of xVFA performance for Lyme disease was achieved by transitioning to a peptide panel assay (Fig. 3d), resulting in 95.5% sensitivity and 100% specificity despite the presence of early-stage Lyme samples within the testing set¹⁰².

Another serological test implemented using the same xVFA platform was used to monitor human immunity levels in response to SARS-CoV-2 infection and vaccination, accurately identifying unprotected, protected, and infected cases¹⁰⁴. The multiplexed design of the xVFA platform allowed for the incorporation of multiple structural proteins of the virus within a single cartridge. The combined responses from these proteins were processed by FCNN models to classify patient immunity levels. Prior to training the final models, the assay was computationally optimized by selecting an optimal subset of the proteins for both IgG and IgM panels, resulting in the highest testing accuracy. The blind testing on 31 serum samples from 8 individuals showed an accuracy of 89.5%, enabling reliable tracking of the immune response dynamics over time. The integration of diverse SARS-CoV-2 proteins into a unified multiplexed panel represents a robust alternative to singleplex tests, with the potential to encompass broader populations, including those vaccinated with different vaccine types and exhibiting diverse immune responses.

Recent work has also demonstrated electrochemical-based VFAs, showcasing an example where the assay signals are processed by a field-effect transistor (FET) and deep learning for cholesterol detection in patient serum¹⁰⁵. In this setup, the cartridge comprises a paper membrane placed over an ion-sensitive sensing electrode, connected to the FET gate. When the sample reaches the membrane, cholesterol-specific enzymes produce protons in response to the cholesterol concentration, which is recorded in FET transfer curves captured repeatedly over a 5-min interval. These transfer curves are further converted into a two-dimensional (2D) heatmap reflecting the kinetic details of the enzymatic reactions. Lastly, a shallow FCNN model optimizes the interval of kinetic data carrying a concentration-specific response and quantifies the cholesterol concentration in serum using the optimal transfer curve signals. Quantification results on 30 blindly tested serum samples exhibited a strong correlation ($R^2 > 0.976$) with a CV of <7% when compared against the ground truth results from a CLIA-certified clinical laboratory. In the same work, the authors have also adopted this VFA platform to immunoassay format, potentially offering a wide range of applications in POCT¹⁰⁵.

As a broader category encompassing VFAs, array-based sensors represent advanced bio/chemical sensing platforms that consist of miniature arrays of distinct sensing elements systematically arranged on a substrate¹⁰⁶. These systems enable high-throughput and multiplexed detection, allowing simultaneous measurement of multiple analytes. VFAs can be considered a subset of array-based sensors, as they also employ reaction compartmentalization or multiple-spot arrays on the sensing membrane. However, a key distinction lies in the reagent delivery mechanism—while array-based sensors typically involve direct exposure of reagents in liquid or gas form to the sensor surface^{107–110}, VFAs rely on a controlled fluid delivery through stacked paper layers to transport samples and reagents to the sensing region. Each microdot within the array can independently capture and detect

specific analytes, producing complex, high-dimensional data outputs ideal for analysis with ML techniques. The combination of multiple dot arrays and ML leverages the array's capability for simultaneous, multi-analyte detection and enhances data interpretation by efficiently processing subtle variations in signal intensity, spatial distribution, or optical changes.

One example of ML-enhanced array-based sensing for POCT is a recent study by Kim et al., which introduced a fluorescent microarray designed for mobile reader-based high-throughput volatile organic compound (VOC) sensing¹¹⁰. This system employed 75 different fluorescent Kaleidolizine derivatives, immobilized on a wax-printed cellulose substrate, where each fluorophore exhibited unique fluorescence intensity and color shifts upon VOC exposure. To decode the complexity of these fluorescence variations, the study utilized a random forest algorithm to analyze hue differences in fluorescence patterns, achieving 97% classification accuracy across five VOCs. This work highlights the necessity of ML in multiplexed sensing applications, as the subtle fluorescence shifts could not be reliably interpreted using conventional threshold-based approaches.

Expanding upon this approach, Yang et al. applied a VOC-based array sensor for bacterial identification, leveraging the distinct metabolic signatures emitted by different bacterial species¹⁰⁹. Their study introduced an ML-enabled paper chromogenic array (PCA), consisting of 23 chromogenic dyes and dye combinations, which underwent colorimetric shifts upon exposure to bacterial VOC emissions. Unlike traditional microbiological techniques requiring extended culturing, this method enabled rapid and non-destructive bacterial detection. A neural network model was trained to classify bacterial strains based on digitized color changes, achieving 91–95% strain-specific identification accuracy. Notably, this system successfully distinguished between viable *E. coli*, pathogenic *E. coli* O157:H7, and *Listeria monocytogenes*, demonstrating its potential for food safety applications. While the imaging step in this study was performed using a desktop scanner, the system can be readily adapted for smartphone-based imaging, presenting strong potential for future POCT applications.

NAATs

NAATs detect the target genetic material of interest by amplifying it using methods such as polymerase chain reaction (PCR)¹¹¹, loop-mediated isothermal amplification (LAMP)¹¹², recombinase polymerase amplification (RPA)¹¹³, and rolling circle amplification (RCA)^{114,115}. These methods have become increasingly prevalent for COVID-19 detection and have been essential for molecular diagnostics over the past decade¹¹⁶, enabling precise detection and quantification of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Importantly, reverse transcription PCR (RT-PCR) is widely adopted as the gold standard for diagnosing viral infections and analyzing gene expression¹¹¹, due to its high accuracy and specificity. Alternative isothermal methods, such as LAMP, RPA, RCA, and CRISPR-mediated¹¹⁷ nucleic acid amplification, eliminate the need for thermal cycling and expensive instrumentation^{116,118}, making them more suitable for bedside or in-field settings¹¹⁹. Moreover, the integration of paper-based LFAs and microfluidic devices further enhances the accessibility and practicality of NAATs¹²⁰.

Despite the merits of NAATs summarized above, these technologies have been hindered by relatively lengthy assay times, low signal intensities at low copies of target sequences, lack of objective and automated result interpretation, and reliance on benchtop equipment¹²¹. To overcome these challenges and streamline the deployment of testing in point-of-care settings, NAATs have been adopted to include four main steps: (i) the collection of sample swabs (e.g., from nasopharyngeal [NP] cavities), (ii) the introduction of samples to portable diagnostic devices, (iii) the amplification of nucleic acids using custom thermal modules, and (iv) the prediction or classification of positive and negative samples through ML-integrated

optical mobile devices^{122,123}. Integration of AI/ML into these portable readout devices has been crucial for advancing NAATs and decentralizing their applications through POCT¹²⁴.

The integration of AI/ML into NAATs can address the limitations mentioned above by significantly reducing assay time, eliminating subjective result interpretation, and maintaining high accuracy¹²⁴. Many POCT NAAT diagnostic tools are equipped with comprehensive AI models such as CNN¹²², long short-term memory (LSTM)²⁰, recurrent neural network (RNN)²¹, transformer¹²⁵, and gated recurrent unit (GRU)¹²⁶. These AI models enable advanced image processing, sample classification, early result prediction, and high-throughput automated screenings¹²⁷. The deployment of these AI models enables the analysis of diagnostic results at lower signal intensity levels, conversion of qualitative metrics into quantitative data for early prediction, and achievement of high accuracy, sensitivity, and specificity within much shorter assay times (e.g., <20 min^{125,126,128}). The specific examples are outlined in Table 3 and in the next paragraphs of this section.

ML-integrated POCT NAAT systems have significantly advanced after the COVID-19 pandemic, driven by the critical need for cost-effective, accurate, sensitive, and high-throughput screening of viral infections¹¹⁸. Recent studies have demonstrated the successful incorporation of AI/ML into these systems with various sensing modalities (e.g., colorimetric^{122,127}, fluorescent^{20,125,126}, electrochemical¹²⁹, and SERS²¹) for SARS-CoV-2 RNA detection (Fig. 4), showcasing the versatility and general learning capabilities of modern AI and ML models¹³⁰. For example, Rohaim et al. developed a hand-held AI-LAMP system for SARS-CoV-2 RNA (*RdRP* gene) detection, which automated image acquisition and processing and reduced assay time and subjectivity of colorimetric LAMP (Fig. 4a)¹²². This system utilized CNN to classify colors more accurately, achieving 98% accuracy and significantly higher sensitivity than the gold standard quantitative RT-PCR (qRT-PCR). Furthermore, Jaroenram et al. developed a dual RT-LAMP assay that utilized both colorimetric analysis and a deep learning detection transformers (DETR)-integrated analysis tool (RT-LAMP-DETR) on a smartphone for ultrasensitive COVID-19 detection (Fig. 4b)¹²⁷. In addition to pH-sensitive indicators that provided a visual indication of results, high-throughput analysis of these colorimetric results was supported by RT-LAMP-DETR through cross-comparison. This result interpretation achieved 100% accuracy, sensitivity, and specificity, validating a scalable method for the screening of COVID-19 suitable for low-resource settings.

Additionally, Tripathi et al. introduced a low-cost lab-on-chip platform that transformed ion-sensitive field-effect transistor (ISFET) data into spectrograms compatible with CNNs to identify nucleic acid amplification (Fig. 4c)¹²⁹. This method, in addition to efficiently identifying infectious diseases and cancer biomarkers, achieved 84.84% accuracy with a 30kB-sized CNN model, facilitating its deployment on edge computing devices. Similarly, Sun et al. developed a portable optoelectronic system interfaced with paper microfluidics and deep learning for the real-time detection of SARS-CoV-2 (Fig. 1f)²⁰. This device transfers real-time data from fluorescent signals of amplified viral sequences to RNN, LSTM, and GRU models, enabling early predictive analysis and reducing assay time by 45%. This model demonstrated robust outcomes for NAATs, with AI-integrated early prediction achieving 98.1% accuracy, 97.6% sensitivity, and 98.6% specificity.

Moreover, a system reported by Sun et al. integrated deep learning with μ PADs for the rapid and accurate detection of SARS-CoV-2 RNA (*ORF1ab* gene), using quantitative PCR (qPCR) data and AI predictive analysis (Fig. 4d)¹²⁶. The analysis was driven by RNN, GRU, and LSTM models to derive qualitative forecasting from real-time PCR analytics of patient samples. The GRU was found to be the most accurate in predicting end-point values and trends of qPCR curves, with a mean absolute percentage error of 2.1%. The accurate predictions made as early as after 13 amplification cycles accelerated the

Table 3 | Applications of ML to NAATs

Ref. no.	Sensing modality	Analytes/targets	Specimens	AI/ML	Reader		Cost	Performance metrics		
					Purpose	Portability		Size and weight	Diagnostic indicators	Analytical indicators
20	Fluorescent (Calcein)	SARS-CoV-2 RNA (ORF1ab, N, and E genes)	n/a ^a	Attention-based RNN, LSTM, Bi-LSTM, and GRU	Result prediction, Assay time reduction	Yes (Custom mobile system with thermal module and reader)	n/a	12.9 × 8.9 × 12 (cm), 0.78 kg	Sensitivity: 97.6%, Specificity: 98.6%, Accuracy: 98.1%	n/a
21	SERS	SARS-CoV-2 RNA	NP swabs	RNN	Result classification	Yes (Portable Raman spectrometer)	n/a	7 kg	Overall accuracy: 98.9%, Accuracy for positives: 97.2%, Accuracy for negatives: 100%	Detection range: 10 ³ –10 ⁵ copies/mL
22	Colorimetric (Streptavidin labeled red particle)	<i>Plasmodium</i> sp. causing malaria	Whole blood	CNN, ResNet50	Result interpretation, Local decision support, Secure data connectivity and management	Yes (Smartphone-based)	~\$20/test	<0.5 kg	Accuracy: 97.83%	n/a
122	Colorimetric (Phenol red)	SARS-CoV-2 RNA (RdRP gene)	NP swabs	CNN	Result classification, Assay time and subjectivity reduction	Yes (Custom mobile system with thermal module and reader)	n/a	14.3 × 10.8 × 6 (cm), 0.78 kg	Accuracy: 98%	LoD: 100 copies/reaction
125	Fluorescent (Calcein)	SARS-CoV-2 RNA (ORF1ab, N, and E genes)	n/a	Vanilla RNN, LSTM, Bi-LSTM, GRU, and Transformer	Result prediction, Assay time reduction	Yes (Custom mobile system with thermal module and reader)	\$0.5 (chip), \$161 (device)	12.9 × 8.9 × 12 (cm), 0.78 kg	Sensitivity: 97.6%, Specificity: 99.1%, Accuracy: 98.6%	n/a
126	Fluorescent (FAM, ROX, HEX)	SARS-CoV-2 RNA (ORF1ab gene)	n/a	RNN, LSTM, GRU	Result prediction, Assay time reduction	No (Benchtop qPCR machine and reader)	~\$0.24 (uPAD)	n/a	Absolute percentage error (MAPE): 2.1%	n/a
127	Colorimetric (Xylenol Orange and lavender green)	SARS-CoV-2 RNA (ORF1ab gene) and human 18 S rRNA	NP swabs	ResNet50, Detection transformer (DETR)	Result Interpretation	No (Benchtop thermal cyclor)	\$8/test	n/a	Sensitivity: 100%, Specificity: 100%, Accuracy: 100%	LoD: 50 copies/reaction
128	SERS	SARS-CoV-2 RNA	NP swabs	ResNet	Enhance sensitivity and specificity, Assay time reduction	No (Benchtop Raman spectrometer)	n/a	n/a	Sensitivity: 100%, Specificity: 100%, Accuracy: 100%	LoD: DNA 1 × 10 ³ CFU/mL (<i>Bruceella ovis</i>), RNA 0.96 × 10 ⁻¹ PFU/mL (SARS-CoV-2 cultures)
129	Electrochemical (H ⁺ sensing via ISFET arrays)	SARS-CoV-2 RNA (N gene), Cancer biomarkers (YAPI, AR-V7 mRNAs, and ESRI DNA)	n/a	Spectrogram-based 2D CNN, 1D CNN, FCN, Inception Time, ResNet, Autoencoder	Classification of chemical signals	Yes (Lab on chip platform)	n/a	n/a	Accuracy: 84.84%, Precision: 85.01%, Recall: 84.43%, F1 score: 84.72% (Inter-biomarker testing, Fold3)	n/a

Table 3 (continued) | Applications of ML to NAATs

Table 3 (continued) Applications of ML to NAATs											
Ref. no.	Sensing modality	Analytes/targets	Specimens	AI/ML	Reader		Cost	Performance metrics			
					Purpose	Portability		Size and weight	Diagnostic indicators	Analytical indicators	Time to result
132	Fluorescent (Cas-Loaded Annotated Micro-Particles, CLAMP)	Human Papilloma-virus DNA (HPV16 and HPV18)	Cervical brushing samples	Mark R-CNN	Object recognition and segmentation	Yes (Smart-phone-based)	n/a	14.8 × 7.2 × 0.8 cm, 0.16 kg	Accuracy: 97.9% F1 score: 97.8%	LoD: 2 aM (1.2 copies/ μ L)	~2 h
134	Bubble-based (PtNPs conjugated with anti-Cas9 antibody)	Zika virus (ZIKV) RNA	Serum	Adaptive adversarial learning, ImageNet	Generate augmented images, Image classification	Yes (Smart-phone-based)	n/a	Motorola MotoX: 12.9 × 6.5 × 1.0 cm, 0.13 kg Apple iPhone 8: 13.8 × 6.7 × 0.7 cm, 0.15 kg	Sensitivity: 100%, Specificity: 100%, Accuracy: 100%	LoD: ~400 aM	~1 h

^an/a indicates not available in the associated reference.

NAAT procedure, reducing the assay time to 12 min and allowing better preparedness for future disease outbreaks.

Similarly, Yang et al. combined SERS sensors with deep learning algorithms for the rapid detection of SARS-CoV-2 RNA in human NP swab specimens (Fig. 1g)²¹. A silver nanorod (AgNR) array sensor functionalized with DNA probes was used for detecting viral RNA. Specifically, SARS-CoV-2 RNA selectively hybridized with complementary DNA sequences immobilized on the AgNR substrate, inducing spectral changes detectable by SERS. These spectral shifts were then analyzed using an RNN model, which classified the results with an accuracy of 97.2% for positive specimens and 100% for negative specimens, providing results within 25 min. As an integration of paper microfluidics with NAATs, Sun et al. introduced an approach that combined paper microfluidics with deep learning and cloud computing for accelerated SARS-CoV-2 RNA analysis¹²⁵. Real-time amplification of synthesized RNA templates was performed on paper materials, and the time-series data were transmitted to a cloud server with preloaded deep learning models for predictive analysis, achieving clinical accuracy, sensitivity, and specificity of 98.6%, 97.6%, and 99.1%, respectively.

Instead of generating numerous copies of target genes to amplify signals, CRISPR-mediated nucleic acid detection leverages the collateral cleavage activity of CRISPR-associated enzymes, where abundant non-target signal reporters are cleaved to produce a detectable signal¹³¹. A notable advancement in this field was reported by Roh et al.¹³², who developed a CRISPR/Cas12-based multiplexed nucleic acid detection platform utilizing spatially encoded hydrogel microparticles (HMPs). This study integrated ML algorithms to automate the analysis of individual HMPs, employing neural networks to recognize/classify coded HMPs and perform segmentation for fluorescence detection. Upon binding of Cas12a/gRNA complexes to human papillomavirus RNA in cervical brushing samples, the HMPs fluoresced and subsequently captured within a microfluidic device, where a Mark R-CNN model automatically identified them. The trained model achieved a high accuracy of 97.9% and an F1 score of 97.8% in the validation tests, successfully differentiating four distinct HMP types. This rapid assay demonstrated an attomolar LoD of 2 aM (equivalent to 1.2 copies/μL), highlighting its high sensitivity.

Beyond CRISPR-mediated approaches, nanoparticle-based virus detection has also gained attention. Draz et al.¹³³ developed a CNN-enabled smartphone system for detecting intact viruses on a microchip using the nanocatalytic activity of platinum nanoparticles (PtNPs). PtNPs catalyze the decomposition of hydrogen peroxide into oxygen gas, generating visual bubble patterns upon immuno-capturing virus particles (e.g., Zika virus [ZIKV], hepatitis B virus, and hepatitis C virus) on a microchip. Building upon these advancements mentioned above, Shokr et al.¹³⁴ integrated CRISPR-based detection with nanoparticle technology by conjugating an anti-Cas9 antibody with PtNPs, enabling ZIKV RNA detection. This system facilitated reliable bubble generation within the microchip, achieving a LoD of 400 aM for ZIKV detection. Furthermore, this platform incorporated adaptive adversarial learning and generative adversarial networks to enhance its adaptability to emerging pathogens, allowing rapid diagnostic reconfiguration without extensive retraining. Additionally, the system augmented real smartphone-taken image datasets to improve generalizability, making it highly suited for low-cost, smartphone-based diagnostics. Collectively, these ML-integrated NAAT advancements provide highly sensitive, portable, and scalable molecular diagnostic solutions, significantly contributing to epidemic preparedness by offering low-cost, rapid, and accurate nucleic acid detection methods.

In addition to SARS-CoV-2 RNA detection, ML-integrated NAAT technologies have been applied to detect other targets such as bacteria DNA²², cancer biomarkers¹³⁵, and λ DNA¹³⁶. For example, Guo et al. proposed a smartphone-based DNA diagnostic tool for malaria that integrated CNN models for local decision support and blockchain

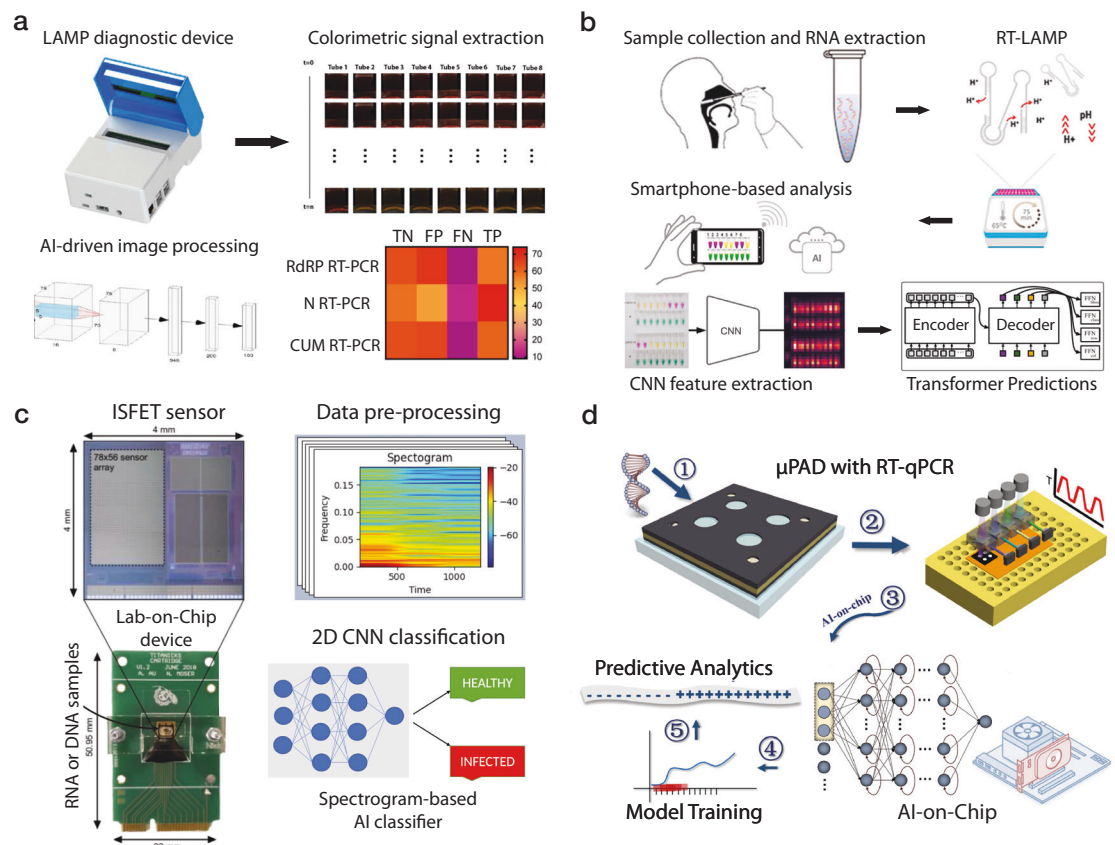


Fig. 4 | Applications of ML to NAATs. **a** A hand-held AI-LAMP device for rapid detection of COVID-19 with AI-based image analysis reduced the sample-to-answer time and improved signal interpretation. **b** A one-step smartphone-based colorimetric RT-LAMP COVID-19 screening method enabled by pH-sensitive dyes and a transformer AI model. **c** A lab-on-chip nucleic acid amplification device that utilized ISFET arrays and a spectrogram-based CNN to classify COVID-19 and three cancer

biomarkers, featuring a compact size and improved accuracy. **d** An AI-aided on-chip μ PAD for COVID-19 detection using three neural network models (i.e., RNN, LSTM, and GRU) for qualitative analysis. **a** This is adapted with permission from ref. 122 by MDPI; **b–d** These are adapted with permission from refs. 126,127,129, respectively, by Elsevier.

technology for secure data management and reporting (Fig. 1h)²². This system used streptavidin-labeled red particles to detect target sequences for a colorimetric readout and employed a CNN model to interpret the results, achieving 97.83% accuracy. A key feature of this system was its blockchain-enabled data security, ensuring tamper-proof record-keeping and controlled access to diagnostic results. Unlike traditional cloud-based methods, blockchain technology provides a low-power, cost-effective alternative for securely transmitting medical data in resource-limited settings. This approach eliminated the need for specialized infrastructure, making it particularly suitable for decentralized healthcare applications. The system was successfully field-tested in rural Uganda, where it exhibited high diagnostic accuracy (correctly identifying over 98% of tested cases), reliable data transfer, and seamless integration into local healthcare workflows. By combining ML-driven image analysis with blockchain-based security, this platform provides a scalable, privacy-preserving framework for real-time disease surveillance and secure diagnostics in underserved regions.

Furthermore, AI/ML technologies have also been used for digital NAATs. Digital NAATs have the unique advantage of extremely high sensitivity (usually 1–10 copies/ μ L) to their target sequences and absolute quantification without the need for standard curve-based calibration¹³⁷. For example, Li et al. developed an all-in-one OsciDrop digital PCR/LAMP system that can perform multiplexed fluorescent quantification of cancer biomarkers (*HER2* and *EGFR* genes)¹³⁵. U-Net and MobileNet models were used for ML-integrated droplet image analysis, enabling highly accurate and consistent nucleic acid

concentration quantifications (CV < 1.5%). Another example of digital NAATs is Fractal LAMP, which utilized a computer vision algorithm to achieve high accuracy for the detection of amplified DNA by recognizing the LAMP byproducts that form fractal structures¹³⁶. A Bayesian model and bootstrapping method facilitated high-accuracy concentration measurements over 3 orders of magnitude. Although these assays and detection procedures may still require benchtop equipment, their potential for miniaturization into portable devices renders them as emerging POCT technologies. In the future, combined progress in device engineering, along with the high sensitivity and absolute quantification ability offered by digital ML-integrated NAATs will substantially expand the application of precision medicine, even in rural areas and low-income countries.

Imaging-based point-of-care sensors

Recent advancements in AI and ML have significantly enhanced image-based diagnostics for various diseases^{138–145}. For example, deep learning models, particularly CNNs and ViTs, have achieved dermatologist-level accuracy in skin cancer detection and lesion classification, leveraging large, annotated datasets to improve diagnostic precision^{138–140}. As another example, AI is being integrated into echocardiography, cardiac magnetic resonance imaging, and coronary computer tomography angiography, enabling automated segmentation, disease detection, and risk stratification with performance comparable to human experts^{143–145}. These advancements and many others underscore the growing role of AI-powered image analysis in enhancing diagnostic accuracy, improving workflow efficiency, and enabling

early disease detection, demonstrating its transformative impact on clinical decision-making. Similarly, AI-driven imaging technologies are also revolutionizing point-of-care diagnostics. Integrating AI with imaging technologies at the point-of-care has significantly transformed diagnostics and treatment within healthcare environments, especially in scenarios where rapid decision-making is essential, such as emergency rooms, rural clinics, and field medical assessments¹⁴⁶. In recent years, point-of-care applications in microbiology and pathology, among others, have been rapidly expanding, benefiting from the integration of AI tools¹⁴⁷. The primary applications of AI in imaging at the point-of-care include shortening of the diagnostic time, image quality enhancement, and increasing the accuracy and portability of the imaging hardware. These advancements not only make imaging technologies more accessible to underserved populations but also reduce the cost associated with conventional benchtop equipment³¹. Some of the specific examples of imaging-based point-of-care sensors are outlined in Table 4.

Bacteria detection. In a recent study, Wang *et al.* proposed a deep learning-based imaging system for rapid detection of bacteria colonies, enabling real-time monitoring and early detection of bacteria species¹⁴⁸. This computational microscopy system periodically captured holographic images of bacteria growth in a petri dish and analyzed these time-lapsed holograms using CNN models, accurately identifying early-stage bacteria colonies (Fig. 5a). In a proof-of-concept demonstration, the device detected three bacteria species, namely *E. coli*, *K. aerogenes*, and *K. pneumoniae*, achieving >95% sensitivity in identifying bacteria colonies and shortening the detection time by >12 h compared to standard Environmental Protection Agency (EPA)-approved methods. In addition, the proposed platform showed a LoD of approximately 1 colony forming unit (CFU)/L within only 9 h of testing, making it an appealing tool for microbiology research and related sensing applications.

Virus detection. Furthermore, a compact label-free live plaque assay was developed to provide a significantly faster plaque-forming unit (PFU) detection for the quantification of viruses²³. A compact lens-free holographic imaging system reconstructed phase images of the target PFUs during the incubation period. The imager scanned the entire area of a six-well plate every hour and utilized a DenseNet neural network-based classifier to convert the phase images of the samples into PFU probability maps, identifying the locations and sizes of the PFUs within the well plate. This stain-free method was capable of automatically detecting the first cell-lysing events caused by vesicular stomatitis virus (VSV) replication as early as 5 h after incubation and achieved a PFU detection rate exceeding 90% in <20 h, providing significant time savings compared to traditional plaque assays, which typically require over 48 h for testing (Fig. 5b).

Parasite imaging. Point-of-care imaging applications have also been developed for diagnosing significant tropical diseases such as schistosomiasis and malaria, which require precise diagnosis typically through microscopy techniques¹⁴⁹. The diagnosis of schistosomiasis relies on bright-field microscopy to identify *Schistosoma haematobium* (*S. haematobium*) eggs in urine samples, where the operator's skill is crucial, especially for mild infections. Similarly, malaria diagnosis involves imaging parasites in blood. Automated miniaturized digital microscopes such as the Schistoscope have been developed to identify *S. haematobium* eggs in real-life biological samples. In recent work, the Schistoscope was applied to automatically detect and quantify *S. haematobium* eggs in urine samples using ML-based algorithms, achieving over 80% sensitivity¹⁵⁰. Furthermore, the EasyScanGo system was utilized for the detection of malaria and employed a CNN-based algorithm for the detection of malaria parasites in blood smears, achieving 91.1% sensitivity and 75.6% specificity, thereby matching the

accuracy of experienced microscopists¹⁵¹. These AI-enhanced tools can be applied for automated screening of tropical diseases in underserved populations and low-income countries that lack highly qualified personnel to perform manual diagnosis.

Another platform for detecting parasites in bodily fluids was developed by Zhang *et al.*¹⁵² Unlike traditional methods, this technique exploited the movement of self-propelling parasites as a natural biomarker and contrast mechanism. The sample was illuminated with a coherent light source, and a CMOS image sensor placed underneath the sample captured the time-lapse holographic speckle patterns. These recorded patterns were analyzed using a custom computational motion analysis (CMA) algorithm, which employed holography to create 3D contrast map highlighting the parasites' movements in the sample. A deep learning-based classifier based on a CNN model was then used to detect and count the parasites from the reconstructed 3D locomotion map. The proposed platform was applied for the detection of trypanosome parasites and showed a strong correlation between detected and spiked parasite concentrations with a LoD of only 3 parasites per 1 mL of biological fluid.

Waterborne parasites, including *Giardia lamblia*, affect 200 million people yearly, causing diarrheal illnesses such as Giardiasis¹⁵³. A study by Göröcs *et al.* introduced a portable, label-free imaging flow cytometer to detect and enumerate *Giardia lamblia* cysts in real-time in liquid samples¹⁵⁴. This device uses lens-free on-chip holographic microscopy to analyze continuously flowing samples at a throughput of 100 mL/h. As the samples flow through the channel, holograms are captured and reconstructed in real-time on a laptop, providing color intensity and phase images. These images are then automatically processed by a CNN model, which digitally sorts and counts the images containing *Giardia lamblia* cysts. This field-portable imaging flow cytometer can detect less than 10 cysts per 50 mL of sample and can be applied for water quality monitoring in low-resource settings and to detect parasites in bodily fluids. The same imaging flow cytometer was further used for the identification of different plankton types in ocean water. (Fig. 5c)¹⁵⁵. Additionally, Koydemir *et al.* developed a mobile phone-based fluorescence microscopy system using a custom ML algorithm based on bootstrap aggregating to detect *Giardia lamblia* cysts¹⁵⁶. This system consisted of a smartphone coupled with an optomechanical attachment weighing only 205 grams (Fig. 5d). It utilized a hand-held fluorescence microscope aligned with the smartphone's camera unit to image custom-designed disposable water sample cassettes. This mobile phone-based microscopy technique had a LoD of 12 cysts per 50 mL of water sample and showed ~84% sensitivity and >76% specificity for the detection of *Giardia* cysts.

Hematology analysis. Some AI-based blood sample imaging techniques for point-of-care settings involve coupling miniaturized microscopes with benchtop computers for sample analysis. For instance, complete blood counting (CBC) faces challenges in differentiating between cell sizes and morphologies due to variations in cell maturity. To address this issue, Sight OLO (Sight Diagnostics, Israel) developed a compact microscopy system for hematology in point-of-care settings, providing differentiated five-part CBC through computer vision and AI. This machine-vision technology differentiates cells by extracting their unique peculiarities through CNN models (Fig. 5e)¹⁵⁷. The platform was compared with a standard Sysmex XN hematology analyzer and demonstrated strong concordance with the ground truth measurements (Pearson's $r > 0.95$) for all major CBC parameters. Another challenge in the field of blood analysis is identifying abnormal erythrocytes, platelets, and leukocytes in the blood smear. For this reason, Chen *et al.* introduced a label-free contrast-enhanced defocusing imaging (CEDI) and machine vision for instant and on-site diagnostic applications¹⁵⁸. The proposed platform identified abnormal blood components through CNN-based algorithms applied to both bright-field and fluorescence microscopy images of the smears (Fig. 5f).

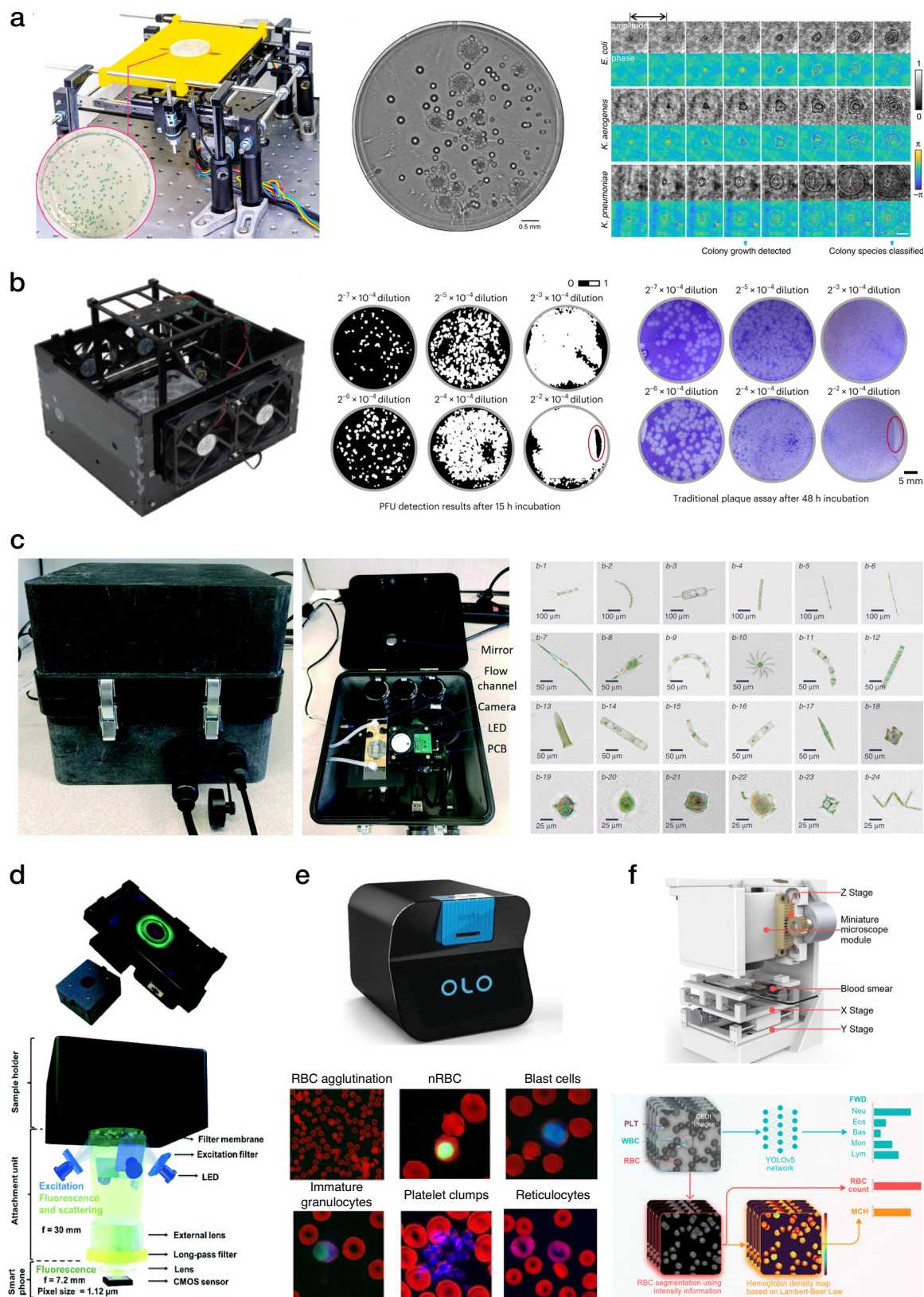
Table 4 | Applications of ML to imaging-based sensors

Ref. no.	Sensing modality	Analytes/targets	Specimens	AI/ML	Purpose	Reader	Performance metrics				
				Method		Portability	Size and weight	Cost	Diagnostic indicators	Analytical indicators	Time to result
23	Lens-free holographic microscopy	Vesicular stomatitis virus (VSV), Herpes simplex virus type 1 (HSV-1), Encephalomyocarditis virus (EMCV)	Virus stock samples	CNN (pseudo-3D DenseNet models)	Detection of viruses, Reduction of detection time	Yes (Custom benchtop lens-free microscope)	-32 × 26 × 28 (cm)	~\$880 (Imaging device)	Sensitivity for PFU detection: 93.7% (VSV), 90.4% (HSV-1), 90.8% (EMSV)	n/a ^a	20 h for VSV, 72 h for HSV-1, 52 h for EMCV
148	Lens-free holographic microscopy	Escherichia coli, Klebsiella aerogenes, Klebsiella pneumoniae	Water	CNN (pseudo-3D DenseNet models), Differential analysis	Detection of bacteria growth, Detection of bacteria colonies, Shortening of detection time	Yes (Custom benchtop lens-free microscope)	~40 × 40 × 150 (cm)	~\$0.6/test	Sensitivity: 90% (7–10 h), >95% (12 h), Precision: 99.2–100% (for detection of bacterial colonies)	LoD: ~11 CFU/L (at 8.5 h) and ~1 CFU/L (at ≤ 9 h).	≤9h
150	Brightfield digital microscopy	S. haematobium eggs	Urine	CNN (UNET), Linear regression	Automated detection of S. haematobium eggs, Counting eggs	Yes (Compact benchtop brightfield microscope)	n/a ^a	~\$700	Sensitivity: ~80%	n/a	n/a
151	Brightfield digital microscopy	Malaria parasite	Whole blood	CNN	Parasite detection and count	Yes (Compact benchtop brightfield microscope)	n/a	n/a	Sensitivity: 91.1%, Specificity: 75.6%	LoD: 100–150 parasite/μL	≤30 min
152	Lens-free holographic microscopy	Trypanosomes, T. vaginalis	Buffer, Whole blood, CSF	CNN, Differential analysis	Detection and count of motile trypanosomes	Yes (Custom benchtop lens-free microscope)	26.4 × 18.3 × 14.1 (cm), 1.69 kg	<\$1850	n/a	LoD (Trypanosomes): 10 parasites/mL (whole blood) and 3 parasites/mL (CSF); LoD (T. vaginalis): ≤3 parasites/mL (Buffer)	~20 min
154	Lens-free holographic microscopy	G. lamblia cysts	Water, Bodily fluids	CNN	Counting of waterborne parasites in real-time	Yes (Custom benchtop lens-free flow cytometer)	19 × 19 × 16 (cm), 1.6 kg	<\$2500	Recovery rate: ~83% (water), ~74% (seawater); Standard deviation: ~ 7% (water), ~6% (seawater)	LoD: <10 cysts/50 mL	1 h per 100 mL, <30 ms per object
155	Lens-free holographic microscopy	Microscopic objects in liquids	Water, Bodily fluids	CNN	Detection of microscopic objects in liquids	Yes (Custom benchtop lens-free flow cytometer)	15.5 × 15 × 12.5 (cm), 1 kg	<\$2500	n/a	LoD: ~700 objects/mL	1 h per 100 mL, ~250 ms per object
156	Mobile-based fluorescent microscopy	G. lamblia cysts	Water	Bootstrap aggregation/bagging	Detection and enumeration of G. lamblia cysts	Yes (Smartphone-based microscope)	14.5 × 8.5 × 7 (cm), 0.205 kg	n/a	Sensitivity: ~84%, Specificity: > 76%	LoD: ~12 cysts/10 mL	1 h
157	Brightfield and fluorescent microscopy	RBC, WBC, Platelets	Whole blood	Morphological segmentation, CNN	Detection of RBC, WBC, and platelets, Feature extraction of RBC and WBC	Yes (Miniaturized microscope)	32.2 × 28.4 × 25.4 (cm), ~9.9 kg	n/a	Pearson's r between proposed method and ground truth: 0.997 (WBC), 0.991	n/a	n/a

Table 4 (continued) | Applications of ML to imaging-based sensors

Ref. no.	Sensing modality	Analytes/targets	Specimens	AI/ML Method	Purpose	Reader		Cost	Performance metrics		
						Portability	Size and weight		Diagnostic indicators	Analytical indicators	Time to result
158	Contrast-enhanced defocusing imaging	RBC, WBC, Hemoglobin	Whole blood	Morphological segmentation, CNN (YOLOv5), UMAP	Cell detection and count, Feature extraction	Yes (Miniaturized microscope)	10.5 × 7.7 × 6.4 (cm), 0.314 kg	~\$297	Pearson's r between proposed method and ground truth: >0.922 (WBC), 0.863 (hemoglobin), and 0.803 (RBC)	n/a	10 min
159	Bright-field microscopy	WBC, Platelets, Platelet aggregates	Blood cell suspension	CNN, random forests, penalized linear discriminant analysis (PLDA) classifier	Classification between WBC, platelets and platelet aggregates, cell morphology, COVID-19 detection	No (FDM microscope)	n/a	n/a	75.3% accuracy for detection of COVID-19	n/a	<5 s per measurement
160	OTS microscopy	Platelet aggregates	Whole blood	CNN	Classification between platelet aggregates based on agonist type	No (OTS microscope)	n/a	n/a	77% accuracy	n/a	n/a
161	Smartphone-based bright-field microscopy	RBC, Sick cells	Whole blood	CNN (U-net)	Image enhancement, Cell detection and count, Classification of SCD (+) and SCD (-) samples	Yes (Smartphone-based microscope)	14.5 × 9 × 8 (cm), 0.35 kg	\$60	Accuracy for SCD diagnostics: 98%, AUC: 0.998	n/a	<7 s
162	Smartphone-based microscopy, colorimetric	AST against <i>Klebsiella pneumoniae</i> bacteria	Clinical bacteria isolates	Statistics-based	Estimation of threshold to determine bacterial growth	Yes (Smartphone-based microscope)	19.5 × 9.8 × 10 (cm), ~0.62 kg	~\$100	Well turbidity detection accuracy: 98.21%, MIC accuracy: 95.12%, Drug susceptibility interpretation accuracy: 99.23%	n/a	≤24 h
163	Colorimetric	AST against <i>Staphylococcus aureus</i> bacteria	Clinical bacteria isolates	FCNN	Prediction of bacteria growth	Yes (Raspberry Pi-based microscope)	17.5 × 45 × 19.2 (cm)	<\$500	Bacteria growth prediction accuracy: 90% (after 7 h), 95% (after 10.5 h)	n/a	≤10 h
164	Colorimetric and fluorescence	<i>E. coli</i> , total coliform	Water, Bodily fluids	Statistics-based	Estimation of threshold to determine bacteria	Yes (Raspberry Pi-based microscope)	1.66 kg	~\$600	n/a	LoD: 1 CFU/100 mL	~16 h

*n/a indicates not available in the associated reference.



Several blood analysis platforms powered by deep learning were developed using flow cytometry techniques^{159,160}. For instance, Zhang et al. utilized a microfluidic imaging flow cytometer coupled to a CNN to detect COVID-19 by analyzing platelet formation in patient blood samples¹⁵⁹. The flow cytometer continuously captured bright-field images of blood cell suspensions, and the CNN further

classified captured samples between single platelets, platelet aggregates, and white blood cells. Subsequently, a random forest model extracted cell features from the images and a linear discriminant analysis-based classifier used these features to distinguish between COVID-19-related and non-COVID-19 thrombosis, achieving an accuracy of 75%.

Fig. 5 | Applications of ML to imaging-based sensors. **a** Image of the holographic microscopy system for bacteria detection (left), whole agar plate image of mixed *E. coli* and *K. aerogenes* colonies (middle), and amplitude and phase images of the individual growing colonies (right). **b** Image of the holographic system for PFU imaging (left), whole plate comparison between the stain-free viral plaque assay after 15 h and the traditional plaque assay after 48 h (right). **c** Image of the field-portable lens-free imaging flow cytometer (left), which can be used for water quality monitoring and to detect parasites in bodily fluids; the reconstructed images of different microplankton species captured using this portable imaging cytometer (right). **d** Image of the smartphone-based fluorescent microscope with the disposable sample cassette (top), and schematic illustration of the emission and

excitation paths (bottom). **e** Schematic illustration of the Sight OLO hematology analyzer (top), and false-colored micrographs of different anomalous cell types and formations captured by OLO, red channel: hemoglobin absorption; green channel: nuclear DNA fluorescence; blue channel: cytoplasmic staining (bottom). **f** Schematic illustration of the miniaturized microscope for automated blood analysis (top), and AI-driven quantification pipeline for FWD, RBC, and MCH count (bottom). **a** This is adopted with permission from ref. 148 by LSA; **b** This is adapted from ref. 23 with permission from Nature BME; **c** This is adapted with permission from ref. 155 by LSA; **d** This is adapted with permission from ref. 156 by Lab on a Chip; **e** This is adapted with permission from ref. 157 by Wiley; **f** This is adapted with permission from ref. 158 by Analyst.

Furthermore, a UNet-based imaging platform for blood analysis was presented by de Haan et al. for automated screening of sickle cells in blood smears using a smartphone-based microscope¹⁶¹. This system comprises two distinct but complementary deep neural networks. The first network enhances and standardizes the blood smear images taken with the smartphone microscope to match the spatial and spectral quality of laboratory-grade microscope images. The second network then uses the enhanced images to perform semantic segmentation, differentiating between healthy and sickle cells in the blood smear. This image processing pipeline takes <7 s per blood smear slide and allows for accurate detection of sickle cells within the sample, achieving ~98% accuracy for the diagnosis of sickle cell disease.

Antimicrobial susceptibility testing. A few AI-driven imaging systems have been proposed for antimicrobial susceptibility testing (AST), which quantifies the efficiency of antibiotics against bacterial infections in patients^{162,163}. AST procedure relies on turbidity measurements of individual wells within a 96-well microplate, often requiring expensive hardware with specialized objectives and mechanical scanning components to capture the entire microplate, limiting applications of AST in resource-limited settings. Recently, fiber optics-based platforms have been proposed as a more cost-effective alternative. These systems utilize LED modules to illuminate the entire well plate and further employ optical fiber bundles to couple transmitted light from the wells to a camera, reducing both the footprint and cost of the device. For instance, Feng et al. proposed a smartphone-based microplate reader that uses a fiber bundle to deliver transmitted light from individual wells to a smartphone camera¹⁶². In this work, an ML-based algorithm was applied to determine the optimal threshold for well turbidity detection and to conduct AST in a blind-testing manner. The reader was tested for AST against *Klebsiella pneumoniae* bacteria, achieving an average well turbidity detection accuracy of 98.21% and a minimum inhibitory concentration accuracy (MIC) of 95.12% when blindly tested on 39 patient isolates. In addition, Brown et al. developed a compact fiber-based system for accelerated AST, reducing the incubation time by at least 8 h compared to the gold standard manual inspection method¹⁶³. The well plate was periodically illuminated by an LED array with a 15-min time interval and a fiber bundle coupled the light transmitted by individual wells to a compact camera, connected to a Raspberry Pi computer. Each well was processed by 21 fibers, capturing the spatial distribution of turbidity within the well, which provided more accurate turbidity estimations compared to configurations with a single fiber per well¹⁶². FCNN models processed signals from all 21 fibers to detect antibiotic response to *Staphylococcus aureus* bacteria over time, achieving 90% detection accuracy after 7 h and 95% accuracy after 10.5 h, considerably faster than the gold standard method, which typically requires 18–24 h. In another application, an automated fiber optics-based microplate reader was applied to accelerate the detection of bacterial growth in water¹⁶⁴. This reader combined fluorescent modality for *E. coli* with colorimetric modality for total coliform and used a Raspberry Pi camera to periodically capture fluorescent and colorimetric images of a 40-well microplate. Bacterial growth was detected

by monitoring the intensity increase over time, based on an empirically optimized threshold, enabling the detection of *E. coli* and total coliform within <16 h, 8 h faster than the conventional approach, with a sensitivity of 1 CFU per 100 mL.

Performance benchmarking of ML models in POCT

Studies to date demonstrate that integrating ML with POCT devices can substantially improve diagnostic performance, but the optimal choice of the ML algorithm to be used often depends on the dataset size, number of features, and task complexity (Supplementary Table S1 in Supplementary Information). When datasets are small, and the feature space is limited, traditional ML models, including random forests, SVMs, and logistic regression, often remain the preferred choice due to their simplicity, faster training times, and reduced risk of overfitting. For instance, Kim et al. demonstrated that random forests outperformed neural networks in classifying prostate cancer using urinary biomarkers in a POCT setting¹⁶⁵. However, as the number of input features increased, neural networks surpassed random forests in accuracy, highlighting their superior ability to model complex, nonlinear relationships. This trend highlights a key principle: while traditional models perform well in simple, low-dimensional applications, deep learning models become more effective as data complexity increases, especially in assays with multiple biomarkers or complex feature interactions.

The relationship between feature complexity and ML model performance also applies to image-based POCT, where data are inherently high-dimensional. In these applications, each pixel can serve as an individual input feature, creating complex, nonlinear relationships that can favor deep learning models over traditional ML approaches. Davis et al. demonstrated that CNNs consistently outperformed simpler models, including random forests, in classifying LFAs from smartphone-captured images¹⁶⁶. Their study found that while random forests performed well on low-resolution images with fewer features, CNNs excelled in high-resolution datasets, where the increased feature space allowed them to extract more nuanced visual patterns. Additionally, CNNs maintained a strong performance even in the presence of Gaussian noise. This highlights the advantage of CNNs in handling real-world imaging variability, making them a preferred choice for external validation and deployment in POCT applications requiring robustness and generalizability.

The advantages of deep learning models become even more apparent in VFAs, where the 2D sensing membrane provides a more complex spatial distribution of information. This trend is well-illustrated in the study by Han et al. on myocardial infarction diagnosis using a chemiluminescent VFA-based cTnI assay¹⁰¹. While both logistic regression and neural networks achieved nearly identical overall accuracy (95.5% in the case of neural networks and 95.6% in logistic regression), a key distinction emerged in their error distributions. The neural network-based approach had no false negatives compared to logistic regression¹⁰¹, making it clinically preferable, as failing to diagnose a myocardial infarction due to a false negative could have severe consequences for patients. These findings underscore the importance of neural network models in applications where diagnostic

sensitivity is crucial, even when traditional models appear to perform at a comparable level. Additionally, random forests had inferior performance compared to neural network models¹⁰¹, suggesting that decision tree-based models may struggle to perform reliable classification near biomarker cut-off values, where subtle variations in signal intensity may have a crucial impact on diagnostics outcomes.

A similar trend can be observed in the work by Eryilmaz et al., which examined ML-based classification of COVID-19 immunity status based on multiplexed serological assays¹⁰⁴. Unlike single-biomarker tests such as troponin detection, serological assessments of immune protection involve multiple antibody titers targeting different viral proteins, such as spike and nucleocapsid, as well as different antibody classes, including IgG and IgM. In this study, logistic regression and random forest both demonstrated suboptimal classification accuracies (83.1% and 77.4%, respectively) compared to the performance of neural networks (89.5%). The superior performance of neural networks highlights their ability to capture nonlinear relationships between biomarker levels and immune status, which traditional ML models struggled to represent accurately.

Taken together, these studies and various others in the literature reinforce a key principle in POCT machine learning model selection: when an assay provides only a small number of input features, traditional ML models such as logistic regression and random forest can perform highly competitively. However, as the dimensionality of the assay increases, whether due to multiplexed detection, image-based inputs, more complex continuous biomarker gradients or test-to-test variations, deep learning models begin to offer clear advantages. Their ability to model nonlinear interactions and extract meaningful features from high-dimensional data allows them to outperform simpler models, particularly in cases where diagnostic sensitivity is critical.

Despite these advancements, ML integration into POCT remains an emerging field, and most existing studies focus on evaluating a single ML algorithm or reporting the best-performing model for their specific test rather than conducting direct comparative analyses across multiple ML techniques. As research in this area continues to expand, a growing number of studies are expected to explore more comprehensive benchmarking methodologies. For instance, meta-analyses comparing ML performance across different POCT modalities, the development of standardized datasets for multi-platform ML evaluation, and domain adaptation studies that assess ML transferability across biomarkers will allow for a more rigorous assessment of ML models in POCT. These efforts will further refine best practices for ML model selection in POCT applications, ultimately improving diagnostic accuracy, robustness, and real-world utility.

Challenges and future directions

Regulatory challenges

One of the foremost challenges to integrating AI and ML into POCT is navigating the complex regulatory landscape¹⁶⁷. Regulatory bodies like the FDA and the European Medicines Agency (EMA) have stringent requirements to ensure the safety and efficacy of AI/ML-enabled medical devices^{168–170}. The FDA categorizes software used for medical purposes under the “software as a medical device” (SaMD) framework¹⁷¹. Within this framework, AI is defined as “a device or product that can identify, analyze, and use big data and large complex data sets from a variety of sources”¹⁷². By 2023, the FDA approved nearly 700 AI/ML-enabled medical devices, with a rapid increase in approvals since 2016¹⁷³. This rapid growth can be attributed to significant advancements in AI algorithms over this period and the increased accessibility of AI-related hardware and software. While further details on the FDA-approved AI/ML-enabled medical devices can be found in previous reports^{168,173}, this subsection will primarily focus on the challenges and limitations of current practices and future steps to address these barriers.

The majority (96.6%) of AI/ML-enabled medical devices have received FDA approval through a Premarket Notification 510(k) process, which demonstrates that a new product is substantially equivalent to a previously approved device (predicate)¹⁷³. While this process reliably ensures the device’s safety and effectiveness without the need for exhaustive clinical trials, the evolving nature of continuously improving AI models imposes concerns about its suitability for AI/ML-enabled medical devices. For instance, as AI algorithms continuously improve, at what point do the changes become significant enough that the device can no longer be considered equivalent to its approved version? What criteria should be used to assess the significance of these changes in the context of regulatory approvals? Addressing these questions is essential to prevent additional regulatory burdens on healthcare companies developing AI/ML-enabled medical devices.

To address these concerns, the FDA recently published a discussion paper proposing a framework to regulate AI/ML-based SaMD, considering the adaptive nature of continuously learning and improving algorithms¹⁷⁴. The paper introduces a total product lifecycle (TPLC) approach, which recommends preparing new premarket submissions to the FDA if software modifications introduce new risks to users. For instance, if software changes create new risks or modify the existing risks in ways that could cause significant harm to users/patients, a new premarket submission is required. In addition, a new premarket submission may be necessary if there are substantial changes to the device’s functionality that significantly affect its clinical performance. The FDA further stratifies these changes/modifications based on their impact on algorithm performance, input data, or intended use: (1) modification of performance without altering inputs or intended use (e.g., updates to the dataset or AI/ML model architecture), (2) modification of the inputs without changes of the intended use (e.g., expanding to new input types), and (3) modification of the intended use (e.g., expanding to diagnose new diseases)¹⁷⁴. The analysis of these modifications will lead to either the submission of a new 510(k) approval request or the documentation of the changes and reanalysis of the risk management files. Given the emerging and highly dynamic regulatory processes surrounding AI-driven POCT devices, researchers and developers should consider multiple strategies to facilitate compliance and ensure robust validation. One critical approach is leveraging Real-World Evidence (RWE) for ML validation. RWE provides postmarket performance data from diverse patient populations, enabling developers to demonstrate the ongoing accuracy and safety of AI models without requiring full re-approval for minor modifications. To incorporate RWE effectively, developers can implement automated monitoring systems for AI drift detection, establish periodic revalidation cycles, and engage with the FDA’s digital health precertification program to streamline oversight for iterative AI updates. By leveraging RWE, POCT developers can ensure their devices remain compliant while continuously improving.

The FDA’s TPLC model provides structured oversight for AI-driven POCT throughout its lifecycle. This approach includes premarket submissions, postmarket performance monitoring, and continuous updates based on real-world data. Researchers and developers can enhance compliance by proactively preparing SaMD Pre-Specifications (SPS) and Algorithm Change Protocols (ACP), which document anticipated AI modifications and outline protocol update procedures. These SPS and ACP documents can be shared with the FDA and other stakeholders (e.g., clinicians, patients, hospitals) to promote the adoption of AI algorithm modifications and maintain transparency throughout product development. This regulatory process covers both premarket evaluation and postmarket monitoring of the product, providing continuous oversight of the device throughout its lifecycle. Additionally, the FDA encourages manufacturers to establish transparency in AI model updates through structured risk assessments, labeling modifications, and regulatory submission triggers when significant changes occur. To maintain postmarket

compliance, AI/ML-enabled POCT developers must also implement robust postmarket monitoring strategies. These include bias and fairness assessment tools to ensure AI predictions remain unbiased across diverse populations, automated performance tracking mechanisms to detect data drift, and regulatory submission workflows for major algorithm modifications. By integrating these strategies, researchers and developers can ensure their AI models remain reliable and clinically effective throughout their deployment.

Similarly to the FDA, the European Union (EU) proposed an AI act on top of the existing General Data Protection Regulation (GDPR), which only partially regulates AI in medical systems¹⁶⁹. The AI Act introduces a risk-based approach, stratifying AI systems into three levels: (1) prohibited risk, (2) high risk, and (3) low or minimal risk. The act directly bans AI practices classified as prohibited, which includes applications that have a direct threat to people's safety, such as discrimination of vulnerable groups (e.g., mentally diseased individuals) or AI systems used for social scoring. High-risk systems currently incorporate eight areas involving biometric identification, employment management (including employment of medical professionals), administration of justice (including medical legislation), and a few other uses that potentially affect large population subgroups. These systems should be registered in an EU-based database and undergo continuous risk analysis to minimize potential harm to users. Low-risk AI systems, by contrast, are not subject to risk management or additional legal obligations. The AI Act further lays down critical issues of ethical AI, such as the diversity of training datasets, human oversight, and transparency of AI models, some of which will also be discussed in the following subsections. While it may take several years for this AI Act to be fully implemented, some aspects may attain legal status earlier.

The deployment of AI-driven POCT in low- and middle-income countries (LMICs) presents unique regulatory challenges compared to high-income regions. Unlike the structured FDA and EMA pathways, many LMICs face multiple challenges, including limited infrastructure, insufficient regulatory expertise, weak enforcement mechanisms, and inequitable access to healthcare technologies. Additionally, the absence of centralized AI governance frameworks necessitates alternative regulatory strategies tailored to the unique needs of these regions. To address these regulatory gaps, WHO's global AI governance framework recommends that LMIC governments collaborate with intergovernmental organizations to formulate AI governance frameworks and regulatory best practices¹⁷⁵. A structured risk-based approval pathway, similar to the FDA's TPLC model, should be adapted for LMICs to ensure AI-driven POCT tools meet safety and efficacy standards while accommodating resource constraints. Governments should also establish AI-focused regulatory units within health ministries and provide training programs for regulators to enhance oversight capabilities.

To facilitate AI validation in LMICs, regulators should focus on context-specific RWE models that reflect local disease prevalence and healthcare delivery conditions, ensuring AI-driven POCT tools perform reliably in low-resource settings. Given infrastructure limitations, regulatory thresholds for AI modifications should account for the need for incremental updates rather than requiring full reapprovals, especially where access to continuous software validation is limited. Additionally, adaptive regulatory pathways should integrate conditional approvals with staged validation processes, allowing AI-driven POCT tools to be introduced in controlled pilot studies before full-scale deployment, aligning with WHO's phased implementation recommendations for LMICs. From a technological perspective, developers should prioritize AI models that function efficiently in low-resource settings by enabling offline operation, reducing reliance on high-performance computing, and leveraging open-source AI models trained on region-specific datasets. AI developers should also implement low-computation models that can function on minimal hardware, addressing infrastructure challenges in LMIC hospitals and clinics. AI deployment in

LMICs should also be guided by strong ethical frameworks, ensuring models are audited for biases, particularly regarding race, gender, and socio-economic status. AI-driven POCT tools should provide transparent decision-making explanations, allowing non-specialist healthcare workers to interpret and validate AI recommendations. Additionally, community engagement initiatives should be implemented to align AI solutions with local cultural and linguistic needs, promoting trust and adoption. Postmarket monitoring should be strengthened by implementing real-time AI performance tracking and clinician feedback mechanisms. LMICs should leverage mobile health platforms for postmarket surveillance, enabling clinicians and patients to report AI performance directly to regulators. Additionally, disease burden shifts should be incorporated into AI monitoring frameworks to ensure long-term clinical relevance. A multi-stakeholder approach—involving governments, global health agencies, industry leaders, and local healthcare providers—is necessary to bridge regulatory gaps and establish a harmonized AI governance framework in both LMIC and non-LMIC settings.

Data limitations

While AI and ML systems are undoubtedly powerful, they are not infallible. These systems are susceptible to failure, confabulation or “hallucination,” generating erroneous or nonsensical outputs¹⁷⁶. In the context of POCT, such errors can lead to incorrect diagnoses or treatment plans, potentially jeopardizing patient safety. Ensuring the reliability and robustness of AI models is critical, requiring the development of comprehensive training datasets and rigorous validation and testing under diverse, real-world conditions^{177,178}. This relies on significant financial and human resources. For example, 11,000 HIV LFAs took 60 well-trained field workers two years to collect⁴⁰. The class balance or imbalance, the distribution of data points to true outcomes, can also affect the size of the training library required and may disproportionately affect LFAs with lower prevalent targets. The reliability of AI models largely depends on the ground truth labels associated with the training data. These labels, typically created by human experts, can sometimes be erroneous due to poor sample quality, ambiguous diagnostics cases, or human error¹⁷⁹. For instance, an LFA outcome might be misinterpreted by a diagnostician when faint test lines or uneven illumination conditions obscure the outcome. Incorrect labels can lead to erroneous decisions by the algorithm and inaccurate diagnostics predictions. In addition, biological specimens are prone to degradation or denaturation depending on storage conditions¹⁸⁰, making strict guidelines for sample collection and storage essential. For example, remnant diagnostic samples obtained from commercial biobanks may be stored at 4 °C for a specified duration, as required by the sample-providing institution, before being released as research samples. This delay can cause degradation of target analytes in the samples, potentially altering the original ground truth values. To ensure accurate labeling, it may be necessary to reserve a portion of the sample immediately after collection for research use, store it under frozen conditions, or perform ground truth re-measurements once the sample transitions to research status. In general, to minimize labeling errors, ground truth labels should be carefully verified using standard laboratory testing methods and cross-checked by multiple human experts before being finalized.

Another source of erroneous outcomes arises from limitations in the dataset used to train ML algorithms. ML algorithm generalization to new types of data, not used during the training process, represents a complex problem, with multiple factors limiting ML models' scalability to blindly tested new or modified datasets¹⁸¹. If certain disease cases or patient subgroups are underrepresented in the training data, the ML model's performance will likely be limited for those cases¹⁸². For example, a model trained for predicting COVID-19 mortality using exclusively U.S. patients had inaccurate predictions on Vietnamese COVID-19 patients¹⁸³. Furthermore, ML models trained to predict

diabetes retinopathy using U.S. patient trials, containing substantially more white patients (>80%) compared to Native Americans (<1%), had an unreliable performance on the Native American population¹⁸⁴. To mitigate potential model biases, training samples must be accurately selected to align with clinical practice, ensuring that the model covers a broad range of diagnostic scenarios. Additionally, to avoid bias towards specific cases, training subgroups should be well-balanced. This can be achieved through data augmentation, such as over-sampling of underrepresented samples. For example, when developing a sensor for myocardial infarction, it is important to include a diverse representation of age, sex, and race, and provide a balanced representation of both healthy cases and patients experiencing MI. Given the acute nature of MI, collecting enough MI-positive samples can be a challenge, requiring close, long-term collaboration between manufacturers and emergency departments to ensure that an adequate number of cases are included in the training and validation datasets.

Furthermore, operational conditions for training and testing samples should be consistent to avoid distributional shifts between datasets, which can limit the model's generalizability – resulting in less accurate predictions on blind testing samples compared to training samples^{181,185,186}. Such discrepancies may arise from subtle differences in sensor batches or variations in readout conditions. For instance, model scalability between samples collected in different research centers may be limited due to variability in the sample preparation protocols, testing equipment, and different data collection protocols. In a pneumonia detection study, CNN models trained to diagnose pneumonia using radiographic images collected in one research center performed poorly when tested on data from another institution within the same country¹⁸⁷. To address this generalizability issue, quality control measures should be implemented to ensure that batch-to-batch and equipment variations do not significantly affect diagnostic outcomes. For instance, when dealing with paper-based LFAs or VFAs, each sensor batch should be validated with calibration samples to confirm that output signals remain within acceptable margins despite lot-to-lot variations of paper materials. To minimize the impact of varying readout conditions, a fixed readout configuration should be used consistently^{18,19,30,62,78,81,82,85,100,104}. Alternatively, varying readout settings can be used during training dataset collection; however, it is critical to ensure they cover the most realistic testing conditions to avoid generalization fallbacks^{16,40,79,80}.

In general, ML algorithms may face challenges when confronted with unfamiliar sample patterns, such as those resulting from non-typical sample compositions (e.g., an abundance of non-specific interferents in a serum/whole blood sample causing strong non-specific sensor responses) or from readout conditions that differ from those used during training. These issues can be mitigated by implementing sample quality assurance mechanisms prior to making diagnostic inferences¹⁸⁸. Quality controls might involve assessing control regions within the sample or analyzing statistical features of the generated patterns. For example, a quality check could compare negative/positive control signals from a tested VFA to the distribution of control signals in the training set, excluding any samples that fall outside the desired confidence interval^{30,104}. In the cases with multiplexed responses, statistical analysis of repeated test signals can be used to identify and eliminate outlier testing channels, ensuring a more accurate sensor output^{19,100}.

To date, a limited number of studies have explored the impact of data diversity on the model performance and substantial efforts, both within the U.S. and internationally, will be needed to develop the infrastructure to fully address the challenges imposed by data disparities between different testing environments and nations¹⁸¹. A shorter-term solution may be to re-calibrate the model through transfer learning approaches by re-training the existing baseline model on a small subset of patients from a new nation/location or a different

testing setting¹⁸⁹. In cases of limited ground truth information available for the blindly tested samples, domain adaptation techniques may be employed to reweigh training data through the importance sampling techniques, emphasizing the most relevant samples based on their “importance” to the tested dataset distribution^{189,190}. Finally, we would like to note that multiple other factors beyond those discussed in this section may affect the ML model scalability, and we direct the reader to other review articles, specifically focusing on this important topic of ML integration and scale-up^{185,189}.

Transparency and explainability

The “black-box” nature of many AI and ML models poses another significant challenge¹⁸⁸. These systems often function in ways that are not transparent or easily interpretable by humans, making it difficult for healthcare professionals to understand the rationale behind specific decisions. This lack of transparency can impede trust and acceptance among clinicians and patients. To address this challenge, there is a growing emphasis on developing explainable AI (XAI) techniques^{191,192}. XAI aims to improve the interpretability of AI/ML models by creating frameworks that allow users to understand the decision-making process and the reasoning behind the model's predictions. These approaches are designed to enhance the transparency of AI systems, making them more understandable and fostering greater trust in their results.

Over the years, several frameworks have been proposed to define model interpretability through XAI methods, with two major approaches emerging: (1) perceptive interpretability and (2) interpretability through mathematical structures¹⁹³. Perceptive interpretability explains a model's performance through the visualization of input features and neurons based on their contribution to the output decision. Some perceptive interpretability methods rely on a set of deterministic “If-Then” rules, which explain the extracted diagnostics decision based on a set of conditions applied to the dataset¹⁹⁴. For instance, Kavya et al. explored allergy diagnostics, applying a set of “If-Then” rules to investigate the relationship between symptoms and allergic disease. They found that symptoms including running nose, coughing or sneezing are more typical in rhinitis patients, while itching and swelling are more prevalent among patients with urticaria¹⁹⁵. Furthermore, Das et al. applied an interpretable AI model for the diagnostics of Alzheimer's disease, utilizing “If-Then” rules to better learn the relationship between diagnostic outcomes and associated biomarkers in patient plasma¹⁹⁶. A lot of perceptive interpretability studies rely on saliency methods, which assign probability values in the form of heatmaps to the input data, highlighting the importance of specific inputs in influencing the model's output¹⁹⁷. Some common saliency methods are SHapley Additive exPlanations (SHAP)¹⁹⁸ and Local Interpretable Model-agnostic Explanation (LIME)¹⁹⁹. Ahmad et al. used LIME-based model interpretability to obtain insights about the likelihood of thyroid cancer recurrence based on patient information such as age, stage of the disease and tumor size, among others²⁰⁰. Sarp et al. utilized LIME to generate a heatmap overlay on the wound images, helping clinicians to interpret the most relevant wound features for the chronic wound classification²⁰¹. In another application, El-Sappagh et al. applied SHAP to evaluate the impact of patient features on Alzheimer's disease progression over time²⁰². Various examples of XAI applications to diagnostics exist, including applications to cancers, surgery and viral diseases, and we would like to direct the readers to other review articles for further details on this topic¹⁹⁴.

The second approach, interpretability through mathematical structures, applies mathematical models to explain the mechanisms within the hidden layers of neural networks. For instance, subspace-related approaches, such as singular vector canonical correlation analysis (SVCCA), identify the most significant directions between model layers by analyzing the space of all inter-neuron connections¹⁹⁷. Although a deeper exploration of XAI techniques is beyond the scope

of this Perspective article, we believe that future work is urgently needed to enable broader adoption of XAI in the field of medical diagnostics. In perceptive interpretability, saliency methods often fail to fully evaluate the input composition in a way that is useful for diagnostics applications. For instance, when interpreting subtle differences between multiple testing regions, medical professionals may want to know which parts of the sensor, beyond the primary test regions, contributed to the diagnostic decision. Depending on the application, various non-trivial input patterns can influence the algorithm's performance^{19,30,100,104}, and clinicians would benefit from identifying these using XAI methods. Furthermore, since interpretable AI/ML is still an evolving field, current feature analysis techniques and interpretability methods, primarily targeting diagnostics accuracy as an evaluation metric, may not provide an objective evaluation of the model. Another limitation of XAI techniques is the small size of datasets used to train and test the models, which might result in overfitting and biases during the model evaluation process. Finally, in terms of XAI evaluation, to date, most research papers focus on XAI methods implementation without providing a reasonable explanation of the XAI impact on the model performance, which should be carefully investigated in the future in a collaboration between AI/ML community and medical experts.

Recently, specific evaluation metrics have been proposed to more comprehensively assess the impact of model inputs on both model predictions and explanations. For instance, Komorowski et al. proposed three metrics, namely (i) faithfulness, (ii) sensitivity, and (iii) complexity, to evaluate the properties of importance attribution maps produced by saliency methods (e.g., LIME)²⁰³. Faithfulness quantifies the correlation between the feature attribution and the model's prediction, sensitivity ensures that similar model input/output pairs lead to the same model explanation, and complexity measures the number of features that are significant to the model's prediction. The same paper also utilizes ViT models, whose self-attention mechanism can be easily visualized for explanation purposes. The authors explore multiple attention heads aggregation mechanisms, including averaging attention across the heads, and taking the maximum or minimum over the heads. When applying the ViT models to COVID-19 diagnostics on Chest X-ray images, the maximum aggregation approach proved optimal as it provided reliable model explanation quality with lower complexity compared to other methods. Overall, ViT based on TransLRP model²⁰⁴ outperformed other ViT algorithms and conventional LIME method, achieving superior performance in both faithfulness and sensitivity.

In general, a close collaboration between AI/ML researchers and medical professionals will be required to develop reliable model interpretation methodologies. Scott et al. proposed a checklist of several questions that clinicians, as non-experts in the AI field, may ask AI/ML researchers prior to the deployment of their model in real clinical practice²⁰⁵. The questions mainly revolve around the context of the algorithm, training data composition, model scalability and potential algorithm harm to patients. These questions will assist clinicians in assessing the ML model readiness and help identify cases when further refinement of the algorithm is needed. We note that model interpretation through XAI techniques does not necessarily provide performance guarantees; instead, these methods describe how the model functions. Therefore, a thorough evaluation of ML models may be only available through careful and rigorous validation across multiple diverse populations and data centers, showcasing stable and accurate performance regardless of the tested patient group and institution²⁰⁶. Historically, extensive validation through randomized controlled trials (RCTs) has been a gold-standard approach to testing various kinds of black-boxes, including drug evaluation. For instance, acetaminophen is a common pain medication, whose action mechanism remains only partially explained²⁰⁷; however, its safety and effectiveness have been confirmed by multiple RCTs. To

conclude, as AI/ML research in medical diagnostics matures and more practical information about the input composition becomes available, our understanding of XAI's opportunities in POCT-related applications will continue to improve.

Other challenges

Most AI/ML models for POCT are trained on thousands of highly sensitive patient data^{40,79}. Issues such as data privacy and informed consent in AI algorithms must be carefully managed. Patient data used to train AI models should be anonymized and handled in compliance with privacy regulations such as GDPR and HIPAA^{208,209}. Establishing ethical frameworks and guidelines is essential to regulate the development and deployment of AI in POCT, ensuring that these technologies are used responsibly and equitably. Even the simplest models for LFAs, such as LFAs for HIV with two result lines, rely on field-acquired images, which may reveal patient diagnoses. As the complexity of AI/ML models grows alongside the quickening development of novel LFAs⁴ training datasets are likely also to expand. Image analysis for many AI/ML models for LFAs is run off-device⁷⁹, requiring the transfer of sensitive data for image classification followed by integration into electronic patient databases or back to the POCT device. Ensuring the security of this data transfer and the AI and ML systems in POCT is crucial. These systems must be protected against cyber threats that could compromise patient data or disrupt the integrity of the diagnostic process. Cybersecurity measures must be robust, incorporating encryption, access controls, and regular security audits. Additionally, to maintain the continued reliability and efficacy of AI-driven POCT devices, continuous updates and retraining with new data are necessary to prevent performance degradation over time.

Ensuring data privacy and cybersecurity is paramount in AI-driven POCT, and a promising approach to achieving this is integrating blockchain technology. Blockchain is a decentralized and tamperproof digital ledger system that enhances security by restricting access to authorized users while maintaining transparency and traceability in data transactions^{210,211}. This technology offers a decentralized and privacy-preserving framework for AI-driven POCT, allowing secure data transmission among patients, clinicians, and electronic health record systems. Unlike traditional cloud-based solutions, blockchain-based architectures provide pre-determined access privileges, enabling different stakeholders to securely retrieve either full datasets or specific data components as needed²². This approach ensures data integrity, transparency, and compliance with privacy regulations such as HIPAA and GDPR, while addressing concerns about AI-driven decision-making in POCT applications. The framework proposed by Guo et al. highlights how blockchain-based access control mechanisms can enhance data security, interoperability, and real-world deployment in AI-assisted diagnostics²². This case study is also covered in Section "NAATs". By integrating blockchain technology, AI-driven POCT platforms can ensure tamperproof record-keeping, controlled data access, and seamless integration into decentralized healthcare networks, ultimately improving trust and security in digital health ecosystems.

Federated learning (FL) is another emerging ML approach for secure data transfer across multiple data centers, providing increased patient privacy and protection^{212,213}. In FL, a single ML model is shared across multiple client servers, utilizing local data from users for training. After training, the results are compiled to a centralized server, however, no data transfer is happening between the servers, enhancing patient privacy. The users update the model iteratively using additional data, continuously sharing an upgraded model with the central server. Since no patient data transfer occurs between the servers, the risk of personal data leakage to an external server is minimized. As a result, FL preserves patient privacy, while providing decentralized training of ML models, potentially enhancing the diversity of training data for better model generalizability across different patient groups.

Although AI/ML-enabled devices are often designed for automated inference of diagnostics decisions, overreliance on these predictions can also introduce biases and potentially reduce diagnostics accuracy²¹⁴. Previous studies have shown that overreliance on algorithmic outputs has decreased accuracy in interpreting electrocardiograms and detecting skin lesions²¹⁵. To uphold high diagnostics standards, manufacturers should collaborate with clinicians to identify rare cases and high-risk areas where AI/ML models may fail. Accordingly, algorithmic outputs should be supplemented with information on the associated risks, and clinicians should perform additional diagnostic procedures and rely on their training and experience, particularly in higher-risk cases, before making final decisions. Both the FDA and EMA provide risk classifications to guide these assessments^{169,174}. Another concern with overreliance on AI is the potential deskilling of medical professionals. Despite the automation provided by AI/ML-enabled medical devices, clinicians must be properly trained to interpret model outcomes and manually inspect samples, especially in borderline or high-risk cases²¹⁶.

In the history of POCT and diagnostics, advancements in point-of-care sensors have often made earlier generations obsolete. For example, with the emergence of hs-cTnI testing for diagnosing MI, older contemporary cTnI assays have been gradually fading away. Between 2019 and 2021, the prevalence of hs-cTnI assays in hospitals increased from 3.3% to 32.6%²¹⁷. Similarly, as AI/ML-enabled POCT platforms are emerging, they are expected to replace traditional diagnostics tools, necessitating proper integration into healthcare practice for wider adoption. For instance, AI-assisted microscopes for detecting schistosomiasis and malaria are now emerging in resource-limited areas, offering higher accuracy and automated analysis compared to conventional manual microscopy techniques^{149,150}. Another example is an ML-based VFA for Lyme disease, which eliminates the need for expensive two-tier laboratory testing¹⁰². Integrating these new platforms into clinical practice requires proper training of medical personnel, generating new risk assessment strategies, and updating liability protocols, among other procedures. To streamline this process, medical device manufacturers should closely collaborate with clinicians, regulatory bodies, and other stakeholders to develop clear integration protocols.

As another important factor for the scalability of AI/ML-enabled medical devices, patent applications on these inventions must adhere to specific requirements to meet patent eligibility criteria. Each patent must clearly describe how AI/ML algorithms are integrated into the device and demonstrate their benefits in terms of technical performance. If the AI/ML concepts are only presented as abstract ideas, they will not be patentable²¹⁸. Furthermore, the proposed AI/ML algorithm must be novel and distinguishable from prior art²¹⁹. For example, if similar ML methods have been used in other applications, the patent must emphasize the unique application of this method for specific medical usage. The patent should highlight how the ML model addresses a particular medical challenge in a non-trivial way, as obvious claims are not patentable. Finally, it is important to detail the technical implementation of the AI/ML algorithm in the patent, including data preprocessing steps, integration of the algorithm with the device hardware, and its impact on device performance²²⁰. This can be achieved by providing both system and method claims within the same patent²²¹. System claims can be used to define the medical device as an integrated system that utilizes ML models to interpret medical data, while method claims can be used to outline the ML-based protocols the device follows to process medical data and deliver diagnostic results.

Finally, ensuring fairness and inclusivity in AI-driven POCT is essential to prevent systemic biases and disparities in diagnostic performance across different patient demographics. One key challenge in medical AI applications, particularly in optical and imaging-based modalities, is variability in accuracy across diverse skin tones^{222,223}. For example, AI models trained predominantly on specific skin tones can

exhibit lower sensitivity and specificity when applied to different skin tones, potentially leading to misdiagnosis and healthcare disparities^{224–226}. To address these concerns, a recent study by Chen et al. introduced an AI-enabled, agnostic imaging approach to measure oxygen saturation, which ensures accurate and unbiased oxygen saturation assessment across all skin tones²²⁷. Traditional pulse oximetry has been reported to exhibit lower accuracy in darker skin tones due to differences in melanin absorption and reflectance, leading to potential biases in decision-making. To mitigate this issue, Chen et al. developed a smartphone-based imaging system that utilizes advanced computational correction models to account for variability in ambient lighting, camera settings, and individual skin pigmentation. Their ML-driven normalization framework effectively helped eliminate discrepancies in oximetry readings, ensuring consistent and equitable diagnostic performance across diverse patient populations. Integrating such fairness-oriented AI models into POCT is crucial for ensuring reliable and unbiased healthcare delivery, particularly in resource-limited or ethnically diverse populations. As AI adoption in POCT expands, future developments must emphasize bias-aware training datasets, domain adaptation strategies, and fairness metrics, ensuring that AI-powered diagnostic tools maintain high performance across all patient groups.

Conclusions and outlook

Several health economics studies have demonstrated the cost-effectiveness of POCT compared to standard lab testing for various diseases, including sexually transmitted infections, cardiovascular diseases, and infections caused by antimicrobial-resistant pathogens^{228–232}. Increased POCT availability raises testing rates, potentially elevating short-term healthcare costs, but this expense is offset in the long term by reduced downstream treatment costs and improved health outcomes. With their connectivity and big data processing capabilities, AI/ML-enabled POCT offers significant potential benefits for healthcare systems. For example, in epidemiological surveillance, AI-driven models can integrate vast amounts of test data collected by a network of point-of-care biosensors with additional sources, such as hospital records, genomic data, and social media posts, to systematically track health data, identify trends, and detect unusual disease patterns^{233,234}. This enables public health systems to proactively prevent and control disease outbreaks. Another example is monitoring antimicrobial-resistant bacterial prevalence in hospitals and the environment. Emerging evidence suggests that healthcare infrastructure, environments, and patient pathways contribute to promoting and sustaining antimicrobial resistance²³⁵. By integrating various types of data—such as laboratory results, clinical information, and social/ecological determinants—AI-driven analyses can produce data-driven insights for healthcare staff and regulators, identifying trends, possible infection sources, and recommended control measures to inform future prevention efforts. Nevertheless, to fully realize these benefits and the widespread adoption of AI within the healthcare community, it is essential to rigorously regulate AI and ML models to ensure safety, fairness, inclusivity, and reliable diagnostics performance in real-life scenarios. This can be achieved by creating diversified training datasets and supporting close collaboration among clinicians, researchers, developers, and regulatory bodies, which will further enhance trust in AI-based POCT tools and promote their acceptance within the diagnostics field.

References

1. Jani, I. V. & Peter, T. F. How point-of-care testing could drive innovation in global health. *N. Engl. J. Med.* **368**, 2319–2324 (2013).
2. Christodouleas, D. C., Kaur, B. & Chorti, P. From Point-of-Care Testing to eHealth Diagnostic Devices (eDiagnostics). *ACS Cent. Sci.* **4**, 1600–1616 (2018).

3. Sikaris, K. A. Enhancing the Clinical Value of Medical Laboratory Testing. *Clin. Biochem. Rev.* **38**, 107–114 (2017).
4. Budd, J. et al. Lateral flow test engineering and lessons learned from COVID-19. *Nat. Rev. Bioeng.* **1**, 13–31 (2023).
5. Wilner, O. I., Yesodi, D. & Weizmann, Y. Point-of-care nucleic acid tests: assays and devices. *Nanoscale* **15**, 942–952 (2022).
6. 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. *Nat. Microbiol.* **4**, 46–54 (2019).
7. Otoo, J. A. & Schlappi, T. S. REASSURED Multiplex Diagnostics: A Critical Review and Forecast. *Biosensors* **12**, 124 (2022).
8. Hu, L. et al. Chapter 11 - Optical smartphone-based sensing: diagnostic of biomarkers In: *The Detection of Biomarkers* (ed. Ozkan, S. A.) 277–302 (Academic Press, 2022).
9. Di Nardo, F., Chiarello, M., Cavallera, S., Baggiani, C. & Anfossi, L. Ten Years of Lateral Flow Immunoassay Technique Applications: Trends, Challenges and Future Perspectives. *Sensors* **21**, 5185 (2021).
10. Kasetsirikul, S., Shiddiky, M. J. A. & Nguyen, N.-T. Challenges and perspectives in the development of paper-based lateral flow assays. *Microfluid. Nanofluidics* **24**, 17 (2020).
11. Choi, J. R. et al. Advances and challenges of fully integrated paper-based point-of-care nucleic acid testing. *TrAC Trends Anal. Chem.* **93**, 37–50 (2017).
12. Jani, I. V. & Peter, T. F. Nucleic Acid Point-of-Care Testing to Improve Diagnostic Preparedness. *Clin. Infect. Dis.* **75**, 723–728 (2022).
13. Boppart, S. A. & Richards-Kortum, R. Point-of-care and point-of-procedure optical imaging technologies for primary care and global health. *Sci. Transl. Med.* **6**, 253rv2 (2014).
14. Frade, S. et al. Malaria RDT (mRDT) interpretation accuracy by frontline health workers compared to AI in Kano State, Nigeria. *VeriXiv*, <https://doi.org/10.12688/verixiv.27.1> (2024).
15. 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. *Malar. J.* **14**, 26 (2015).
16. Lee, S. et al. Sample-to-answer platform for the clinical evaluation of COVID-19 using a deep learning-assisted smartphone-based assay. *Nat. Commun.* **14**, 2361 (2023).
17. Lee, S. et al. Rapid deep learning-assisted predictive diagnostics for point-of-care testing. *Nat. Commun.* **15**, 1695 (2024).
18. Joung, H.-A. et al. Point-of-Care Serodiagnostic Test for Early-Stage Lyme Disease Using a Multiplexed Paper-Based Immunoassay and Machine Learning. *ACS Nano* **14**, 229–240 (2020).
19. Ballard, Z. S. et al. Deep learning-enabled point-of-care sensing using multiplexed paper-based sensors. *NPJ Digit. Med.* **3**, 66 (2020).
20. Sun, H. et al. Paper microfluidics with deep learning for portable intelligent nucleic acid amplification tests. *Talanta* **258**, 124470 (2023).
21. Yang, Y. et al. Rapid Detection of SARS-CoV-2 RNA in Human Nasopharyngeal Specimens Using Surface-Enhanced Raman Spectroscopy and Deep Learning Algorithms. *ACS Sens.* **8**, 297–307 (2023).
22. Guo, X. et al. Smartphone-based DNA diagnostics for malaria detection using deep learning for local decision support and blockchain technology for security. *Nat. Electron.* **4**, 615–624 (2021).
23. Liu, T. et al. Rapid and stain-free quantification of viral plaque via lens-free holography and deep learning. *Nat. Biomed. Eng.* **7**, 1040–1052 (2023).
24. Koydemir, H. C., Coulibaly, J. T., Tseng, D., Bogoch, I. I. & Ozcan, A. Design and validation of a wide-field mobile phone microscope for the diagnosis of schistosomiasis. *Travel Med. Infect. Dis.* **30**, 128–129 (2019).
25. Isozaki, A. et al. AI on a chip. *Lab Chip* **20**, 3074–3090 (2020).
26. Yigci, D., Ahmadpour, A. & Tasoglu, S. AI-Based Metamaterial Design for Wearables. *Adv. Sens. Res.* **3**, 2300109 (2023).
27. Tezzezen, E., Yigci, D., Ahmadpour, A. & Tasoglu, S. AI-Based Metamaterial Design. *ACS Appl. Mater. Interfaces* **16**, 29547–29569 (2024).
28. Ahmadpour, A., Yetisen, A. K. & Tasoglu, S. Piezoelectric Metamaterial Blood Pressure Sensor. *ACS Appl. Electron. Mater.* **5**, 3280–3290 (2023).
29. Song, Y. et al. 3D-printed epifluidic electronic skin for machine learning-powered multimodal health surveillance. *Sci. Adv.* **9**, eadi6492 (2023).
30. Goncharov, A. et al. Deep Learning-Enabled Multiplexed Point-of-Care Sensor using a Paper-Based Fluorescence Vertical Flow Assay. *Small* **19**, e2300617 (2023).
31. Najjar, R. Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging. *Diagnostics* **13**, 2760 (2023).
32. Bhaiyya, M., Panigrahi, D., Rewatkar, P. & Haick, H. Role of Machine Learning Assisted Biosensors in Point-of-Care-Testing For Clinical Decisions. *ACS Sens.* **9**, 4495–4519 (2024).
33. Cui, F., Yue, Y., Zhang, Y., Zhang, Z. & Zhou, H. S. Advancing Biosensors with Machine Learning. *ACS Sens.* **5**, 3346–3364 (2020).
34. Tran, N. K. et al. Evolving Applications of Artificial Intelligence and Machine Learning in Infectious Diseases Testing. *Clin. Chem.* **68**, 125–133 (2021).
35. Wang, B. et al. Smartphone-based platforms implementing microfluidic detection with image-based artificial intelligence. *Nat. Commun.* **14**, 1341 (2023).
36. Ballard, Z., Brown, C., Madni, A. M. & Ozcan, A. Machine learning and computation-enabled intelligent sensor design. *Nat. Mach. Intell.* **3**, 556–565 (2021).
37. Salim, A. & Lim, S. Review of recent metamaterial microfluidic sensors. *Sensors* **18**, 232 (2018).
38. Nguyen, D. T. et al. Ambient health sensing on passive surfaces using metamaterials. *Sci. Adv.* **10**, eadj6613 (2024).
39. HIV rapid diagnostic test market landscape. *News from World Health Organization* https://cdn.who.int/media/docs/default-source/hq-hiv-hepatitis-and-stis-library/eic-hiv-market-landscape-report_june2023.pdf?sfvrsn=6072a75b_5 (2023).
40. Turbé, V. et al. Deep learning of HIV field-based rapid tests. *Nat. Med.* **27**, 1165–1170 (2021).
41. Wong, N. C. K. et al. Machine learning to support visual auditing of home-based lateral flow immunoassay self-test results for SARS-CoV-2 antibodies. *Commun. Med.* **2**, 78 (2022).
42. What is machine learning (ML)? *IBM webpage*, <https://www.ibm.com/topics/machine-learning> (2024).
43. Seymour, C. W. et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA* **321**, 2003–2017 (2019).
44. Jackson, H. W. et al. The single-cell pathology landscape of breast cancer. *Nature* **578**, 615–620 (2020).
45. Campanella, G. et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **25**, 1301–1309 (2019).
46. Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F. & Campbell, J. P. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl. Vis. Sci. Technol.* **9**, 14 (2020).
47. Robitaille, B., Marcos, B., Veillette, M. & Payre, G. Quasi-Newton methods for training neural networks. *WIT Trans. Inf. Commun. Technol.* **2**, 13 (1993).

48. Sampson, G., Rumelhart, D. E., McClelland, J. L. & Group, T. P. R. Parallel Distributed Processing: Explorations in the Microstructures of Cognition. *Language* **63**, 871 (1987).
49. Hecht-Nielsen, R. Theory of the Backpropagation Neural Network. In: *Neural Networks for Perception* (ed. Wechsler, H.) 65–93 (Academic Press, 1992).
50. Chaturvedi, N. et al. Advances in point-of-care optical biosensing for underserved populations. *TrAC Trends Anal. Chem.* **175**, 117731 (2024).
51. Sarvamangala, D. R. & Kulkarni, R. V. Convolutional neural networks in medical image understanding: a survey. *Evol. Intell.* **15**, 1–22 (2022).
52. Bahadır, E. B. & Sezgentürk, M. K. Lateral flow assays: Principles, designs and labels. *TrAC Trends Anal. Chem.* **82**, 286–306 (2016).
53. Parolo, C. et al. Tutorial: design and fabrication of nanoparticle-based lateral-flow immunoassays. *Nat. Protoc.* **15**, 3788–3816 (2020).
54. Yu, S., Nimse, S. B., Kim, J., Song, K.-S. & Kim, T. Development of a Lateral Flow Strip Membrane Assay for Rapid and Sensitive Detection of the SARS-CoV-2. *Anal. Chem.* **92**, 14139–14144 (2020).
55. Lee, D., Ozkaya-Ahmadov, T. & Sarioglu, A. F. Chemically Amplified Multiplex Detection of SARS-CoV-2 and Influenza A and B Viruses via Paint-Programmed Lateral Flow Assays. *Small* **19**, e2208035 (2023).
56. Brangel, P. et al. A Serological Point-of-Care Test for the Detection of IgG Antibodies against Ebola Virus in Human Survivors. *ACS Nano* **12**, 63–73 (2018).
57. Han, G.-R., Jang, H., Ki, H., Lee, H. & Kim, M.-G. Reagent Filming for Universal Point-of-Care Diagnostics. *Small Methods* **5**, e2100645 (2021).
58. Kim, S. et al. Highly sensitive pregnancy test kit via oriented antibody conjugation on brush-type ligand-coated quantum beads. *Biosens. Bioelectron.* **213**, 114441 (2022).
59. Pereira, C. et al. Paper-based biosensors for cancer diagnostics. *Trends Chem* **4**, 554–567 (2022).
60. Quesada-González, D., Jairo, G. A., Blake, R. C., Blake, D. A. & Merkoçi, A. Uranium (VI) detection in groundwater using a gold nanoparticle/paper-based lateral flow device. *Sci. Rep.* **8**, 16157 (2018).
61. Bergua, J. F. et al. Lateral flow device for water fecal pollution assessment: from troubleshooting of its microfluidics using bioluminescence to colorimetric monitoring of generic *Escherichia coli*. *Lab Chip* **21**, 2417–2426 (2021).
62. Han, G.-R., Koo, H. J., Ki, H. & Kim, M.-G. Paper/Soluble Polymer Hybrid-Based Lateral Flow Biosensing Platform for High-Performance Point-of-Care Testing. *ACS Appl. Mater. Interfaces* **12**, 34564–34575 (2020).
63. Han, G.-R. & Kim, M.-G. Highly Sensitive Chemiluminescence-Based Lateral Flow Immunoassay for Cardiac Troponin I Detection in Human Serum. *Sensors* **20**, 2593 (2020).
64. Gupta, R. et al. Ultrasensitive lateral-flow assays via plasmonically active antibody-conjugated fluorescent nanoparticles. *Nat. Biomed. Eng.* **7**, 1556–1570 (2023).
65. Miller, B. S. et al. Spin-enhanced nanodiamond biosensing for ultrasensitive diagnostics. *Nature* **587**, 588–593 (2020).
66. Hong, L. et al. High performance immunochromatographic assay for simultaneous quantitative detection of multiplex cardiac markers based on magnetic nanobeads. *Theranostics* **8**, 6121–6131 (2018).
67. Miller, B. S. et al. Sub-picomolar lateral flow antigen detection with two-wavelength imaging of composite nanoparticles. *Biosens. Bioelectron.* **207**, 114133 (2022).
68. Loynachan, C. N. et al. Platinum Nanocatalyst Amplification: Redefining the Gold Standard for Lateral Flow Immunoassays with Ultrabroad Dynamic Range. *ACS Nano* **12**, 279–288 (2018).
69. Renzi, E., Piper, A., Natri, F., Merkoçi, A. & Lombardi, A. An Artificial Miniaturized Peroxidase for Signal Amplification in Lateral Flow Immunoassays. *Small* **19**, e2207949 (2023).
70. Han, G.-R., Ki, H. & Kim, M.-G. Automated, Universal, and Mass-Produced Paper-Based Lateral Flow Biosensing Platform for High-Performance Point-of-Care Testing. *ACS Appl. Mater. Interfaces* **12**, 1885–1894 (2020).
71. Lee, D. et al. Capillary flow control in lateral flow assays via delaminating timers. *Sci. Adv.* **7**, eabf9833 (2021).
72. Raiko, K. et al. Improved sensitivity and automation of a multi-step upconversion lateral flow immunoassay using a 3D-printed actuation mechanism. *Anal. Bioanal. Chem.* **416**, 1517–1525 (2024).
73. Liu, J. et al. Naked-Eye Readout Distance Quantitative Lateral Flow Assay Based on the Permeability Changes of Enzyme-Catalyzed Hydrogelation. *Anal. Chem.* **95**, 8011–8019 (2023).
74. Xiao, Y. et al. Distance-based lateral flow biosensor for the quantitative detection of bacterial endotoxin. *Chin. Chem. Lett.* **35**, 109718 (2024).
75. Mudanyali, O. et al. Integrated rapid-diagnostic-test reader platform on a cellphone. *Lab Chip* **12**, 2678–2686 (2012).
76. Ozcan, A. Mobile phones democratize and cultivate next-generation imaging, diagnostics and measurement tools. *Lab Chip* **14**, 3187–3194 (2014).
77. Wood, C. S. et al. Taking connected mobile-health diagnostics of infectious diseases to the field. *Nature* **566**, 467–474 (2019).
78. Tong, H. et al. Artificial intelligence-assisted colorimetric lateral flow immunoassay for sensitive and quantitative detection of COVID-19 neutralizing antibody. *Biosens. Bioelectron.* **213**, 114449 (2022).
79. Arumugam, S. et al. Rapidly adaptable automated interpretation of point-of-care COVID-19 diagnostics. *Commun. Med.* **3**, 91 (2023).
80. Bermejo-Peláez, D. et al. A Smartphone-Based Platform Assisted by Artificial Intelligence for Reading and Reporting Rapid Diagnostic Tests: Evaluation Study in SARS-CoV-2 Lateral Flow Immunoassays. *JMIR Public Heal. Surveill.* **8**, e38533 (2022).
81. Zhang, S. et al. A Quantitative Detection Algorithm for Multi-Test Line Lateral Flow Immunoassay Applied in Smartphones. *Sensors* **23**, 6401 (2023).
82. Carrio, A., Sampedro, C., Sanchez-Lopez, J. L., Pimienta, M. & Campoy, P. Automated Low-Cost Smartphone-Based Lateral Flow Saliva Test Reader for Drugs-of-Abuse Detection. *Sensors* **15**, 29569–29593 (2015).
83. Yan, S. et al. SERS-based lateral flow assay combined with machine learning for highly sensitive quantitative analysis of *Escherichia coli* O157:H7. *Anal. Bioanal. Chem.* **412**, 7881–7890 (2020).
84. Mendels, D.-A. et al. Using artificial intelligence to improve COVID-19 rapid diagnostic test result interpretation. *Proc. Natl. Acad. Sci.* **118**, e2019893118 (2021).
85. Wang, W., Chen, K., Ma, X. & Guo, J. Artificial intelligence reinforced upconversion nanoparticle-based lateral flow assay via transfer learning. *Fundam. Res.* **3**, 544–556 (2023).
86. Yan, W. et al. Machine Learning Approach to Enhance the Performance of MNP-Labeled Lateral Flow Immunoassay. *Nano Micro Lett.* **11**, 7 (2019).
87. Consortium, T. L. A. et al. Machine learning for determining lateral flow device results for testing of SARS-CoV-2 infection in asymptomatic populations. *Cell Rep. Med.* **3**, 100784 (2022).

88. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 770–778 (IEEE, 2016).
89. Jiang, N. et al. Lateral and Vertical Flow Assays for Point-of-Care Diagnostics. *Adv. Healthc. Mater.* **8**, e1900244 (2019).
90. Lei, R., Wang, D., Arain, H. & Mohan, C. Design of Gold Nano-particle Vertical Flow Assays for Point-of-Care Testing. *Diagnostics* **12**, 1107 (2022).
91. Chen, R. et al. Vertical flow assays based on core-shell SERS nanotags for multiplex prostate cancer biomarker detection. *Analyst* **144**, 4051–4059 (2019).
92. Puravankara, V. et al. Surface-Enhanced Raman spectroscopy for Point-of-Care Bioanalysis: From lab to field. *Chem. Eng. J.* **498**, 155163 (2024).
93. Yu, X. et al. Defect-Engineered Coordination Compound Nano-particles Based on Prussian Blue Analogues for Surface-Enhanced Raman Spectroscopy. *ACS Nano* **18**, 30987–31001 (2024).
94. Jiao, Y. et al. 3D vertical-flow paper-based device for simultaneous detection of multiple cancer biomarkers by fluorescent immunoassay. *Sens. Actuators B: Chem.* **306**, 127239 (2020).
95. Cheng, Y. et al. Dual-signal readout paper-based wearable biosensor with a 3D origami structure for multiplexed analyte detection in sweat. *Microsyst. Nanoeng.* **9**, 36 (2023).
96. Chen, R. et al. Vertical Flow Assay for Inflammatory Biomarkers Based on Nanofluidic Channel Array and SERS Nanotags. *Small* **16**, e2002801 (2020).
97. Duan, S. et al. Deep learning-assisted ultra-accurate smartphone testing of paper-based colorimetric ELISA assays. *Anal. Chim. Acta* **1248**, 340868 (2023).
98. Lee, W. et al. Thread/paper- and paper-based microfluidic devices for glucose assays employing artificial neural networks. *Electrophoresis* **39**, 1443–1451 (2018).
99. Tay, D. M. Y. et al. Accelerating the optimization of vertical flow assay performance guided by a rational systematic model-based approach. *Biosens. Bioelectron.* **222**, 114977 (2023).
100. Han, G.-R. et al. Deep Learning-Enhanced Paper-Based Vertical Flow Assay for High-Sensitivity Troponin Detection Using Nanoparticle Amplification. *ACS Nano* **18**, 27933–27948 (2024).
101. Han, G.-R. et al. Deep Learning-Enhanced Chemiluminescence Vertical Flow Assay for High-Sensitivity Cardiac Troponin I Testing. *Small* **21**, e2411585 (2025).
102. Ghosh, R. et al. Rapid single-tier serodiagnosis of Lyme disease. *Nat. Commun.* **15**, 7124 (2024).
103. Jung, H.-A. et al. Paper-based multiplexed vertical flow assay for point-of-care testing. *Lab Chip* **19**, 1027–1034 (2019).
104. Eryilmaz, M. et al. A Paper-Based Multiplexed Serological Test to Monitor Immunity against SARS-CoV-2 Using Machine Learning. *ACS Nano* **18**, 16819–16831 (2024).
105. Jang, H.-J. et al. Deep Learning-Based Kinetic Analysis in Paper-Based Analytical Cartridges Integrated with Field-Effect Transistors. *ACS Nano* **18**, 24792–24802 (2024).
106. Yang, J. et al. Machine learning-assisted optical nano-sensor arrays in microorganism analysis. *TrAC Trends Anal. Chem.* **159**, 116945 (2023).
107. Li, Z. et al. A machine learning approach-based array sensor for rapidly predicting the mechanisms of action of antibacterial compounds. *Nanoscale* **14**, 3087–3096 (2022).
108. Wang, X. et al. Metabolism-triggered sensor array aided by machine learning for rapid identification of pathogens. *Biosens. Bioelectron.* **255**, 116264 (2024).
109. Yang, M. et al. Machine learning-enabled non-destructive paper chromogenic array detection of multiplexed viable pathogens on food. *Nat. Food* **2**, 110–117 (2021).
110. Kim, H. et al. Kaleidoscopic fluorescent arrays for machine-learning-based point-of-care chemical sensing. *Sens. Actuators B Chem.* **329**, 129248 (2021).
111. Zhu, H. et al. PCR past, present and future. *Biotechniques* **69**, 317–325 (2020).
112. Notomi, T., Mori, Y., Tomita, N. & Kanda, H. Loop-mediated isothermal amplification (LAMP): principle, features, and future prospects. *J. Microbiol.* **53**, 1–5 (2015).
113. Daher, R. K., Stewart, G., Boissinot, M. & Bergeron, M. G. Recombinase Polymerase Amplification for Diagnostic Applications. *Clin. Chem.* **62**, 947–958 (2016).
114. Yao, C., Zhang, R., Tang, J. & Yang, D. Rolling circle amplification (RCA)-based DNA hydrogel. *Nat. Protoc.* **16**, 5460–5483 (2021).
115. Cherkaoui, D., Huang, D., Miller, B. S., Turbé, V. & McKendry, R. A. Harnessing recombinase polymerase amplification for rapid multi-gene detection of SARS-CoV-2 in resource-limited settings. *Biosens. Bioelectron.* **189**, 113328 (2021).
116. Kang, T., Lu, J., Yu, T., Long, Y. & Liu, G. Advances in nucleic acid amplification techniques (NAATs): COVID-19 point-of-care diagnostics as an example. *Biosens. Bioelectron.* **206**, 114109 (2022).
117. Wang, M., Zhang, R. & Li, J. CRISPR/cas systems redefine nucleic acid detection: Principles and methods. *Biosens. Bioelectron.* **165**, 112430–112430 (2020).
118. Narasimhan, V. et al. Nucleic Acid Amplification-Based Technologies (NAAT)—Toward Accessible, Autonomous, and Mobile Diagnostics. *Adv. Mater. Technol.* **8**, 2300230 (2023).
119. Cherkaoui, D. et al. CRISPR-assisted test for *Schistosoma haematobium*. *Sci. Rep.* **13**, 4990 (2023).
120. Carrell, C. et al. Beyond the lateral flow assay: A review of paper-based microfluidics. *Microelectron. Eng.* **206**, 45–54 (2019).
121. García-Arroyo, L. et al. Benefits and drawbacks of molecular techniques for diagnosis of viral respiratory infections. Experience with two multiplex PCR assays. *J. Med. Virol.* **88**, 45–50 (2016).
122. Rohaim, M. A. et al. Artificial Intelligence-Assisted Loop Mediated Isothermal Amplification (AI-LAMP) for Rapid Detection of SARS-CoV-2. *Viruses* **12**, 972 (2020).
123. Niemz, A., Ferguson, T. M. & Boyle, D. S. Point-of-care nucleic acid testing for infectious diseases. *Trends Biotechnol.* **29**, 240–250 (2011).
124. Lee, S. H. et al. Emerging ultrafast nucleic acid amplification technologies for next-generation molecular diagnostics. *Biosens. Bioelectron.* **141**, 111448 (2019).
125. Sun, H. et al. Integrated smart analytics of nucleic acid amplification tests via paper microfluidics and deep learning in cloud computing. *Biomed. Signal Process. Control* **83**, 104721 (2023).
126. Sun, H. et al. AI-aided on-chip nucleic acid assay for smart diagnosis of infectious disease. *Fundam. Res.* **2**, 476–486 (2021).
127. Jaroenram, W. et al. One-step colorimetric isothermal detection of COVID-19 with AI-assisted automated result analysis: A platform model for future emerging point-of-care RNA/DNA disease diagnosis. *Talanta* **249**, 123375–123375 (2022).
128. Kim, M. G. et al. Deep Learning Assisted Surface-Enhanced Raman Spectroscopy (SERS) for Rapid and Direct Nucleic Acid Amplification and Detection: Toward Enhanced Molecular Diagnostics. *ACS Nano* **17**, 18332–18345 (2023).
129. Tripathi, P. et al. Classification of nucleic acid amplification on ISFET arrays using spectrogram-based neural networks. *Comput. Biol. Med.* **161**, 107027 (2023).
130. Sejnowski, T. J. The unreasonable effectiveness of deep learning in artificial intelligence. *Proc. Natl. Acad. Sci.* **117**, 30033–30038 (2020).

131. Patchsung, M. et al. Clinical validation of a Cas13-based assay for the detection of SARS-CoV-2 RNA. *Nat. Biomed. Eng.* **4**, 1140–1149 (2020).
132. Roh, Y. H. et al. CRISPR-Enhanced Hydrogel Microparticles for Multiplexed Detection of Nucleic Acids. *Adv. Sci.* **10**, 2206872 (2023).
133. Draz, M. S. et al. Virus detection using nanoparticles and deep neural network-enabled smartphone system. *Sci. Adv.* **6**, eabd5354 (2020).
134. Shokr, A. et al. Mobile Health (mHealth) Viral Diagnostics Enabled with Adaptive Adversarial Learning. *ACS Nano* **15**, 665–673 (2021).
135. Li, C. et al. All-In-One OsciDrop Digital PCR System for Automated and Highly Multiplexed Molecular Diagnostics. *Adv. Sci.* **11**, 2309557 (2024).
136. Muñoz, H. E. et al. Fractal LAMP: Label-Free Analysis of Fractal Precipitate for Digital Loop-Mediated Isothermal Nucleic Acid Amplification. *ACS Sens.* **5**, 385–394 (2020).
137. Ye, S. et al. OsciDrop: A Versatile Deterministic Droplet Generator. *Anal. Chem.* **94**, 2918–2925 (2022).
138. Jeong, H. K., Park, C., Henao, R. & Kheterpal, M. Deep Learning in Dermatology: A Systematic Review of Current Approaches, Outcomes, and Limitations. *JID Innov.* **3**, 100150 (2023).
139. Wu, Y. et al. Skin Cancer Classification With Deep Learning: A Systematic Review. *Front. Oncol.* **12**, 893972 (2022).
140. Naqvi, M., Gilani, S. Q., Syed, T., Marques, O. & Kim, H.-C. Skin Cancer Detection Using Deep Learning—A Review. *Diagnostics* **13**, 1911 (2023).
141. Ewerlöf, M., Strömberg, T., Larsson, M. & Salerud, E. G. Multi-spectral snapshot imaging of skin microcirculatory hemoglobin oxygen saturation using artificial neural networks trained on in vivo data. *J. Biomed. Opt.* **27**, 036004–036004 (2022).
142. Cabanas, A. M. et al. Evaluating AI Methods for Pulse Oximetry: Performance, Clinical Accuracy, and Comprehensive Bias Analysis. *Bioengineering* **11**, 1061 (2024).
143. Ouyang, D. et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* **580**, 252–256 (2020).
144. Jafari, M. et al. Automated diagnosis of cardiovascular diseases from cardiac magnetic resonance imaging using deep learning models: A review. *Comput. Biol. Med.* **160**, 106998 (2023).
145. Lin, A. et al. Deep learning-enabled coronary CT angiography for plaque and stenosis quantification and cardiac risk prediction: an international multicentre study. *Lancet Digit. Heal.* **4**, e256–e265 (2022).
146. Martinelli, M., Moroni, D., Prochazka, A. & Strojnik, M. Editorial: Artificial intelligence in point of care diagnostics. *Front. Digit. Heal.* **5**, 1236178 (2023).
147. Gomes, R. F. T. et al. Use of Deep Neural Networks in the Detection and Automated Classification of Lesions Using Clinical Images in Ophthalmology, Dermatology, and Oral Medicine—A Systematic Review. *J. Digit. Imaging* **36**, 1060–1070 (2023).
148. Wang, H. et al. Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning. *Light Sci. Appl.* **9**, 118 (2020).
149. Choi, H. L. et al. Landscape analysis of NTD diagnostics and considerations on the development of a strategy for regulatory pathways. *PLoS Neglected Trop. Dis.* **16**, e0010597 (2022).
150. Oyibo, P. et al. Schistoscope: An Automated Microscope with Artificial Intelligence for Detection of Schistosoma haematobium Eggs in Resource-Limited Settings. *Micromachines* **13**, 643 (2022).
151. Das, D. et al. Field evaluation of the diagnostic performance of EasyScan GO: a digital malaria microscopy device based on machine-learning. *Malar. J.* **21**, 122 (2022).
152. Zhang, Y. et al. Motility-based label-free detection of parasites in bodily fluids using holographic speckle analysis and deep learning. *Light Sci. Appl.* **7**, 108 (2018).
153. Hooshyar, H., Rostamkhani, P., Arbabi, M. & Delavari, M. Giardia lamblia infection: review of current diagnostic strategies. *Gastroenterol. Hepatol. Bed Bench* **12**, 3–12 (2019).
154. Göröcs, Z. et al. Label-free detection of Giardia lamblia cysts using a deep learning-enabled portable imaging flow cytometer. *Lab Chip* **20**, 4404–4412 (2020).
155. Göröcs, Z. et al. A deep learning-enabled portable imaging flow cytometer for cost-effective, high-throughput, and label-free analysis of natural water samples. *Light Sci. Appl.* **7**, 66 (2018).
156. Koydemir, H. C. et al. Rapid imaging, detection and quantification of Giardia lamblia cysts using mobile-phone based fluorescent microscopy and machine learning. *Lab Chip* **15**, 1284–1293 (2015).
157. Bachar, N. et al. An artificial intelligence-assisted diagnostic platform for rapid near-patient hematology. *Am. J. Hematol.* **96**, 1264–1274 (2021).
158. Chen, D. et al. Multiparameter mobile blood analysis for complete blood count using contrast-enhanced defocusing imaging and machine vision. *Analyst* **148**, 2021–2034 (2023).
159. Zhang, C. et al. Real-time intelligent classification of COVID-19 and thrombosis via massive image-based analysis of platelet aggregates. *Cytometry* **103**, 492–499 (2023).
160. Zhou, Y. et al. Intelligent classification of platelet aggregates by agonist type. *eLife* **9**, e52938 (2020).
161. de Haan, K. et al. Automated screening of sickle cells using a smartphone-based microscope and deep learning. *npj Digit. Med.* **3**, 76 (2020).
162. Feng, S., Tseng, D., Carlo, D. D., Garner, O. B. & Ozcan, A. High-throughput and automated diagnosis of antimicrobial resistance using a cost-effective cellphone-based micro-plate reader. *Sci. Rep.* **6**, 39203 (2016).
163. Brown, C. et al. Automated, Cost-Effective Optical System for Accelerated Antimicrobial Susceptibility Testing (AST) Using Deep Learning. *ACS Photonics* **7**, 2527–2538 (2020).
164. Tok, S. et al. Early detection of E. coli and total coliform using an automated, colorimetric and fluorometric fiber optics-based device. *Lab Chip* **19**, 2925–2935 (2019).
165. Kim, H. et al. Noninvasive Precision Screening of Prostate Cancer by Urinary Multimarker Sensor and Artificial Intelligence Analysis. *ACS Nano* **15**, 4054–4065 (2021).
166. Davis, A. M. & Tomitaka, A. Machine Learning-Based Quantification of Lateral Flow Assay Using Smartphone-Captured Images. *Bio-sensors* **15**, 19 (2025).
167. Harvey, H. B. & Gowda, V. Regulatory Issues and Challenges to Artificial Intelligence Adoption. *Radiol. Clin. North Am.* **59**, 1075–1083 (2021).
168. Zhu, S., Gilbert, M., Chetty, I. & Siddiqui, F. The 2021 landscape of FDA-approved artificial intelligence/machine learning-enabled medical devices: An analysis of the characteristics and intended use. *Int. J. Med. Inform.* **165**, 104828 (2022).
169. Meszaros, J., Minari, J. & Huys, I. The future regulation of artificial intelligence systems in healthcare services and medical research in the European Union. *Front. Genet.* **13**, 927721 (2022).
170. Benjamens, S., Dhunoo, P. & Meskó, B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digit. Med.* **3**, 118 (2020).
171. Software as a Medical Device (SaMD), U.S. Food & Drug Administration webpage, <https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd> (2018).
172. Digital Health Terms, U.S. Food & Drug Administration webpage, <https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-terms#:~:text=Advanced%20Analytics,Required%20for%20artificial%20intelligence%20devices> (2022).
173. Joshi, G. et al. FDA-Approved Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices: An Updated Landscape. *Electronics* **13**, 498 (2024).

174. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD), U.S. Food & Drug Administration webpage, <https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf> (2019).
175. Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization webpage, <https://www.who.int/publications/i/item/9789240029200> (2021).
176. Hatem, R., Simmons, B. & Thornton, J. E. A Call to Address AI “Hallucinations” and How Healthcare Professionals Can Mitigate Their Risks. *Cureus* **15**, e44720 (2023).
177. Yang, J. et al. Application of Artificial Intelligence to Advance Individualized Diagnosis and Treatment in Emergency and Critical Care Medicine. *Diagnostics* **14**, 687 (2024).
178. Evans, H. & Snead, D. Why do errors arise in artificial intelligence diagnostic tools in histopathology and how can we minimize them? *Histopathology* **84**, 279–287 (2024).
179. Mincu, D. & Roy, S. Developing robust benchmarks for driving forward AI innovation in healthcare. *Nat. Mach. Intell.* **4**, 916–921 (2022).
180. Wu, A. H. B. et al. Short- and Long-Term Cardiac Troponin I Analyte Stability in Plasma and Serum from Healthy Volunteers by Use of an Ultrasensitive, Single-Molecule Counting Assay. *Clin. Chem.* **55**, 2057–2059 (2009).
181. Celi, L. A. et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. *PLOS Digit. Health* **1**, e0000022 (2022).
182. Zahlan, A., Ranjan, R. P. & Hayes, D. Artificial intelligence innovation in healthcare: Literature review, exploratory analysis, and future research. *Technol. Soc.* **74**, 102321 (2023).
183. Kaushal, A., Altman, R. & Langlotz, C. Geographic Distribution of US Cohorts Used to Train Deep Learning Algorithms. *JAMA* **324**, 1212–1213 (2020).
184. Arcadu, F. et al. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *npj Digit. Med.* **2**, 92 (2019).
185. Finlayson, S. G. et al. The Clinician and Dataset Shift in Artificial Intelligence. *N. Engl. J. Med.* **385**, 283–286 (2021).
186. Chen, R. J. et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat. Biomed. Eng.* **7**, 719–742 (2023).
187. Zech, J. R. et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med.* **15**, e1002683 (2018).
188. Zhang, J. & Zhang, Z. Ethics and governance of trustworthy medical artificial intelligence. *BMC Med. Inform. Decis. Mak.* **23**, 7 (2023).
189. Subbaswamy, A. & Saria, S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Bio-statistics* **21**, 345–352 (2020).
190. Huang, J., Smola, A. J., Gretton, A., Borgwardt, K. M. & Scholkopf, B. Correcting sample selection bias by unlabeled data. In: *Proceedings of the 20th International Conference on Neural Information Processing Systems*, 601–608 (Neural Information Processing Systems Foundation, Inc., 2006).
191. Hassija, V. et al. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cogn. Comput.* **16**, 45–74 (2024).
192. Nasarian, E., Alizadehsani, R., Acharya, U. R. & Tsui, K.-L. Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. *Inf. Fusion* **108**, 102412 (2024).
193. Tjoa, E. & Guan, C. A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE Trans. Neural Netw. Learn. Syst.* **32**, 4793–4813 (2021).
194. Zhang, Y., Weng, Y. & Lund, J. Applications of Explainable Artificial Intelligence in Diagnosis and Surgery. *Diagnostics* **12**, 237 (2022).
195. Kavya, R., Christopher, J., Panda, S. & Lazarus, Y. B. Machine Learning and XAI approaches for Allergy Diagnosis. *Biomed. Signal Process. Control* **69**, 102681 (2021).
196. Das, D., Ito, J., Kadowaki, T. & Tsuda, K. An interpretable machine learning model for diagnosis of Alzheimer’s disease. *PeerJ* **7**, e6543 (2019).
197. Borys, K. et al. Explainable AI in medical imaging: An overview for clinical practitioners – Saliency-based XAI approaches. *Eur. J. Radiol.* **162**, 110787 (2023).
198. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777 (2017).
199. Zhao, X., Huang, X., Robu, V. & Flynn, D. BayLIME: Bayesian Local Interpretable Model-Agnostic Explanations. In: *37th Conference on Uncertainty in Artificial Intelligence*, 887–896 (2021).
200. Ahmad, M. A.-S. & Haddad, J. An Explainable AI Model for Predicting the Recurrence of Differentiated Thyroid Cancer. *arXiv*; <https://doi.org/10.48550/arxiv.2410.10907> (2024).
201. Sarp, S., Kuzlu, M., Wilson, E., Cali, U. & Guler, O. The Enlightening Role of Explainable Artificial Intelligence in Chronic Wound Classification. *Electronics* **10**, 1406 (2021).
202. El-Sappagh, S., Alonso, J. M., Islam, S. M. R., Sultan, A. M. & Kwak, K. S. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer’s disease. *Sci. Rep.* **11**, 2660 (2021).
203. Komorowski, P., Baniecki, H. & Biecek, P. Towards Evaluating Explanations of Vision Transformers for Medical Imaging. In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 3726–3732 (IEEE, 2013).
204. Chefer, H., Gur, S. & Wolf, L. Transformer Interpretability Beyond Attention Visualization. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 782–791 (IEEE, 2021).
205. Scott, I., Carter, S. & Coiera, E. Clinician checklist for assessing suitability of machine learning applications in healthcare. *BMJ Heal. Care Inform.* **28**, e100251 (2021).
206. Ghassemi, M., Oakden-Rayner, L. & Beam, A. L. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit. Health* **3**, e745–e750 (2021).
207. Kirkpatrick, P. New clues in the acetaminophen mystery. *Nat. Rev. Drug Discov.* **4**, 883–883 (2005).
208. Amini, M. M. et al. Artificial Intelligence Ethics and Challenges in Healthcare Applications: A Comprehensive Review in the Context of the European GDPR Mandate. *Mach. Learn. Knowl. Extr.* **5**, 1023–1035 (2023).
209. Gerke, S. & Rezaeikhonakdar, D. Privacy aspects of direct-to-consumer artificial intelligence/machine learning health apps. *Intell. Based Med.* **6**, 100061 (2022).
210. Attaran, M. Blockchain technology in healthcare: Challenges and opportunities. *Int. J. Healthc. Manag.* **15**, 70–83 (2022).
211. Altay, A., Learney, R., Güder, F. & Dincer, C. Sensors in blockchain. *Trends Biotechnol.* **40**, 141–144 (2022).
212. Nguyen, D. C. et al. Federated Learning for Smart Healthcare: A Survey. *ACM Comput. Surv.* **55**, 1–37 (2022).
213. Rahman, A. et al. Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges and open issues. *Cluster Comput.* **26**, 2271–2311 (2023).
214. DeCamp, M. & Lindvall, C. Latent bias and the implementation of artificial intelligence in medicine. *J. Am. Med. Inform. Assoc.* **27**, 2020–2023 (2020).
215. Felmingham, C. M. et al. The Importance of Incorporating Human Factors in the Design and Implementation of Artificial Intelligence for Skin Cancer Diagnosis in the Real World. *Am. J. Clin. Dermatol.* **22**, 233–242 (2021).
216. Macrae, C. Managing risk and resilience in autonomous and intelligent systems: Exploring safety in the development, deployment,

- and use of artificial intelligence in healthcare. *Risk Anal.* 1–18 <https://onlinelibrary.wiley.com/doi/10.1111/risa.14273> (2024).
217. McCarthy, C. et al. Implementation of High-Sensitivity Cardiac Troponin Assays in the United States. *J. Am. Coll. Cardiol.* **81**, 207–219 (2023).
 218. 35 U.S.C. 101 Inventions patentable. US Patent and Trademark Office webpage, <https://www.uspto.gov/web/offices/pac/mpep/s2104.html> (2022).
 219. 35 U.S.C. 102 Conditions for patentability; novelty. US Patent and Trademark Office webpage, <https://www.uspto.gov/web/offices/pac/mpep/s2156.html> (2024).
 220. Regulations under the PCT, Rule 5 – The Description, World Intellectual Property Organization webpage, <https://www.wipo.int/pct/en/texts/rules/r5.html> (2024).
 221. Method vs. system claims, PatentAssociate.com webpage, <https://patentassociate.com/2022/07/11/differences-between-method-and-system-claims/> (2022).
 222. Benčević, M., Habijan, M., Galić, I., Babin, D. & Pižurica, A. Understanding skin color bias in deep learning-based skin lesion segmentation. *Comput. Methods Programs Biomed.* **245**, 108044 (2024).
 223. Martin, D. et al. Effect of skin tone on the accuracy of the estimation of arterial oxygen saturation by pulse oximetry: a systematic review. *BJA Br. J. Anaesth.* **132**, 945–956 (2024).
 224. Daneshjou, R. et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. *Sci. Adv.* **8**, eabq6147 (2022).
 225. Groh, M. et al. Evaluating Deep Neural Networks Trained on Clinical Images in Dermatology with the Fitzpatrick 17k Dataset. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 1820–1828 (IEEE, 2021).
 226. Chiu, M. C., Wang, Y., Kuo, Y. J. & Chen, P. Y. DDI-CoCo: A Dataset for Understanding the Effect of Color Contrast in Machine-Assisted Skin Disease Detection. In: *2024 IEEE International Conference on Acoustics, Speech and Signal Processing*, 6905–6909 (IEEE, 2024).
 227. Chen, J. et al. Inclusive and Accurate Clinical Diagnostics Using Intelligent Computation and Smartphone Imaging. *ACS Sens.* **9**, 5342–5353 (2024).
 228. Eckman, M. H., Reed, J. L., Trent, M. & Goyal, M. K. Cost-effectiveness of Sexually Transmitted Infection Screening for Adolescents and Young Adults in the Pediatric Emergency Department. *JAMA Pediatr.* **175**, 81–89 (2021).
 229. Mital, S. et al. Estimated cost-effectiveness of point-of-care testing in community pharmacies vs. self-testing and standard laboratory testing for HIV. *AIDS* **37**, 1125–1135 (2023).
 230. El-Osta, A. et al. Does use of point-of-care testing improve cost-effectiveness of the NHS Health Check programme in the primary care setting? A cost-minimisation analysis. *BMJ Open* **7**, e015494 (2017).
 231. Lingervelder, D., Koffijberg, H., Kusters, R. & IJzerman, M. J. Health Economic Evidence of Point-of-Care Testing: A Systematic Review. *Pharmacoecoon. Open* **5**, 157–173 (2021).
 232. Tolley, A., Bansal, A., Murerwa, R. & Dicks, J. H. Cost-effectiveness of point-of-care diagnostics for AMR: a systematic review. *J. Antimicrob. Chemother.* **79**, 1248–1269 (2024).
 233. Ming, D. et al. Connectivity of rapid-testing diagnostics and surveillance of infectious diseases. *Bull. World Heal. Organ.* **97**, 242–244 (2019).
 234. Wiemken, T. L. & Kelley, R. R. Machine Learning in Epidemiology and Health Outcomes Research. *Annu. Rev. Public Health.* **41**, 21–36 (2020).
 235. Cocker, D. et al. Healthcare as a driver, reservoir and amplifier of antimicrobial resistance: opportunities for interventions. *Nat. Rev. Microbiol.* **22**, 636–649 (2024).
- G.-R.H. acknowledges the Basic Science Research Program through the National Research Foundation of South Korea (NRF) funded by the Ministry of Education (NRF-2021R1A6A3A14039885). E.R acknowledges support from the Medical Research Council (MR/W006774/1). R.A.M. acknowledges support of the i-sense EPSRC IRC in Early Warning Sensing Systems for Infectious Diseases (EP/K031953/1), The i-sense EPSRC IRC in Agile Early Warning Sensing Systems for Infectious Diseases and AMR (EP/R00529X/1), m-Africa MRC Global Challenge Research Fund. R.A.M., E.R. and D.G. acknowledge the EPSRC Digital Health Hub for AMR (EP/X031276/1). K.G. acknowledges support from the MEXT Quantum Leap Flagship Program (JPMXS0120330644) and the JST ASPIRE Program (JPMJAP2316) and is grateful to Serendipity Lab for collaboration opportunities.

Author contributions

A.O., D.D., S.T., K.G., and R.A.M. supervised the research. G.-R.H., A.G., and A.O. conceived the research, coordinated the writing process, wrote parts of the paper, and edited all sections. M.E., S.Y., R.G., B.P., D.Y., F.L., E.R., and D.G. wrote parts of the manuscript and edited all sections. All the authors contributed to the preparation of the manuscript. A.O. initiated the research.

Competing interests

A.O. and D.D. are inventors of issued patents and pending patent applications on computational POC sensors. K.G. is a shareholder of CYBO, LucasLand, and FlyWorks. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-58527-6>.

Correspondence and requests for materials should be addressed to Aydogan Ozcan.

Peer review information *Nature Communications* thanks Jeong Hoon Lee, who co-reviewed with Seungmin Lee, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

Acknowledgements

The authors acknowledge the support of the US National Science Foundation (NSF) PATHS-UP Engineering Research Center (NSF #1648451).