Smartphone-based DNA diagnostics for malaria detection using deep learning for local decision support and blockchain technology for security

Xin Guo[1,4], Muhammad Arslan Khalid[1,4], Ivo Domingos[2], Anna Lito Michala [2], Moses Adriko[3], Candia Rowel[3], Diana Ajambo[3], Alice Garrett[1], Shantimoy Kar[1], Xiaoxiang Yan[1], Julien Reboud [1], Edridah M. Tukahebwa[3] and Jonathan M. Cooper [1]✉

In infectious disease diagnosis, results need to be communicated rapidly to healthcare professionals once testing has been completed so that care pathways can be implemented. This represents a particular challenge when testing in remote, low-resource rural communities, in which such diseases often create the largest burden. Here, we report a smartphone-based end-to-end platform for multiplexed DNA diagnosis of malaria. The approach uses a low-cost paper-based microfluidic diagnostic test, which is combined with deep learning algorithms for local decision support and blockchain technology for secure data connectivity and management. We validated the approach via field tests in rural Uganda, where it correctly identified more than 98% of tested cases. Our platform also provides secure geotagged diagnostic information, which creates the possibility of integrating infectious disease data within surveillance frameworks.

There remains a substantial burden from infectious disease in low-resource rural communities, not least as a consequence of malaria. After two decades of decline in prevalence, the disease is now increasing in 13 countries, with around 228 million malaria cases and 405,000 deaths globally each year. More than 90% of these cases are in Africa[1]. Diagnostic testing continues to underpin control and prevention strategies, primarily through the use of rapid, point-of-care, lateral flow immunoassays, which are affordable, sensitive, specific, user-friendly, rapid and robust, equipment-free and deliverable devices, meeting the World Health Organization (WHO) ASSURED criteria[2].

Despite recent successes in testing, the 2019 WHO World Malaria Report[1] highlights that a considerable proportion of populations living in rural areas still do not have access to prompt diagnosis, and emphasizes the need for rapid integrated diagnostic testing that can inform treatment and underpin elimination strategies. Current reports also point to the need for infectious disease diagnostics to become embedded in regional or national case management systems, with improved digital connectivity, enabling local access to surveillance data in remote communities[3].

One challenge is that such digital connectivity must be compatible with the different levels of medical record keeping within low- and middle-income countries (LMICs), which may range from sophisticated internet-enabled systems in urban settings to more distributed and often fragmented infrastructure in low-resource rural districts (where paper records and registries are often considered the norm)[4]. Providing methods to enable connectivity between such rural communities and centralized medical facilities is particularly important[1], as this information drives the flow of healthcare resources from governments, healthcare systems and charities[5].

A second challenge is that data collected in rural settings are often reported and re-recorded as they travel through administrative structures, from village communities, local district offices and regional administrative centres to the health ministries of national governments. Thus, improving connectivity in decision making (as well as embedding trust in the recording, transfer and reporting of data) is important, as it informs the timing of national campaigns, such as regional mass drug administrations, when treating reservoirs of infectious diseases in local communities[6].

Linking the collection of patient healthcare diagnostic data from testing with geospatial information in communities is also increasingly important across all healthcare systems, allowing disease prevalence to be mapped in real time and interventions to be focused rapidly (whether this be for endemic diseases such as malaria in low-resource communities or for seasonal flu in high-resource healthcare settings[7]). The development of systems that can be readily adapted to accommodate existing reporting mechanisms and can ensure that information is transmitted securely will increase trust in the recorded data that underpin intervention, treatment and prevention strategies.

Within LMICs, sub-Saharan Africa is at the forefront of developing and adopting digital technologies to improve healthcare[8]. For example, mobile phones are already widely used in the logistics and implementation of diagnostics, surveillance, prevention and treatment programmes and have the potential to be fully integrated into geotagged information systems. Examples of mobile health applications using artificial intelligence (AI) also already exist, including those for eye diagnostics in rural settings[9]. Furthermore, mobile phone-enabled data connectivity has been used for imaging and cloud-based analysis to standardize malaria detection using rapid diagnostic test (RDT) lateral flow immunoassays, most commonly detecting the histidine-rich protein-2/3 (*hrp*2/3) antigen present in the malaria parasite[10].

However, problems with malaria diagnosis using RDTs have recently emerged. It was reported in 2016 that eight commercial RDTs gave sensitivities to detect malaria parasites of only

[1]Division of Biomedical Engineering, The James Watt School of Engineering, University of Glasgow, Glasgow, UK. [2]School of Computing Science, University of Glasgow, Glasgow, UK. [3]Vector Control Division, Ministry of Health, Kampala, Uganda. [4]These authors contributed equally: Xin Guo, Muhammad Arslan Khalid. ✉e-mail: jon.cooper@glasgow.ac.uk
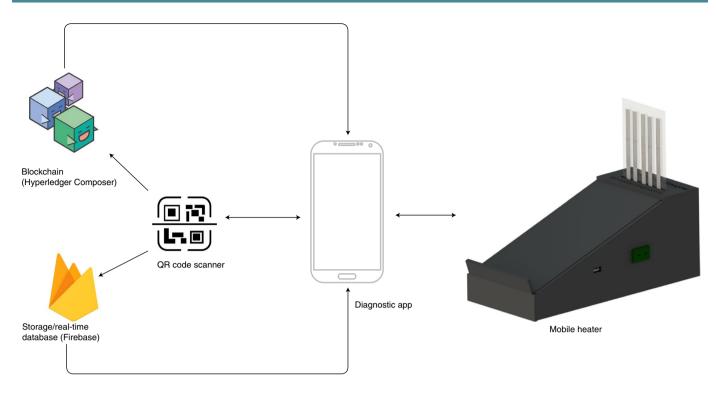
Fig. 1 | System architecture. Schematic of the system architecture, which includes a mobile heater, an Android app to control the heater and manage the diagnostic (paper-based microfluidic) assay (including start/stop), as well as a backend engine comprising a blockchain network for secure and trusted connections and a deep learning model for decision support.

~75%, with a negative predictive value of ~75% compared to the gold-standard laboratory-based polymerase chain reaction (PCR)[11]. The WHO[12] has recently attributed the poor performance of the current immunodiagnostic RDTs in part to the increasing prevalence of parasites with *hrp*2/3 gene deletions, causing a high prevalence of false-negative RDT immunodiagnostic results among symptomatic malarial patients[13]. The WHO now classifies *hrp*2/*hrp*3 deletions as a threat to malaria control and elimination and has called for disease monitoring using DNA-based assays[14,15]. However, these technologies require bulky and complex instrumentation and often need training in the interpretation of results where multiple test and control outputs are measured as part of a care pathway.

In this Article, we report a smartphone-based system that uses deep learning algorithms to provide local decision support for a multiplexed paper-based microfluidic lateral flow DNA molecular assay for *Plasmodium* sp. (Fig. 1). Our approach includes blockchain end-to-end connectivity to enable local healthcare staff to securely interpret and report the outcomes of diagnostic readouts. The system offers high diagnostic accuracy in the collection, interpretation and reporting of results, while improving the trustworthiness of data collection and transfer and providing end-to-end diagnostics for low-resource, rural settings. We illustrate its capabilities via field testing in rural communities in East Uganda.

Standard web-based security approaches are currently not sufficient to support the transfer of safety-critical and sensitive data including medical diagnostic information over wireless networks[16]. Alternative security systems that provide data provenance and management often require specialized equipment and trained personnel, adding an increased burden to resource-limited settings[17]. By contrast, blockchain provides a low-power and low-cost approach to incorporate digital security into governed processes, improving interoperability while supporting immutability and high levels of trust by allowing access for only 'endorsed' transactions. In such methods, an individual's information is stored in a tamperproof digital ledger, secured by a unique digital signature. Copies of this

ledger can be held locally by healthcare workers in a blockchain network, which ensures that it remains accessible and consistent, and that each change to the network is verified by a consensual mechanism. Such methods are widely used in financial transactions and have recently found application in geospatial tracking of individuals' interactions during the COVID-19 pandemic[7]. They have also been used before for medical data-sharing schemes in well-resourced settings to alleviate security and privacy issues[18-20].

The paper-based microfluidic diagnostic test we developed differentiates the endemic malaria-causing parasitic species in East Africa, *Plasmodium falciparum*, from all other parasitic species that cause malaria[21], enabling informed species-specific therapy or, in the future, surveillance. Our species-specific DNA-based diagnostic devices closely align with the ASSURED criteria of the WHO[2] and have been designed to be capable of being integrated with wireless communication systems through existing cellular networks, without additional requirements. By using a common, secure protocol GitHub OAuth[22], which is open-source and independent of the manufacturer, we can provide information exchange that is flexible (including images and metadata) and capable of integration into existing testing and reporting systems (so as to be accessible securely through devices such as smartphones, basic computers and mobile apps). Our system can be readily adapted to other sources of data input, including those for other different infectious and chronic disease co-morbidities. It can also be used to input information into existing digital health management systems being used either locally or nationally[23]. One potential example could involve linking our mobile platform into DHIS2, an open-source platform used extensively for healthcare-related digital information.

Diagnostic system
The key components of our diagnostic system (Fig. 1) are a mobile phone-controlled heater for DNA amplification through isothermal heating, a paper microfluidic chip for DNA testing[21], a blockchain architecture and an AI component for decision support (Fig. 2).
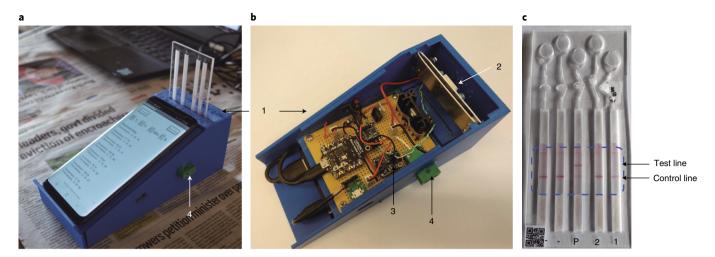
Fig. 2 | System design. a, The assembled device, showing the phone used to supply power, control the assay conditions (on/off, start/stop and temperature), collect results, communicate with the cloud, analyse data and provide geotagging. The diagnostic chip is shown, inserted into the heating element. The whole device, including the mobile phone, is lightweight (<500 g) and can be held in one hand, with the potential to enable diagnostics to be delivered anywhere (without the need to transport equipment, for example). **b**, Open section view of the device and associated circuit. The numbered parts, respectively, are (1) the casing and main body of the device, (2) the aluminium band for receiving the diagnostic device and conducting heat for the nucleic acid amplification assay, (3) circuit components including a microcontroller, heater controller and power supply unit, and (4) the external port for thermal calibration. **c**, The plastic cartridge including a microfluidic circuit with chambers for the LAMP reaction and lateral flow strips for readout, as well as the QR code for traceability. The dashed lines outline the cropped area for analysis by AI, with the test and control lines indicated (see Supplementary Fig. 6 for details).
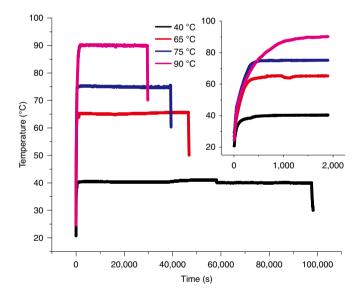

Fig. 3 | Mobile heater characterization. The temperature was recorded for different target temperatures, 90 °C (purple), 75 °C (blue), 65 °C (red) and 40 °C (black). The temperature decreased when the battery power became limited. The lifetime of the phone's batteries was dependent upon whether other functionalities were being used (for example, Wi-Fi connectivity). If required, the battery lifetime of the mobile phone was extended by use of a hand-held battery pack. Inset: zoom-in on the temperature 'ramping' up, demonstrating the effectiveness of the control of the proportional, integral and derivative algorithm (axes are the same as in the main panel). Heating to 65 °C took 10 min (600 s), providing the ability to run a full LAMP assay in <1 h (including sample processing[21]).

The heater's performance was characterized using a thermocouple to validate the temperatures sensed and applied, with different target temperatures. Figure 3 shows that the errors

between the real temperature and the target temperatures were all within ±0.5 °C (standard deviation) at 40, 65, 75 and 90 °C. A 10,000-mAh power bank could be used to provide more than the 9 h of the phone's battery life, if required, in the absence of mains power supply.

The performance of a blockchain network (Fig. 4), including latency and maximum throughput, can influence user experience when uploading diagnostic tests onto the cloud for data provenance and long-term preservation. Evaluation was targeted at these two main functions of the blockchain network and was measured against the benchmark Hyperledger Caliper 0.2.0. The test environment was an Ubuntu 18.04 virtual machine with 4-GB RAM and a four-core processor. The test was carried out with a Caliper 2-organization-1-peer test model and the whole process included 12 rounds (6 rounds for each transaction) with different numbers of transactions and send rates. Table 1 shows that the maximum throughput of the blockchain was ~10 transactions per second, and the system did not lose any data when tested under conditions involving high send rates (Supplementary Table 1 provides the system resource usage during the test).

A dataset with five categories containing 92 test images was used for training the AI decision support tool. These images comprised examples collected from the loop-mediated isothermal amplification (LAMP) diagnostic tests performed in the laboratory and included 11 '1N2P' (one negative 'N' and two positive 'P' test lanes), 13 '1P2N', 23 'double-positive', 15 'negative' and 30 'invalid' tests, as defined later). This library of results was used to test the accuracy of the convolutional neural network (CNN) model (classification credentials are provided in Supplementary Table 2). The test dataset and the training dataset were independent of each other and generated randomly (Methods). The sparse categorical cross-entropy loss function, which can be presented as equation (1), was used to evaluate the performance of the tool:
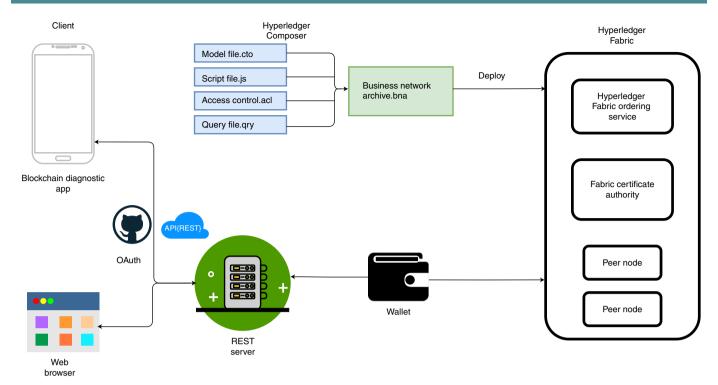
$$L_i = -\sum_j t_{i,j} \log(p_{i,j}) \tag{1}$$

Fig. 4 | System architecture of the blockchain network. The business network archive (BNA) file was packed with the model file, the script file, the access control file and the query file, and was deployed to a Fabric Runtime. Users' access to the blockchain network was from a web browser on a standard desktop or laptop computer or through the mobile app. OAuth 2.0 provided the authentication service. The users' network cards containing private and public keys were stored in their wallet.

where i indicates the samples and j the class (1–5), enabling the loss value L_i to be calculated using $_{pi,j}$ as the likelihood of prediction and $t_{i,j}$ the true value. The whole training process included 20,000 steps (equivalent to 500 epochs).

Figure 5a,b demonstrates the efficiency of training convergence (with an accuracy of 97.83%) with low loss (0.16 loss). The confusion matrix (Fig. 5c) shows that three of the diagnostic categories demonstrate 100% accuracy (1N2P, invalid and negative), while 8% of 1P2N were wrongly classified as invalid and only 4% of the double-positive cases were mistakenly predicted as 1N2P.

In a diagnostic context, any invalid classification has only minor repercussions if it is identified immediately, as the test is being performed at the point of care and can simply be repeated. Such an event translates into only minor delays (as the assay is rapid, at <1 h from 'sample to answer') and marginal increases in costs. For example, the mislabelling of double-positive tests, where a patient with a *P. falciparum* infection is detected but does not obtain a positive Pan test, will result in a prompt for the operator to repeat the test, with limited impact on the patient's outcome (the Pan primers cover all *Plasmodium* species, including *P. falciparum*, so any patients positive for *P. falciparum* should be positive for Pan). Alternatively, in cases where a test is misclassified as invalid or a double positive is misclassified as 1N2P, the patient would still be given the correct treatment. In all cases, as long as the test result is immediately available and the healthcare practitioners are informed by the decision support tool, any repetition of the assay will not result in a delay to treatment of >1 h. Thus, in both cases, the system's trustworthiness is not negatively impacted.

When compared to previously demonstrated approaches in decision support[10], our CNN model is able to efficiently and accurately provide outputs that can be trusted. Furthermore, it does so using smartphone edge computing that does not rely upon connectivity to the cloud, thus making it more suitable for use in rural settings and inherently more secure. Interpretation of the AI output provides a prompt that supports the practitioners' ability to understand what care pathway or treatment/action is suitable for each result in all possible eventualities. There is no further interpretation required by the practitioner to estimate possible misclassification probabilities or errors, enforcing high explainability with no need for transparency in the decision taken by the algorithm, so providing dimensions that translate to accountable reliable and trustworthy (ART) principles[24].

Compared to the state of the art in malaria cloud-based diagnostics[10], our approach greatly increases accuracy to 97.83%. It should also be noted that the use of AI decision support architectures allows the training to be continuously enhanced, thus providing further improved decision support capabilities over time. This would be beneficial for new tests that do not attain the same level of accuracy using laboratory-based training sets. In this context, the use of blockchain technology also ensures that images and results can be used for future training while abiding by stringent privacy constraints.

To validate the applicability of the platform, we performed field testing in a rural community in Uganda (as described in the Methods and elsewhere[21]). Information on the diagnostic device manufacturing and logging were recorded on the mobile phone in the UK, before arrival in Uganda, linked via a QR code, printed on each device. The operator in Uganda first scanned the QR code, before entering patient information (most commonly as an anonymized ID number), and then again before performing the sample nucleic acid amplification in the heater (Supplementary Fig. 1 and Supplementary Video 1). The analyst also scanned the QR code to record the results manually (as part of our internal validation protocols).

To further demonstrate the versatility of the blockchain/AI interface in enabling different AI strategies to be applied to the platform, the system was modified to detect each individual DNA-based lateral flow result, assign one of three classes (positive if two lines are

Table 1 | Blockchain performance evaluation results

Transaction name	Succeed	Failed	Send rate (TPS)	Average latency (s) (min–max)	Throughput (TPS)
ProduceDevice	50	0	5.1	0.41 (0.19–0.60)	5.0
	100	0	10.1	1.71 (0.42–3.04)	9.2
	200	0	20.1	10.97 (0.77–14.30)	10.2
	300	0	30.1	22.12 (6.02–23.10)	10.3
	400	0	40.1	28.57 (4.40–33.22)	10.3
	500	0	50.1	36.99 (9.49–46.43)	10.3
DoTheTest	50	0	5.1	0.42 (0.22–0.63)	4.8
	100	0	10.1	2.52 (0.27–4.36)	8.7
	200	0	20.1	10.87 (0.97–13.60)	10.2
	300	0	30.1	18.44 (1.17–23.73)	10.6
	400	0	40.1	26.70 (3.36–33.30)	10.6
	500	0	50.1	36.02 (17.05–43.02)	10.3

TPS, transactions per second.

found, negative if only one is read, invalid if none is present) and then combine these outcomes to reveal an overall readout, providing a decision support prompt for the result of the test as information on the detection or not of *Plasmodium* (Supplementary Fig. 2).

The ability of the platform to support different AI systems was validated by comparing an SSD ResNet50 neural net to obtain diagnostic test results with our CNN model. Both showed excellent performance in giving diagnostic predictions. The SSD ResNet50 and our CNN model have different advantages in analysing the diagnostic results. The CNN is both simpler and faster, while the SSD ResNet50 can provide more information, such as the result on each strip. Figure 5d provides the precision–recall curves for our CNN model and for the SSD ResNet50.

A total of 40 tests were carried out on 40 school children in a village setting. Only one test was incorrectly labelled by the model, which was a test on which the positive control (human gene) was labelled as negative (it should have been positive and thus valid). The model was able to correctly label 11 tests as invalid where experimental outcomes had been compromised (there were three reasons for the tests to be categorized as 'invalid', most commonly when channels were blocked and no sample reaching the strips, or when test strips did not show the control lane or when a control assay did not provide the expected result).

In future device iterations, when mass manufacturing processes such as moulding can be used, we would expect these issues to be greatly reduced. Importantly, this 'invalid' test outcome validates that our decision support system can identify such errors and inform the user if a test needs to be repeated. This can be performed readily in practice, at the point of care, so contributing to an enhancement of 'trust' in the technology by determining the efficacy of the test and distinguishing between failures and valid cartridges, so further improving the explainability of the model's decision.

All results were unblinded retrospectively at The University of Glasgow against independent manual recording to avoid potential bias, but also to compare the results with the gold-standard PCR assay results (Supplementary Methods). Of the 28 tests that were correctly assigned and valid, 16 were true positives (positive for the manually recorded test, the blockchain records and real-time PCR), six were true negatives, three were false negatives and three were false positives (with respect to the gold standard). The results of all tests are provided in Supplementary Data 1. The blockchain implementation ensured the security of transactions, opening up the possibility for integration into surveillance databases, while maintaining the required safety around data privacy.

Integration and trustworthiness

Globally, the number of smartphone users has grown from 2.5 billion to 3.2 billion from 2016 to 2019, and this number has been predicted to rise to around 3.8 billion in 2021[25] worldwide. It is also anticipated that sub-Saharan Africa will remain the fastest-growing region, with growth of ~5% and an additional 167 million subscribers over the period to 2025. Smartphone penetration in sub-Saharan African countries continues to rise in the general population, with Uganda reaching 23% and the local Ministry of Health using smartphone apps to provide frontline health workers with access to patient healthcare records[26]. As the number of individual mobile connections increases, so will Internet of Things (IoT) connections (now approaching nine billion devices globally) also rapidly expand. IoT is already predicted to be one of the principal vehicles driving improvements in global healthcare provision[8].

It is also predicted that increasing the versatility of smartphones will greatly reduce the costs of digital interventions when compared with traditional methods. Mobile health innovations are actively being developed for a wide variety of use cases[5,27]. However, there is still much room for improvement in terms of latency-tolerant solutions that do not depend on continuous network support, as there are still many geographic 'dead zones' with limited or no cellular service. We integrated this latter concern in the design of our system, which we ensured had a facility to enable all transactions to be stored in the mobile phone until a cellular service was available. Stakeholders in sub-Saharan Africa, including telecommunications industries and government agencies, continue to demonstrate a keen interest in this activity, with an overarching aim to address cost and infrastructure challenges.

By combining smartphones into this IoT context, we have demonstrated the capability, capacity and opportunity for edge computing to be used in such remote, rural geographical environments, where local conditions of connectivity to the internet (and thus the cloud) can be intermittent—so advancing the state of the art compared to existing cloud-based diagnostics such as the IoT solution for e-health[28]. Importantly, our system enables the integration of geotagged infectious disease prevalence and treatment information, endorsed through blockchain communication, to be introduced within digital data healthcare management systems, such as those being developed on the open-source platform DHIS2[23].

We have used machine learning approaches to classify images. In contrast to existing methods that have previously been implemented on a standard smartphone to assess data under a variety of conditions (for example, background colour, blur and light conditions) by
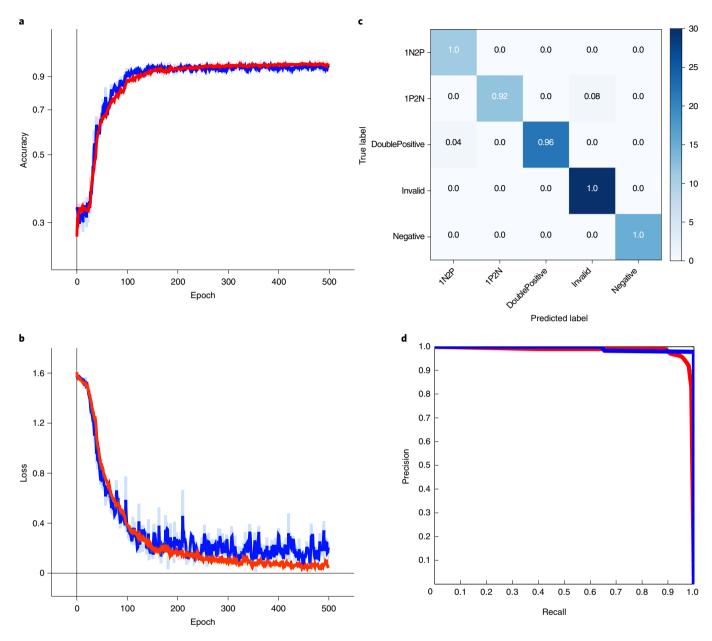
Fig. 5 | AI performance. a,b, Accuracy (between 0–1) (**a**) and loss (defined in equation (1)) (**b**) during the AI training process during epochs (the red trace provides results for the training dataset and the blue trace for the validation dataset). **c,** Confusion matrix of the test results for the CNN model, representing the predicted label and the actual label of every test image. The background colour of each grid of the matrix represents the number of images that were classified into that case (darker indicating more images), and the number on each grid is the relative success of predictions in that case (where 1.0 represents 100%). **d,** Precision–recall curves for our CNN (blue) and the SSD ResNet50 (red) to compare their predictive abilities, with recall measured as TP/(TP + FN) and precision as TP/(TP + FP), where TP are the true-positive predictions, FN the false-negative predictions and FP the false-positive predictions. If the confidence level of a prediction exceeds a threshold (for example, 0.8), the result is deemed a positive case; if not, it is a negative case. If the prediction matches the true label of the input then the output is true; if not, it is false. The area under the curve (AUC) of the CNN and ResNet50 curves, respectively, are 0.993 and 0.983.

using a CNN, we have obtained better, more accurate results when suitably trained with sufficient data. Necessarily, the size of the dataset is a key factor to be considered in the selection of a classification algorithm, particularly when environment conditions are varied[29]. Our results show that the CNN performs well, even when trained with a small dataset, and was sufficiently able to adjust to any lighting/background conditions encountered during testing in the rural community settings. Importantly, CNNs have very low demands in terms of memory and central processing unit in the inference stage,

making them suitable for deployment on mobile phones and IoT scenarios in under-served community settings[30,31].

We also note that, when compared to traditional computational vision approaches, which need developers to define with a high degree of granularity which features are important in every image, the CNN requires no expert analysis and adjustment in image classification tasks[30-32]. In this case, we trained our CNN with a dataset that only included images and corresponding labels, although we also note that our CNN model is retrainable, so it could be updated

with different datasets and used for different diagnoses or different multiplex number, without having to redesign the algorithm.

Trust in healthcare delivery is perceived through trust in the data recorded and generated through the particular technology/system[33]. Interestingly, the use of AI in neural networks has recently been challenged in terms of trustworthiness[34], particularly in safety-critical outputs, thus presenting a potential barrier in its acceptability in healthcare. Despite recent approvals of AI-as-medical-devices[35], when diagnosis accuracy is the critical priority, trust is still linked directly to the ease of identification of false predictions and their subsequent effect. Through the implementation of blockchain, we can improve data provenance and enable standardization, thus improving trustworthiness in our overall system.

To achieve this, we developed end-to-end trustworthiness, which was enabled through the system itself, with layers of trust mapped hierarchically onto the device, while using the mobile app to implement the CNN inference and the blockchain network. By applying this architecture, we were able to integrate three distinct layers of trustworthiness within our diagnostic system: trust in data accuracy at the sensor layer, supporting the accountability and reliability of AI; trust in the decisions generated from the data, within the application layer, supporting accountability and explainability; trust in the security of the whole system, within the networking and processing layer, addressing confidentiality, integrity and availability. By doing so, we propose that our system addresses the full scope of trust, required in AI systems, as defined by the ART principles[24].

We also addressed trustworthiness through authenticity and data privacy in the provisions of blockchain, with data integrity being addressed through the identification of valid/invalid tests within the CNN. Data availability was also supported by the blockchain architecture and the edge computing app, providing both global provenance and local diagnostic decision support. As a consequence, within this framework, the diagnostic data quality was ensured through the model's accuracy in evaluation, including, for example, in discarding invalid tests. Device interoperability was itself fundamentally supported by the blockchain architecture, and data freshness was ensured by the point-of-care collection and processing of the diagnostic test. Such features are closely aligned with the WHO recommendations to support the benefits of digital health involving the use of decision support systems for healthcare workers[36].

Finally, we note that the issue of trust, including that associated with cloud-enabled diagnostic testing, raises ethical concerns, including the capability and capacity to transmit personal identifiable information. Privacy preservation of private and individual data is of paramount importance, motivating the need for privacy preservation frameworks including the recently proposed BeepTrace networks[7]. Developing secure, trusted data transmission that can be endorsed is needed to overcome the concerns of privacy, security and data ownership. This is particularly important in decentralized diagnostics in low-resource areas, where data are collated and used by multiple agencies (including government, charities and universities). It is often the case that stakeholders involved in different aspects of the implementation of screening and treatment programmes within these 'care systems' are from different national states and may be operating under different legal frameworks. For example, data centres and servers, upon which information may be stored, may be subject to different data protection laws depending on their own location and jurisdiction, leading to potential issues not just over their ability to securely store data, but also over its ownership[37].

We believe that our approach provides a framework that could inform an open-source connectivity standard for disease surveillance, treatment and ultimately for elimination programmes to build upon this resource. It provides a secure mechanism to connect the actionable information from rural and urban healthcare infrastructures with governments and healthcare agencies to implement and improve care pathways and outcomes. By way of example, in the case given in this study, permissions were needed both from central and regional offices from within the Ministries of Education and Health to both collect and use the data, while including a duty to report epidemiological findings directly to government.

Limitations

Following guidelines from the WHO, we have developed an ASSURED system that provides point-of-care testing[2]. The approach is not fully equipment-free as it requires a mobile phone and a heater device, although we argue that this is the minimum necessity for assay testing and an improvement in overall vertical cost. We note that the requirement for the use of smartphones (over earlier models with non-smartphones) represents a potential barrier and should be further evaluated in-country. However, as discussed earlier, such phones are becoming increasingly pervasive and ubiquitous, with widespread use in diagnostic applications in sub-Saharan Africa.

In terms of AI trustworthiness, we have evaluated the possible misclassification errors and proven that, in all cases, any errors would lead to a prompt, requesting repetition of the test. Any unintended biases introduced through the training set did not demonstrate a negative effect. This is evidenced through both the accuracy of the model and the specific misclassifications. The misclassification cases were effectively leading to appropriate actions and thus to a fully trustworthy system, in accordance with the ART principles[24].

Our approach bases its safety and security provision on the implementation of blockchain features, although we recognize that recent studies have demonstrated that blockchain is not fully tamperproof, and that an approach using distributed ledger and a combination of classic cryptographic security for both the local data held in the app and the website application may be required in the future[35]. We also recognize that information is temporarily stored on the phone (for later propagation to the cloud) and that this, including similarly locally stored information on web browsers, could be another point with a security risk.

According to recent regulation, anonymization is no longer enough to preserve privacy[38], and strategies that include obscuring the exact geolocation while preserving the locality are being proposed in the literature[39]. In our approach, we have focused on primary anonymization, using IDs to protect sensitive information. In the future, further steps might be required to ensure privacy in the data collected on the cloud database. However, the geolocation information that we collect relates only to the place the test was conducted (for example, local clinic or community centre), which does not in itself individually identify the people visiting this facility (thus satisfying the privacy requirements in terms of geolocation obscurity).

Finally, and as stated above, our CNN system allows for models to be incrementally retrained, either as more data are accumulated in the cloud or as new diagnostic opportunities arise, and, in the future, the effect of this on accuracy must be evaluated (such that if higher-accuracy models are developed, the improved model can be updated to all phone users).

Conclusions

We have reported a smartphone-based end-to-end platform for multiplexed DNA-based lateral flow diagnostic assays in remote, low-resource settings. Our decision support tool provides automated detection of the results and their analysis, supporting human expertise, and transactions involved in data handling are secured, trusted and endorsed using blockchain technology. In anticipation of future AI guidelines for healthcare[40], we designed our platform so that it supports the following functionalities: explainability[41],

accuracy to enable AI decision trust, ethical use of data through privacy-preserving blockchain networks, interoperability to enable wider connectivity with divergent standards and policies, and data formatting for standardization and provenance. In the future, we will improve user-friendliness for practitioners in different sub-Saharan countries.

Methods
The diagnostic platform comprises both hardware and software. The hardware includes a three-dimensional (3D) printed mobile heater for LAMP-based diagnostics[42] as well as a mobile phone and a low-cost disposable sensor cartridge, while the software includes an Arduino program, an Android app and a Hyperledger blockchain network. These are described in detail below.

Hardware. *Circuit design for the diagnostic instrumentation.* The instrument was designed as a low-cost diagnostic device, controlled by an Android mobile phone. Integral to the delivery of the DNA-based diagnostic assay is LAMP, performed by integrating paper microfluidics within low-cost disposable cartridges, as described and validated in our previous publication[21]. To implement the assay, the sample needs to be heated to a constant temperature of 65 °C, which is enabled either using the mobile phone On-To-Go (OTG) functionality or with a back-up battery power pack (through a micro-USB port, a two-way switch and a voltage regulator (LM317T)). The temperature is maintained using a control circuit (the circuit diagram is provided in Supplementary Fig. 3), a microcontroller unit, two temperature sensors (one acting as reference, one measuring the cartridge temperature) and a heating unit. This forms a small, lightweight, low-cost and long-lasting instrument (Fig. 2a and Supplementary Fig. 4).

The microcontroller uses a Bluno Beetle Arduino board (DF Robot). The heating unit comprises a thermoelectric generator (TEG; Peltier module, 0.76 W, 600 mA, 2.5 V, 15 × 15 mm, RS) and an n-channel MOSFET (IRLB8721). The temperature sensor unit uses two AD-8495 analogue k-type thermocouple amplifiers (Adafruit). The TEG serves as the heat source, controlled by the MOSFET. Through interaction with a mobile phone, the state of the heater (and thus the function of the device) can be both monitored and managed, including switching and cycling heaters 'on' or 'off' to maintain constant isothermal heating. For all assays, temperature profiles during LAMP are recorded as part of the quality assurance process.

Device design. The casing of the heater was designed with Autodesk Inventor 2019 and 3D-printed (Stratasys F170) with acrylonitrile butadiene styrene co-polymer (Fig. 2a,b). The heater includes an aluminium band around the LAMP reaction chambers in the cartridge (numbered 2 in Fig. 2a and 4 in Supplementary Fig. 2) to enhance thermal transfer and ensure homogeneity of temperature across the device.

The heater enables control of the temperature of the LAMP reaction chambers embedded within the plastic microfluidic chip. This uses a proportional, integral and derivative (PID) control mechanism in the Arduino code, adjusting the duty cycle of the output pulse width modulated (PWM) signal. Supplementary Tables 3 and 4 provide indicative manufacturing and operating costs, on a laboratory scale, illustrating the suitability for resource-limited settings.

Blockchain network. The blockchain network, shown in Fig. 4, was based on the open development toolkit Hyperledger Composer and Hyperledger Fabric blockchain network, which were hosted on a Google Cloud server. The core of the Hyperledger Composer blockchain network is a business network archive (BNA), including a model file, script file, access control file and query file. The BNA is deployed to an existing Hyperledger Fabric Runtime (including fabric ordering service, certificate authority and peer nodes). Users need to use a peer card, which contains the public key and their private key, to obtain access to the blockchain network.

The support of the representational state transfer (REST) application program interface (API) and GitHub Oauth authentication allows users to access the blockchain from a web browser on a standard desktop computer or from the mobile phone using a purpose-built bespoke Android app. The database service in the cloud allows a central point to collect information, enabling later analysis on geotagged disease propagation in communities, with a secure point accessible by healthcare providers across the hierarchy of the healthcare system. Anonymization of information in this database ensures privacy, while trust in the recorded data is always maintained, greatly improving the endorsement and privacy aspects compared with either manual or email transfer of records.

The asset (in our case the diagnostic device), participants (manufacturers), operators (the healthcare workers involved in the delivery of the diagnostic assays and their analysis) and transactions (as connections) are all defined in the BNA file (Supplementary Fig. 5). The diagnostic device is addressed with a unique identifier (ID), and the related information (including, for example, the date of manufacture) is printed as a QR code on the device (Fig. 2c). Participants have their own ID and username stored on the chaincode (the ledger). The role they can play is limited

by access control, although they can create a new device record or update device information.

The algorithms are as follows:

Algorithm 1. ProduceDevice
    Input: device ID, test name, manufacturer, date of manufacture, expire date, bench number, production place, status
    Result: Add new device record to the chaincode
    If *device exists* **then**
    return
    else
    set test name, participant (manufacturer), date of manufacture, expire date, bench number, production place, status to device attribute
    get asset registry
    emit 'produce device' event

Algorithm 2. DoTheTest
    Input: device ID, status, operator, test date, patient ID, gender, weight, URL (link to image of device after test), result, geo-location
    Result: Add test information to existing device
    If *device does not exist* **then**
    return
    else
    update status
    set operator, test date, patient ID, gender, weight, URL, result, test place to device attribute
    emit 'do the test' event

Deep learning. We implemented two different neural networks to analyse the results as images of the devices after the diagnostic test[21], an example of which is shown in Fig. 2c. We developed a CNN model based on Keras TensorFlow 2.0, using an object-detection model within a Python program for data analysis as a fast classifier, helping local healthcare workers to test the results after each diagnosis. The functions include loading images from Firebase cloud storage and validating and comparing the test results with the records stored on the blockchain network.

An object detection model based on a faster region-based convolutional neural network (R-CNN) ResNet50 model was also developed alongside the CNN but was not implemented in our app. Instead, it was used as a gold-standard reference for post analysis to independently validate the results. Both methods were developed based on TensorFlow 2.0 Keras.

Classification network. The CNN model was developed and integrated into a mobile app to classify the images of the paper-based microfluidic diagnostic tests, automatically. The five-plex DNA diagnostic strips, including species-specific diagnostics for *Plasmodium* sp. as well as controls, were used as designed previously[21], based on lateral (capillary) flow with a control line and a test line (Fig. 2c). These comprise two test strips for detecting *Plasmodium falciparum* (Pf) and *Plasmodium* pan (Ppan), which cause malaria, one positive control channel (using a *BRCA1* human gene) and two negative control strips (one for each species) (see below).

The test strip in each channel has three possible outcomes: negative, positive and blank (invalid). Thus, using all combinations of results across the five lateral flow strips gives 243 different possible result scenarios, including operator errors. To reduce the complexity of classification, the outputs are subdivided into five categories, as described in Supplementary Table 2 ('+' for positive, '−' for negative, '/' for invalid). Supplementary Fig. 6 shows an example of the result in each class.

The training datasets were obtained by carrying out targeted tests on synthetically prepared samples. Positive samples were obtained from LAMP of a Pf target sequence (the WHO DNA standard obtained from the National Institute for Biological Standards) at 10^5 copies per reaction. Negative samples were obtained by LAMP using Pf primers and probes without any target (in this case using deionized water). Both networks were initially trained in the cloud. The trained network was then incorporated in the Android app for edge-computed decision support.

To increase the range of intensities in the bands available for training, amplicons were used at different amplification times (5, 10 and 15 min), leading to 100 images in each class. To reduce the training time and improve the accuracy, the images were cropped in the app to a small 16:9 picture that contained results (Fig. 2c). All training images were resized to 128 × 128 × 3 before sending to the model. An image generator (TensorFlow) was used to address the overfitting issue caused by the small dataset. During the training process, the image generator randomly adjusted the parameters of the input images, such as brightness, contrast, zoom range and orientation, at the beginning of every training step.

The CNN was based on the TensorFlow Keras sequential model, which is 'a plain stack of layers where each layer only has one input tensor and one output tensor'[43]. The structure of the sequential model is simple, allowing us to build it in a shorter time with Keras API, which generated computationally lightweight models suitable for smartphone deployment[44].

The structure of the CNN was fine-tuned by adjusting the number of layers and parameters such as the number of nodes, batch size and learning rate, through an iterative process to give our final CNN. To balance the model between overfitting and underfitting, several models with different structures were trained with the

same training dataset and tested with the same test dataset (the training and test sets are independent of each other) to increase accuracy and lower loss.

Our model hosted 16 layers—four convolution layers, four max-pooling layers, a flatten layer, three dropout layers and four dense layers. The model structure is shown in Supplementary Fig. 7. The convolutional layers extract features from the input images by scanning the input with a weighted matrix (convolution kernel). The process of generating a single feature map can be presented as

$$A_j = f\left(\sum_{i=1}^{N} I_i \times K_{i,j} + B_j\right) \tag{2}$$

Every input matrix I_i is convolved with kernels $K_{i,j}$, and a bias B_j is added to every element in the sum of convoluted matrices. The nonlinear activation function $f(x)$ is applied to the matrix. All convolution layers use the activation function ReLU (rectified linear unit) to improve the learning speed and nonlinearity of the model, setting all negative values of input matrices to zero. The max-pooling layer reduces the dimension of the output matrices of the previous convolution layer, using a 2×2 kernel with stride 2 to scan its input and taking the largest number from four adjacent elements.

To extract sufficient features and detail while reducing the number of parameters in the training process, four convolutional layers were implemented, with a pooling layer following each convolutional layer. The output of the last pooling layer was flattened to a 1D tensor and sent to the fully connected dense layers by the flatten layer. As the training dataset was relatively small and only had three categories, the model needed to have more fully connected (FC) layers and relatively fewer neurons[45]. Consequently, three dense layers (size 128) and one dense layer (size 5) were used to obtain better accuracy. Between every dense layer, a dropout layer was utilized to prevent overfitting. The first three dense layers also used ReLU as their activation function, with the last dense layer, which was the output layer, using SoftMax ($S(x_i)$) as its activation function to provide the predictions and their probability:

$$S(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \tag{3}$$

Object detection network. The same dataset without cropping was used and labelled with LabelImg for training the object detection model. There are five labels in the label map—negative, positive, empty, device and QR code—where negative, positive and empty indicate the outcome of each strip. After labelling, the images were divided into two separate subsets, 90% for training and 10% for testing, and corresponding tf-record (a format of the TensorFlow dataset) files were created.

Android app. The Android app was designed with Android Studio v3.5+ in Java. The minimum requirements of using this app are Android version >5.0 and Bluetooth 4.0. It provides different functions and screens for different users (screenshots are provided in Supplementary Fig. 1 and three clips of use in Supplementary Video 1 as the different participants). Manufacturers can add new device records to the blockchain from the app.

The operators also use the app to control the LAMP heater via Bluetooth. After each diagnostic test, the healthcare worker only needs to take a picture of the device and enter any metadata, as a text file, about the test. The app connects the image to cloud storage and updates the device information on the blockchain. Device information can be viewed simply by scanning the QR code on the manufactured device or manually entering the device ID.

The location API provided by Google Play is used to collect geographic information. When the app is launched, it requests permission for using the location data. Once a new device record is created or a diagnostic test is complete, the app obtains the location information using Wi-Fi, a cellular tower or GPS (depending on availability and battery charge levels). This geographic data, including latitude and longitude, are uploaded with other information to the blockchain network.

Multiplex LAMP system. The primer sets used for the LAMP assay were based on previously published[21] primer sequences for *P. falciparum*, and a *BCRA1* gene fragment serves as positive control. The primers were all purchased from Eurofins Genomics. The sequences were provided by Reboud et al.[21], and the reactions were amplified for no more than 45 min at 65 °C. For training the CNN, lateral-flow test strips were obtained from devices with experiments carried out with artificial Pf templates as described above.

Field testing. Field testing was carried out in Uganda and followed the same protocol as previously used[21] to demonstrate the functionality of the platform in the field. Briefly, we tested blood samples collected from 40 school children from Kocoge Primary School in Tororo District. This study was conducted as part of the activity carried out by the Vector Control Division (VCD) of the Ministry of Health (MoH) in Kampala, Uganda on neglected tropical diseases, and was approved by the VCD MOH Research and Ethics Committee (VCDREC/078) and Uganda National Council for Science and Technology (HS 2193). Anonymized pupil details

were computerized and tagged using ID numbers. The participants were 5–12 years old and gender balanced. No personal data were revealed to the investigators. Written informed consent of the children's parents and the head teacher were also obtained (see protocols in the previous study[21] for details). All samples were retrospectively retested by PCR in the UK[21].

Ethics approval was carried out on the basis of presumed positive, given the high prevalence of the disease, which is endemic in the region. All individuals were treated accordingly following testing under MoH Uganda guidance. Analyses were performed in the children's classrooms. For each individual, a finger-prick (~5 µl) of whole blood was used, with sample processing including sample lysis, DNA extraction and amplification performed using the paper 'origami' protocol, as previously published[21] (further details are provided in the Supplementary methods).

For each sample, the person running the test scanned the QR code of the device to be used (already 'created' by the manufacturer) and entered the required information on the test (Supplementary Video 1) before inserting the device into the heater, controlled by the phone for amplification. The QR code was scanned and a picture of the results taken, directly in the field without any specific control (that is, without any 'reader'), generating images within a changing environment. The phone then returned the results for interpretation by the 'analyst', who could provide decision support to the person in charge of treatment. All testing steps (including results) were also recorded manually to ascertain the validity of the results presented. When network connectivity was not available, the transactions were stored in the phone until connectivity was restored.

All analyses were double-blinded between on-site field testing and reference tests, performed retrospectively using real-time PCR in the laboratories at the University of Glasgow as a gold standard, as described in detail in the Supplementary methods. Data analysis was performed using Microsoft Excel for Mac (v16.44) and Origin (OriginLabs, v2016). After testing at the local community school in Uganda, all used paper devices and small plastic consumables were incinerated by burning, while glass slides and RDTs used as reference techniques were stored in a biohazard container for safe disposal at the VCD, Kampala.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study for Figs. 3 and 5 are available in the University of Glasgow's Enlighten repository with the identifier https://doi.org/10.5525/gla.researchdata.1106.

Code availability

The code for the Android app, the CNN and blockchain architecture is available through an open Zenodo repository (https://doi.org/10.5281/zenodo.4429293), which includes a GitHub repository for the code (https://github.com/XGuoo/BlockchainDiagnostics), and is licensed under an Open Source GNU GPLv3 Licence. The repository README.md markdown file describes the dataset and includes installation and demo instructions.

References

1. *World Malaria Report 2019* (WHO, 2019); https://www.who.int/news-room/feature-stories/detail/ world-malaria-report-2019
2. Mabey, D., Peeling, R. W., Ustianowski, A. & Perkins, M. D. Diagnostics for the developing world. *Nat. Rev. Microbiol.* **2**, 231–240 (2004).
3. *The Roll Back Malaria Strategy for Improving Access to Treatment Through Home Management of Malaria* (archived) (WHO, 2014); https://apps.who.int/iris/handle/10665/69057
4. Yukich, J. O. et al. A description of malaria sentinel surveillance: a case study in Oromia Regional State, Ethiopia. *Malar. J.* **13**, 88 (2014).
5. Wood, C. S. et al. Taking connected mobile-health diagnostics of infectious diseases to the field. *Nature* **566**, 467–474 (2019).
6. Alabdulatif, A., Khalil, I., Forkan, A. R. M. & Atiquzzaman, M. Real-time secure health surveillance for smarter health communities. *IEEE Commun. Mag.* **57**, 122–129 (2019).
7. Xu, H. et al. BeepTrace: blockchain-enabled privacy-preserving contact tracing for COVID-19 pandemic and beyond. *IEEE Internet Things J* . **8**, 3915–3929 (2021).
8. Holst, C. et al. Sub-Saharan Africa—the new breeding ground for global digital health. *Lancet Digit. Health* **2**, e160–e162 (2020).
9. Bastawrous, A. Increasing access to eye care. There's an app for that. Peek: smartphone technology for eye health. *Int. J. Epidemiol.* **45**, 1040–1043 (2016).
10. Scherr, T. F., Gupta, S., Wright, D. W. & Haselton, F. R. Mobile phone imaging and cloud-based analysis for standardized malaria detection and reporting. *Sci. Rep.* **6**, 28645 (2016).

11. Wanja, E. W. et al. Field evaluation of diagnostic performance of malaria rapid diagnostic tests in western Kenya. *Malar. J.* **15**, 456 (2016).

12. *Response Plan to pfhrp2 Gene Deletions Global Malaria Programme* (WHO, 2019).

13. Nolder, D. et al. Failure of rapid diagnostic tests in *Plasmodium falciparum* malaria cases in UK travellers: identification and characterisation of the parasites. *Int. J. Infect. Dis.* https://doi.org/10.1016/j.ijid.2021.05.008 (2021).

14. Grignard, L. et al. A novel multiplex qPCR assay for detection of *Plasmodium falciparum* with histidine-rich protein 2 and 3 (pfhrp2 and pfhrp3) deletions in polyclonal infections. *EBioMedicine* **55**, 102757 (2020).

15. Kreidenweiss, A. et al. Monitoring the threatened utility of malaria rapid diagnostic tests by novel high-throughput detection of *Plasmodium falciparum* hrp2 and hrp3 deletions: a cross-sectional, diagnostic accuracy study. *EBioMedicine* **50**, 14–22 (2019).

16. Esposito, C., De Santis, A., Tortora, G., Chang, H. & Choo, K. K. R. Blockchain: a panacea for healthcare cloud-based data security and privacy? *IEEE Cloud Comput.* **5**, 31–37 (2018).

17. Perakslis, E. D. Using digital health to enable ethical health research in conflict and other humanitarian settings. *Confl. Health* **12**, 23 (2018).

18. Farouk, A., Alahmadi, A., Ghose, S. & Mashatan, A. Blockchain platform for industrial healthcare: vision and future opportunities. *Comput. Commun.* **154**, 223–235 (2020).

19. Cheng, X., Chen, F., Xie, D., Sun, H. & Huang, C. Design of a secure medical data sharing scheme based on blockchain. *J. Med. Syst.* **44**, 52 (2020).

20. Vazirani, A. A., O'Donoghue, O., Brindley, D. & Meinert, E. Blockchain vehicles for efficient medical records management. *npj Digit. Med.* **3**, 1 (2020).

21. Reboud, J. et al. Paper-based microfluidics for DNA diagnostics of malaria in low resource underserved rural communities. *Proc. Natl Acad. Sci. USA* **116**, 4834–4842 (2019).

22. *Authorizing OAuth Apps. GitHub Developer Guide* (GitHub); https://developer.github.com/ apps/building-oauth-apps/ authorizing-oauth-apps/

23. Dehnavieh, R. et al. The District Health Information System (DHIS2): a literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. *Health Inf. Manag. J.* **48**, 62–75 (2019).

24. Vourganas, I., Stankovic, V. & Stankovic, L. Individualised responsible artificial intelligence for home-based rehabilitation. *Sensors* **21**, 2 (2020).

25. *Smartphone Users Worldwide* (Statista, 2020); https://www.statista.com/statistics/330695/number-of- smartphone-users-worldwide/

26. Okeleke, K. *Uganda: Driving Inclusive Socio-Economic Progress Through Mobile-Enabled Digital Transformation* (GSM Association, 2019); https://www.gsma.com/ mobilefordevelopment/wp-content/ uploads/2019/03/Uganda-Report- Driving-inclusive-socio-economic-progress-through-mobile-enabled- digital-transformation.pdf

27. Scherr, T. F., Moore, C. P., Thuma, P. & Wright, D. W. Evaluating network readiness for mHealth interventions using the Beacon mobile phone app: application development and validation study. *JMIR mHealth uHealth* **8**, e18413 (2020).

28. Kumar, S., Srivastava, R., Pathak, S. & Kumar, B. in *Intelligent Data Security Solutions for e-Health Applications* (eds Singh, A. K. & Elhoseny, M.) 219–235 (Elsevier, 2020).

29. Zhong, B. et al. A comparative study of image classification algorithms for *Foraminifera* identification. In *Proc. 2017 IEEE Symposium Series on Computational Intelligence* (SSCI) 1–8 (IEEE, 2017); https://doi.org/10.1109/SSCI.2017.8285164

30. Wang, P., Fan, E. & Wang, P. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognit. Lett.* **141**, 61–67 (2021).

31. Wang, L. et al. Comparative analysis of image classification methods for automatic diagnosis of ophthalmic images. *Sci. Rep.* **7**, 41545 (2017).

32. O'Mahony, N. et al. Deep learning vs. traditional computer vision. In *Advances in Computer Vision CVC 2019 Advances in Intelligent Systems and Computing* (eds Arai, K. & Kapoor, S.) Vol. 943 https://doi.org/10.1007/978-3-030-17795-9_10 (Springer, 2019).

33. Jaigirdar, F. T., Rudolph, C. & Bain, C. Can I trust the data I see?: a physician's concern on medical data in IoT health architectures. In *Proc. Australasian Computer Science Week Multiconference* Article no. 27, 1–10 (ACM, 2019); https://doi.org/10.1145/3290688.3290731

34. Cheng, M., Nazarian, S. & Bogdan, P. There is hope after all: quantifying opinion and trustworthiness in neural networks. *Front. Artif. Intell.* **3**, 54 (2020).

35. Cammarota, R. et al. Trustworthy AI inference systems: an industry research view. Preprint at https://arxiv.org/abs/2008.04449 (2020).

36. *Recommendations on Digital Interventions for Health System Strengthening* (WHO, 2019).

37. *Global Strategy on Digital Health 2020–2025* (WHO, 2021); https://cdn.who.int/media/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf

38. *General Data Protection Regulation (GDPR)—Official Legal Text* (European Parliament and Council of the European Union, 2018); https://gdpr-info.eu/

39. Acharya, J., Bonawitz, K., Kairouz, P., Ramage, D. & Sun, Z. Context-aware local differential privacy. In *Proc. 37th International Conference on Machine Learning* Vol. 119, 52–62 (PMLR, 2020).

40. Sun, T. Q. & Medaglia, R. Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare. *Gov. Inf. Q.* **36**, 368–383 (2019).

41. Díez, J., Pérez-Núñez, P., Luaces, O., Remeseiro, B. & Bahamonde, A. Towards explainable personalized recommendations by learning from users' photos. *Inf. Sci.* **520**, 416–430 (2020).

42. Notomi, T. et al. Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* **28**, 63 (2000).

43. Chollet, F. The sequential model (Keras); https://keras.io/guides/sequential_model/

44. TensorFlow Lite. ML for mobile and edge devices (TensorFlow); https://www.tensorflow.org/lite

45. Shabbeer Basha, S. H., Ram Dubey, S., Pulabaigari, V. & Mukherjee, S. Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing* **378**, 112–119 (2019).

Acknowledgements

Author contributions

X.G. and M.A.K. developed and implemented the decision support system and, with I.D., implemented it on mobile platforms. M.A.K., A.G., S.K. and X.Y. performed experiments in the laboratory in Glasgow. A.G., M.A.K., M.A., J.R., E.M.T. and J.M.C. designed the field study. M.A.K., A.G., C.R., D.A., J.R. and J.M.C. carried out the field study. X.G., A.L.M., J.R. and J.M.C. analysed the data and wrote the manuscript. All authors edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41928-021-00612-x.

Correspondence and requests for materials should be addressed to J.M.C.

Peer review information *Nature Electronics* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

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

nature research

Corresponding author(s):   Jonathan M. Cooper

Last updated by author(s):   Jul 4, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data was collected using the software developed for the study, available here https://github.com/XGuoo/BlockchainDiagnostics, developed using TensorFlow 2.X (https://www.tensorflow.org/), Hyperledger Composer v0.19/0.20 (https://hyperledger.github.io/), Hyperledger Composer rest server v0.19/0.20 (https://hyperledger.github.io/), Hyperledger Composer playground v0.19/0.20 (https://hyperledger.github.io/), and Hyperledger Fabric v1.1/1.2 (https://hyperledger.github.io/), as well as Android Studio v3.5+ (https://developer.android.com/studio)
Data analysis	Data was analysed using Microsoft Excel for Mac (v16.44) and Origin (OriginLabs, v2016)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study for Figures 3-5 are available in the University of Glasgow's Enlighten repository with the identifier DOI http://dx.doi.org/10.5525/gla.researchdata.1106.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The study was not statistically-powered. We sampled 40 subjects as a proof-of-concept of the performance of the app and blockchain approach to perform in the field. The number was chosen in line with ethics requirement to suit the length of the study. We anticipated that 40 subjects would be enough to provide convincing arguments on the reproducibility of the platform.
Data exclusions	No data was excluded.
Replication	No replication was required as no conclusion on the clinical outcomes were sought.
Randomization	In the field study, participants were selected randomly by teachers. Participants were not allocated to groups as no clinical analysis was performed.
Blinding	The results of the field study were benchmarked in a double-blind fashion with retrospective PCR.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	no covariate is relevant to the study, but we worked with a gender balanced cohort, age 5-12.
Recruitment	Participants were selected randomly by teachers. Any bias would only impact the ratio of positives to negatives, which is not an outcome of the study. Biaises can include the fact that teachers would like the students to behave well during the study, to reflect well on the School, but also they may favour fragile children as they will be in contact with the nurse who may provide additional care. However this does not impact the results of our study.
Ethics oversight	Vector Control Division (MOH) Research and Ethics Committee, VCDREC/078 and Uganda National Council for Science and Technology (HS 2193)

Note that full information on the approval of the study protocol must also be provided in the manuscript.