



Guo, X. et al. (2021) Smartphone-based DNA diagnostics for malaria detection using deep learning for local decision support and blockchain technology for security. *Nature Electronics*, 4(8), pp. 615-624.
(doi: [10.1038/s41928-021-00612-x](https://doi.org/10.1038/s41928-021-00612-x))

There may be differences between this version and the published version.
You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/244276/>

Deposited on 17 June 2021

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

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

Authors list

Xin Guo^{1,±}, Muhammad Arslan Khalid^{1,±}, Ivo Domingos², Anna Lito Michala², Moses Adriko³, Candia Rowell³, Diana Ajambo³, Alice Garrett¹, Shantimoy Kar¹, Xiaoxiang Yan¹, Julien Reboud¹, Edridah M. Tukahebwa³, Jonathan M. Cooper^{1,*}

Affiliations

¹ Division of Biomedical Engineering, The James Watt School of Engineering, University of Glasgow, G12 8LT Glasgow, United Kingdom;

² School of Computing Science, University of Glasgow, Glasgow, G12 8RZ, UK

³ Vector Control Division, Ministry of Health, Kampala, Uganda

[±]These authors contributed equally

* corresponding author: jon.cooper@glasgow.ac.uk

Abstract

In infectious disease diagnosis, results need to be rapidly communicated to doctors once testing has been completed, in order for care pathways to be implemented. This is a 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 Internet-of-Things platform for multiplexed DNA malaria diagnosis. The approach uses a low-cost paper-based microfluidic diagnostic test, combined with deep learning algorithms for local decision support and blockchain technology to provide 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, providing the opportunity for the integration of infectious disease data within surveillance frameworks.

Main Text

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 its prevalence, the disease is now increasing in 13 countries, with ~228 million malaria cases and 405,000 deaths globally each year. More than 90% of these cases are in Africa¹. Diagnostic testing continues to underpin control and prevention strategies, primarily through the use of rapid, point-of-care, lateral flow immunoassays, as affordable, sensitive, specific, user-friendly, rapid and robust, equipment-free and deliverable devices, meeting the World Health Organization (WHO) ASSURED criteria².

Despite recent successes in testing, the 2018 WHO World Malaria Report¹ highlights that a considerable proportion of populations living in rural areas still do not have access to prompt diagnosis, and emphasises the need for rapid integrated diagnostic testing that can inform treatment and underpin elimination strategies. Current reports also point to the particular 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 communities³.

One of the major challenges is that solutions for 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, present 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 centralised medical facilities is particularly important,¹ as this information drives the flow of healthcare resources from governments, healthcare systems and charities⁵.

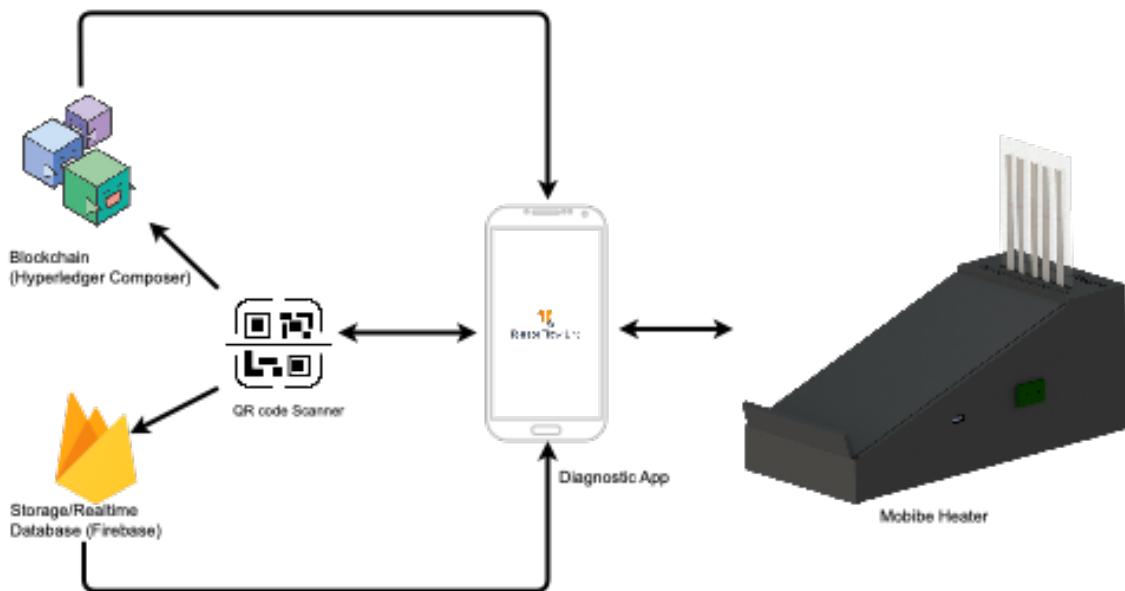


Fig. 1. System architecture: The figure shows the system's architecture including a mobile heater, an android app to control the heater and manage the diagnostic (paper-based microfluidic) assay (including start/stop), as well as a backend engine comprising a blockchain network for secure and trusted connections and a deep learning model, for decision support.

A second challenge is that data collected in rural settings is often reported and re-recorded as it travels through the 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, but also 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 communities⁶.

Linking the collection of patient healthcare diagnostic data from testing with geospatial information in communities is also becoming increasingly important across all healthcare systems, enabling real-time mapping of disease prevalence and interventions to be rapidly focussed (whether this be for endemic diseases such as malaria in low resource communities, or for seasonal flu or respiratory illnesses in high-resource healthcare settings⁷). The development of systems that can be readily adapted to accommodate existing reporting mechanisms, and which can ensure that information is transmitted securely, will increase trust in the recorded data that underpins intervention, treatment and prevention strategies.

Within LMICs, sub-saharan Africa is now at the forefront of developing and adopting digital technologies to improve healthcare⁸. For example, mobile phones are already widely used in the logistics and implementation of diagnostics, surveillance, prevention and treatment programmes and thus have the potential to be fully integrated into geotagged information systems. Examples of mobile health applications using artificial intelligence also already exist, including those for eye diagnostics in rural settings, where patient burden often takes its greatest toll⁹. Other early examples of mobile phone enabled data connectivity have used imaging and cloud-based analysis for the standardisation of malaria detection based upon rapid diagnostic test (RDT) lateral flow immunoassays, detecting the Histidine Rich Protein-2/3 (*hrp2/3*) antigen present in the malaria parasite¹⁰.

However, recently there have been significant failings in RDTs in malaria diagnosis, which first emerged in the last five years, when Wanja *et al.*¹¹, showed that eight commercial RDTs gave sensitivities to detect malaria parasites of only ~75%, with a negative predictive value of ~75%, compared to the “gold-standard” laboratory-based polymerase chain reaction (PCR). The WHO¹² has recently attributed the poor performance of the current immunodiagnostic RDTs in part to the

increasing prevalence of parasites with *hrp2/3* gene deletions, causing a high prevalence of false-negative RDT immunodiagnostic results amongst symptomatic malarial patients¹³. The WHO now classifies *hrp2/hrp3* deletions as a threat to malaria control and elimination and has called for disease monitoring using DNA based assays^{14,15}. However, these technologies currently require significant 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 Internet-of-Things (IoT) system (Figure 1) comprising edge compute nodes, that uses deep learning algorithms to provide local decision support for a multiplexed 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 read-outs. 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 in field testing undertaken in rural 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 networks¹⁶. Alternative security systems that provide data provenance and management often require specialised equipment and trained personnel, adding an increased burden to resource-limited settings¹⁷. In 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 individuals’ 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 pandemic⁷. 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 use differentiates the endemic malaria-causing parasitic species in East Africa, *Plasmodium falciparum*, from all other parasitic species that cause malaria²¹, 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 WHO² 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 OAuth²² that is open-source and independent of manufacturer, we can provide information exchange that is flexible (including images and metadata) and capable of integration into existing testing and reporting systems (such as to be accessible securely through devices such as smart phones, basic computers and mobile apps). Our system can not only be readily adapted to other sources of data input, including other different infectious and chronic disease co-morbidities, but can also be used to input information into existing digital health management systems, being used either locally, or nationally²³. One example could involve, in future, 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, a blockchain architecture and a AI component for decision support (pictured in Figure 2). The heater’s performance was characterised 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

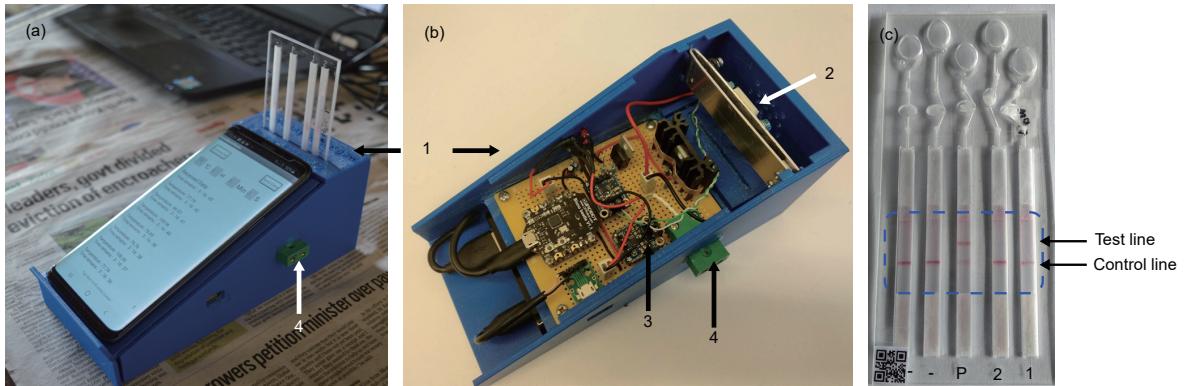


Fig. 2. System design: (a) The assembled device showing the phone used to supply power, control the assay conditions (On/Off, Start/Stop and temperature), collect results, communicate with the cloud, analyse data and provide geotagging – as an integrated IoT-edge computing diagnostic device. The diagnostic chip is shown, inserted into the heating element. The whole device, including the mobile phone is lightweight (<500g) 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; 4 external port for thermal calibration; (c) The plastic cartridge including microfluidic circuit with chambers for LAMP amplification reaction and lateral flow strips for readout, as well as QR code for traceability. The dashed lines represent the cropped area for analysis by AI, with the test and control line shown (see Supplementary Figure S6 for details).

target temperatures were all within $\pm 0.5^\circ\text{C}$ (standard deviation) at 40°C , 65°C , 75°C and 90°C . A 10,000 mAh power bank could be used to provide more than the 9 hours of the phone's battery life, if this is required, in the absence of mains power supply.

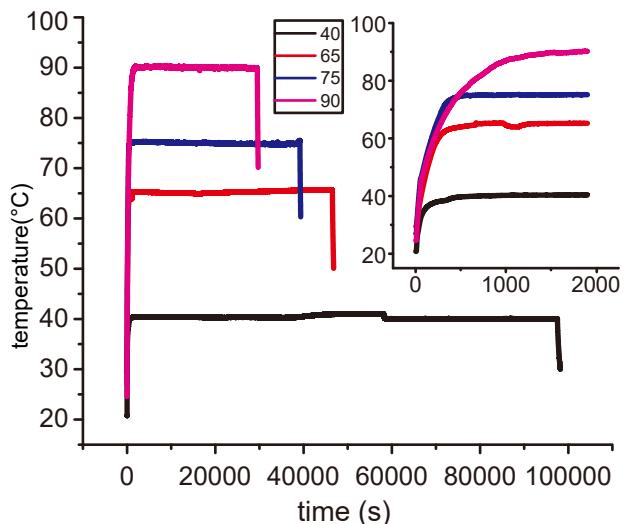


Fig. 3 Mobile heater characterisation. The temperature ($^\circ\text{C}$) was recorded over a period of 24 hours for different target temperatures, e.g. 90°C (purple), 75°C (blue), 65°C (red) and 40°C (black). The temperature decreased when the battery power became limited, indicating the capability of stable heating for up to 12h at 65°C (the temperature most commonly used for LAMP) from a single charge of a mobile phone. The lifetime of the phone's batteries was however dependent upon whether other functionalities were being used (e.g. Wi-Fi connectivity). Inset (detail) shows the temperature “ramping” up, demonstrating the effectiveness of the control of the proportional, integral, and derivative (PID) algorithm. Heating to 65°C took 10 min (600s), providing the ability to run a full LAMP assay in <1h (including sample processing²¹).

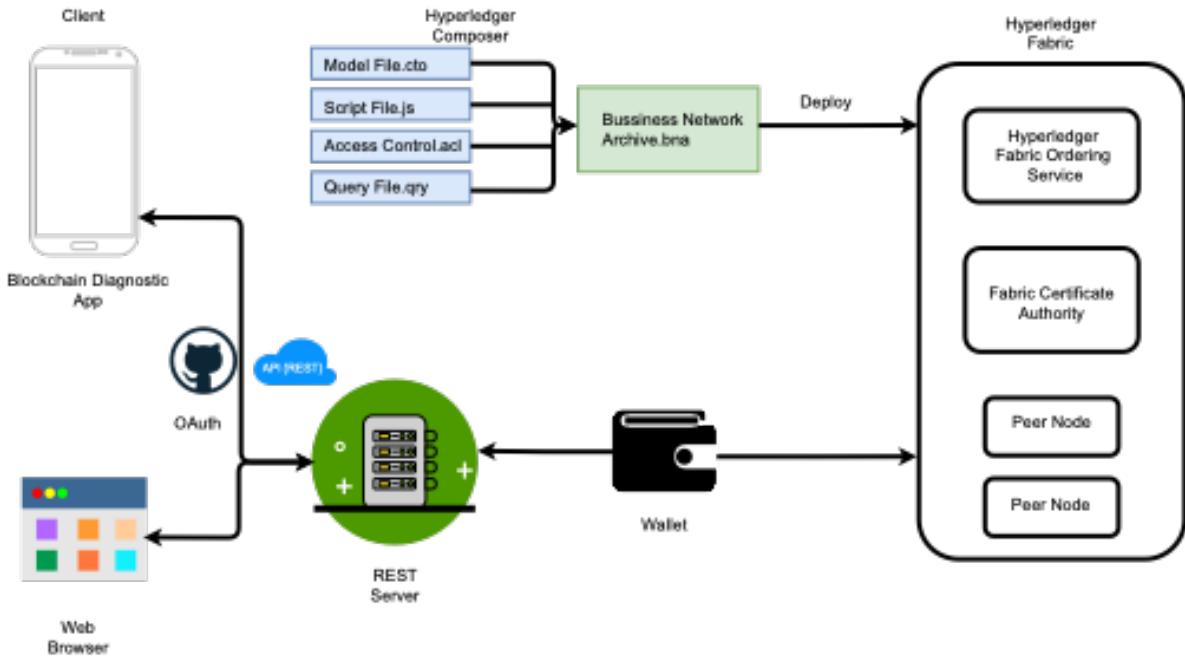


Fig. 4. System architecture of the blockchain network. The business network archive (BNA) file was packed with the Model file, the script file, the access control file and the query fil, and was deployed to a Fabric runtime. Users access to the blockchain network was from a web browser on a standard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

The performance of a blockchain network (Fig. 4), including latency and maximum throughput, can both influence user experience when uploading diagnostic tests onto the cloud for data provenance and long term preservation. The performance evaluation was targeted at the 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 transaction and send rates. Table 1 show that the maximum throughput of the blockchain was ~10 transitions per second in processing transactions, and the system did not lose any data when tested under conditions involving high send rates (Supplementary Table S1 provides the system resource usage during the test).

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 |

A dataset with five categories containing 92 test images was used for training the AI decision support tool. These comprised examples that were collected from the 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 CNN model (classification credentials are included in the Supplementary Table S2). The test dataset and the training dataset were independent of each other and generated randomly – see Methods Section). The sparse categorical cross-entropy loss function, which can be present as Equation (1), was used to evaluate the performance of the tool, accordingly.

$$L_i = - \sum_j t_{i,j} \log (p_{i,j}) \quad (1)$$

Where i indicates the samples and j the class (1-5), enabling the loss value L_i to be calculated using $p_{i,j}$ as the likelihood of prediction and $t_{i,j}$ the true value. The whole training process included 20,000 steps.

Figure 5a-b demonstrates the efficiency of training convergence (with an accuracy of 97.83%) with low loss (0.16 loss). The confusion matrix (Figure 5c) shows that three of the diagnostic categories demonstrated 100% accuracy (1N2P, double-positive, invalid and negative), whilst 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 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 hr from “sample-to-answer”) and marginal increases in costs. For example, the mislabelling of double-positive tests, where a patient with a *Plasmodium 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 Pf, so that any patients positive for Pf should be positive for Pan). Alternatively, in cases when a test is either misclassified as invalid, or when 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 was immediately available, and the healthcare practitioners was informed by the decision support tool, any repetition of the assay would not result in a delay to treatment of >1h. 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 was able to efficiently and accurately provide outputs that can be trusted. Furthermore it does so using smartphone edge computing that does not rely on 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 which translate to Accountable Reliable and Trustworthy (ART) principles²⁴.

Compared to the state of the art in malaria cloud-based diagnostics¹⁰, our approach significantly 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 which 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 whilst abiding to stringent privacy constraints.

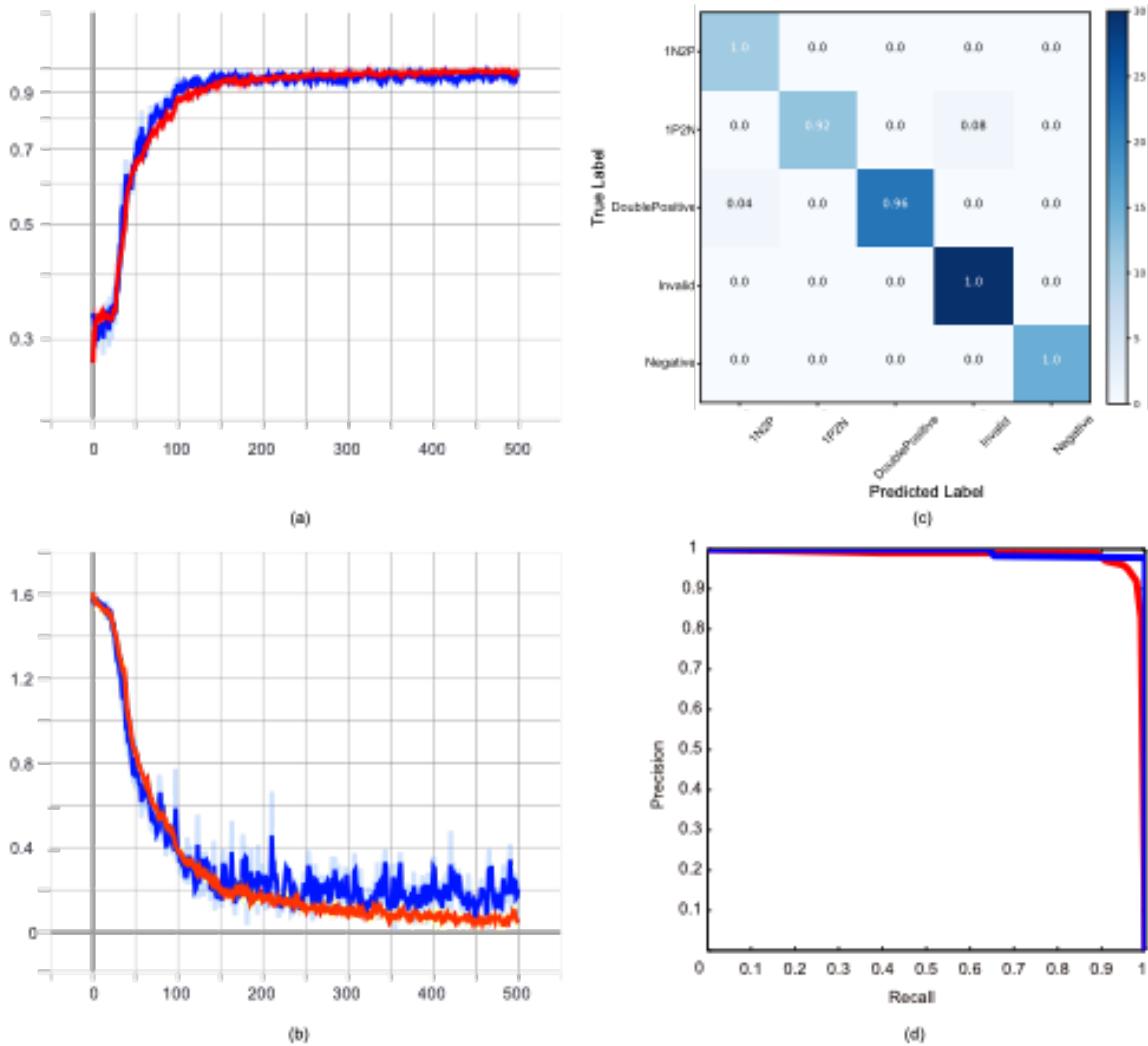


Fig. 5. AI performance. (a) Accuracy and (b) the loss during the AI training process during epochs (the blue trace provides results for the training dataset, red trace is for the validation dataset); (c) shows a confusion matrix of the test results of the CNN model, representing the predicted label and 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 there were more images), whilst the number on each grid is the relative success of predictions in that case (where 1.0 represents 100%). (d) The precision-recall curves for our CNN (blue) and the SSD ResNet 50 (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 is false negative predictions and FP are false positive predictions. If the confidence level of a prediction exceeds a threshold (e.g. 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 AUC of the CNN and ResNet50 respectively are 0.993 and 0.983.

In order to validate the applicability of the platform, we performed field testing in a rural community in Uganda, as previously described in Methods and elsewhere²¹. Information on device manufacturing and logging were recorded on the mobile phone in the UK, prior to arrival in Uganda, linked via a QR code, printed on each device. The Operator in Uganda first scanned the QR code, prior to entering patient information (most commonly, as an anonymised ID number) and then again, prior to performing the sample nucleic acid amplification in the heater (Supplementary Figure S1 and Movie M1). 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 found, negative if only one is read, invalid if none are present) and then combine these outcomes to reveal an overall read-out, providing a decision support prompt for the result of the test (as information on the detection or not of Plasmodium), Supplementary Figure S2.

The ability of the platform to support different AI systems was validated by comparing an SSD ResNet50 neural net for the diagnostic test results, with our CNN model, both showing 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 could provide more information, such as the result on each strip. Figure 5d provides the precision-recall curves of our CNN model and 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 for which 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 molding can be used, we would expect these issues to be reduced significantly. 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.

The results were unblinded at The University of Glasgow retrospectively against independent manual recording to avoid potential bias, but also to compare the results with the gold standard PCR assay results, see Supplementary Methods. Of the 28 tests that were correctly assigned and valid, 16 were true positives (positive for both the manually recorded test, the blockchain records and real-time PCR), 6 were true negatives, whilst 3 were false negatives and 3 were false positives (with respect to the gold standard). The results of all tests are available in Supplementary DataFile 1. The blockchain implementation ensured the security of transactions, opening up the possibility for integration into surveillance databases, whilst 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 2021²⁵ 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% where the local Ministry of Health use smartphone apps to provide frontline health workers with access to patient healthcare record²⁶. As the numbers of individual mobile connections increases, so IoT connections, which now approach 9 billion devices globally, will also rapidly expand. IoT is already predicted to be one of the principle vehicles to drive improvements in global healthcare provision⁸.

It is also predicted that increasing the versatility of smartphones will greatly reduce 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 upon 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 IoT-edge 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 demonstrated the capability, capacity and opportunity for edge computing to be used in such 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-health²⁸. Importantly, our system enables the integration of geotagged infectious disease prevalence and treatment information, that is endorsed through blockchain communication, to be introduced within digital data healthcare management systems, such as those being developed on the open source platform DHIS2²³.

We used machine learning approaches to classify images. In contrast to existing methods which have previously been implemented on a standard smartphone to assess data under a variety of conditions (e.g. background colour, blur and light conditions) by using a CNN, we 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 varied²⁹. 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 that were encountered during testing in the rural community settings. Importantly, CNNs have very low demands in memory and CPU 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 as to 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, such that 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/system³³. Interestingly, the use AI in neural networks has recently been challenged in terms of *trustworthiness*³⁴ particularly in safety-critical outputs, so presenting an potential barrier in its *acceptability* in healthcare. Despite recent approvals of AI-as-medical-devices³⁵, 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 standardisation 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 IoT device whilst 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, namely: trust in data accuracy at the sensor layer, supporting Accountability and Reliability of AI; trust in the decisions generated from the data, within the application layer, supporting Accountability and Explainability; as well as trust in the security of the whole system, within the networking & 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 principles²⁴.

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 of 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 WHO recommendations to support the benefits of digital health involving the use of decision support systems for healthcare workers³⁶.

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 of privacy preservation frameworks including the recently proposed BeepTrace networks⁷. 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 decentralised diagnostics in low-resource areas, where data is collated and used by multiple agencies (including government, charities, 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, datacentres and servers, upon which information may be stored, may be subject to different data protection laws, depending upon their own location and jurisdiction, leading to potential issues not just over their ability to securely store data, but also over its ownership³⁷.

We believe that this study 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 rural and urban healthcare infrastructures with governments and healthcare agencies of the actionable information needed 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, whilst including a duty to report epidemiological findings directly to government.

Limitations of the study

Following guidelines from WHO, we have presented an ASSURED IoT system that provides point-of-care testing³⁸. The approach is not fully equipment-free as it requires a mobile phone and the 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 of the use of smartphones (over earlier models of non-smart mobile phones) represents a potential barrier and should be further evaluated in-country. However, as discussed earlier, the smartphone is 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 miss-classification errors and proven that in all cases, any errors would lead to a prompt, requesting repetition of the test. Any unintended biases that are 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 principles²⁴.

Our approach bases its safety and security provision on the implementation of blockchain features, although we recognise that recent studies have demonstrated that blockchain is not fully tamperproof, and that an approach utilising 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 future³⁵. Further we recognise 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 attachment.

According to recent regulation, anonymisation is no longer enough to preserve privacy³⁹ and strategies which include obscuring the exact geolocation whilst preserving the locality are being proposed in literature⁴⁰. In our approach, we have focused on primary anonymisation, using IDs to protect sensitive information. Further steps might, in future, 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 (e.g. local clinic or community centre) which does not in itself individually identifying the people visiting this facility (and thus satisfying the privacy requirements in terms of geolocation obscurity).

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

Conclusions

We have reported a mobile phone based end-to-end IoT edge integrated platform for multiplexed 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 healthcare⁴¹, we designed our platform such that it supports the following functionalities: explainability⁴²; 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 standardisation and provenance. In the future, we will look to improve user-friendliness for the practitioners in different sub-Saharan countries.

References

1. World malaria report 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. *Nature Reviews Microbiology* vol. 2 231–240 (2004).
3. WHO | The Roll Back Malaria strategy for improving access to treatment through home management of malaria (archived). *WHO* (2014).
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* doi:10.1038/s41586-019-0956-2.
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. (2020).
8. Holst, C. *et al.* Sub-Saharan Africa—the new breeding ground for global digital health. *The Lancet Digital Health* vol. 2 e160–e162 (2020).
9. Bastawrous, A. Increasing access to eye care.. There's an app for that. Peek: Smartphone technology for eye health. *International Journal of Epidemiology* vol. 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**, 1–9 (2016).
11. Wanja, E. W. *et al.* Field evaluation of diagnostic performance of malaria rapid diagnostic tests in western Kenya. *Malar. J.* **15**, 1–10 (2016).
12. Health Organization, W. *Response plan to pfhrp2 gene deletions Global Malaria Programme*.
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.* (2021)

- doi:10.1016/j.ijid.2021.05.008.
- 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**, (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 Chesmal Siriwardhana and Donal O'mathuna. *Conflict and Health* vol. 12 1–8 (2018).
 - 18. Farouk, A., Alahmadi, A., Ghose, S. & Mashatan, A. Blockchain platform for industrial healthcare: Vision and future opportunities. *Computer Communications* vol. 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**, 1–11 (2020).
 - 20. Vazirani, A. A., O'Donoghue, O., Brindley, D. & Meinert, E. Blockchain vehicles for efficient Medical Record management. *npj Digital Medicine* vol. 3 1–5 (2020).
 - 21. Reboud, J. *et al.* Paper-based microfluidics for DNA diagnostics of malaria in low resource underserved rural communities. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 4834–4842 (2019).
 - 22. Authorizing OAuth Apps | GitHub Developer Guide. <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 Information Management Journal* vol. 48 62–75 (2019).
 - 24. Vourgas, I., Stankovic, V. & Stankovic, L. Individualised Responsible Artificial Intelligence for Home-Based Rehabilitation. *Sensors* **21**, 2 (2020).
 - 25. Smartphone users worldwide 2020 | Statista. <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>.
 - 26. Kenechi Okeleke, *Uganda: Driving inclusive socio-economic progress through mobile-enabled digital transformation*. (2019). <https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/03/Uganda-Report-Driving-inclusive-socio-economic-progress-through-mobile-enabled-digital-transformation.pdf> [accessed 02/06/2021]
 - 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**, (2020).
 - 28. Kumar, S., Srivastava, R., Pathak, S. & Kumar, B. Cloud-based computer-assisted diagnostic solutions for e-health. in *Intelligent Data Security Solutions for e-Health Applications* 219–235 (Elsevier, 2020). doi:10.1016/b978-0-12-819511-6.00012-1.
 - 29. Zhong, B. *et al.* A Comparative Study of Image Classification Algorithms for Foraminifera Identification.
 - 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**, 1–11 (2017).
 - 32. O' Mahony, N. *et al.* Deep Learning vs. Traditional Computer Vision.
 - 33. Can I Trust the Data I See? | Proceedings of the Australasian Computer Science Week Multiconference. <https://dl.acm.org/doi/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*. (2020).
36. *recommendations on digital interventions for health system strengthening*.
37. WHO. *Draft global strategy on digital health 2020– 2025*.
38. Draft global strategy on digital health 2020– 2025. https://www.who.int/docs/default-source/documents/gs4dhdcaa2a9f352b0445bafbc79ca799dce4d.pdf?sfvrsn=f112ede5_42.
39. General Data Protection Regulation (GDPR) – Official Legal Text. <https://gdpr-info.eu/>.
40. Acharya, J., Bonawitz, K., Kairouz, P., Ramage, D. & Sun, Z. Context-Aware Local Differential Privacy. (2019).
41. 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).
42. Díez, J., Pérez-Núñez, P., Luaces, O., Remeseiro, B. & Bahamonde, A. Towards explainable personalized recommendations by learning from users' photos. *Inf. Sci. (Ny)*. **520**, 416–430 (2020).
43. Notomi, T. *et al.* Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res.* **28**, 63 (2000).
44. The Sequential model. https://keras.io/guides/sequential_model/.
45. TensorFlow Lite | ML for Mobile and Edge Devices. <https://www.tensorflow.org/lite>.
46. 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. (2019) doi:10.1016/j.neucom.2019.10.008.

Acknowledgements

The authors are grateful to Dr Poppy Lamberton for help with ethical clearance and facilitation, Epigem Ltd for help with device manufacturing, Erin O'shaughnessy for device assembly, Omer Tufayl for initial app GUI development, as well as Alon Atuhire for sampling and treatment in the field in Uganda.

The study was supported by the UK Global Challenges Research Fund, the Scottish Funding Council, and Engineering and Physical Sciences Research Council (EPSRC) Institutional Support Fund (Grant EP/R512813/1), as well as by EPSRC EP/R01437X/1 also supported by the National Institute for Health Research, and EP/T029765/1. AG acknowledges support from EPSRC studentship EP/N509668/1.

Authors contributions

X.G., M.A.K. developed and implemented the decision support system, and with ID 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.

Material and methods

The diagnostic platform comprises both hardware and software. The hardware included a 3D printed mobile heater for Loop-mediated isothermal amplification (LAMP)-based diagnostics⁴³ as well as mobile phone and a low-cost disposable sensor cartridge, whilst 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 was LAMP amplification, performed by integrating paper microfluidics within low cost disposable cartridges, as described and validated in our previous publication²¹. In order to implement the assay, the sample needs to be heated to a constant temperature of 65°C, which was 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 was maintained using a control circuit (circuit diagram provided in Supplementary Figure S3), a micro-controller unit, two temperature sensors (one acting as reference, one measuring the cartridge temperature) and a heating unit. This formed a small, light-weight, low-cost and long-lasting instrument (Figure 2a and Supplementary Figure S4).

The micro-controller used a Bluno Beetle Arduino board (DF Robot). The heating unit comprised a thermoelectric generator, TEG, (Peltier Module, 0.76 W, 600 mA, 2.5 V, 15 x 15 mm, RS) and an n-channel MOSFET (IRLB8721). The temperature sensor unit used two AD-8495 analogue k-type thermocouple amplifiers (Adafruit). The TEG served 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) could be both monitored and managed, including switching/cycling heaters “on” or “off” to maintain constant isothermal heating. For all assays, temperature profiles during the LAMP amplification were 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 (ABS) co-polymer (Figure 2a-b). The heater included an aluminium band around the LAMP reaction chambers in the cartridge (numbered 2 in Figure 2a and 4 in Supplementary Figure S2), to enhance thermal transfer and ensure homogeneity of temperature across the device.

The heater enabled to control the temperature of the LAMP reaction chambers embedded within the plastic microfluidic chip. This used a proportional, integral, and derivative (PID) control mechanism in the Arduino code, adjusting the duty cycle of the output PWM signal. Supplementary Tables S3 and S4 provide indicative manufacturing and operating costs, on a laboratory scale, illustrating the suitability for resource-limited settings.

Blockchain network

The blockchain network, shown in Figure 4, was based upon 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 was a business network archive (BNA), including a model file, script file, access control file and query file. The BNA was deployed to an existing Hyperledger Fabric runtime (including fabric ordering service, certificate authority and peer nodes). Users needed 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 REST API and GitHub Oauth authentication allowed 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 allowed a central point to collect information, enabling for later analysis on geo-tagged disease propagation in the communities, with a secure point

accessible by healthcare providers across the hierarchy of the healthcare system. Anonymisation of information in this database ensured privacy, while the trust in the recorded data was always maintained, greatly improving the endorsement and privacy aspects, compared with either the manual or email transfer of records.

The asset (which in our case is the diagnostic device), the participants (manufacturers), operators (as the healthcare workers involved in the delivery of the diagnostic assays and their analysis) and the transactions (as connections) were all defined in the BNA file (Supplementary Figure S5). The diagnostic device was addressed with a unique identifier (ID), and the related information (including e.g. date of manufacture), were printed as a QR code on the device (Figure 2c). Participants had their own ID and username stored on the chaincode (i.e. the ledger). The role that they can played was limited by the access control, although they could either create a new device record or update a device information.

The algorithms show as follows:

| Algorithm 1. Produce the device | Algorithm 2. Perform the test |
|--|--|
| <p>Input: device ID, test name, manufacturer, date of manufacture, expire date, bench number, production place, status</p> <p>Result: Add new device record to the chaincode</p> <pre> If device exists then return else set test name, participant (manufacturer), date of manufacture, expire date, bench number, production place, status to device attribute </pre> <p>get asset registry emit 'produce device' event</p> | <p>Input: device ID, status, operator, test date, patient ID, gender, weight, URL (link to image of device after test), result, geo-location</p> <p>Result: Add test information to existing device</p> <pre> 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 </pre> <p>emit 'do the test' event</p> |

Deep Learning

We implemented two different neural networks to analyse the results as images of the devices after the diagnostic test²¹, an example of which is shown in Figure 2c. We developed a convolutional neural network (CNN) model based upon 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 results after each diagnosis. The functions included loading images from Firebase cloud storage, validating and comparing the test results with the records stored on the blockchain network.

An object detection model based upon 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 previously²¹, based on lateral (capillary) flow showing a control line and a test line (Figure 2c). This comprised two test strips for detecting *Plasmodium falciparum* (Pf) and *Plasmodium pan* (Ppan) that

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 had three possible outcomes which were negative, positive and blank (invalid). Thus, using all combinations of results across the five lateral flow strips gave 243 different possible result scenarios, including operator errors. To reduce the complexity of classification, the outputs were sub-divided into five categories which are described in Table S1 (“+” for positive, “-” for negative, “/” for invalid), while Supplementary Figure S6 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 the LAMP amplification of a Pf target sequence (WHO DNA standard obtained from National Institute for Biological Standards) at 10^5 copies/reaction. Negative samples were obtained by LAMP amplification using Pf primers and probes without any target (in this case using de-ionised 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, 15 mins), 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 (Figure 2c). All training images were resized to 128*128*3 before sending to the model. An image generator (TensorFlow) was used to address the over-fitting 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”⁴⁴. The structure of the sequential model was simple, allowing us to build it in a shorter time with Keras API which generated computationally light-weight models suitable for smartphone deployment⁴⁵.

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 over-fitting and under-fitting, 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 sixteen layers including four convolution layers, four max-pooling layers, a flatten layer, three dropout layers and four dense layers. The model structure is shown in Supplementary Figure S7. The convolutional layers extracted features from the input images by scanning its input with a weighted matrix (convolution kernel). The process of generating a single feature map could be presented as Equation 2:

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

Every input matrix I_i was convolved with kernels $K_{i,j}$, and a bias B_j was added to every element in the sum of convoluted matrices. The non-linear activation function $f(x)$ was applied to the matrix. All convolution layers used the activation function ReLU to improve the learning speed and non-linearity of the model, setting all negative values of input matrices to zero. The Max-Pooling layer reduced 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.

In order to extract sufficient features and detail whilst 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 then 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 neurons⁴⁶. 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, see Equation 3.

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

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 which are 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 sub-sets, 90% for training and 10% for testing, and corresponding tf-record (a format of 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 (the screenshots could be found in Figure S1 and three clips of use in Movie M1 as the different participants). Manufacturers could 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 simply only needed to take a picture of the device and enter any metadata as a text file, about the test. The app connected the image to cloud storage and updated 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.

We used the location API provided by Google Play to collect geographic information. When the App was launched, it requested permission for using the location data. Once a new device record was created or a diagnostic test was complete, the App obtained the location information using Wi-Fi, a cellular tower or GPS (depending upon availability and battery charge levels). This geographic data, included latitude and longitude and was uploaded with other information to the blockchain network.

Multiplex LAMP system

The primer sets used for the LAMP assay were based on previously published²¹ primer sequences for *P. falciparum*, whilst a BCRA1 gene fragment served as a positive control. The primers were all purchased from Eurofins Genomics. The sequences are provided by Reboud et al.²¹, and the reactions were amplified for no more than 45 min at 65°C. For the training of 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 used²¹ 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). Anonymised pupil details were computerised 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 previous study for details). All samples were re-tested by PCR in the UK retrospectively.²¹

Ethics approval was upon the basis of presumed positive, given the high prevalence of the disease, which is endemic in the region. All individuals were treated accordingly following the test under MoH Uganda guidance. Analyses were performed in the children 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 published²¹ (further details are provided in 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 (see Movie M1) 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 (i.e. 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 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.

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>.

Code availability

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

Supplementary Data

Supplementary Methods

The following methods are described in more detail in a previous publication¹⁷, and are summarised here for ease of understanding.

The test was based upon an integrated device that has two component parts each performing different functions, namely: (i) a paper “origami” sample processing unit (which were fabricated without the need for either specialised facilities or a clean room by using a hot wax printer, Xerox ColorQube 8570); and (ii) a paper-based microfluidic lateral flow device (designed in the University of Glasgow and fabricated by Ustar, China) which was mounted in plastic cassettes (manufactured by CNC machining at Epigem Ltd) containing blister pack finger pumps and a microfluidic network of chambers (in which LAMP amplification was carried out, prior to diagnostic result visualisation on the lateral flow strips).

The two component parts were manually assembled prior to testing, providing a sample-to-answer solution, in which patient sample was introduced into the “origami” device, and following a series of processing steps, enabled by successive folding of the paper¹⁷, the sample was enriched and purified, introducing parasite DNA into the microfluidic lateral flow devices where it was amplified using an isothermal LAMP protocol at 65°C, before the answer was visualized on the lateral flow device.

For each individual person tested, a finger-prick blood sample was taken and was split into four aliquots accordingly:

1. A drop of blood was added at the community setting in Uganda to a CareStart™ Malaria rapid immunodiagnostic test (RDT), provided by the Ministry of Health using One step Malaria HRP-II (*P.falciparum*) and pLDH (*P. pan*) Antigen Rapid Test (SD Bioline Malaria Ag). 5µL of blood was placed on the RDT reception area and four drops of CareStart™ assay diluent were added. The RDT was allowed to stand for 20 minutes during which time capillary flows carry to the sample along the lateral flow strip to enable immunodiagnosis. The RDT provided a field-based reference technique, helping with the validation of the nucleic acid-based diagnostic (see 3 below).
2. A drop of blood was added at the community setting in Uganda to a glass slide to produce a thick smear, which was stained using Giemsa dye and visualized using optical microscopy by trained technicians working for the Vector Control Division of the Ministry of Health, Kampala. The drop of blood from the finger-prick was placed at the center of a glass slide and spread as a smear. The slide was dried for 20 mins and was then stained with 10% Giemsa stain for a further 10 mins before being examined under a compound microscope using a 100x oil immersion objective lens (Olympus Corporation Tokyo, Japan). Malaria parasites were counted against white blood cells to give a parasitaemia/µL of blood, as a reference method.
3. The same blood sample was used at the community setting in Uganda immediately (before clotting <2min) for the microfluidic paper DNA test. An aliquot of 10 µL of lysis buffer containing magnetic beads was added to 5 µL of blood. After 5 mins incubation, a further aliquot of binding buffer was added and the DNA was captured on the beads, which were spotted onto the sample zone of the origami paper device. All residues were eluted away by dispensing a volume-independent aliquot of a washing buffer onto the sample reception area of the origami device, with bead-bound parasitic DNA being retained within the paper matrix. This was origami-folded against distributing channels of the milled plastic cassette. The sample preparation processing required two manipulations including the addition of washing and elution buffers, a procedure readily performed within the community setting by the local healthcare workers.

The device was sealed with an acetate film to prevent evaporation and LAMP amplification was performed at 65°C for 40 minutes using the mobile phone controlled 3D machined heating unit. *Plasmodium* DNA was amplified with the amplicons were generated with both biotin and FITC-labelled

primers, as a complex construct bearing two potential ligand binding sites. The running buffer chamber blister/finger pumps were manually pressed to enable the buffer to push the LAMP amplified, labelled amplicons onto the lateral flow device. The sample migrated along the lateral flow strip by capillarity and appropriate/selective DNA binding occurred at different zones, depending upon the ligands present. Strong colors were generated by the plasmon interactions from colloidal gold labels within the appropriate visualization zones. The result was imaged after 40 mins using the mobile phone camera.

4. A drop of blood was also added in the community setting in Uganda to an FTA card (for retrospective analysis with PCR amplification assays). Finger-prick whole blood samples were spotted onto FTA classic cards (Whatman) which were air-dried. They were shipped to the UK where they were stored at -20°C until used for the “gold-standard” PCR reference analysis. Sample discs from the dried sample spot were punched out and placed in a clean tube into which phosphate buffered saline (PBS), pH 7.4, was added before incubation at room temperature for 2 hours. An aliquot of 20 µL protein K was added to the sample and it was incubated at ~55°C for 1 hr before processing using MagaZorb® DNA Mini-Prep Kit (Promega) and elution with proprietary elution buffer.

PCR was run on double-blinded samples by a second (different) person to the one who had performed the DNA extraction. The extracted samples were coded by a third person and real-time PCR was performed in an ABI 7500 fast real-time PCR instrument (Applied Biosystems) as the gold-standard reference for assessing the diagnostic performance of the paper-based microfluidic device. The results were unblinded by the third person.

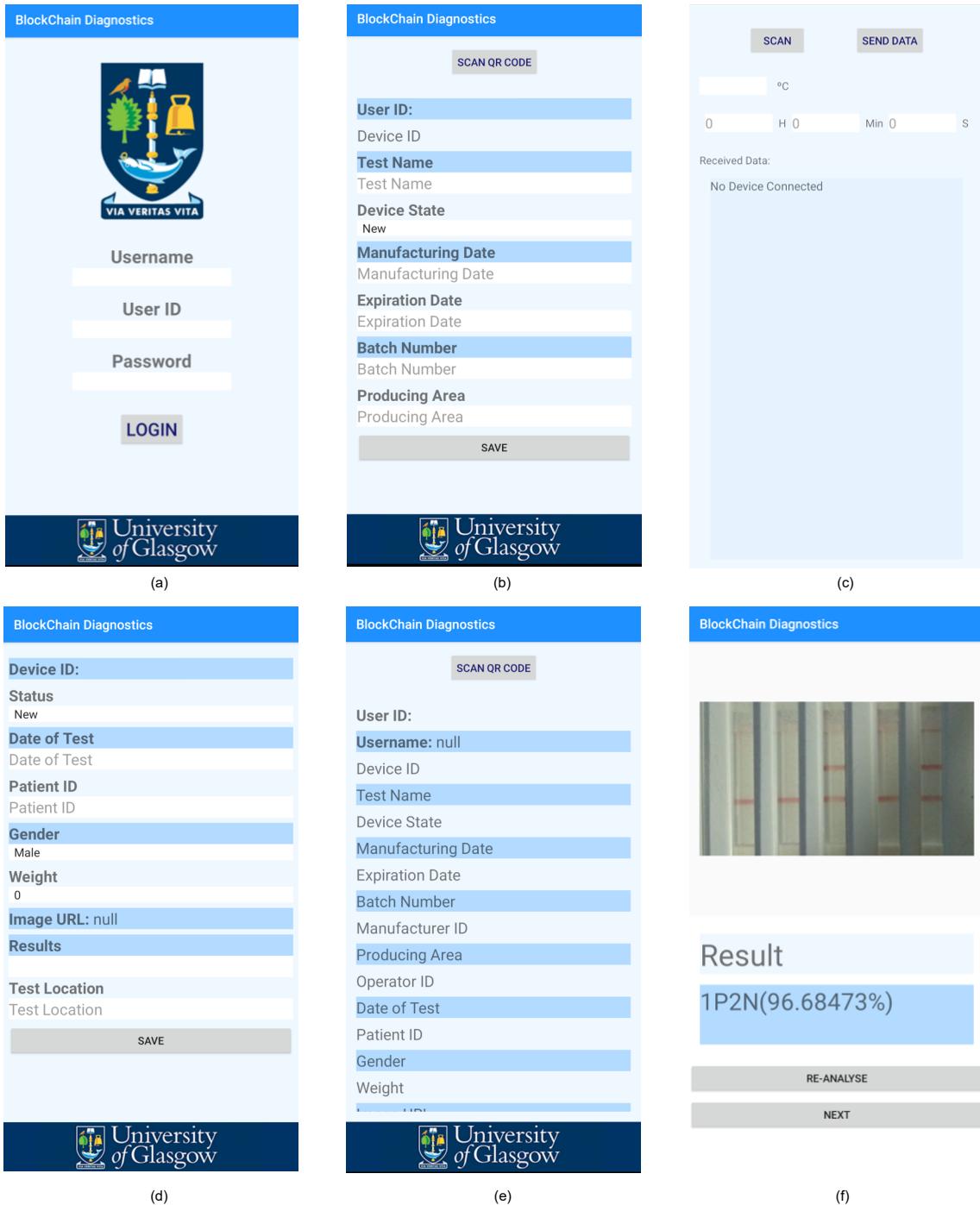


Fig. S1. Screenshots of app. A story-board showing the screens within the mobile phone app, comprising: (a) The login page of the app; (b)The screen for manufacturers to add new device information to the system; (c) The heater control page for operator; (d) The screen for operators to upload the test information; (e) The screen for analysts to read device information; (f) The screen for result analysis. **Movie M1** demonstrates the full run-through.

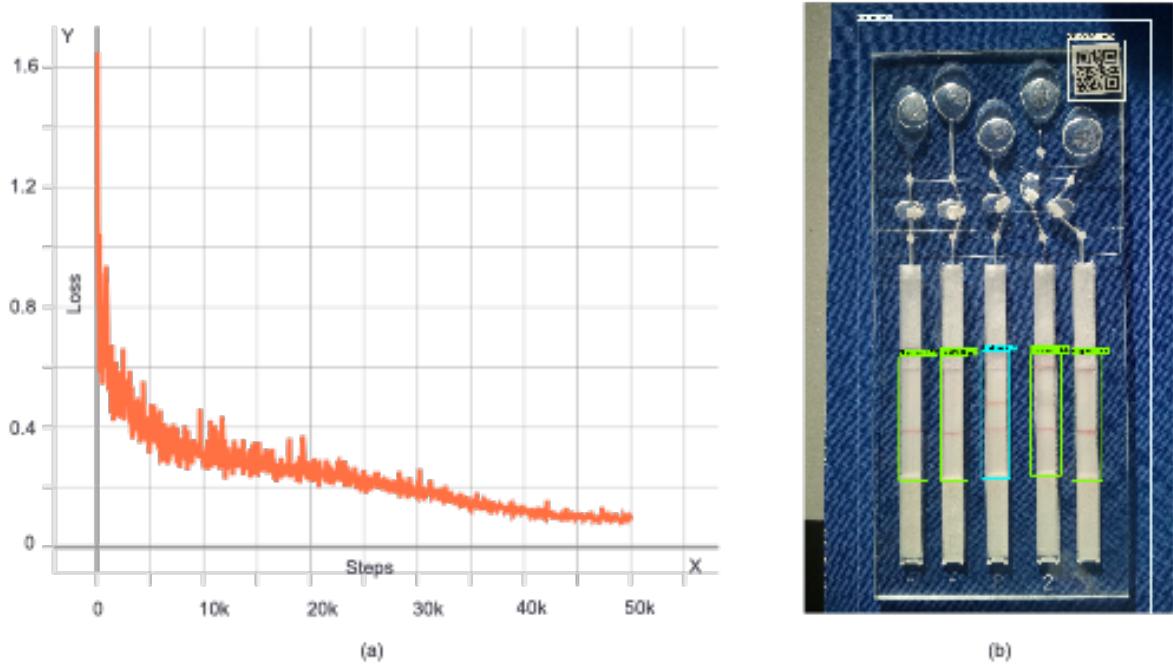


Fig. S2. Performance of the object detection model. (a) The loss curve of the training process of object detection model. The model was based on SSD resnet50, with the training steps (x axis) and the losses (y axis); and (b) a sample image with predictions.

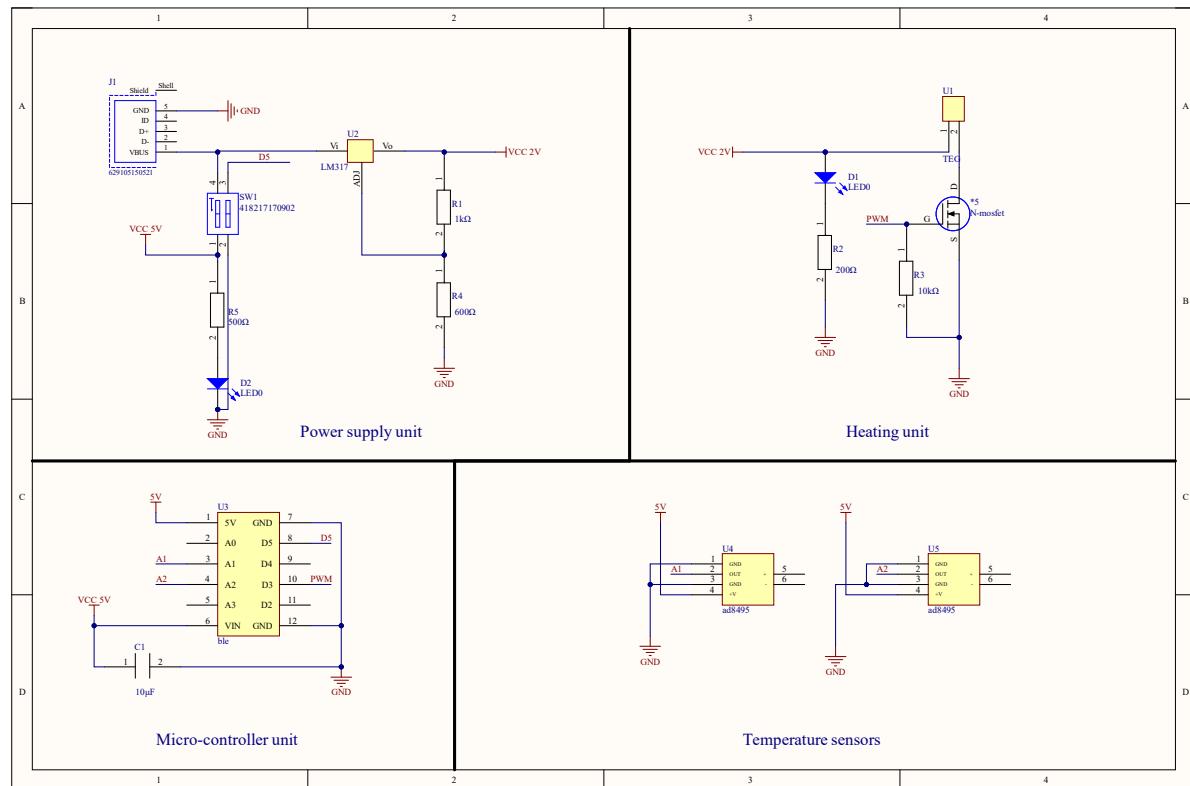


Fig. S3. Circuit schematic diagram including the microcontroller unit which controls the temperature setting for heating the LAMP reaction. Details of the components used are described in the Methods.

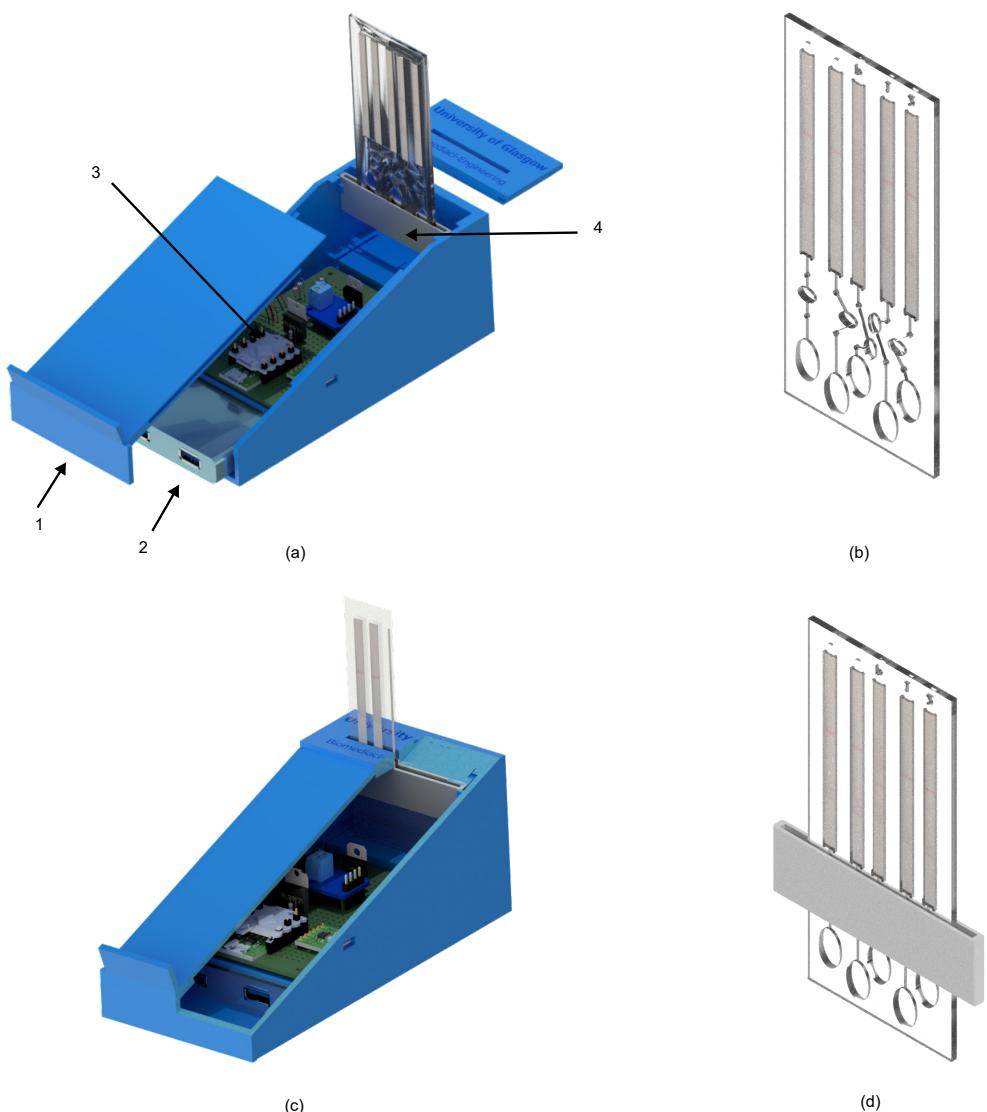


Fig. S4. Heater design. A Render showing: (a) the exploded view of the heater. The numbered parts are: 1 the cover of the heater with mobile phone holder; 2 the power bank; 3 the internal circuit; and 4 the aluminium loop board; (b) ¾ View of the heater within the casing; (c) Multiplex diagnostic LAMP device; (d) the multiplex diagnostic LAMP device housed within the aluminium loop (labelled as item (4) in (a) above).

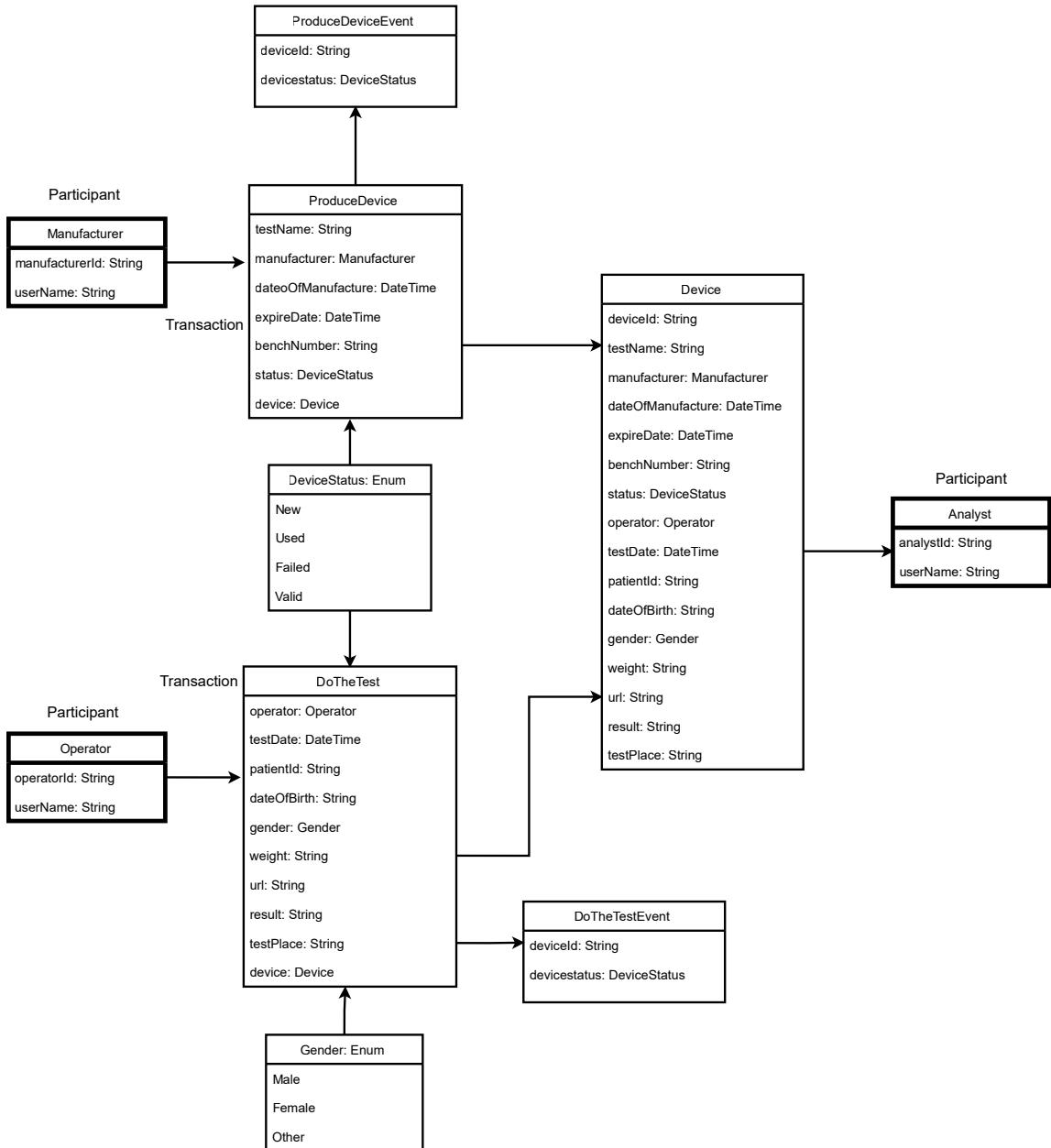


Fig. S5. Workflow of the blockchain network. The information of manufacturer, operator and analysts includes their ID and username, which was recorded with other data required by transactions to the diagnostic device properties. Each block provided the information of data required and its type in different classes.

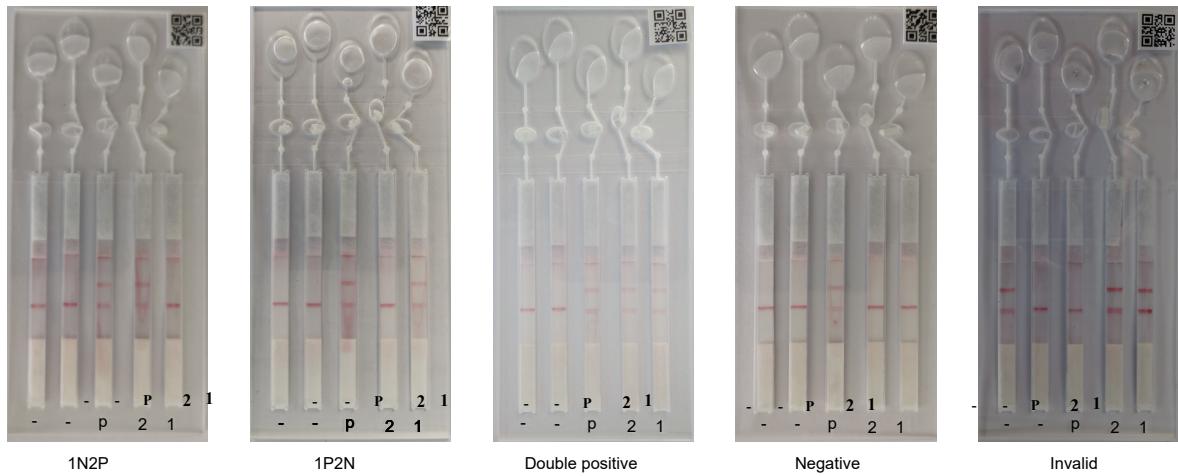


Fig. S6. Examples images from different categories. The channel marked with "1" and "2" are the test channels, the channel marked with "p" is the positive control channel and channels marked with "-" are negative control channels.

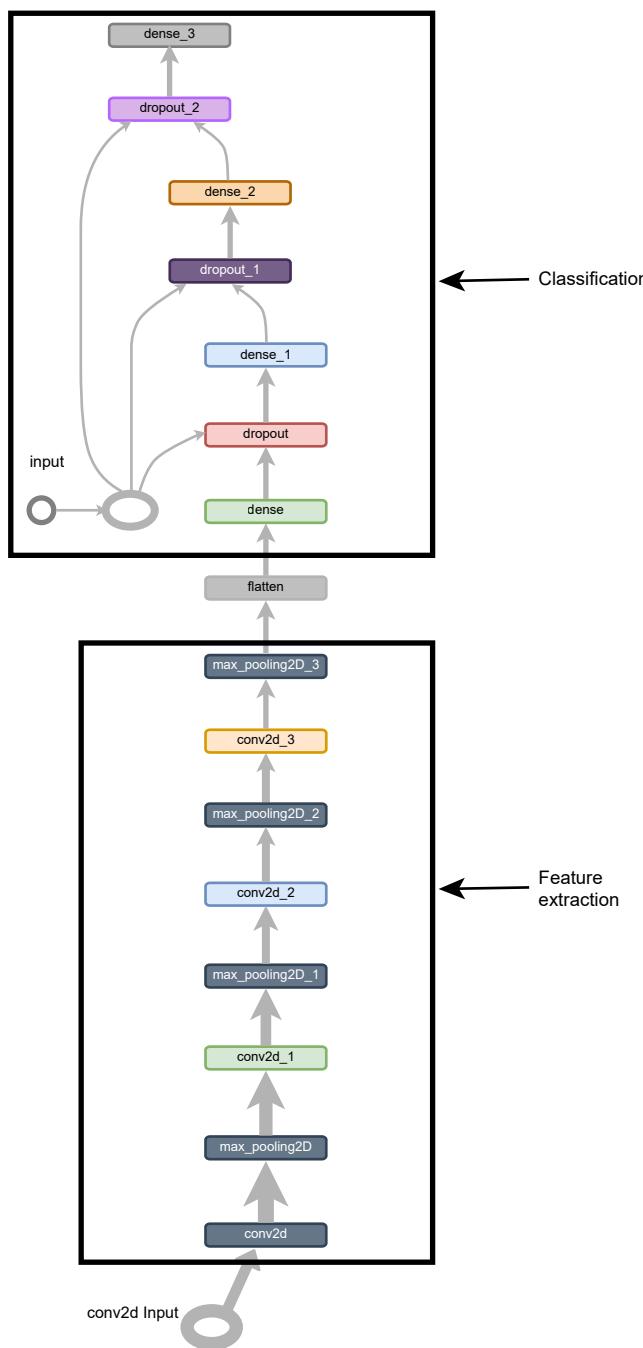


Fig. S7. Schematic showing the structure of the Convolution Neural Network model.

Table S1. System resource usage during the blockchain performance test

| Node Name | Memory (avg) MB | CPU% (avg) | Traffic In (MB) | Traffic Out (MB) | Disc Read (KB) | Disc Write (MB) |
|---|--------------------|---------------|--------------------|------------------|-------------------|--------------------|
| dev-peer0.org1.example.com-assay-tracking-network-0.1.0 | 118.9 | 42.8 | 9.7 | 4.5 | 0.0 | 0.0 |
| dev-peer0.org2.example.com-assay-tracking-network-0.1.0 | 174.5 | 9.5 | 1.8 | 1.3 | 0.0 | 0.0 |
| peer0.org2.example.com | 363.8 | 6.9 | 5.3 | 11.0 | 380.0 | 1.4 |
| peer0.org1.example.com | 318.4 | 26.0 | 21.2 | 26.2 | 16.0 | 1.4 |
| orderer.example.com | 53.8 | 0.3 | 0.7 | 1.4 | 0.0 | 0.9 |
| couchdb.org1.example.com | 185.0 | 58.5 | 4.6 | 13.3 | 128.0 | 2.7 |
| ca.org2.example.com | 17.9 | 0.01 | 0.003 | 0.003 | 0.0 | 0.0 |
| ca.org1.example.com | 7.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| couchdb.org2.example.com | 174.0 | 18.82 | 1.00 | 3.1 | 48.0 | 2.7 |

Table S2. Classification criteria.

| Channel Class | Negative control 1 | Negative control 2 | Positive control | Test 1 (Pf) | Test 2 (Ppan) |
|-----------------|--------------------|--------------------|------------------|-------------|---------------|
| 1N2P | - | - | + | + | - |
| 1P2N | - | - | + | - | + |
| Double positive | - | - | + | + | + |
| Negative | - | - | + | - | - |
| Invalid | +, / | +, / | -, / | / | / |

Table S3. The cost of the assay. This estimation of costs was performed based upon the laboratory purchase price; however, this can be significantly reduced if the production was mass manufactured, through economies of scale. For example, the plastic components are currently machined but could be embossed in reel-to-reel processes or injection molded, leading to significant cost savings.

| Components | Unit price (£) | Cost per test (£) |
|---|-----------------------|-------------------|
| Plastic sheet | -- | 2 |
| Lateral flow strip | 0.8 | 4 |
| PCR sealing film | 195 (100 sheets) | 1.95 |
| Paper origami device | 48.6 (100 sheets) | 0.05 |
| Mastermix | 346.5 (300 reactions) | 5.8 |
| Primers | -- | 1.25 |
| Extraction reagents (Promega) | 221 (200 preps) | 0.2 |
| Total | | 18.65 |
| Device | | |
| Power bank | | £10 |
| Electronic components (Thermoelectric generator, resistors, etc.) | | £20 |

Table S4. Operating costs, based on the following suppliers and technologies: in UK a 4G plan with 10 GB data costs £10-£20 per month. Operating costs are expected to fluctuate based on usage and cellular network charges in different countries

| | |
|-------------------------|---|
| Compute engine (google) | £ 31.27 per instance per month (\$200 free credits) |
| Network (cellular) | £ 12/12GB-month (Depends on local telecom operator) |
| Total | £ 50-60 |