ARTICLE OPEN

Check for updates

Deep learning-enabled point-of-care sensing using multiplexed paper-based sensors

Zachary S. Ballard[1,2], Hyou-Arm Joung[1,3], Artem Goncharov[1], Jesse Liang[2,3], Karina Nugroho[3], Dino Di Carlo[2,3], Omai B. Garner[4] and Aydogan Ozcan[1,2,3 ✉]

We present a deep learning-based framework to design and quantify point-of-care sensors. As a use-case, we demonstrated a low-cost and rapid paper-based vertical flow assay (VFA) for high sensitivity C-Reactive Protein (hsCRP) testing, commonly used for assessing risk of cardio-vascular disease (CVD). A machine learning-based framework was developed to (1) determine an optimal configuration of immunoreaction spots and conditions, spatially-multiplexed on a sensing membrane, and (2) to accurately infer target analyte concentration. Using a custom-designed handheld VFA reader, a clinical study with 85 human samples showed a competitive coefficient-of-variation of 11.2% and linearity of $R^2 = 0.95$ among blindly-tested VFAs in the hsCRP range (i.e., 0–10 mg/L). We also demonstrated a mitigation of the hook-effect due to the multiplexed immunoreactions on the sensing membrane. This paper-based computational VFA could expand access to CVD testing, and the presented framework can be broadly used to design cost-effective and mobile point-of-care sensors.

INTRODUCTION

Computation has great potential for improving diagnostics. By identifying complex and nonlinear patterns from noisy inputs, computational tools present an opportunity for automated and robust inference of medical data. For example, several studies have shown deep learning as a method to automatically identify tumors from an image, potentially enabling diagnostics in low-resource settings that lack a trained diagnostician[1–3]. Additionally, computational solutions have been demonstrated earlier in the diagnostics pipeline to virtually stain pathology slides and enhance image resolution through the use of convolutional neural networks[4–6]. Though much of this recent success is within the field of imaging, diagnostics that rely on biosensing can similarly leverage computational tools to improve sensing results and design future systems.

Point-of-care (POC) testing can especially benefit from computational sensing approaches. Due to their low-cost materials, compact designs, and requirement for rapid and user-friendly operation, POC tests are often less accurate when compared to traditional laboratory tests and assays[7–12]. For example, paper-based immunoassays such as rapid diagnostic tests (RDTs) offer an affordable and user-friendly class of POC tests which have been developed for malaria, HIV-1/2, and cancer screening, among other uses[13–17]. However, these RDTs lack the sensitivity and specificity needed for certain diagnostic applications largely due to issues of reagent stability, fabrication, and operational variability, as well as matrix effects present in complex samples such as blood[15,16,18]. Additionally, a well-known competitive binding phenomenon called the hook-effect can lead to false reporting of results, specifically in instances where the sensing analyte can be present over a large dynamic range[19–24]. Therefore, computational tools alongside portable and cost-effective assay readers present a unique opportunity to compensate for some of these constraints[25–34]. By quantifying the signals generated on paper-based substrates, machine learning algorithms have the potential to significantly improve the performance of POC sensors, without a significant hardware cost or increased complexity to the assay protocol.

As a demonstration of this emerging opportunity at the intersection of computational sensing and machine learning, we report a computational paper-based vertical flow assay (VFA) for cost-effective high-sensitivity C-reactive protein (hsCRP) testing, also referred to as cardiac CRP testing (cCRP) (see Fig. 1)[35]. Here, we implement a deep learning-based computational sensing framework to jointly develop the CRP quantification algorithm with the multiplexed sensing membrane of the VFA (Fig. 1b), selecting the most robust subset of sensing channels via feature selection methods in order to accurately infer the CRP concentration. Recent work by our group investigated the use of neural networks for POC Lyme disease diagnostics using a VFA format, achieving competitive results compared to the gold-standard clinical testing.[32] However, in contrast to this previous report, here we uniquely demonstrate (1) precise *quantification* of a protein biomarker as opposed to a binary (positive/negative) decision, (2) the incorporation of the test fabrication information into the learning model to improve quantitative sensing performance, and (3) a significantly extended sensing dynamic range through computational analysis of multiplexed immunoreaction spots, all targeting the same analyte in uniquely different ways. Taken together, this unique deep-learning based POC sensing method enables our low-cost and rapid (<12 min) VFA to successfully address the unmet clinical need of CRP quantification in the high-sensitivity range (i.e., 0–10 mg/L), as well as to identify samples outside of this range despite the presence of the hook-effect.

CRP is a general biomarker of inflammation, however slightly elevated CRP levels in blood can be an indicator of atherosclerosis, and have been shown to be a predictor for heart attacks, stroke, and sudden cardiac death for patients with and without a history

[1]Department of Electrical and Computer Engineering, University of California, Los Angeles, CA, USA. [2]California NanoSystems Institute, University of California, Los Angeles, CA, USA. [3]Department of Bioengineering, University of California, Los Angeles, CA, USA. [4]Department of Pathology and Medicine, University of California, Los Angeles, CA, USA. ✉email: ozcan@ucla.edu
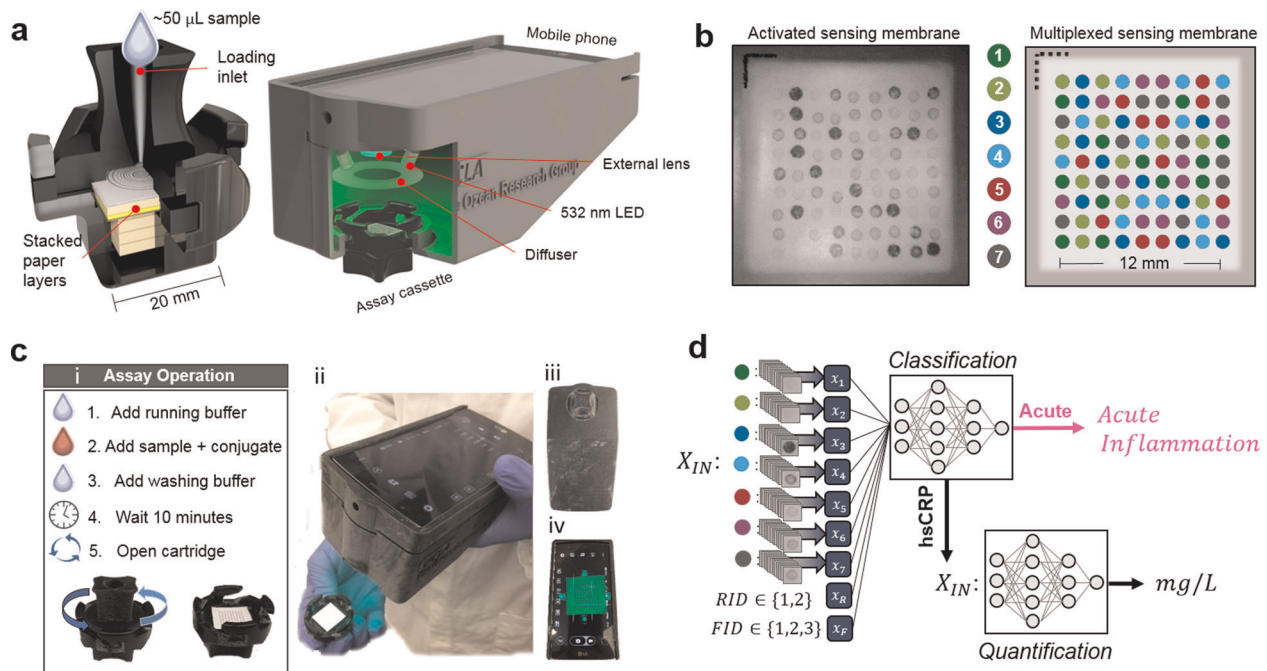
Fig. 1 Overview of the multiplexed vertical flow assay. a The VFA cassette cross-section and mobile-phone reader with the inserted VFA cassette to be tested. **b** The multiplexed sensing membrane contained within the VFA cassette. The algorithmically determined immunoreaction spot layout (right) contains seven unique spotting conditions, each of which uniquely reacts with the sensed analyte and the signal-forming Au NPs. A raw image of an activated sensing membrane taken with the mobile-phone reader is shown to the left. **c** (i) The VFA assay operation protocol (ii) The VFA cassette and mobile phone reader after the assay completion. The VFA cassette inserted into the mobile phone reader from the (iii) bottom and (iv) top view. **d** Block diagram of the computational analysis, showing the input features X_{IN} which contain, the average signals from like-spotting conditions along with the reagent batch ID (RID) and the fabrication batch ID (FID).

of CVD[36–39]. Therefore, the hsCRP test is a quantitative test commonly ordered by cardiologists to stratify certain patients into low, intermediate, and high risk groups for CVD based on clinically defined cut-offs: below 1 mg/L is considered low risk, between 1 and 3 mg/L is intermediate risk, and above 3 mg/L is high-risk[40]. As a result, the hsCRP test requires a high degree of accuracy and precision, especially around the clinical cut offs, putting it out-of-reach of traditional paper-based systems[41]. Additionally, in the presence of infection, tissue injury, or other acute inflammatory events, CRP levels can rise nearly three orders of magnitude, making hsCRP testing with immuno- and nephelometric- assays vulnerable to the hook-effect[37,39,42]. As a result, samples with greatly elevated CRP levels can be falsely reported as within the hsCRP range (i.e., <10 mg/mL), and therefore wrongly interpreted for CVD risk stratification.

To address these existing challenges of POC hsCRP testing, we implemented the aforementioned computational sensing methods with our paper-based multiplexed sensor and performed a clinical study with 85 patient serum samples and >250 VFA tests created over multiple fabrication batches, and compared the sensor performance to an FDA-approved assay and a nephelometric reader (Dimension Vista System, Siemens). Our blind testing results yielded an average coefficient of variation (CV) of 11.2% and a coefficient of determination (R^2) of 0.95 over an analytical measurement range of 0 mg/L to 10 mg/L. It is important to note that although there is no FDA-approved POC hsCRP sensor, various systems have been demonstrated in the literature[43–47]. However, the tests which report accurate quantification in the high-sensitivity range employ fluorescent-based chemical assays and benchtop readers to overcome the performance limits of their traditional colorimetric counterparts. In contrast, this work uniquely demonstrates a new data-driven sensor design and read-out framework, powered by deep learning, for improving POC testing. We applied this machine

learning-enabled sensing framework to a colorimetric paper-based multiplexed test for quantification of hsCRP as a use-case, and demonstrated its competitive quantitative performance using a mobile reader, without the need for more advanced and sensitive molecular assays and their corresponding benchtop read-out systems.

We believe that the presented POC hsCRP sensor platform could provide a rapid and cost-effective means to obtain valuable diagnostic and prognostic information for CVD, expanding access to actionable health information, especially for at-risk populations that often go underserved. Broadly, our results also highlight computational sensing as an emerging opportunity for iterative assay and sensor development. Given a training data set, machine learning-based feature selection algorithms can be implemented to determine the most robust sensing channels for a given multiplexed system such as protein micro-array, well-plate assay, or multi-channel fluidic device, among others. This can therefore lead to optimized and cost-effective implementations of multiplexed bio-sensing systems for future POC diagnostic applications.

RESULTS

Optimization of VFA spots and conditions using machine learning
The multiplexed sensing membrane of the VFA contains up to 81 spatially-isolated immunoreaction spots that are each defined by a 'spotting condition' which refers to the capture protein and the associated buffer dispensed onto the nitrocellulose (NC) sensing membrane prior to the assembly and activation. Machine learning-based optimization and feature selection of the immunoreaction spots was therefore performed in two distinct steps to reveal the optimal VFA configuration: spatial spot selection and condition selection, illustrated in Fig. 2a, b, respectively. For the spot selection process, a cost function, $j_{m,p}$, was defined per sensing spot to represent the normalized distance from the
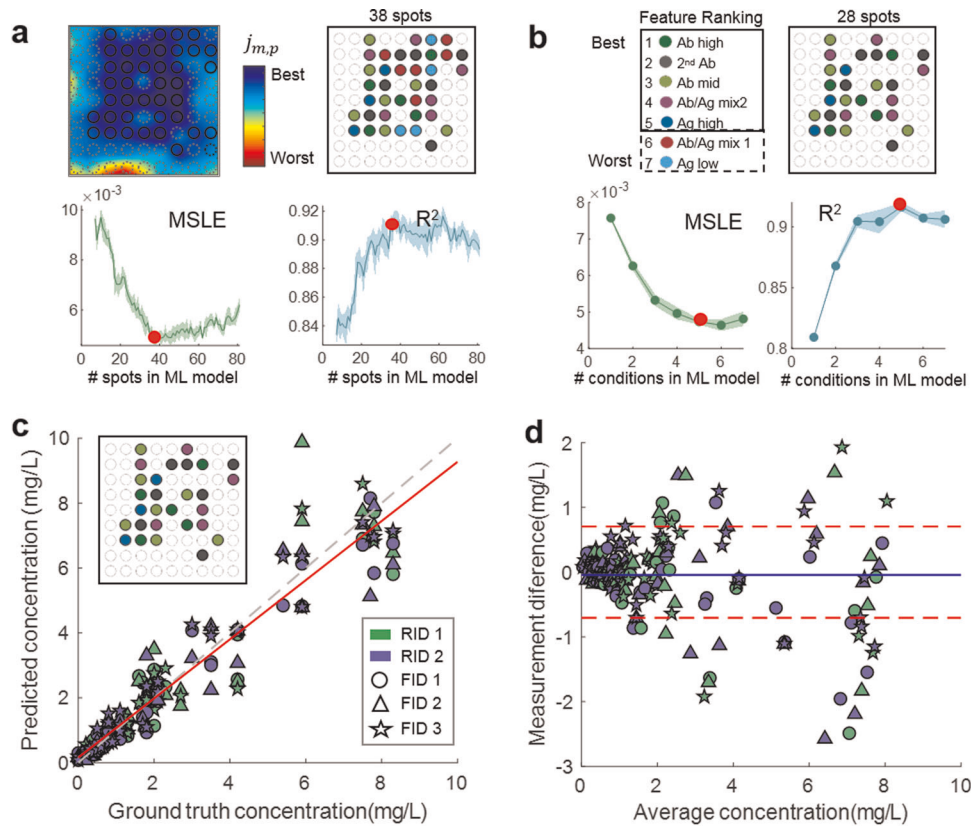
Fig. 2 Cross-validation and feature selection analysis using the training data set of clinical samples ($N_{train} = 209$). a The spot selection process. A heat-map (top left) is generated by plotting the cost function $j_{m,p}$ across the sensing membrane. The cross-validation performance, both MSLE and the coefficient of determination (R^2), is then plotted against the number of spots selected based on $j_{m,p}$ (bottom). The optimal subset of spots (top right) is then selected based off the optimal quantification performance indicated by the solid red marker. **b** The condition selection process. Conditions are ranked based on an iterative elimination method (top left), and the cross-validation performance is plotted against the number of conditions input into the quantification network. The optimal subset of conditions (top right) is then selected based off the optimal quantification performance indicated by the solid red marker. **c** The cross validation results using the selected features, where the ground truth CRP concentration is plotted against the predicted CRP concentration. The marker color and shape represent the different reagent batch ID (RID) and the fabrication batch ID (FID), respectively. **d** Bland-Altman plot of the same cross-validation results, where the dashed red lines represent the ± standard deviation of the measurement difference from the tested VFAs.

mean of like-spots (i.e., spots that share the same condition) averaged over the samples in the training set,

$$j_{m,p} = \sum_{n=1}^{N_{Train}} \frac{\left| s'_{m,n,p} - \overline{s'}_{m,n} \right|}{\overline{s'}_{m,n}} \qquad (1)$$

where $s'_{m,n,p}$ is the normalized absorbance signal of a given immunoreaction spot (defined in Methods, Equation 2) with the added index n indicating the nth sample in the training set. $\overline{s'}_{m,n}$ is the spot signal averaged over each condition within a single test, i.e. $\overline{s'}_{m,n} = \frac{1}{P_m} \sum_{p=1}^{P_m} s'_{m,n,p}$.

The heat map in Fig. 2a, which is interpolated from a 9 × 9 matrix of the cost function defined at each spot of the VFA, visualizes the statistically robust active areas of the VFA sensing membrane. To select a subset of spots from the 9 × 9 grid configuration, we then performed a k-fold ($k = 5$) cross-validation. The cross validation was performed over 75 iterations where the input to the neural network, X_{IN}, was defined by incrementally smaller subsets of the original 81 spots for each iteration. The spot with the maximum cost $j_{m,p}$ was eliminated at each iteration, resulting in the last iteration containing a subset of 7 spots, each corresponding to a different condition. The mean-squared logarithmic error (MSLE) from the cross validation was then plotted for every iteration to visualize the trade-off between the number of spots and the error of the network inference (Fig. 2a). Due to the random training process of the neural network, there is

noise associated with this curve, however, a clear performance benefit can be seen after the elimination of the first 30–40 spots corresponding to the highest $j_{m,p}$. It is also clear that further reducing the number of spots results in substantial increase in quantification error. Therefore, the approximate minimum of the MSLE curve was used to define a subset of 38 spots for subsequent analysis.

After this initial spot selection (Fig. 2a), this subset of 38 spots was further subject to a condition selection step to further optimize the performance of our computational VFA for hsCRP. This second phase of the feature selection aims to select the most robust sensing channels as defined by the unique chemistry attributed to the different spotting conditions. To this end, we performed a second iterative k-fold ($k = 5$) cross-validation analysis, eliminating one spotting condition each iteration and tracking the cross-validation error as a result of each elimination. This process was repeated for incrementally smaller subsets of conditions defined by the minimum MSLE result from the previous iteration. Resulting from this analysis, Fig. 2b reports the MSLE and coefficient of determination as function of the number of spotting conditions, suggesting that eliminating the Ab/Ag Mixture 1 (Mix 1) and the low concentration Antigen condition (Ag-low) can lead to slightly better or equivalent performance when compared to the inclusion of all the original spotting conditions.

Taken together, this machine learning-based optimization of the VFA leads to the statistical selection of the best combination of

spots and conditions (Fig. 2c inset) that can computationally determine the analyte concentration. The cross-validation results, compared to the gold standard hsCRP measurements, are also reported in Fig. 2c, d. Here the inputs to the neural network, X_{IN}, are defined by the optimal spot configuration as determined by the spot and condition selection (see Fig. 2c inset), and also include two additional integer features which correspond to the reagent batch ID ($RID \in \{0, 1\}$) and the fabrication batch ID, ($FID \in \{1, 2, 3\}$).

After this feature selection and cross-validation analysis reported in Fig. 2, the final CRP quantification algorithm was trained using the entire training set ($N_{train} = 209$) and the optimal spot configuration (Fig. 2c inset). In addition to the CRP quantification algorithm, a second classification algorithm was trained to identify the CRP samples representing an acute inflammation event, with a CRP concentration of > 10 mg/L ($\hat{N}_{train} = 6, \hat{N}_{test} = 6$) (Fig. 1d). Next, we report our blind testing results using this optimized hsCRP VFA platform and trained algorithms.

Validation of computational VFA performance for CRP measurements

Our computational VFA results from the blind testing set ($N_{test} = 57$) correlated well to the quantification results of the gold-standard hsCRP Flex cartridge run on the Dimension Vista System (see Fig. 3). These samples were analyzed using only the pixel information contained within the computationally determined subset of 28 spots and 5 conditions (Fig. 3a). The x_m signals (see Methods, Eq. 3) along with the fabrication batch ID and reagent batch ID of each test sample were first classified by an initial neural network to determine if the test was in the hsCRP range (<10 mg/L) or the acute inflammation range (>10 mg/L); we achieved 100% classification accuracy, and correctly classified 6 samples as acute and the rest (51 samples) as in the hsCRP range. The samples classified in the hsCRP range were then routed to a quantification neural network, whereas the acute samples were simply reported as acute along with a confidence score, as summarized in Fig. 3c.

The quantification accuracy of the hsCRP samples using our computational VFA was characterized by a direct comparison to the gold-standard values (Fig. 3b, c). With 51 tests quantified in the hsCRP range, the R^2 value was found to be 0.95, with a slope and intercept of the linear best-fit line being 0.98 and 0.074, respectively. The overall average CV of the blind testing data was found to be 11.2% with the average CV for the low risk, intermediate-risk, and high-risk stratified samples quantified as 11.5%, 10.1%, and 12.2%, respectively. As a reference point, the FDA review criteria for hsCRP testing state an acceptance criterion of ≤20% overall CV, with a specific CV of ≤ 10% for samples in the low risk category (i.e., <1 mg/L)[41].
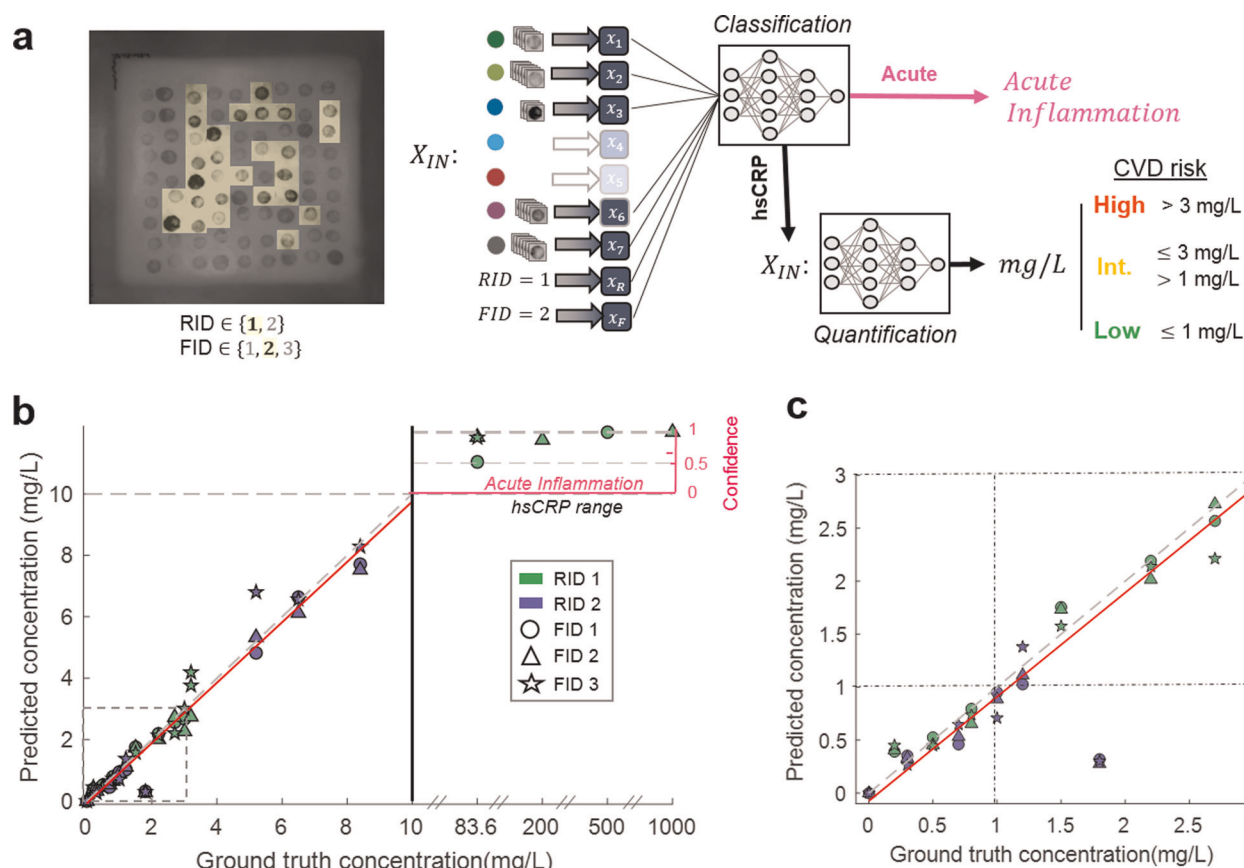
Fig. 3 Blind testing results of clinical samples ($N_{test} = 57$). a The features selected from the cross-validation analysis are extracted from a blind testing image and input into the neural network-based processing which infers the final CRP concentration. The clinical cutoffs for stratifying patients in terms of cardiovascular disease (CVD) risk are shown on the right. **b** The ground truth CRP concentration plotted against the VFA predicted CRP concentration (left y-axis) from blindly tested clinical samples. The dotted line represents a perfect match ($y = x$) and the red line represents the linear best fit. The confidence score is plotted (right y-axis) for the samples classified as acute. The marker color and shape represent the different reagent batch ID (RID) and fabrication batch ID (FID), respectively. **c** The blind testing results for the low and intermediate CVD risk regimes, where the dotted lines represent the clinical cutoffs at 1 and 3 mg/L.

DISCUSSION

Our VFA-based hsCRP test benefits from machine learning in several ways. Firstly, using neural networks to select optimal spots and infer analyte concentration from the highly multiplexed sensing channels greatly improves our quantification accuracy when compared to e.g., a standard multi-variable regression, which yields an average %CV of 47% with an R^2 value of 0.79 (see Supplementary Fig. 1). Deep learning algorithms such as the fully-connected network architecture used in this work, contain a much larger number of learned/trained coefficients along with multiple layers of linear operations and non-linear activation functions when compared to standard linear regression models. These added degrees of freedom enable neural networks to converge to robust models which can learn non-obvious patterns from a confounding set of variables, making them a powerful computational tool for assay interpretation and calibration. However, one concern with deep learning approaches is the possibility of overfitting to the given training set, especially in the instance of limited data. To mitigate this issue, we incorporated regularization terms in the hyper-parameter search (both L2 regularization and dropout), and found via cross-validation that the lowest error model employed the maximum degree of dropout regularization (i.e., 50%)[50,51]. However, we observed better quantification results in the blindly tested samples when compared to the cross validation analysis, suggesting that our model appropriately generalized over the operational range of the hsCRP sensor.

Secondly, by incorporating fabrication information using RID and FID input features, the neural network was able to learn from batch-specific patterns and signals. This resulted in a 12.9% reduction in the blindly tested MSLE when compared to the performance of a network trained without these fabrication batch input features. Similarly, incorporating the fabrication information reduced the overall %CV from 16.6% to 11.2% and increased R^2 value from 0.92 to 0.95. It is important to note that these VFA tests ($N = 273$) were fabricated without the use of industry-grade production equipment such as humidity and temperature controlled chambers, and in addition, several fabrication steps involved manual assembly. Taken together, these simple input features can benefit the performance and quality assurance of future computational POC tests following the methodology of this work. For example, the fabrication information could be included for each test in the form of a Quick Response (QR) code or could alternatively be logged into a GUI by the user before the measurement data are sent to the quantification network (running on a local or remote computer).

Another benefit of our computational VFA platform is the mitigation of false sensor response due to the hook effect. The VFA format importantly enables rapid computational analysis of highly multiplexed immunoreaction spots with minimal cross talk or interference among spots, which is inevitable for the case of standard lateral flow assays or RDTs. The multiplexed information reported by the different spotting conditions therefore allows for unique combinatorial signals to be generated over a large dynamic range (see Fig. 3b). The hook effect is clearly seen in our raw sensor data, exhibited by the capture antibody (Ab) condition (see Supplementary Figs. 2 and 3), illustrating how this condition alone can lead to false reporting of high analyte concentrations, i.e. in the case of acute inflammation. Therefore, without the incorporation of the monotonically responsive CRP antigen (Ag) spotting condition as one of the multiplexed channels in the VFA, high-concentration CRP samples can be falsely reported as low concentration due to the hook effect. This conclusion would still be true even if we trained another neural network that used a limited number of conditions as input; for example, by re-training the classification network using only the Ab and Secondary Ab spotting conditions as inputs, we found that the 83.6, 200, and 1000 mg/L samples are falsely reported as

having CRP concentrations of 7.81, 7.34, and 3.84 mg/L, respectively. In the case of analyzing only the Ab channel, all of the high-concentration CRP samples would have been falsely reported as having concentrations below 10 mg/L. These results highlight the importance of multiplexed sensing in our computational VFA platform to mitigate the limitations induced by the hook effect in order to algorithmically enhance the dynamic range of our sensor.

A comparison of the computational xVFA to other commercially available hsCRP tests is shown in Supplementary Table 1, highlighting comparable performance along with some major advantages of our platform such as its portability, low-cost, low sample-volume, and significantly extended dynamic range. It is important to note that though there are commercially available tests, no FDA approved POC test exists for hsCRP.

Computational sensing broadly refers to the joint design and optimization of sensing hardware and software, and as implemented in this study, provides a framework for data-driven assay development where the diagnostic or quantification algorithm informs the multiplexed sensor design and vice versa. As detailed in the Methods section, the computational sensing approach begins with the selection of a neural network architecture and associated cost function. This first step is paramount to the computational sensor design, as it defines the model and error metric with which the subsequent feature selection is performed. The determination of the cost function, therefore, poses an interesting question for future computational sensors and diagnostic tests: because the selection of the cost function defines the training of a neural network, what are the most clinically appropriate error functions with which one should design a computational sensing system? For example, in the case of cardiovascular risk stratification with the hsCRP test, an error of ±0.1 mg/L is more problematic for samples that are in the range of the clinically defined cutoffs (i.e., 1 and 3 mg/L) when compared to samples with relatively higher CRP concentrations, such as 8 mg/L. Therefore, a traditional cost function for regression such as the mean-squared-error may not be as appropriate as the mean-squared-logarithmic-error or mean-absolute-percentage error, which take into account the relative ground-truth concentration for each error calculation. Therefore, special consideration must be given to the cost functions employed, and custom cost functions defined jointly by physicians/clinicians and engineers should be considered.

Feature selection and machine learning based optimization can similarly be used to inform the sensing membrane design. POC sensors can especially benefit from feature selection to circumvent noise borne out of their low-cost materials (such as paper used in our VFA) and operational variations. For example, the heat-map in Fig. 2a very well reveals how the immunoreaction spots closest to the edges of the sensing membrane contain the most variation in their normalized signals. This most likely results from the position-dependent vertical flow variations inherent in the inexpensive VFA format, which uses paper materials totaling < $0.2 per hsCRP test (Table S1). These areas can, therefore, be avoided in future iterations of the sensor development, saving reagent costs and fabrication time, while also preserving robust sensing channels. Furthermore, identifying these areas of statistical variation can also inform the fabrication process. For example, Fig. 2a also shows that the top edge of the VFA sensing membrane is statistically more robust than the bottom and sides of the sensing membrane. Therefore, this spot selection analysis indicates a unidirectional fabrication bias in the lateral alignment of the sensing membrane within the VFA stack, which can be addressed in future iterations of the batch fabrication process.

Complementing the spot selection, the statistical condition selection process investigates the efficacy of the sensing channels and the unique immunoreactions defined by their spotting condition. Inherent complexities of the underlying chemistry such as the stochastic arrangement of the capture proteins within the

porous NC membrane, as well as the effects of steric hindrance, pH, humidity, and temperature can obscure intuition behind the selection of spotting conditions for a given sensing application. Therefore, computational sensing systems can benefit from data-driven selection of sensing channels. For example, Fig. 2b shows that the quantification performance improves slightly upon the out-right elimination of the Mix 1 and Ag-low conditions. This suggests that their signal response is redundant or less stable when compared to the other conditions, and is confirmed by the poor repeatability of the Ag signal between the reagent and fabrication batches (see Supplementary Fig. 3). Such a feature selection procedure in a highly multiplexed format like the VFA could, therefore, be used to computationally screen spotting conditions from a large number of differing capture chemistries including, but not limited to, different structures of capture antibodies/antigens (i.e., polyclonal vs. monoclonal) as well as varying buffer conditions and reagent concentrations. Conditions which do not empirically benefit sensor performance can be replaced by new conditions in another iteration of the development phase, or be replaced by additional redundancies of effective conditions in order to benefit from signal averaging.

Additionally, this statistical feature selection and optimization process can inform cost-performance trade-offs to help design the most robust and cost-effective implementations of POC assays. For example, the reagent cost for the immunoreaction spots contained in the hsCRP VFA test is reduced by 62%, from $2.61 to $0.97 per test, by implementing only the computationally selected chemistries. Additionally, certain spotting conditions might have an optimal capture protein concentration due to steric hindrance effects or higher degrees of non-specific binding. Therefore, in a computational sensor, reagent costs can be significantly reduced without sacrificing assay performance by employing these statistically optimized capture-protein concentrations. One should also note here that these reagent costs per test would be significantly reduced under large scale manufacturing, benefiting from economies of scale, which is expected to bring the total cost per test (including all the materials and reagents) to <$0.5. Importantly, given the vertical flow format and the absence of cross-talk among immunoreaction spots, future sensor designs with only the optimal sub-set of spots should yield the same sensing results as demonstrated here since the vacancy or presence of the unneeded spots have no effect on the remaining spatially-isolated parallel immunoreactions.

Taken together, we showed a data-driven sensor design and read-out framework, enabled by deep learning, for improving POC tests. As a use-case scenario, we demonstrated hsCRP testing with a colorimetric paper-based multiplexed VFA and clinical samples covering a large dynamic range. The multiplexed sensing membrane contained in the VFA was jointly developed with a quantification algorithm based on a fully-connected neural network architecture. First, a training data-set was formed by measuring human serum samples with the VFA. Then, through cross-validation of the training set, the most robust subset of sensing channels was selected from the multiplexed sensing membrane and used to train a CRP quantification network. The network was then blindly tested with additional clinical samples and compared to the gold standard CRP measurements, showing very good agreement in terms of quantification accuracy and precision. Additionally, the multiplexed channels and computational analysis helped us overcome limitations to the operational range of the CRP test borne out of the hook-effect. Our results demonstrate how a computational sensing framework and multiplexed sensor design can be used to engineer robust and cost-effective POC tests that have the potential to democratize diagnostics and expand access to care.

METHODS

Multiplexed VFA overview

The multiplexed VFA platform is comprised of functional paper layers stacked within a 3D-printed plastic cassette. These layers contain different paper materials and wax printed structures which have been optimized to support uniform vertical flow of serum across a two-dimensional nitrocellulose (NC) sensing membrane (Fig. 1a, Supplementary Table 2). Similar to conventional paper-based immunoassays, the VFA works by immobilizing a target analyte onto a paper substrate through binding to a complimentary capture antigen or antibody previously adsorbed within the porous structure[32,48]. Gold nanoparticles conjugated with a secondary antibody are then introduced and bound to the immobilized analyte in a sandwich structure, resulting in a color signal on the sensing membrane. The operation of our VFA test involves three sequential injection steps: (1) the running buffer, (2) the sample serum and nanoparticle conjugate, and (3) the washing buffer (Fig. 1c). After a 10-minute wait-period, the assay is complete and the VFA cassette is opened by twisting apart the top and bottom case, revealing the multiplexed sensing membrane on the top layer of the bottom case (Fig. 1c). This bottom case is then inserted into a custom-designed mobile-phone reader. An image of the activated multiplexed sensing membrane is subsequently captured and analyzed via a fully-automated image processing and deep learning-based CRP quantification algorithm (Fig. 1d). This article was previously published in pre-print[49].

Multiplexed sensing membrane fabrication and VFA assembly

The multiplexed sensing membrane contains up to 81 spatially-isolated immunoreaction spots that are each defined by a 'spotting condition' which refers to the capture protein and the associated buffer dispensed onto the NC sensing membrane prior to assembly and activation. Therefore, to design the multiplexed sensing membrane for computational analysis, a custom spot-assignment algorithm was developed to generate a 'spot map' within the active area of the sensor. Based on a given grid spacing and number of spotting conditions, the assignment algorithm distributes spotting conditions such that no single spotting condition is disproportionately positioned near the center or the edge of the sensing membrane. Because the vertical flow rate can vary radially across the sensing membrane, leading to variations of each immunoreaction across the sensor area, this step mitigates a potential bias on any given spotting condition. With seven spotting conditions (see Supplementary Table 3) in a 9×9 grid format (1.3 mm periodicity), the spot-assignment algorithm produced the map shown in Fig. 1c, which was implemented as the initial design for this study.

An automated liquid dispenser (MANTIS, Formulatrix®) was used to deposit 0.1 μL of the different protein conditions directly onto an NC membrane in the algorithmically determined pattern shown in Fig. 1c. During the spotting process, up to 24 NC sensing membranes were produced on a single connected sheet, constituting one fabrication batch, and up to three batches were produced on a given day. In order to evaluate batch-to-batch variations, we intentionally produced sensing membranes over multiple fabrication batches as well as with two reagent batches (i.e., sets of reagents which had unique storage times and/or lot numbers). Each sensing membrane was therefore tagged with a corresponding fabrication batch ID (FID, e.g., 1, 2 or 3,) and reagent batch ID (RID, e.g., 1 or 2).

Following the automated spotting procedure, the NC sheets were incubated at room temperature for 4 h after which they were submerged in 1% BSA blocking solution and allowed to incubate at room temperature for 30 min. The NC sheets were then dried in an oven at 37 °C for 10 min, after which they were cut into individual sensing membranes (1.2 × 1.2 cm) using a razor. The remaining paper materials contained in the VFA were produced following the methods outlined in a previous publication[48]. All the paper materials, including the NC sensing membrane were then assembled within the top and bottom cases of a 3-D printed VFA cassette, with foam tape holding together the paper stack (see Supplementary Table 2).

hsCRP assay procedures

Each hsCRP measurement with our VFA test is performed as follows: first 5 μL of serum sample is diluted 10 times in a running buffer (3% tween 20, 1.6% BSA in PBS) resulting in a 50 μL sample solution. Then 200 μL of running buffer is injected into the VFA inlet and allowed to absorb. After absorption into the VFA paper-stack (~ 30 s), 50 μL of sample solution is mixed with

50 µL of the gold-nanoparticle (Au NP) conjugate solution (see Supplementary Information for synthesis), and the mixture is pipetted into the inlet and allowed to absorb. Lastly, after absorption of the sample solution, 400 µL of the running buffer is added to wash away the nonspecifically bound proteins and Au NPs. After a 10-minute reaction time, the VFA cassette is then opened, and inserted into the bottom of the mobile-phone reader (Fig. 1a). This mobile reader images the multiplexed sensing membrane using the standard Android camera app (ISO: 50, shutter at 1/125, autofocused), and saves a raw image of the VFA sensing membrane (.dng file) for subsequent processing and quantification of the CRP concentration.

Data processing

Custom image processing software was developed to automatically detect and segment the immunoreaction spots in each mobile-phone image of the activated VFA cassette (see Supplementary Fig. 4). After segmentation, the pixel average of each spot is calculated and subtracted by the pixel average of a locally defined background containing BSA blocked NC membrane. Each background-subtracted spot signal is then normalized to the sum of all the spots on the sensing membrane. The final spot signal $s'_{m,p}$ is, therefore, described by,

$$s'_{m,p} = \frac{s_{m,p} - b_{m,p}}{\sum_p \sum_m (s_{m,p} - b_{m,p})} \tag{2}$$

where m represents the spotting condition, and the p represents the pth redundancy on the VFA per condition. $s_{m,p}$ is the pixel average of a given segmented spot, and $b_{m,p}$ is the local background signal. The final VFA signal per condition can then be calculated as:

$$x_m = \frac{1}{P_m} \sum_{p=1}^{P_m} s'_{m,p} \tag{3}$$

where P_m is the number of redundancies for a given spotting condition. The normalization step in Equation 2 helps us to account for sensor-to-sensor variations borne out of pipetting errors, fabrication tolerances, as well as operational variances.

Clinical testing

We procured remnant human serum samples in compliance with ethical standards of the UCLA Institutional Review Board (i.e., the ethical approval was granted by UCLA IRB #19-000172) for hsCRP testing using our VFA platform. Patient consent was waived in UCLA IRB approval since these are pre-existing remnant specimen that were collected independent of this research project. Each clinical sample was previously measured within the standard clinical workflow as part of the UCLA Health System using the CardioPhase hsCRP Flex® reagent cartridge (Cat. No. K7046, Siemens) and Dimension Vista System (Siemens). In total, we measured 85 clinical samples in triplicate with our VFA sensors. All but one sample was within the standard hsCRP range of 0–10 mg/L, with the outlier having a concentration of 83.6 mg/L. In addition to testing these clinical samples, nine CRP-free serum samples (Fitzgerald Industries International, 90R-100) were measured as well as nine artificial samples created by spiking 200, 500, and 1000 mg/L CRP into CRP-free serum samples. These artificial samples were tested to simulate serum samples from patients undergoing acute inflammatory events. Though relatively rare in the context of hsCRP testing, such high concentration samples can be falsely reported as having a low CRP concentration due to the hook-effect. Therefore, these samples were included to test if our multiplexed computational VFA could avoid such false reporting. Among different batches of 273 fabricated VFA sensors, we removed one VFA test from the data-set due to a fabrication error (misalignment, see Supplementary Fig. 5a), and removed two triplicates due to abnormally high levels of non-specific binding, which was immediately obvious in the low signals on the sensing membrane and unusual pink color observed on the top case (Supplementary Fig. 5b).

Computational VFA sensor analysis

After the clinical study was completed the image data from the activated VFA tests were partitioned into a training set ($N_{train} = 209$) and testing set ($N_{test} = 57$). This data partition was structured to ensure that the testing samples would be distributed linearly over the hsCRP range, and that samples were pulled proportionally from the different fabrication batches within each cardiovascular risk stratification group. The raw background-subtracted pixel average values are shown in Supplementary Fig. 3, where the marker color and shape indicate the FID and the RID, respectively.

The training set was then analyzed via a k-fold cross-validation ($k = 5$) to determine the optimal learning algorithm for quantification of CRP concentration from the inputs X_{IN}. We evaluated different fully connected networks through a random hyper-parameter search, where the number of nodes, layers, regularization, dropout, batch-size, and cost-function were each randomly selected from a user-constrained list. A tiered neural network architecture (Supplementary Fig. 6) with a cost function of mean-squared logarithmic error (MSLE) yielded the best performance over the random iterations of the cross-validation. Within each tier of the network architecture is a fully-connected network with two hidden layers (512 and 64 nodes, respectfully) each with a ReLu (Rectified Linear Unit) activation function and 0.5 dropout. As an alternative, a single neural network with multiple hidden layers, in contrast to the tiered structure, could also be used in providing an accurate and generalizable model.

Reporting summary

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

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

CODE AVAILABILITY

The codes that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Litjens, G. et al. A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017).
2. Bejnordi, B. E. et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* **318**, 2199–2210 (2017).
3. Hu, L. et al. An observational study of deep learning and automated evaluation of cervical images for cancer screening. *J. Natl Cancer Inst.* https://doi.org/10.1093/jnci/djy225 (2019).
4. Rivenson, Y. et al. Deep learning microscopy. *Opt., Opt.* **4**, 1437–1443 (2017).
5. Rivenson, Y. et al. PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning. *Light.: Sci. Appl.* **8**, 23 (2019).
6. Rivenson, Y. et al. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat. Biomed. Eng.* https://doi.org/10.1038/s41551-019-0362-y (2019).
7. Shaw, J. L. V. Practical challenges related to point of care testing. *Practical Lab. Med.* **4**, 22–29 (2016).
8. Anastassova Dineva, M., Mahilum-Tapay, L. & Lee, H. Sample preparation: a challenge in the development of point-of-care nucleic acid -based assays for resource-limited settings. *Analyst* **132**, 1193–1199 (2007).
9. Wang, S. et al. Advances in addressing technical challenges of point-of-care diagnostics in resource-limited settings. *Expert Rev. Mol. Diagnostics* **16**, 449–459 (2016).
10. Schito, M. et al. Opportunities and challenges for cost-efficient implementation of new point-of-care diagnostics for HIV and tuberculosis. *J. Infect. Dis.* **205**, S169–S180 (2012).
11. Yager, P., Domingo, G. J. & Gerdes, J. Point-of-care diagnostics for global health. *Annu. Rev. Biomed. Eng.* **10**, 107–144 (2008).
12. Kozel, T. R. & Burnham-Marusich, A. R. Point-of-care testing for infectious diseases: past, present, and future. *J. Clin. Microbiol.* **55**, 2313–2320 (2017).
13. López-Marzo, A. M. & Merkoçi, A. Paper-based sensors and assays: a success of the engineering design and the convergence of knowledge areas. *Lab Chip* **16**, 3150–3176 (2016).
14. Martinez, A. W., Phillips, S. T., Whitesides, G. M. & Carrilho, E. Diagnostics for the developing world: microfluidic paper-based analytical devices. *Anal. Chem.* **82**, 3–10 (2010).
15. Mahato, K., Srivastava, A. & Chandra, P. Paper based diagnostics for personalized health care: Emerging technologies and commercial aspects. *Biosens. Bioelectron.* **96**, 246–259 (2017).

16. Smith, S., Korvink, J. G., Mager, D. & Land, K. The potential of paper-based diagnostics to meet the ASSURED criteria. *RSC Adv.* **8**, 34012–34034 (2018).

17. Paper Diagnostics Market Worth $10.50 Billion by 2025 | CAGR: 8.0%. https://www.grandviewresearch.com/press-release/global-paper-diagnostics-market.

18. Primiceri, E. et al. Key Enabling Technologies for Point-of-Care Diagnostics. *Sensors (Basel)* **18** (2018).

19. Hoofnagle, A. N. & Wener, M. H. The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry. *J. Immunol. Methods* **347**, 3–11 (2009).

20. Amarasiri Fernando, S. & Wilson, G. S. Studies of the 'hook' effect in the one-step sandwich immunoassay. *J. Immunol. Methods* **151**, 47–66 (1992).

21. Jassam, N., Jones, C. M., Briscoe, T. & Homer, J. H. The hook effect: a need for constant vigilance. *Ann. Clin. Biochem.* **43**, 314–317 (2006).

22. Rey, E., O'Dell, D., Mehta, S. & Erickson, D. Mitigating the hook effect in lateral flow sandwich immunoassays using real-time reaction kinetics. *Anal. Chem.* **89**, 5095–5100 (2017).

23. Oh, J. et al. A hook effect-free immunochromatographic assay (HEF-ICA) for measuring the C-reactive protein concentration in one drop of human serum. *Theranostics* **8**, 3189–3197 (2018).

24. Kyoung, Oh,Y. et al. A three-line lateral flow assay strip for the measurement of C-reactive protein covering a broad physiological concentration range in human sera. *Biosens. Bioelectron.* **61**, 285–289 (2014).

25. Berg, B. et al. Cellphone-based hand-held microplate reader for point-of-care testing of enzyme-linked immunosorbent assays. *ACS Nano* **9**, 7857–7866 (2015).

26. McRae, M. P., Simmons, G., Wong, J. & McDevitt, J. T. Programmable bio-nanochip platform: a point-of-care biosensor system with the capacity to learn. *Acc. Chem. Res.* **49**, 1359–1368 (2016).

27. Xu, X. et al. Advances in smartphone-based point-of-care diagnostics. *Proc. IEEE* **103**, 236–247 (2015).

28. Zhu, H. et al. Optical imaging techniques for point-of-care diagnostics. *Lab Chip* **13**, 51–67 (2013).

29. Ballard, Z. S. et al. Computational sensing using low-cost and mobile plasmonic readers designed by machine learning. *ACS Nano* **11**, 2266–2274 (2017).

30. Ozcan, A. Mobile phones democratize and cultivate next-generation imaging, diagnostics and measurement tools. *Lab Chip* **14**, 3187–3194 (2014).

31. Mudanyali, O. et al. Integrated rapid-diagnostic-test reader platform on a cellphone. *Lab Chip* **12**, 2678–2686 (2012).

32. Joung, H.-A. et al. Point-of-care serodiagnostic test for early-stage lyme disease using a multiplexed paper-based immunoassay and machine learning. *ACS Nano* https://doi.org/10.1021/acsnano.9b08151 (2019).

33. Qin, Q. et al. Algorithms for immunochromatographic assay: review and impact on future application. *Analyst* **144**, 5659–5676 (2019).

34. Yan, W. et al. Machine learning approach to enhance the performance of MNP-labeled lateral flow immunoassay. *Nano-Micro Lett.* **11**, 7 (2019).

35. Ridker, P. M. A test in context: high-sensitivity c-reactive protein. *J. Am. Coll. Cardiol.* **67**, 712–723 (2016).

36. Lloyd-Jones, D. M. et al. Framingham risk score and prediction of lifetime risk for coronary heart disease. *Am. J. Cardiol.* **94**, 20–24 (2004).

37. Adukauskienė, D. et al. Clinical relevance of high sensitivity C-reactive protein in cardiology. *Medicina* **52**, 1–10 (2016).

38. Koenig, W. et al. C-Reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men: results from the MONICA (Monitoring Trends and Determinants in Cardiovascular Disease) Augsburg Cohort Study, 1984 to 1992. *Circulation* **99**, 237–242 (1999).

39. Shrivastava, A. K., Singh, H. V., Raizada, A. & Singh, S. K. C-reactive protein, inflammation and coronary heart disease. *Egypt. Heart J.* **67**, 89–97 (2015).

40. 2013 ACC/AHA Guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults. *Circulation* https://www.ahajournals.org/doi/abs/10.1161/01.cir.0000437738.63853.7a (2014).

41. Health, C. for D. and R. Guidance Documents (Medical Devices and Radiation-Emitting Products)—Review Criteria for Assessment of C Reactive Protein (CRP), High Sensitivity C-Reactive Protein (hsCRP) and Cardiac C-Reactive Protein (cCRP) Assays—Guidance for Industry and FDA Staff. https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm077167.htm.

42. Blake, G. J. & Ridker, P. M. Inflammatory bio-markers and cardiovascular risk prediction. *J. Intern. Med* **252**, 283–294 (2002).

43. Dong, M. et al. Rapid and low-cost CRP measurement by integrating a paper-based microfluidic immunoassay with smartphone (CRP-Chip). *Sensors* **17**, 684 (2017).

44. Wu, R. et al. Quantitative and rapid detection of C-reactive protein using quantum dot-based lateral flow test strip. *Analytica Chim. Acta* **1008**, 1–7 (2018).

45. Cai, Y. et al. Development of a lateral flow immunoassay of C-reactive protein detection based on red fluorescent nanoparticles. *Anal. Biochem.* **556**, 129–135 (2018).

46. Oh, S. W. et al. Evaluation of fluorescence hs-CRP immunoassay for point-of-care testing. *Clin. Chim. Acta* **356**, 172–177 (2005).

47. Joung, H.-A., Oh, Y. K. & Kim, M.-G. An automatic enzyme immunoassay based on a chemiluminescent lateral flow immunosensor. *Biosens. Bioelectron.* **53**, 330–335 (2014).

48. Joung, H.-A. et al. Paper-based multiplexed vertical flow assay for point-of-care testing. *Lab. Chip* https://doi.org/10.1039/C9LC00011A (2019).

49. Ballard, Z. et al. Deep learning-enabled point-of-care sensing using multiplexed paper-based sensors. bioRxiv Preprint at https://doi.org/10.1101/667436 (2019).

50. Baldi, P. & Sadowski, P. J. *Understanding Dropout. in Advances in Neural Information Processing Systems 26* (eds. Burges, C. J. C. et al.) 2814–2822 (Curran Associates, Inc., 2013).

51. Srivastava, N. et al. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

Z.S.B. optimized and fabricated VFA sensing platform, performed the clinical testing measurements, and developed the computational sensing framework and data analysis. H.-A.J. optimized and fabricated the VFA sensing platform, performed the clinical testing measurements including the reagent handling and synthesis. A.G. wrote the automated image processing code for sensor analysis. J.L. worked to optimize and develop the protein spotting process. K.N. optimized and fabricated the VFA sensors. O.B.G., D.D.C. and A.O. oversaw and supervised the research. A.O. initiated and conceptualized the computational sensing project.

COMPETING INTERESTS

Z.B., H.J., A.G., D.D., O.G. and A.O. have a pending patent application on the contents of this work.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at https://doi.org/10.1038/s41746-020-0274-y.

Correspondence and requests for materials should be addressed to A.O.

Reprints and permission information is available at http://www.nature.com/reprints

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