

## Supporting Information

# Deep learning-enhanced paper-based vertical flow assay for high-sensitivity troponin detection using nanoparticle amplification

*Gyeo-Re Han,<sup>1</sup> Artem Goncharov,<sup>1</sup> Merve Eryilmaz,<sup>1,2</sup> Hyou-Arm Joung,<sup>1</sup> Rajesh Ghosh,<sup>2</sup> Geon Yim,<sup>3</sup> Nicole Chang,<sup>2</sup> Minsoo Kim,<sup>2</sup> Kevin Ngo,<sup>2</sup> Marcell Veszpremi,<sup>1</sup> Kun Liao,<sup>1</sup> Omai B. Garner,<sup>4</sup> Dino Di Carlo,<sup>2,5\*</sup> and Aydogan Ozcan<sup>1,2,5,6\*</sup>*

<sup>1</sup>Electrical & Computer Engineering Department, <sup>2</sup>Bioengineering Department, <sup>3</sup>Department of Chemistry and Biochemistry, <sup>4</sup>Department of Pathology and Laboratory Medicine, <sup>5</sup>California NanoSystems Institute (CNSI), <sup>6</sup>Department of Surgery, University of California, Los Angeles, CA 90095 USA.

\*Corresponding Authors: [dicarlo@ucla.edu](mailto:dicarlo@ucla.edu), [ozcan@ucla.edu](mailto:ozcan@ucla.edu)

## Contents

**Supplementary Note 1.** Signal amplification reaction.

**Supplementary Note 2.** Assay optimization.

**Supplementary Note 3.** Comparison of two outlier analysis (OA) models.

**Supplementary Note 4.** Limitations of cTnI quantification with a single neural network model.

**Supplementary Note 5.** Power-fitting function.

**Table S1.** Detailed information of the clinical sample test dataset.

**Table S2.** MAE scores for the blind testing samples processed by the cTnI quantification neural networks.

**Table. S3.** Correlation coefficients and CVs for different cTnI quantification models.

**Table. S4.** Cost evaluation.

**Figure S1.** Graphical user interface of the Raspberry Pi-based reader.

**Figure S2.** Reagent flow mechanism during the amplification reaction.

**Figure S3.** Solution-based characterization of Au-ion reduction chemistry.

**Figure S4.** Optimization of the assay.

**Figure S5.** Optimization of the reagent concentrations for Au-ion reduction-based signal amplification reaction.

**Figure S6.** Representative images of sensing membranes after conducting assays with varying concentrations of cTnI spiked in human serum.

**Figure S7.** Architecture of the optimal classification neural network ( $DNN_{\text{Classification}}$ ).

**Figure S8.** Architectures of the optimal quantification neural networks.

**Figure S9.** Performance of the optimized neural network models on the validation sets.

**Figure S10.** Predictions of cTnI concentrations in clinical samples from non-optimized quantification models.

**Figure S11.** Evaluation of the imaging-based signal quantification capacity of the Raspberry Pi-based portable reader.

**Figure S12.** Evaluation of the impact of the number of testing spots on the assay performance.

**Figure S13.** Detailed configuration of the paper layers of the hs-VFA.

**Movie S1.** Real-time response of the signal amplification reaction on hs-VFA (time-lapsed).

### Supplementary Note 1. Signal amplification reaction.

The colorimetric signal amplification reaction occurred as a result of the chemical reduction of  $\text{Au}^{3+}$  on the surface of the AuNP conjugates. The reagent solution, which is a mixture of  $\text{Au}^{3+}$  and  $\text{H}_3\text{NO}$ , is injected into the 2<sup>nd</sup> top case and initially flows downward into the absorption pads through the sensing membrane. After removing the 2<sup>nd</sup> top case after 3 min, the amplification reaction continued with the reagent flowing back up from the absorption pads to the sensing membrane by evaporation (**Figure S2**). As spectroscopic evidence of the amplification reaction, we observed a significant redshift (from 520 nm to 560 nm) of the absorption peak, broadening of the peak, and darkening of the aqueous solution, which supports signal amplification by particle growth (see **Figure S3**).

### Supplementary Note 2. Assay optimization.

We optimized assay steps to enhance the signal-to-noise ratio in the hs-VFA system. Our tests regarding cartridge compressibility, which determines the degree of paper layer compression when assembled, showed that the best signal-to-noise ratio occurred at 25% compressibility, where the signal intensity increased and the non-specific intensity decreased (**Figure S4a**). We attribute this to improved contact efficiency between paper layers, enhancing the washing of remaining conjugates in the 1<sup>st</sup> top case and free conjugates in the sensing membrane. Compressibility beyond 25% was not considered due to the relatively thick paper layers, making it difficult to assemble the top and bottom cases. Using 15 nm AuNPs led to a 50% increase in signal intensity compared to 40 nm AuNPs (**Figure S4b**). This enhancement is due to the superior membrane permeability of the 15 nm AuNPs and the higher particle quantity ( $1.40 \times 10^{12}/\text{mL}$ ) at the same optical density (OD) concentration, approximately 15.5 times more than 40 nm AuNPs ( $9.00 \times 10^{10}/\text{mL}$ ). In the optimized running buffer composition and combination, we observed an average improvement in signal intensity of approximately 44–68% (**Figure S4c**).

The sample volume we used in this study, 50  $\mu\text{L}$ , is experimentally optimized. Increasing the sample volume resulted in higher signal intensity. However, once the volume reached 50  $\mu\text{L}$ , the increase in intensity was minimal (**Figure S4d**). Additionally, above 50  $\mu\text{L}$ , the absorption time of the sample by the top case significantly increased, resulting in longer assay times and greater variability (**Figure S4e**). Therefore, we concluded that the ideal volume for optimal signal intensity and minimal assay time was 50  $\mu\text{L}$ .

We also studied the dependence on the concentration of both the capture and the conjugate. We used a fixed amount (10  $\mu\text{L}$ , 1 mg/mL, sufficient to saturate the AuNP's surface) of the detection antibodies for conjugate synthesis. We also altered the conjugate concentration (AuNP-detection antibody complex) in the optimization study. The results of the optimization experiments indicated that the optimal concentration for the capture antibody was 1 mg/mL (**Figure S4f**) and that for the conjugate was 2.5 OD (**Figure S4g**), which resulted in the highest signal-to-noise ratio. Therefore, these concentrations were used throughout the final assay testing.

For amplification, we evaluated signal intensity and inter-assay uniformity at the testing spots with various combinations of  $\text{Au}^{3+}$  and  $\text{H}_3\text{NO}$  concentrations and selected the condition that ranked highest in both criteria ( $\text{Au}^{3+}$ : 10 mM and  $\text{H}_3\text{NO}$ : 10 mM, **Figure S5**).

### Supplementary Note 3. Comparison of two outlier analysis (OA) models.

While integrating the OA algorithm into our process, we compared the statistics-driven approach (statistical OA model) with a differential OA method (OA-D), which utilized differential changes in raw

intensities between the time-lapse images (see the “Computational analysis of hs-VFA signals” subsection in the Methods for more details about OA and OA-D). This comparison enabled us to evaluate the impact of different OA methods on assay uniformity (intra-assay) and reproducibility (inter-assay). When analyzing assay data sets ( $N=330$ ) using OA algorithms, the statistics-based analysis retained an average of  $6.5 \pm 1.5$  out of the original 10 test spots, whereas the OA-D method retained  $4.8 \pm 2.1$  spots (**Figure 3b, inset i**). In the statistical OA model, we observed higher exclusion rates for spots located towards the outer edges of the sensing membrane (i.e., No. 1, 2, 5, 10 in **Figure 3b, inset ii**). This is mainly due to minor differences in flow homogeneity between the inner and outer spots, within a CV of 5%. When comparing the intra-assay CV before and after implementing the OA, we observed a substantial improvement in spot-to-spot assay uniformity (**Figure 3b, inset iii**) and assay reproducibility (**Figure 3b, inset iv**). Specifically, with the use of the statistical OA model, the assay uniformity improved from a CV of 4.1% to a CV of 1.1%, and assay reproducibility improved from a CV of 8.2% to a CV of 5.1%, which was superior to the differential model (OA-D). Therefore, we chose the statistical OA model for subsequent analysis as it offered better generalizability (by retaining more test spots) and achieved superior data refinement compared to other exclusion methods.

#### **Supplementary Note 4.** Limitations of cTnI quantification with a single neural network model.

Quantitative predictions of cTnI concentrations from  $DNN_{\text{Quantification}}$  for 14 blind tested samples were  $< 40$  pg/mL, contradicting the predictions from  $DNN_{\text{Classification}}$  (i.e., these samples were predicted as  $\geq 40$  pg/mL by  $DNN_{\text{Classification}}$ ). Ground truth cTnI concentrations in all of these samples were in the 50–200 pg/mL range, whereas quantitative outputs from  $DNN_{\text{Quantification}}$  were equal to 0 pg/mL. One of the reasons for such false negative outputs from  $DNN_{\text{Quantification}}$  for the lower concentration samples can be the limited amount of training samples. Only 12 out of 64 training samples fell into the 50–200 pg/mL concentration range and  $DNN_{\text{Quantification}}$  might be overfitting to higher concentrations considering the large total dynamic range of the samples used to train the model (i.e., 40–40,000 pg/mL). The overfitting issue can be minimized by including more samples from all relevant clinical ranges into the training set<sup>1-3</sup>. However, even with large training sets and minimal model overfitting, a single quantification network (i.e.,  $DNN_{\text{Quantification}}$ ) might still return contradicting results with respect to  $DNN_{\text{Classification}}$ , especially for the borderline samples (i.e., samples close to the cut-off level of the equipment used to measure ground truth concentrations, 40 pg/mL in this work). Quantification errors on such samples can be attributed to the proximity between the assay signals for borderline samples and samples from the  $< 40$  pg/mL range, as well as the interference of the noise factors from the sample matrix and the low-cost nature of hs-VFA cartridges. Keeping an adjudicative neural network (i.e.,  $DNN_{\text{Low}}$ ) to process samples with contradicting predictions between  $DNN_{\text{Quantification}}$  and  $DNN_{\text{Classification}}$  would still be beneficial to achieve optimal cTnI concentration accuracy in the low concentration ranges and improve sensitivity for more reliable high-sensitivity cTnI detection.

#### **Supplementary Note 5.** Power-fitting function.

The performance of the optimized quantification neural networks was compared with a rule-based method where the time-lapse response of hs-VFA was related to cTnI concentration through a power-fitting function. The input to the function was the time-lapse signal from hs-VFA ( $I_{\text{Time-lapse}}$ ) defined by equation (1). The power-fitting function form is defined by the power law, and the explicit function form is outlined below:

$$y' = a * I_{\text{Time-lapse}}^b, \quad a = 4.359, \quad b = 3.636, \quad (\text{S1})$$

$y'$  is the predicted concentration. The performance of this power-fitting model on blindly tested clinical samples is shown on **Figure S10a**.

## References

1. Alwosheel, A.; Cranenburgh, S. van; Chorus, C. G. Is Your Dataset Big Enough? Sample Size Requirements When Using Artificial Neural Networks for Discrete Choice Analysis. *J. Choice Model.* **2018**, *28*, 167–182. DOI: 10.1016/j.jocm.2018.07.002.
2. Ajiboye, A. R.; Abdullah-Arshah, R.; Qin, H.; Isah-Kebbe, H. Evaluating the effect of dataset size on predictive model using supervised learning technique. *Int. J. Comput. Syst. Softw. Eng.* **2015**, *1*, 75–84. DOI: 10.15282/ijsecs.1.2015.6.0006.
3. Kokol, P.; Kokol, M.; Zagoranski, S. Machine Learning on Small Size Samples: A Synthetic Knowledge Synthesis. *Sci. Prog.* **2022**, *105*, 00368504211029777. DOI: 10.1177/00368504211029777.

**Table S1.** Detailed information of the clinical sample test dataset.

Sample No.	Ground Truth Level (cTnI, pg/mL)	1. Before data-processing		2. After data-processing (Virtual control indicator and outlier analysis)		3-1. Deep learning-based analysis (Classification)			3-2. Deep learning-based analysis (Quantification)		
		Iw/o Normalized, Time-lapse (a.u.) (n=2)	CV (%)	Iw/o Normalized, Time-lapse (a.u.) (n=2)	CV (%)	Category	Classification prediction (1st repeat) (2nd repeat)	Note (1st repeat) (2nd repeat)	Category	Quantification prediction (cTnI, pg/mL) (n=2)	CV (%)
1	< 40	0.0447	15.35	0.0467	5.55	Training/validation set	.	.	Training/validation set	.	.
2	< 40	0.0298	17.69	0.0291	7.92	Training set	.	.	Training set	.	.
3	< 40	0.0367	1.00	0.0336	5.95	Training set	.	.	Training set	.	.
4	< 40	0.0373	5.25	0.0360	3.39	Training set	.	.	Training set	.	.
5	< 40	0.0379	18.00	0.0310	9.44	Training set	.	.	Training set	.	.
6	< 40	0.0284	7.07	0.0264	5.42	Training/validation set	.	.	Training/validation set	.	.
7	< 40	0.0457	0.85	0.0477	1.10	Training/validation set	.	.	Training/validation set	.	.
8	< 40	0.0395	1.65	0.0415	0.83	Training/validation set	.	.	Training/validation set	.	.
9	< 40	0.0464	2.78	0.0455	2.04	Blind testing	0.05303 0.19758	.	Not included	.	.
10	< 40	0.0545	4.58	0.0525	0.49	Blind testing	0.40918 0.49816	.	Not included	.	.
11	< 40	0.0525	7.89	0.0478	1.60	Blind testing	0.15789 0.08529	.	Not included	.	.
12	< 40	0.0460	1.86	0.0408	0.78	Blind testing	0.11780 0.11351	.	Not included	.	.
13	< 40	0.0643	4.66	0.0610	1.40	Blind testing	0.52127 0.69452	False positive False positive	Blind testing	53.76	17.55
14	< 40	0.0446	7.01	0.0438	0.73	Blind testing	0.10597 0.10888	.	Not included	.	.
15	< 40	0.0457	2.77	0.0451	1.36	Training/validation set	.	.	Training/validation set	.	.
16	< 40	0.0438	2.97	0.0430	0.05	Training/validation set	.	.	Training/validation set	.	.
17	< 40	0.0430	3.23	0.0463	0.66	Training/validation set	.	.	Training/validation set	.	.
18	< 40	0.0438	9.76	0.0385	1.14	Training/validation set	.	.	Training/validation set	.	.
19	< 40	0.0419	11.17	0.0436	0.83	Training/validation set	.	.	Training/validation set	.	.
20	50	0.0427	2.19	0.0422	0.64	Blind testing	0.63668 0.61239	.	Blind testing	47.70	12.32
21	50	0.0351	15.18	0.0315	7.95	Training/validation set	.	.	Training/validation set	.	.
22	60	0.0397	4.06	0.0389	2.44	Training/validation set	.	.	Training/validation set	.	.
23	60	0.0399	2.04	0.0387	2.01	Blind testing	0.82452 0.76070	.	Blind testing	49.92	4.07
24	60	0.0368	3.11	0.0360	1.30	Training/validation set	.	.	Training/validation set	.	.
25	70	0.0559	7.99	0.0554	1.25	Blind testing	0.52977 0.76158	.	Blind testing	120.62	9.16
26	80	0.3565	3.76	Excluded by virtual control indicator		.	.	.	.	.	.
27	100	0.0463	2.84	0.0470	1.20	Blind testing	0.85223 0.89543	.	Blind testing	86.82	2.41
28	110	0.0685	7.32	0.0659	2.96	Blind testing	0.70074 0.69257	.	Blind testing	134.93	0.68
29	120	0.0572	12.42	0.0513	2.12	Blind testing	0.78632 0.72014	.	Blind testing	89.53	5.17
30	130	0.0425	5.14	0.0489	3.48	Training/validation set	.	.	Training/validation set	.	.
31	150	0.0718	5.29	0.0670	0.65	Blind testing	0.79812 0.84434	.	Blind testing	128.04	1.43
32	180	0.0670	0.45	0.0659	0.24	Training/validation set	.	.	Training/validation set	.	.
33	240	0.0513	3.86	0.0577	1.58	Blind testing	0.53131 0.17662	False negative	Blind testing	338.76	N/A*
34	340	0.0777	2.59	0.0784	0.99	Blind testing	0.73111 0.80259	.	Blind testing	414.28	0.17
35	360	0.0807	4.72	0.0830	2.39	Training/validation set	.	.	Training/validation set	.	.
36	410	0.0609	8.38	0.0600	0.15	Blind testing	0.85008 0.76469	.	Blind testing	298.73	10.59
37	440	0.2273	1.89	Excluded by virtual control indicator		.	.	.	.	.	.
38	660	0.0938	3.01	0.0887	1.91	Blind testing	0.83459 0.78099	.	Blind testing	806.83	8.15
39	1190	0.0715	0.01	0.0700	1.88	Training/validation set	.	.	Training/validation set	.	.
40	1270	0.0982	0.84	0.0956	1.41	Blind testing	0.86999 0.91340	.	Blind testing	1303.87	9.61
41	1570	0.1114	6.27	0.1293	0.69	Blind testing	0.64959 0.74180	.	Blind testing	1924.96	6.04
42	1740	0.0167	0.68	Excluded by virtual control indicator		.	.	.	.	.	.
43	1910	0.0202	5.33	Excluded by virtual control indicator		.	.	.	.	.	.
44	2340	0.1275	2.74	0.1269	2.42	Blind testing	0.90919 0.96594	.	Blind testing	2550.50	2.89
45	2500	0.1000	3.52	0.0998	0.31	Blind testing	0.86076 0.95967	.	Blind testing	2417.62	2.63
46	2660	0.0925	9.58	0.0970	1.78	Blind testing	0.88875 0.81604	.	Blind testing	1473.54	1.79
47	4290	0.1587	4.78	0.1651	6.58	Blind testing	0.71109 0.94713	.	Blind testing	4324.45	10.02
48	4920	0.1225	1.73	0.1245	0.62	Training/validation set	.	.	Training/validation set	.	.
49	6050	0.1163	0.69	0.1109	1.56	Training set	.	.	Training set	.	.
50	6560	0.1073	1.57	0.1063	1.38	Training set	.	.	Training set	.	.
51	12050	0.1212	2.17	0.1221	1.74	Training set	.	.	Training set	.	.
52	13380	0.2028	2.23	0.2007	1.69	Training set	.	.	Training set	.	.
53	14250	0.1855	2.13	0.1829	1.96	Training set	.	.	Training set	.	.
54	39500	0.2720	2.75	0.2706	4.97	Training set	.	.	Training set	.	.
		<b>Mean of CV(%)</b>	<b>4.97</b>	<b>Mean of CV(%)</b>	<b>2.25</b>					<b>Mean of CV(%)</b>	<b>6.16</b>

\*N/A: Not available. Only one repeat from the duplicate testing of those samples was identified as positive ( $\geq 40$  pg/mL).

**Table S2.** MAE scores for the blind testing samples processed by the cTnI quantification neural networks. MAE scores are calculated between duplicate serum sample measurements for each patient.

Sample No.	Ground truth (pg/mL)	MAE (%)
9	< 40	12.4
20	50	8.7
23	60	2.9
25	70	6.5
27	100	1.7
28	110	0.5
29	120	3.7
31	150	1.0
33	240	79.4
34	340	0.1
36	410	7.5
38	660	5.8
40	1270	6.8
41	1570	4.3
44	2340	2.0
45	2500	1.9
46	2660	1.3
47	4290	7.1

**Table S3.** Correlation coefficients and CVs for different cTnI quantification models.

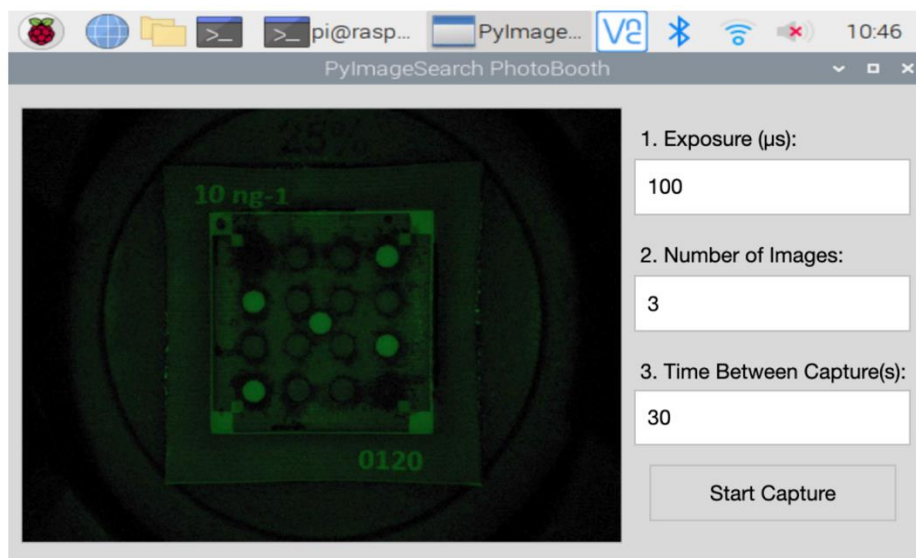
Models	Pearson's <i>r</i> (Entire range)	Pearson's <i>r</i> (< 1,000 pg/mL)	CV (%) (Entire range)	CV (%) (< 1,000 pg/mL)	Refs.
1 Optimal model (Time-lapse & After OA)	0.965	0.959	6.2	6.5	Fig. 4c
2 Power fitting model (Time-lapse & After OA)	0.861	0.776	10.5	10.5	Fig. S10a
3 Time-lapse & Before OA	0.951	0.921	8.8	8.7	Fig. S10b
4 End-point & After OA	0.956	0.806	5.3	4.6	Fig. S10c

**Table S4.** Cost evaluation. (a) hs-VFA cartridge. (b) Portable reader.

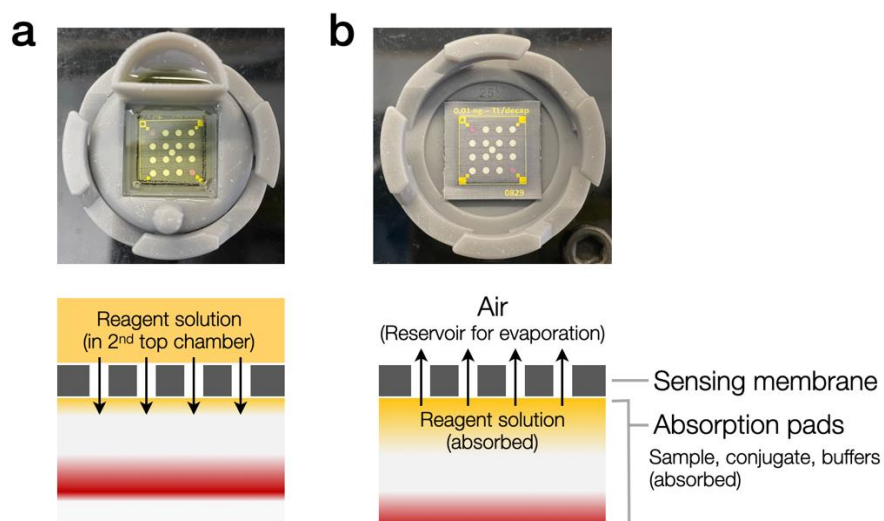
a. Cost for VFA cartridge			
No.	Contents	Category	Cost/test
1	AuNP		\$0.15
2	Antibodies	Assay reagents	\$1.95
3	Chemicals and buffers		\$0.24
4	Nitrocellulose		\$0.09
5	Asymmetric membrane		\$0.20
6	CF7 pads (Interpad)	Paper materials	\$0.02
7	Others (Absorption pad, form tape, blocking reagents, etc.)		\$0.24
8	3D printed cartridge		\$0.96
9	Others (acrylic plastic, glue, etc.)	Plastic cartridge	\$0.01
<b>Total cost/test</b>			<b>\$3.86</b>

b. Cost for portable reader			
No.	Contents	Category	Cost
1	Raspberry Pi 3	Processor and display	\$53.00
2	Touch screen display		\$50.00
3	Camera module		\$15.00
4	Macro lens	Optical components	\$27.50
5	LEDs		\$1.92
6	Housing, cassette tray, and lens/LED holder	3D printed components	\$11.00
7	Wires, resistors, screws, and power supply.	Others	\$11.96
<b>Total cost</b>			<b>\$170.38</b>

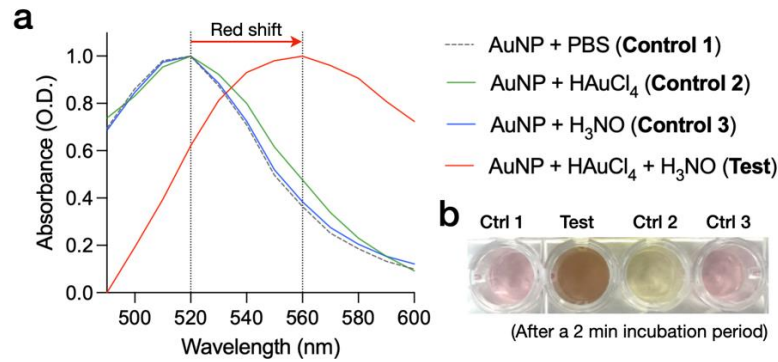




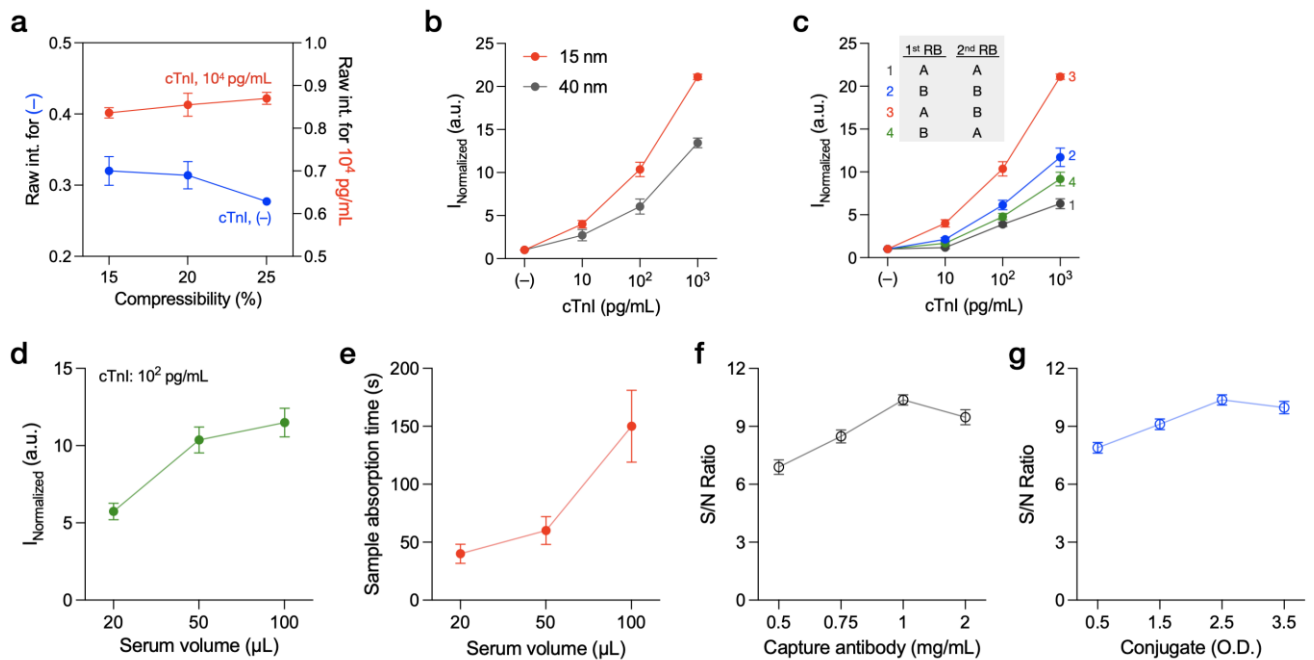
**Figure S1.** Graphical user interface of the Raspberry Pi-based reader.



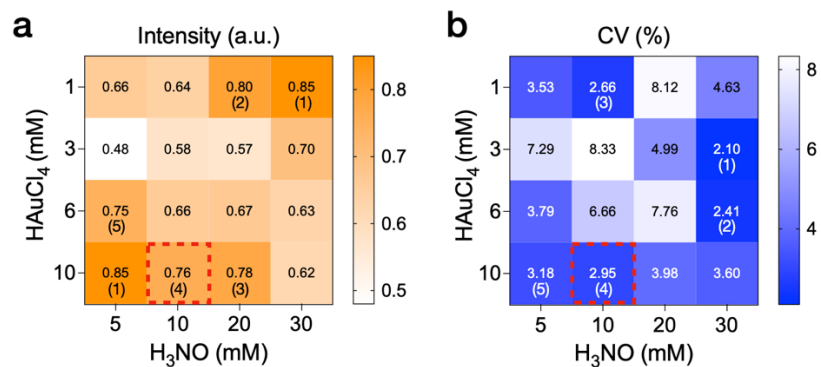
**Figure S2.** Reagent flow mechanism during the amplification reaction. (a) Reagent transport from the reagent chamber of the 2<sup>nd</sup> top case to absorption pads. (b) The reversed flow of reagent from the absorption pads to the sensing membrane after decapping the 2<sup>nd</sup> top case.



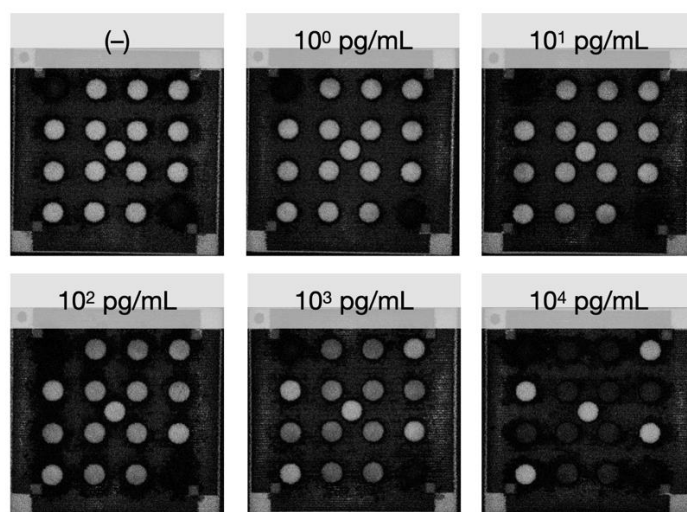
**Figure S3.** Solution-based characterization of Au-ion reduction chemistry. (a) Analysis of the absorption peak shift. (b) Comparison of color changes in the solutions.



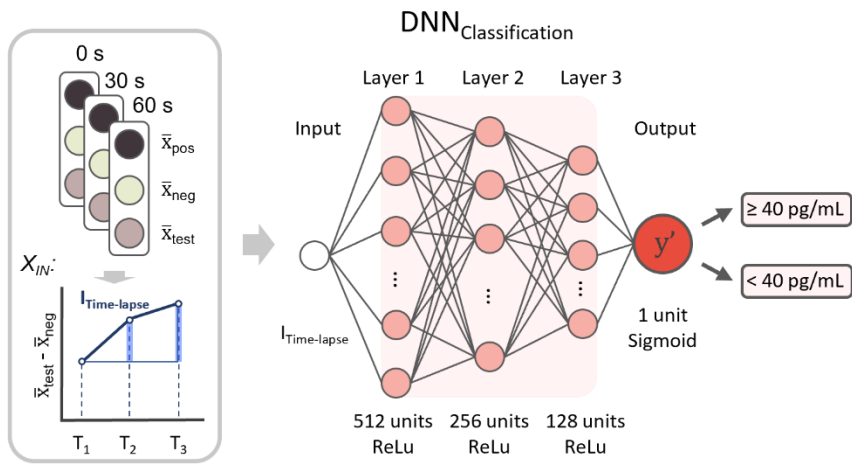
**Figure S4.** Optimization of the assay. (a) Evaluation of the effect of cassette compressibility. (b) Impact of the size of AuNP on assay performance. (c) Selection of the running buffer combination. Buffer A contains 1% v/v Triton X-100 and 1% v/v BSA in PBS. Buffer B contains 3% v/v Tween 20, 1% v/v albumin, 0.5% w/w protein saver, and 1% w/w trehalose in PBS. Effect of different sample volumes on (d) signal intensity and (e) sample absorption time. Effect of (f) capture antibody concentration and (g) conjugate concentration on assay result. For this test,  $10^2$  pg/mL and 0 pg/mL of cTnI spiked in cTnI-free serum were used. Signal-to-noise (S/N) ratio was calculated using the signal intensities obtained by assaying these two concentrations. Each data point in (a), (b), (c), (e), (f), and (g) represents the mean of triplicates  $\pm$  SD. Data points in (d) represent the mean of 50 repeats  $\pm$  SD.



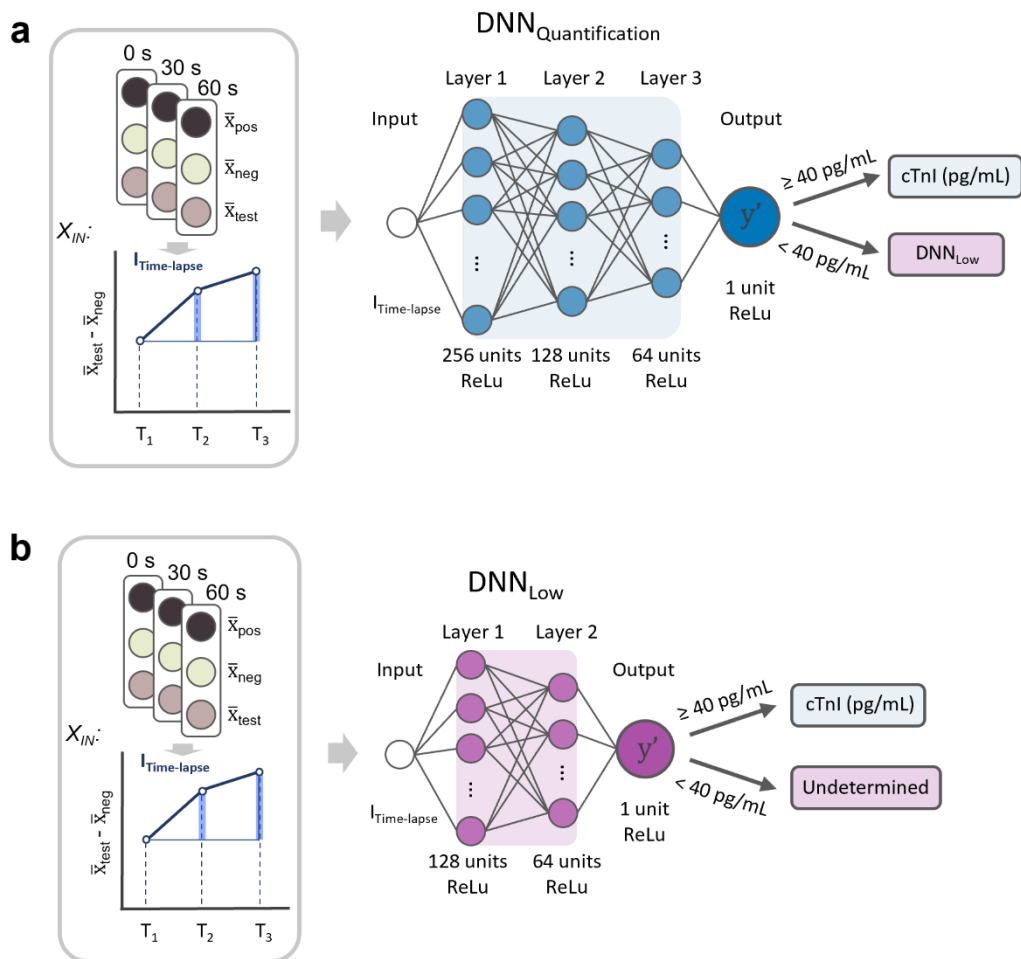
**Figure S5.** Optimization of the reagent concentrations for Au-ion reduction-based signal amplification reaction. (a) Signal intensity analysis. (b) Reaction uniformity test. A 1 ng/mL cTnI was assayed triplicate. The intensity in (a) represents the mean of the three measurements. CV(%) in (b) represents the mean of inter-assay CV(%) from the triplicates. The numbers in the bracket indicate the rank.



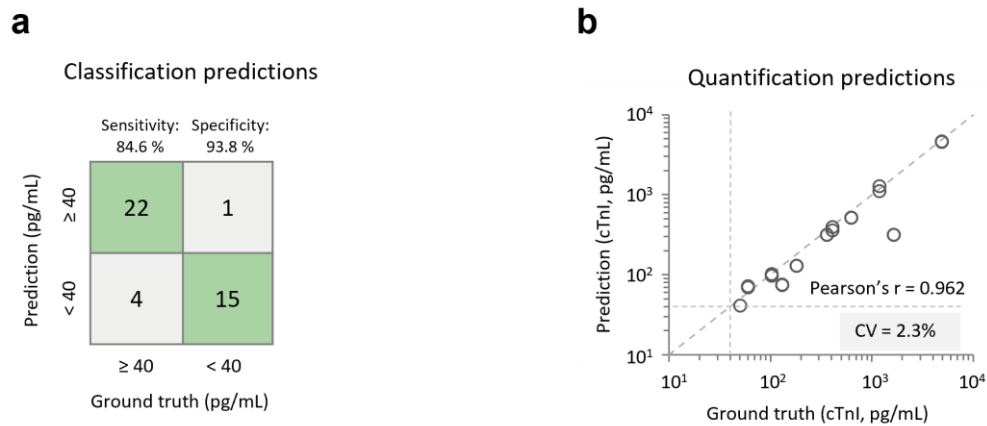
**Figure S6.** Representative images of sensing membranes after conducting assays with varying concentrations of cTnI spiked in human serum. Images were captured at  $t = 60$  s within the time-lapse sequence.



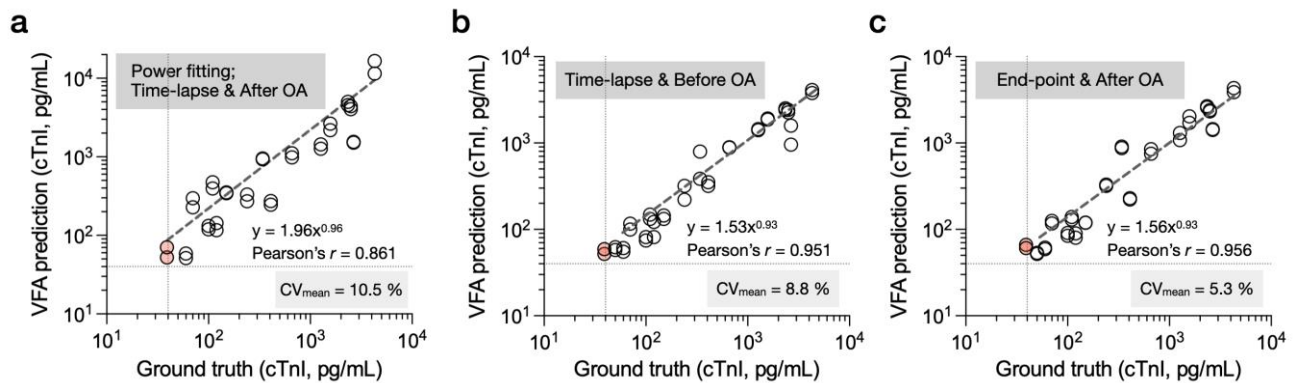
**Figure S7.** Architecture of the optimal classification neural network ( $DNN_{\text{Classification}}$ ).



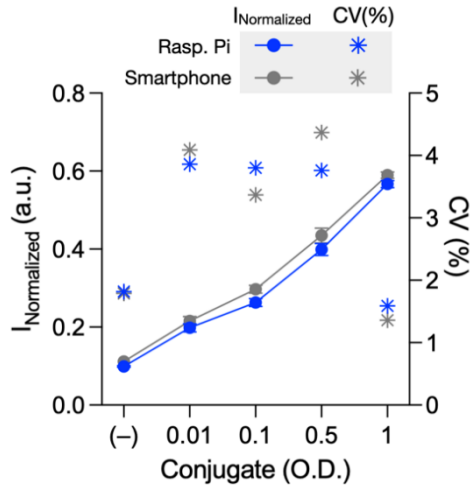
**Figure S8.** Architectures of the optimal quantification neural networks. (a)  $DNN_{\text{Quantification}}$  and (b)  $DNN_{\text{Low}}$ .



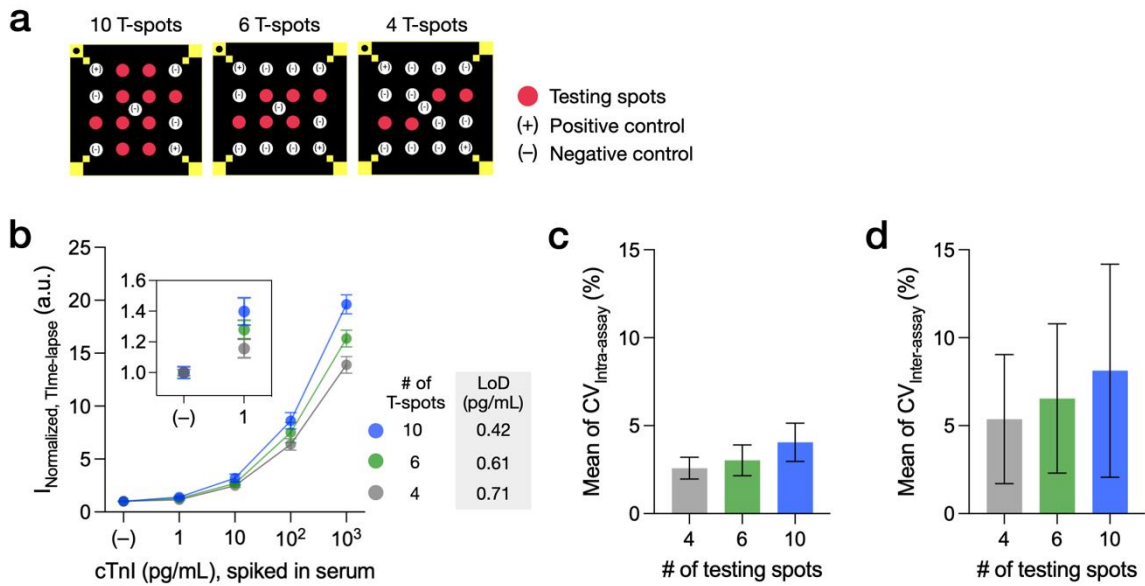
**Figure S9.** Performance of the optimized neural network models on the validation sets. (a) Classification predictions on the validation set. (b) Quantification predictions on the validation set.



**Figure S10.** Predictions of cTnI concentrations in clinical samples from non-optimized quantification models. (a) Power-fitting model with the time-lapse input data processed by OA. (b) Optimized neural network models (i.e.,  $DNN_{\text{Quantification}}$  and  $DNN_{\text{Low}}$ ) with the time-lapse input data not processed by OA. (c) Optimized neural network models with the end-point input data processed by OA. Dashed lines on x and y axis indicate 40 pg/mL.

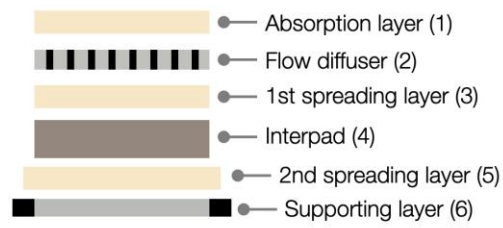
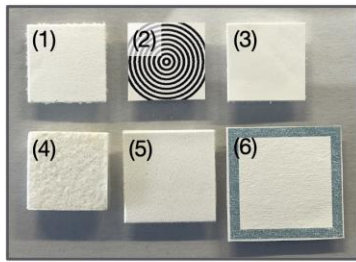


**Figure S11.** Evaluation of the imaging-based signal quantification capacity of the Raspberry Pi-based portable reader. Each data point represents the mean of triplicates  $\pm$  SD.

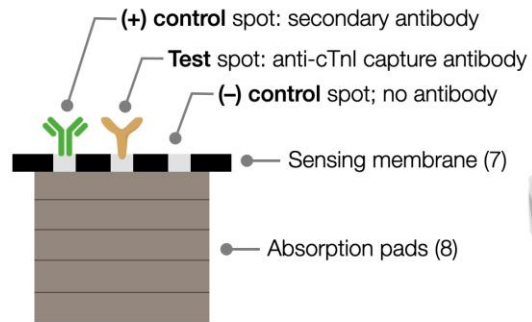
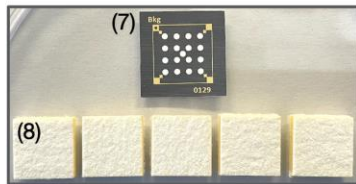


**Figure S12.** Evaluation of the impact of the number of testing spots on the assay performance. (a) Spotting map, (b) Sensitivity, (c) Intra-assay uniformity, and (d) Inter-assay repeatability. Each data point represents the mean of triplicates  $\pm$  SD without outlier exclusion.

**a** Components of the 1st top

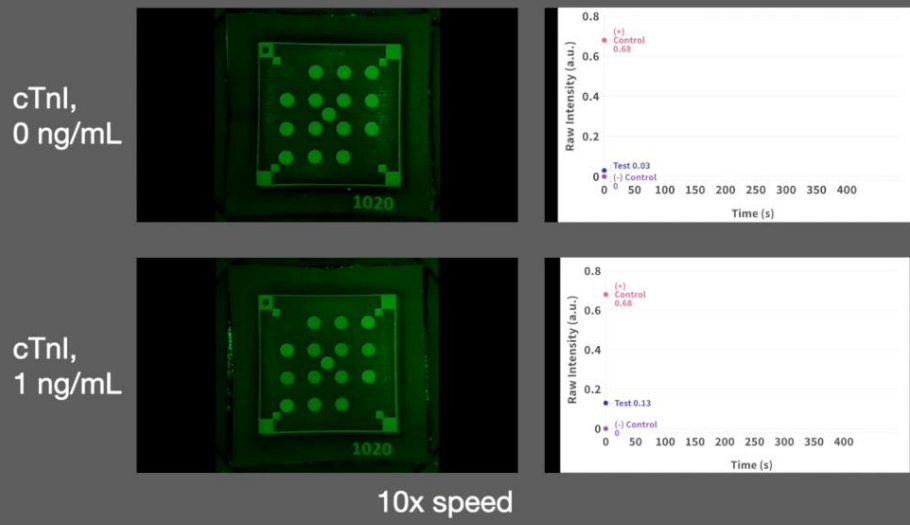


**b** Components of the bottom case



**Figure S13.** Detailed configuration of the paper layers of the hs-VFA. (a) Components of the 1<sup>st</sup> top case. (b) Components of the bottom case.

## Real-time signal amplification response (Time-lapsed)



**Movie S1.** Real-time response of the signal amplification reaction on hs-VFA (time-lapsed).