Supporting Information

Smartphone-Based Colorimetric Detection System for Portable Health Tracking

Samira Balbach,¹ Nan Jiang,²* Rosalia Moreddu,³ Xingchen Dong,^{1,4} Wolfgang Kurz,¹ Congyan Wang,¹ Jie Dong,¹ Yixia Yin,⁵ Haider Butt,⁶ Martin Brischwein,⁷ Oliver Hayden,⁷ Martin Jakobi,¹ Savas Tasoglu,⁸ Alexander W. Koch,¹ Ali K. Yetisen,^{1,4*}

¹ Institute for Measurement Systems and Sensor Technology, Technical University of

Munich, Munich 80333, Germany

- ² West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu 610041, China
- ³ Istituto Italiano di Tecnologia, 16163 Genova, Italy
- ⁴ Department of Chemical Engineering, Imperial College London, London SW7 2AZ, UK
- ⁵ State Key Laboratory of Advanced Technology for Materials Synthesis and Processing, Wuhan University of Technology, Wuhan 430070, China
- ⁶ Department of Mechanical Engineering, Khalifa University, Abu Dhabi 1277788, UAE
- ⁷ Department of Electrical and Computer Engineering, TranslaTUM Campus, Technical University of Munich, Munich 80290, Germany
- ⁸ Department of Mechanical Engineering, University of Connecticut, CT 06268, USA
 *Email: jiangnansophia@scu.edu.cn; a.yetisen@imperial.ac.uk



Figure S1. pH sensor characterization in the color space. A) Concentration-dependent RGB characterization of

the pH sensor. B) Concentration-dependent HSV characterization of the pH sensor.



Figure S2. Protein sensor characterization in the color space. A) Concentration-dependent RGB characterization

of the protein sensor. B) Concentration-dependent HSV characterization of the protein sensor.



Figure S3. Glucose sensor characterization in the color space. A) Concentration-dependent RGB characterization of the glucose sensor. B) Concentration-dependent HSV characterization of the glucose sensor.



Figure S4. Comparison on the glucose sensors chromaticity in buffer solution and human urine samples. A)

Glucose-dependent RGB components in buffer solution and human urine. Left to right: red, green, blue. B) Glucose-dependent HSV components in buffer solution and human urine. Left to right: hue, saturation, value.

Methods of computational color evaluation and concentration measurement

Starting with the detected and sorted rectangles of the reference chart and test strip image, to detect the colors of the rectangles, each bounding rectangle is saved as a single sub-matrix, called submat. All submats are sent successively to the respective color determination function that calculates their color. By the row and column number of the rectangle, the order in which the submats are sent to this function is defined. In this way, the detected colors are assigned to the corresponding rectangle. Following this, the color of each reference rectangle is compared with the color of each test rectangle per analyte. Afterwards, the color distances are calculated and sorted, resulting in the test evaluation of the biomarker values.

1. CIE-xy-based color evaluation

Color detection and evaluation of the CIE-xy-based method comprises color space transformation and color distance calculation. The final evaluation takes place in the CIE-xy color spectrum. Therefore, the chromaticity values of the colored rectangles are compared.

1.1. Color space transformation

For the CIE-xy-based evaluation, a color determination function for the submats was developed. To this function, the RGB submats of the bounding rectangles are sent. The number of pixels npix of a submat is its number of rows multiplied by its number of columns. Each pixel consists of a vector with the three different color values red (R), green (G), and blue (B).

First, it is necessary to find the mean value for all three color values out of all pixels of one

submat. Therefore, all color values of a certain color out of all pixels are added. This means that the red values of all pixels are added together, resulting in R_{total} . The same applies to green and blue, resulting in G_{total} and B_{total} . These values build the vector (R, G, B) total. This vector is divided by the overall number of pixels npix (Eq. S1). The resulting vector (R, G, B) mean comprises the mean color values. In other words, the mean value for each of the three colors out of all pixels of the submat is determined.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix}_{mean} = \frac{1}{npix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{total}$$
 (Eq. S1)

Since the image data is provided as sRGB, 8 bit, and non-linearized by the camera, the color space conversion from the sRGB values to the CIE-XYZ values is conducted to acquire device-independent data. Hereby, the needed vector (R, G, B) for the transformation is the vector (R, G, B) mean. After scaling, linearization, and conversion of the sRGB data to CIE-XYZ data, the 2D chromaticity values x and y are calculated to compensate for further disruptive effects, e.g., unequal camera to object distances or exposure times. These steps are executed for every submat, i.e., for every colored rectangle. Thus, an x and a y chromaticity value for every colored rectangle of the reference chart and the test strip are obtained. These resulting chromaticity values are necessary to determine the color distances between the colors of the reference chart and the test strip, as explained in the following.

1.2 Color distance calculation

To evaluate a urine test, for each biomarker, the colors of the rectangles of the test strip are compared to the colors of the rectangles of the reference chart. Comparing two colors means determining the color difference, which is expressed as the distance between two points of color in a certain color space.¹ Since the sRGB color space is non-uniform and therefore, a calculated

distance in this color space is not representative for the human color perception, the determination of color distances is conducted in the uniform CIE-XYZ color space.² The first part of the comparison is done by calculating the distances of the colors that have to be compared. The Euclidean distance dis between two colors col1 and col2 is calculated as follows:

$$dis_{col1, col2} = \sqrt{\left[(x_{col1} - x_{col2})^2 + (y_{col1} - y_{col2})^2 \right]}$$
(Eq. S2)

This equation is used to calculate all necessary distances, starting with the first biomarker, pH. The "col1" in (Eq. S2) stands for the test strip color, while "col2" is successively assigned to each color of the reference chart. The distance between the pH test strip color TSpH and the color of the first rectangle of the reference chart for pH RCpH1 is calculated as:

$$dis_{TSpH, RCpH1} = \sqrt{\left[(x_{TSpH} - x_{RCpH1})^2 + (y_{TSpH} - y_{RCpH1})^2 \right]} \quad (Eq. S3)$$

Subsequently, the remaining distances for all biomarkers are calculated the same way.

The second part of the comparison is done by sorting the color distances. The goal is to find the smallest one, i.e., the smallest distance between the test strip rectangle color and a reference chart rectangle color of the same biomarker, stating that these two colors are the most similar ones. The algorithm for this task sorts the color distances increasingly for each biomarker. Hereby, the original order of the reference chart is saved as indices together with the new, sorted order. If, for example, the color distance to the third reference rectangle is the shortest, the algorithm will remember a number for the order of the distance, here 1, and a number for the original order, here 3. In other words, a table is generated that associates the rank in terms of distance to the "rectangle ID", i.e. the order number inside its row. Thus, the identifying ordering number or position is kept. Being aware of what position the reference rectangle has

in the original order is important for the evaluation when assigning analyte concentrations to the test areas. Every reference rectangle has a specific biomarker value to which the ordering number corresponds.

Since it is very unlikely that a color distance is zero, stating the test rectangle and a reference rectangle of a specific analyte have the exact same color, in all probability, the test color is in between of two reference colors. This implicates that these two reference colors have the two shortest distances to the test color. For further proceeding, the two reference rectangles which correspond to those two reference colors and their corresponding original position need to be determined. The reference chart is built that the colors for the increasing analyte concentrations are arranged from long to short wavelengths concerning the visible light spectrum. For that reason, the original positions of the two reference rectangles of interest are next to each other. The test and reference rectangle colors for a specific analyte are figured as points in the CIE-xy color space (Figure S 1a). The test color value is defined as point C. The reference color value closest to the test color value is point A. The corresponding and also shortest Euclidean color distance is defined by [CA]. The reference color value second closest to the test color value is point B. To C, it has the second shortest Euclidean color distance, [CB]. The line segment from A to B is [AB]. Because the test color may not be on the same curve as the two reference colors, it is projected to the baseline built by the reference colors. To reach this, the perpendicular is created from C to [AB]. The foot of the perpendicular, called O, is the point of intersection of the perpendicular and the line segment. Point O depicts the projected color value of C onto [AB] (Figure S 1a). To determine the analyte concentration of the test color, the following distances are calculated: The distances between the projected color value and the two reference color values, i.e., [AO] and [BO], and the distance [AB]. First, based on the general equation (Eq. S4) with the generic angle γ and the vectors \vec{A} (Eq. S5) and \vec{B} (Eq. S6) of a triangle, the cosine of the angles α and β is determined (Figure S 1a). The developed C++ function for angle calculations is dependent on pt1, pt2, and pt0. "pt1" is the vector comprising the x- and y-value of the test point C, "pt2" is the vector comprising the x- and y-value of the reference point at which the cosine is not calculated, and "pt0" is the vector comprising the xand y-value of the reference point at which the cosine is calculated.

$$\cos(\gamma) = \frac{\vec{A}\vec{B}}{|\vec{A}||\vec{B}|} = \frac{dx_1dx_2 + dy_1dy_2}{\sqrt{\left[(dx_1dx_1 + dy_1dy_1)(dx_2dx_2 + dy_2dy_2) + 10^{-10}\right]}}$$
(Eq. S4)

$$\vec{A} = \begin{bmatrix} dx_1 \\ dy_1 \end{bmatrix} = \begin{bmatrix} pt_1 \cdot x - pt_0 \cdot x \\ pt_1 \cdot y - pt_0 \cdot y \end{bmatrix}$$
(Eq. S5)

$$\vec{B} = \begin{bmatrix} dx_2 \\ dy_2 \end{bmatrix} = \begin{bmatrix} pt_2.x - pt_0.x \\ pt_2.y - pt_0.y \end{bmatrix}$$
(Eq. S6)

To obtain \vec{A} and \vec{B} for calculating the cosine of α and β , denoted as \vec{A}_{α} (Eq. S7), \vec{B}_{α} (Eq. S8), \vec{A}_{β} (Eq. S9) and \vec{B}_{β} (Eq. S10), the following calculations are performed:

$$\vec{A}_{\alpha} = |C - A| = \begin{bmatrix} C.x - A.x \\ C.y - A.y \end{bmatrix}$$
(Eq. S7)

$$\vec{B}_{\alpha} = |B - A| = \begin{bmatrix} B.x - A.x \\ B.y - A.y \end{bmatrix}$$
(Eq. S8)

$$\vec{A}_{\beta} = |C - B| = \begin{bmatrix} C.x - B.x\\ C.y - B.y \end{bmatrix}$$
(Eq. S9)

$$\vec{B}_{\beta} = |A - B| = \begin{bmatrix} A.x - B.x \\ A.y - B.y \end{bmatrix}$$
(Eq. S10)

Thus, the cosines of the angles α (Eq. S11) and β (Eq. S12) result in

$$\cos\left(\alpha\right) = \frac{\vec{A}_{\alpha}\vec{B}_{\alpha}}{\left|\vec{A}_{\alpha}\right|\left|\vec{B}_{\alpha}\right|}$$
(Eq. S11)

$$\cos\left(\beta\right) = \frac{\bar{A}_{\beta}\bar{B}_{\beta}}{\left|\bar{A}_{\beta}\right|\left|\bar{B}_{\beta}\right|}$$
(Eq. S12)

The distances [AO] (Eq. S13) and [BO] (Eq. S14) can now be determined as:

$$dis_{A0} = |dis_{CA}cos(\alpha)|$$
(Eq. S13)

$$dis_{BO} = |dis_{CB}cos(\beta)|$$
(Eq. S14)

With these results, the distance [AB] is built by adding the sub distances as:

$$dis_{AB} = dis_{A0} + dis_{B0}$$
(Eq. S15)

By making use of these results, it is now possible to find the analyte values, completing the evaluation. Several cases are observed for this step. First, it is of importance to know whether the evaluation takes place linearly or non-linearly. Based on the given reference values on the provided reference chart, pH is evaluated linearly (Figure S 1b), while protein and glucose are not (Figure S 1c,d). In a config header file, the linearity for each analyte is defined as a Boolean variable, i.e., true if linear or false if not. If the evaluation happens to be non-linear, all reference parameters of the corresponding analyte are sent to a function that calculates the piecewise Hermite interpolating polynomial. This polynomial is based on the given reference parameters, which are the supporting values for the interpolation algorithm. After the linearity of the evaluation process of each analyte is clarified, another two cases need to be contemplated to remain in the interval defined by the nearest reference point A and the second nearest reference point B. In the CIE-xy color space, the reference points for the analytes start with a higher xvalue that decreases with a rising analyte concentration. In the first case, the index of A is higher than the index of B. In this case, the distance of the measured color point C mapped to AB has to be added to A, i.e., from the higher index in direction to the lower index B. In the other case,

the mapped distance has to be subtracted from the lower index A and thus go into the direction of B. By these operations, the "real-valued" index of the above-mentioned point O located between A and B is obtained. The real-valued index O now serves either as the x-value for the piecewise Hermite interpolating polynomial in the non-linear case or as the x value of the linear function defined by the corresponding analyte values of A and B. The result of these functions is the actual analyte value. This procedure is carried out in a loop, where the next analyte is evaluated in the next run. The resulting analyte values are collected in a vector and returned as a string to the Java/Android part of the program.

Concerning the non-linear evaluation functions, regular polynomial functions were used at first. However, it showed that these are, on the one hand, insufficient since the variances of the reference points to the according points of the closest polynomial function are too big. On the other hand, the coefficients would have to be determined beforehand, resulting in high effort. Next, cubic spline interpolation was used for the evaluation functions. As it turned out later, these functions were still not accurate enough. Therefore, piecewise Hermite interpolating polynomials are used for the evaluation. The comparison of the latter two evaluation options is illustrated in **Figure S1c,d**.



Figure S5: CIE-xy-based evaluation functions. (a) Exemplary triangle in the CIE-xy color space illustrating the relation of the color values with the angles α and β . C: The test color point. A: The reference color point nearest to C. B: The reference color point second nearest to C. O: C projected onto the line segment of A and B. D and E: Further reference points. (b) Linear function for pH. (c) Non-linear functions for protein. (d) Non-linear functions for glucose. The points mark the reference values of the reference chart that serve as supporting values for the functions. For the non-linear functions, two options are compared. The green graphs show the piecewise Hermite interpolating polynomials. The red dashed graphs show the cubic spline interpolations.

2. HSV-based color evaluation

The HSV-based color detection and evaluation takes place in the HSV color spectrum. Therefore, the H-values of the colored rectangles are compared.

2.1. Color space transformation

An HSV-based color determination function was developed. To this function, the OpenCVobtained HSV submats are sent. The submats consist of the color values hue (H), saturation (S), and "value" (V) (Figure S 2a). For each of the three parameters, the average is calculated analogously to R, G, and B of the CIE-xy-based color calculation. Since the H-value is based on 360°, different shades of the red color range approximately from 340° to 360° and from 0° to 30°. In OpenCV, the range of the H-value comprises 0 to 180 since two consecutive degrees are consolidated. If a submat contains H-values near to the left side and right side of the 360° or 0° border, the calculated average is incorrect. This occurs especially, or rather only for the reddish pH rectangles. Therefore, a test is introduced before the calculation of the average. If the measured H-value of a pixel of the submat is higher than 160 and less than 180, then 180 are subtracted from it. This projects all H-values to the same side of the border. Therefore, that side of the border was chosen where more H-values were detected. Another test to prevent anomalies is performed by checking whether the H-, S-, and V-value of a single pixel are all zero at the same time. If this is the case, the pixel is not included in the calculation of the averages. Subsequent to the calculation of the averages of H, S and V, it is checked whether the average H-value is negative. If this is the case, 180 is added to this value.

2.2. Color comparison and calculation

After obtaining the averages of the H-, S- and V-value for each colored rectangle of the reference chart and test strip, the average H-values are necessary for further proceeding. These average H-values of the reference colors for a respective analyte serve as x-values for the evaluation function. Assigned to these x-values are the reference analyte concentrations, serving as y-values. By these x-y-reference-pairs, the piecewise Hermite interpolating

evaluation polynomials are created for each analyte. The average H-value of a test color is plugged in the corresponding function, yielding the analyte concentration assigned to this H value.

At first, cubic spline interpolation was used for the non-linear evaluation functions. As it turned out later, these functions were not accurate enough. Therefore, piecewise Hermite interpolating polynomials are used for the evaluation. The comparison of these two evaluation options for pH, protein, and glucose is illustrated in Figure S 2b-d.



Figure S6: HSV-based evaluation functions. (a) HSV color space with the ranges for hue (H): [0, 360], saturation (S): [0 %, 100 %], value (V): [0 %, 100 %]. (b) Linear function for pH. (c) Non-linear functions for protein. (d) Non-linear functions for glucose. The points mark the reference values of the reference chart that serve as supporting values for the functions. For the non-linear functions, two options are compared. The green graphs

show the piecewise Hermite interpolating polynomials. The red dashed graphs show the cubic spline interpolations.

3. Testing accuracies

The average testing accuracies of testing with buffer solutions under real conditions for several concentrations per analyte are portrayed in Table 1. These values correspond to the data of Figures 4 to 6. Overall, a testing accuracy of about 98% is yielded.

Table S1: Average testing accuracies for the data given in Figure 4 to 6. Piecewise Hermite interpolating polynomials were used for either CIE-xy based or HSV based evaluation. Testing with buffer solutions under real conditions.

Analyte	Evaluation	Reference value	Testing accuracy [%]
рН	CIE-xy based	5	100
		6.5	99.64
		7.5	99.00
		9	100
	HSV based	5	100
		6.5	99.97
		7.5	99.77
		9	100
Protein	CIE-xy based	30 mg/dL	96.72
		100 mg/dL	95.43
		200 mg/dL	96.79
		500 mg/dL	100
	HSV based	30 mg/dL	99.46
		100 mg/dL	95.38
		200 mg/dL	97.67
		500 mg/dL	99.99
Glucose	CIE-xy based	50 mg/dL	92.54
		100 mg/dL	98.91
		200 mg/dL	94.6
		500 mg/dL	98.78
		800 mg/dL	98.41
		1000 mg/dL	100
	HSV based	50 mg/dL	94.13
		100 mg/dL	97.63
		200 mg/dL	98.69
		500 mg/dL	99.6
		800 mg/dL	98.68
		1000 mg/dL	100

References

- R. Yang, W. Cheng, X. Chen, Q. Qian, Q. Zhang, Y. Pan, P. Duan, P. Miao. Color Space Transformation-Based Smartphone Algorithm for Colorimetric Urinalysis. *ACS Omega* 2018, 3, 12141-12146.
- 2. S. Akraa, A. P. T. Tam, H. Shen, Y. Tang, B. Z. Tang, J. Li, S. Walker. A smartphone-based point-of-care quantitative urinalysis device for chronic kidney disease patients. *J Netw Comput Appl* **2018**, 115, 59-69.