Electronic Supplementary Information (ESI)

Instrument-Free Single-Step Direct Estimation of Plasma Glucose Level from one Drop of Blood Using Smartphone-Interfaced Analytics on a Paper Strip

Sujay K Biswas[a], Subhamoy Chatterjee[b], Sampad Laha[c], Victor Pakira[d], Nirmal K Som[e], Satadal Saha[a,f,g], Suman Chakraborty[c,d,*]

[a]*School of Medical Science and Technology, Indian Institute of Technology Kharagpur, Kharagpur, 721302, India.*

[b]*Electronics and Electrical Communication Engineering, Indian Institute of Technology Kharagpur, Kharagpur, 721302, India*

[c]*Department of Mechanical Engineering, Indian Institute of Technology Kharagpur, Kharagpur,721302, India.*

[d]*Advanced Technology Development Centre, Indian Institute of Technology Kharagpur, Kharagpur,721302, India.*

[e]*B C Roy Technology Hospital, Indian Institute of Technology Kharagpur, Kharagpur,721302, India.*

[f]*BC Roy Institute of Medical Science and Research, Indian Institute of Technology Kharagpur, Kharagpur, 721302, India.*

[g]*JSV Innovations Pvt. Ltd, Kolkata, 700025, India.*

Assay protocol of the plasma glucose test kit

a) Take a paper strip from the kit box and place it onto clean area
b) Set a new lancet needle from the given kit. Set the depth of the needle (in between 2-5.5). Wipe the finger using 70% alcohol/swab.
c) Prick the finger. Wipe the first drop of the blood using cotton.
d) Draw the blood by touching the finger at the bottom of the paper channel. Wait till the reaction pad is fully wet by the plasma. No separate sample metering arrangement is required.
e) Wait for few minutes to generate the color at the reaction pad
f) Insert the strip through the slit of the plastic box.
g) Switch on the plastic box. Start the smartphone app. This will start the camera, capture images at the pre-set interval and display the result after analyzing the images without manual interpretation needed.

Image Pre-processing

Step 1: Image de-noising. Because of the small size low cost CMOS image sensor, non-diffused light and non-uniform illumination and random nature of imaging, Smartphone captured images contain significant noise. Generally, the noise present in imaging is modelled as additive Gaussian white noise (AWGN). If the original uncorrupted image pixel intensity, $y(i, j)$ is corrupted by AWGN $\eta(i, j)$, then the measured image pixel intensity at the (i,j) location is given by

$$z(i, j) = y(i, j) + \eta(i, j) \tag{S1}$$

The aim of any image de-noising technique is to estimate the true pixel intensity $y(i, j)$ from the noisy measurement pixel value $z(i, j)$ and some prior knowledge about noise statistics.

Block matching and 3D collaborative filtering (BM3D) is a very well-known image de-noising technique with known noise parameter (variance). The basic assumption of BM3D algorithm is the existence of various self-similar blocks within an image. Any difference in intensity within these blocks is attributed due to noise. Hence, first the image is segmented into various non-overlapping blocks of similar size. Then, the similar blocks are grouped in 3D array or groups. Collaborative filtering is a special ML procedure developed to deal with these 3D groups. Due to the similarity between the grouped fragments, the transform can achieve a highly sparse representation of the true signal so that the noise can be well separated by shrinkage. The initial noise estimate for this algorithm is provided by the minimum variance of these similar blocks.

Step 2: Image segmentation: First the de-noised image is resized from (3456 x 4608) into (500×500) pixels to reduce the computational load. Here, the background is uniform and homogeneous and thus can be easily separated from the colorful reaction pad or foreground by any thresholding algorithm. We used Otsu thresholding for segmenting the region of interest

(ROI) from background. The algorithm exhaustively searches for an automatic threshold that minimizes the intra class variance of both background and foreground regions. The total intra-class variance is expressed as the weighted sum of variances of foreground and background. Let $\sigma_0^2(t)$ and $\sigma_1^2(t)$ be the variances for the foreground and background pixel intensity, respectively. Accordingly, the intra class variance is given by

$$\sigma_w^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \tag{S2}$$

The weights are obtained from the image histogram as probabilities of two classes. Thus, an optimum threshold is obtained by minimizing the intra-class variance.

To tackle the non-uniform color generation, the segmented ROI image is converted into HSV color space and the saturation plane is again thresholded by the Otsu algorithm to keep the colorful region as foreground and any non-colorful region in the reaction pad due to non-uniformity in color generation as background.

Step 3: Morphological operations: Here, we have performed a set of theoretical, mathematical and morphological operations to filter out any kind of unwanted shape attached erroneously with the desired foreground. The segmented image is subjected to a series of morphological operations of erosion, dilation, opening and closing with a disk structuring element. The erosion of a binary image A by the structuring element B is defined by

$$A \bullet B = \{z \in E \mid B_z \subset A\} \tag{S3}$$

The dilation of a binary image A by the structuring element B is given by

$$A \oplus B = \bigcup_{b \in B} A_b \tag{S4}$$

Erosion is set theoretic subtraction which aims to erode any unwanted shape detached from the main foreground. On the other hand, dilation is set theoretic summation, which aim to fill any gap in the foreground. Opening of the binary image A by structuring element B is a series of operations, starting with the erosion of A by B, followed by the dilation the result by B and is given by the following equations:

$$A \times B = (A \bullet B) \oplus B \tag{S5}$$

The closing operation is just opposite to the opening operation. It first performs dilation of binary image A by structuring element B then erosion of the result by structuring element B and is given by the following equation:

$$A * B = (A \oplus B) \bullet B \tag{S6}$$

Limit of detection (LOD)

The limit of detection (LOD) is determined experimentally at 40 mg/dl. Below this glucose concentration, the results become inconsistent. In parallel, we estimated the LOD via quantitative approach using the following method. The limit of blank (LoB) is defined by the highest

apparent analyte concentration when the region of interest contains no analyte and is quantified by the following equation:

$$LoB = mean_{blank} + 1.645(SD_{blank}) \quad \text{(S7)}$$

where $mean_{blank}$ represents the mean value of the apparent analyte concentration of three blank images considered for this quantification as reference. SD_{blank} is the standard deviation of these three apparent concentrations calculated from blank. Finally, the limit of detection (LOD) is calculated by utilizing both the measured LoB and the lowest concentration of the analyte and is given by the following equation:

$$LOD = LoB + 1.645(SD_{low\ concentration\ sample}) \quad \text{(S8)}$$

$SD_{low\ concentration\ sample}$ is the standard deviation for three glucose values obtained quantitatively for three repetitions of the experiment with lowest possible analyte concentration.

For the LOD quantification, we considered experimental images of three tests for both 30 mg/dl and 40 mg/dl standard samples and three images of blank reaction spot without any analytes. The $mean_{blank}$ was found to be 36.93 mg/dl and SD_{blank} 3.29. This provides the LoB value of 42.33 mg/dL.

Next, we calculated the standard deviation of 3.29 for 30 mg/dl and 2.99 for 40 mg/dl glucose concentration. Finally, the LOD was estimated to be 47.25 mg/dl for the standard glucose concentration of 40 mg/dl, 49.49 mg/dl for the standard glucose concentration of 30 mg/dl. The quantitatively estimated LOD of 47.27 mg/dl served as a reliable lower bound as the experimentally observed LOD turned out to be 40 mg/dl.

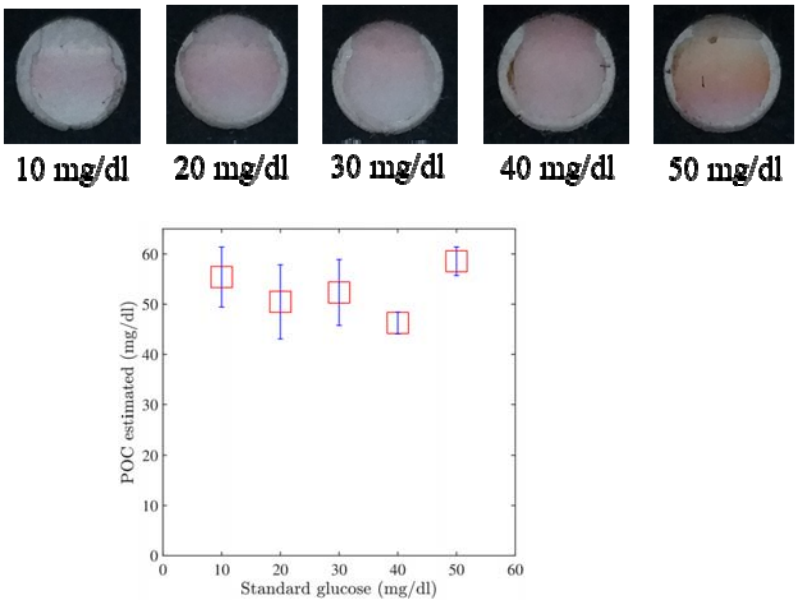


Fig. S1. Limit of detection (LOD) of the glucose assay (experimental). Standard glucose sample used for the determination of the LOD by invoking the calibration curve produced for the glucose assay. A series

of standard glucose concentrations from 50 mg/dl to 10 mg/dl was considered and tested for the evaluation of LOD with accurate quantification. For each data point, the test was performed thrice to check the accuracy and repeatability of the result.

(a) (b)

Fig. S2. LOD optimization using quantitative method. (a) Region of interest (ROI) of the blank reaction spot (b) ROI of the reaction spot after color generation for glucose concentration of 30 mg/ dl

Normalization of the inter-smartphone variation

We used two different smartphones (MI Redmi Y1 and Samsung Galaxy A6+) for illustrating the normalization procedure on the result of the plasma glucose assay with different source-specifications of the training images. The normalization methodology is described in the main manuscript text. Results illustrating the outcome for a gold-standard validated glucose concentration of 160 mg/dl are presented in the Table S1.

Table S1: Result of the glucose assay (160 mg/dl) with the MI Redmi Y1 and Samsung Galaxy A6+ smartphones

Image id	Smartphone	Predicted by regression without normalization (mg/dl)	Predicted by least square-based normalization (mg/dl)
MI1	MI Redmi Y1	148.2342524	158.8972808
MI2	MI Redmi Y1	150.8377969	161.7991038
MI3	MI Redmi Y1	151.0634622	162.0427478
MI4	MI Redmi Y1	150.3945934	161.3155347
MI5	MI Redmi Y1	151.0139822	161.962329
S1	SAMSUNG Galaxy A6+	162.8697018	162.8697018
S2	SAMSUNG Galaxy A6+	160.993312	160.993312
S3	SAMSUNG	160.6682431	160.6682431

	Galaxy A6+		
S4	SAMSUNG Galaxy A6+	160.6801957	160.6801957
S5	SAMSUNG Galaxy A6+	160.9392226	160.9392226

RReliefF algorithm for feature selection

For the correct selection of the desired statistical features from a large number of primary and secondary features extracted by processing colored images captured for the estimation of plasma glucose, we employed an iterative rank assignment algorithm RReliefF on the selected features. Generally, there is a risk of erroneous selection and also unknowingly discarding statistically important features by manual selection from a large number of handcrafted feature sets. In its automated feature selection process, on the contrary, the RReliefF algorithm assigns weight to all the extracted features based on a random sample of observations of features and computes the corresponding dependent variable (plasma glucose concentration) from all the features and variable sets.

The above algorithm works on the principle of penalizing the features with large distance but nearest response values (plasma glucose concentration) and awarding features with small or negligible distance and nearest response values (plasma glucose concentration). The key steps of the analysis are summarized below.

Let Y be the plasma glucose concentration, f_1 and f_2 be two random features picked from a feature set. Let O_1 and O_2 be the two nearest neighbor observations. Let W_{dy} be the weight function for having significant differences in Y (plasma glucose concentration). W_{dj} represents the weight for having significant differences in the predicted Y. Let $W_{dy \bigcap dj}$ be the weight function for having significant differences in both the actual values of Y and the predicted Y. W_u denotes an iteratively updated weight given to the concerned feature after some specified numbers of iterations. The algorithm advances from an initial choice of the parameters till the final convergence via the following major steps:

1. Initialize $W_{dy}, W_{dj}, W_{dy \bigcap dj}$ and W_u as 0.
2. For iteration 1 to K:
   For an iteration i:

$$W^{i}_{dy} = W^{i-1}_{dy} + \Delta_y(O_1, O_2).d_{12}$$

$$W^{i}_{dj} = W^{i-1}_{dj} + \Delta_j(O_1, O_2).d_{12}$$

$$W^{i}_{dy \bigcap dj} = W^{i-1}_{dy \bigcap dj} + \Delta_y(O_1, O_2).d_{12}$$

$\Delta_y(O_1, O_2)$ is the difference in the values of the plasma glucose concentration for two random observations O_1 and O_2.

$$\Delta_y(O_1, O_2) = \frac{|Y_1 - Y_2|}{\max(Y) - \min(Y)}$$

$$\Delta_j(O_1, O_2) = \frac{|O_1 - O_2|}{\max(F_1) - \min(F_2)}$$

where, F_1 and F_2 are the predictors of the two observations O_1 and O_2.

$$d_{12} = \frac{d^{"}_{12}}{\sum_{l=1}^{n} d^{"}_{1l}}$$

where, $d^{"}_{12} = \exp(-rank(1,2)/\sigma)^2$

Here, $rank(1,2)$ is the position of the observation O_2 among the nearest neighbour of the observation O_1.

3. After the iteration K, the algorithm calculates the predicted and updated weight W_u by the following expression:

$$W_u = \frac{W_{dy \bigcap dj}}{W_{dy}} - \frac{W_{dj} - W_{dy \bigcap dj}}{K - W_{dy}}$$

4. Sort the set of all W_u in descending order and select the most decisive features for the regression.

Using the algorithm RReleifF, we thus identified three highest rank secondary features of color saturation (S), ratio of (B/R) and average ratio (B/R and G/R) as the most significant features.

Detection video of the glucose assay (See the uploaded Supplementary Video).

The detection video delineates the operational steps of our POC device for estimation of plasma glucose level using GlucoMED smartphone app.

Step-1: Extraction of finger-prick blood using a lancing device

Step-2: Drawing of the blood drop using the paper-strip

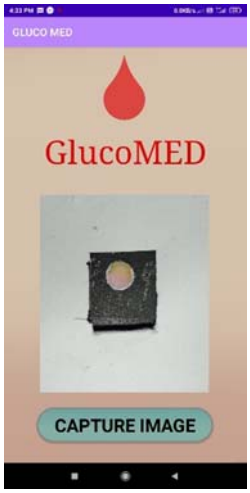
Step-3: Color generation in the reaction pad

Step-4: Insertion of paper-strip into the device and image capturing

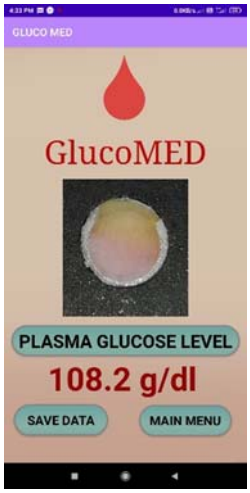
Step-5: Image analysis and result display using GlucoMED app

(a)

(b)



(c)



(d)

Fig. S3. Step-wise operation of the GlucoMED app. Pictorial representation of step-by-step operation of the "GlucoMED" app for quantitative estimation of plasma glucose. (a) app icon on the smartphone screen (b) start new test or previous test results (c) capture image (d) quantitative result display.

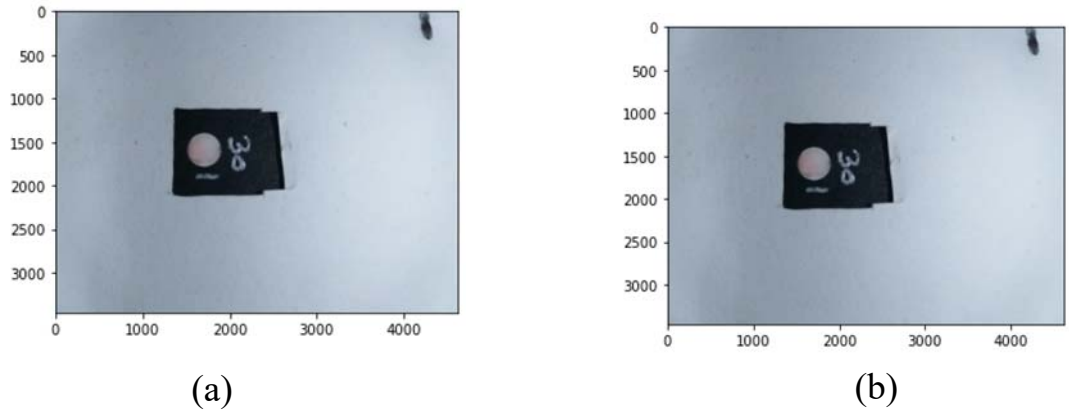


Fig. S4. Image preprocessing. (a) original image captured using smartphone camera (b) de-noised image after the operation by BM3D method

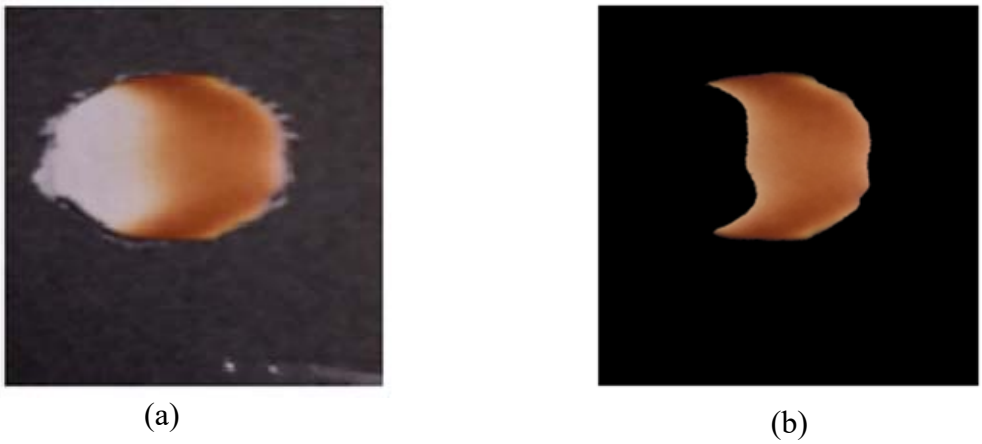


Fig. S5. ROI segmentation on a captured image. (a) original image of colored reaction spot for glucose concentration 300 mg/dl and (b) ROI after segmentation process

Reagent stability

The reagent stability results are summarily captured in Fig. S6and S7. Accompanying discussion is included in the main text.

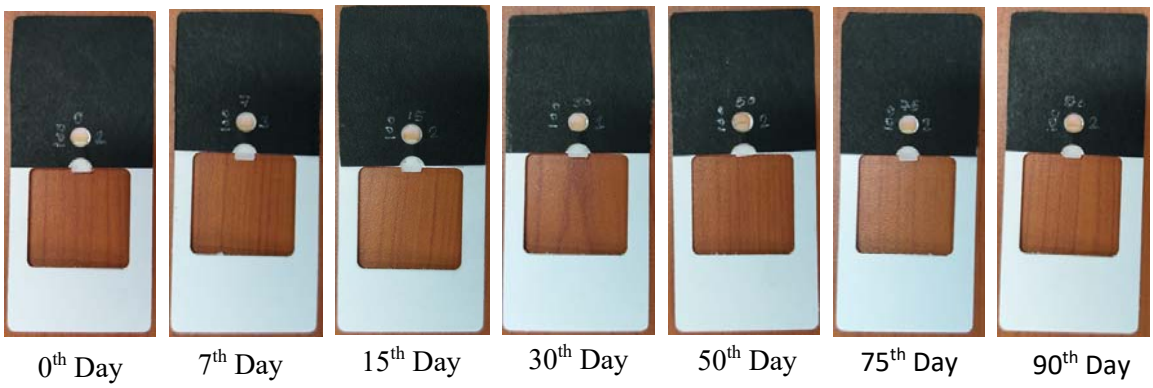


Fig. S6. Experimental images of the paper strip with color generated at the reaction pad. These images were taken during the reagent stability test with prescheduled date as mentioned.

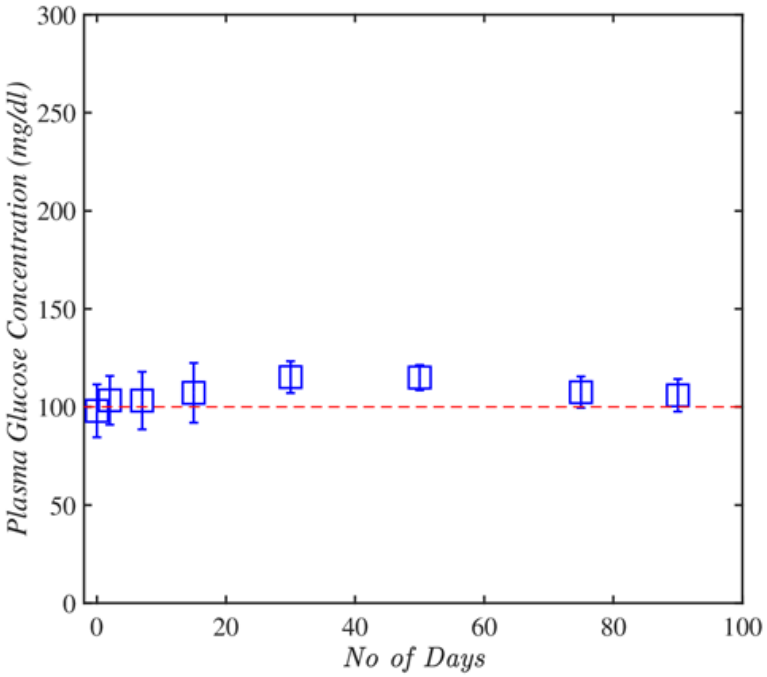


Fig. S7. Reagent stability test. Results of reagent stability test of off-the-shelves test strips with 100 mg/dl glucose concentration. Tests were performed with prescheduled dates (0th, 2nd, 7th, 15th, 30th, 50th, 75th and 90th) after lyophilisation of the reagent onto the reaction pad. Error bar represents the result of the test repeated five times with five strips stored for a particular date using same glucose concentration (100 mg/dl).

Table S2: Manufacturing cost-analysis per paper strip sensor

Component	Quantity	Cost of unit (USD)	Cost (USD)
GOD POD reagent (ARKRAY Inc)	5 µl	8.83 (450 ml)	0.000010
Potassium Iodide (KI)	0.2 mg	94.65 (100 gm)	0.00019
Whatman nitrocellulose filter paper Grade 1	5 mm diameter	12.62 (100 pc. 125 mm dia)	0.000076
Plasma separation membrane (LF1, GE)	6 mm	58.81(50 m)	0.0071
Wax coated paper	35 mm × 35 mm	350 (100 m)	0.0015
1 mm paper sheet for bottom support	35 mm × 80 mm	200 (100 m)	0.0020
Ancillary cost	-	-	0.0063
Total Cost (technical)			~ 0.02 USD
Additional Cost (Business components) @ 50% per test			0.01 USD
Total Cost (per test)			0.03 USD

Table S3: Manufacturing cost-analysis of the portable plastic box

Component	Quantity	Cost (USD)
Plastic box made of 3D printed filament	1	1.9
Arduino UNO	1	6.96
LED light	1	0.13

Electric wires, connectors, adhesives etc	-	0.025
Ancillary cost	-	0.63
Total Cost (USD)		9.07

Table S4. Competitive landscape of our assay with other glucose meters

Key Parameters	**Accu-Chek Active**	**GlucoOne**	**Dr. Trust**	**AccuSure**	**Control D**	**Contour Plus**	**On Call Plus**	**Our Device**
Detection Technology	Reflectance photometry	Electrochemistry	Electrochemistry	Electrochemistry	Electrochemistry	Electrochemistry	Electrochemistry	**Colorimetric**
Test method	Whole Blood Glucose	Whole Blood Glucose	Whole Blood Glucose	Whole Blood Glucose	Whole Blood Glucose	Whole Blood Glucose	Whole Blood Glucose	**Direct Plasma Glucose**
Cost of the Test Strip (USD)	0.25	0.21	0.25	0.32	0.50	0.29	0.28	**0.03**
Accuracy with Hct variation	Error (10-20%)	Error (10 - 20%)	Error (10-40%)	Error (10 - 40%)	Error (10 - 20%)	Error (10 - 40%)	Error (10 - 40%)	**Error < 10%**