Machine learning based rapid diagnostic- test reader for albuminuria using smartphone

Ritambhara Thakur, Prateek Maheshwari, Sudip Kumar Datta, Satish Kumar Dubey, Chandra Shakher

Abstract: Albuminuria is an excellent marker for early diagnosis of kidney and cardiovascular disease. Urine dipsticks that are widely used for rapid screening of albumin, lack sensitivity and specificity in lower concentrations (<300 mg/dL) which is clinically very significant for early diagnosis and often provide qualitative or semi-quantitative results. Precise quantification of lower concentrations is based on urinary analyzers that are not portable and cannot be used in point-of-care (PoC) settings. Here, the feasibility of an accessory free analytical device has been demonstrated using a smartphone. Amalgamation of a smartphone with a dipstick enables rapid and inexpensive diagnosis. It estimates not only the standard five concentrations used in dipstick method, but ten different concentrations. This enables accurate detection and quantification of albumin at lower concentration, clinically significant in the early diagnosis of kidney disease. In order to mitigate ambient light conditions and shadow of the smartphone, images of strips were taken in a smartphone camera with "Flash ON" mode. Machine learning algorithms were used to classify ten different albumin concentrations, corresponding to normal, micro albuminuria and macro albuminuria conditions. The study was performed under varying illumination conditions using multiple smartphones. Random Forest algorithm yields an accuracy of 92% in constant illumination and variable smartphone conditions. In variable smartphone and illumination condition, it yields 82% accuracy on the test data. The detection limit of the proposed method is 7.8125 mg / dL.

Keyword**:** Microalbumin, Point-of-care, Smartphone, Machine learning, proteinuria, dipstick

I. INTRODUCTION

Chronic kidney disease(CKD) is the 12th leading cause of death and is a major health problem worldwide[1][2]. Globally, the prevalence of CKD is 9.1% with 700 million cases, according to Global Disease burden study[3]. CKD is a non-communicable disease that occurs due to various causes including diabetes and hypertension. Albuminuria is used as a diagnostic screening tool for undiagnosed CKD. Elevated albumin (Microalbuminuria between 30 mg/dL and 300 mg/dL) excretion in urine is an early sign of kidney damage and is correlated with the progression of CKD and cardiovascular disease [4]. Furthermore, Glomerular filtration rate (GFR), a tool to measure kidney function, is often normal or partially elevated in the early stages (Stage1 and 2). Often, the symptoms of kidney dysfunction in patients are evident only in the later stages of the disease (stage 4 and 5) and until this stage, patients are generally unaware of kidney dysfunction. Albuminuria can play an excellent role in the early detection of the kidney damage in individuals. Also, prediction of abnormal urinary albumin excretion is one of the early manifestations of the insulin resistance syndrome. Hence, as early progression of CKD is slow, it should be monitored regularly by assessment of albuminuria.

At present, various analytical methods are used for measurement of albumin concentration in blood and urine: Dye binding, immunochemical assay, HPLC, electrophoresis, nephelometry, turbidimetry and near infrared spectroscopy [5]. Although these techniques are specific and sensitive to the measurements at lower concentration, they require well-equipped laboratory settings and cannot be employed in PoC settings.

In recent years, global access to the internet has made smartphone applications a viable alternative to the laboratory-based testing procedures. With the use of image processing algorithms, they are becoming more computationally efficient, and can be used for quantification of biochemical assays[6], paper based microfluidics[7], and lateral flow assays [8][9][10]. They are being considered as a viable healthcare solution in the developing world as they do not need skilled manpower to conduct such basic tests in PoC settings. Quantification of analytes using smartphone based PoC device is mainly influenced by the following factors: 1) Ambient light conditions,2) Spectral sensitivity of the camera sensor and 3) Settings of the smartphone camera .In the last decade, various solutions with mobile accessories (add-on modules that can be used as smartphone attachments) have come up [11] [12] [13] [14] [15] to mitigate ambient lighting conditions. Also, every smartphone model comes with different camera sensors (CMOS sensor) whose spectral sensitivity varies with model and Make. Camera settings in the smartphone are also adjusted to have the best image appearance and hence do not give the true value of the spectral change. Such camera settings are auto-adjusted according to ambient lighting conditions, yielding spurious spectral values for minute variations in external conditions.

Smartphone based PoC methods to measure albumin concentration are broadly classified into two categories: 1) Quantitative [16][17][18], and 2) semi-quantitative [19][6][20][21][22][23][24][12][25][26][27].Quantitative methods are cuvette based and primarily operate on Beer-Lambert's Law. A summary of this approach, proposed by

R. Thakur, P. Maheshwari, S. K. Dubey and C Shakher are with SeNSE, Indian Institute of technology, Delhi, 110016, India (e-mail:

S. K Datta is with Dept. of Laboratory Medicine, All India Institute of Medical Sciences, New Delhi, 110029, India.

researchers in recent years, have been provided in Table I. However, these quantitative methods have few shortcomings. Although, these methods provide fully quantitative estimation of the albumin, they require preprocessing of the sample. In addition, each smartphone comes with different hardware design. The position of the camera in smartphones vary with brand and Make. This necessitates the customization in hardware architecture of the PoC system for different smartphone models, posing limitation on its usage. Also, attachment to the smartphone only deals with ambient light condition, however there will be significant variation in color values with change in smartphone model.

Alternatively, in general practice, paper-based urine dipstick method, based on dye binding is used for the screening of albuminuria. Although, these dipsticks are highly specific (90-100%); they are semi-quantitative and lack sensitivity at lower concentrations (32-46%) and hence not suitable for quantification of microalbuminuaria(<300mg/dL) [28] [29]. Further, estimation of colorimetric change in dipstick becomes difficult as there is a very minute change in the color with concentration of analyte. Also, the interpretation of these dipsticks are operator dependent and hence subjective in nature. In recent years, researchers have come up with different PoC based systems that take care of such errors. In addition, smartphone-based approaches have been introduced that take care of variation in ambient light conditions and exhibit good inter-phone repeatability. They include both; 1) Add-on devices [20][21], and 2) and accessory-free[27][22]. Moreover, they employ algorithm-based color quantification approaches with or without reference chart [6][19]. Table II summarizes the recent development in dipstick based semi-quantitative methods to estimate various analytes. All these approaches primarily dealt with standard dipstick, microfluidic assay, and, paper-based devices which work on colorimetry principle. In dipsticks, researchers have introduced different methods to classify five standard concentration values of albumin. Different techniques adopted to calibrate these system have been illustrated in Table II. They include 1) Use of Control object (reference color chart) along with image processing algorithms [20], 2) Color space transformation [17][26],and 3) Machine learning classifiers[24][22].Researchers have opined the use of color object for calibration, to make these systems accessory-free. However, these systems require calibration at every step, for every lighting condition, and for every smartphone. Also, illumination is not uniform over the strip and the control object. Control object can have different mean color values as it is generally positioned at a certain distance from the strip. Also, as the illumination around the strip and the control object is not uniform, shadow of smartphone will result in different mean color values for both, control object and the strip.

To estimate the concentration of the analyte in the sample, images of the strips are captured using smartphone, and, different color spaces; Red-Green-Blue (RGB), Hue-Saturation-Value (HSV) and CIE L*a*b* color space; are employed to quantify the colorimetric changes [8-18] as summarized in Table II. In one of the studies, Mutlu *et al* used mean RGB values to quantify and classify pH values using machine learning [22]. Mathaweesansurn *et al* employed 'Hue' parameter of HSV color space to quantify albumin in urine sample by carrying out reaction between albumin and tetrabromophenolphthalein ethyl ester (TBPE) in the presence of Triton X-100[17]. *Tong et al* used color correction method, where RGB values were calculated using background and testing spots for cell counting in microfluidic assay [20]. Colemn *et al* quantified p24 antigen of HIV, using large dataset and showed that saturation parameter of HSV color space outperforms RGB color space to quantify different concentrations of p24 antigen [27]. In most of the studies, researchers have tried to generalize a specific color space for quantification of analytes. However, it varies from analyte to analyte and therefore, cannot be generalized.

In addition, owing to various independent input variables (camera spectral sensitivity, camera settings), it is a challenging task to generalize one color space parameter for the quantification of images acquired using a smartphone due to variation in inter-and intra-smartphone models. To address these variabilities, machine learning algorithms can play a vital role. The machine learning based models can statistically distinguish different parameters and their reliance on the concentration in which the model is pre-trained with all variable parameters [22][25]. These models can be used to identify colorimetric values and can also address the issue of interphone variability. The advantages and limitations of recent PoC systems for albumin concentration measurement, employed in cuvette and paper-based smartphone settings, is summarized in Table III.

In this work, a smartphone-based feasibility study for the quantitative estimation of albumin concentration in urine dipstick test strips, over a wide range, has been demonstrated using machine learning algorithms. Instead of five standard concentration values of albumin reported using commercially available dipsticks, ten concentrations were measured to arrive at more significant diagnostic screening process. Instead of external illumination, flashlight of smartphone was utilized to compensate the discrepancies in the light intensity arising due to variation in illumination condition and shadow of the smartphone. Images of the test strips were acquired using smartphone in variable lighting conditions. Specific smartphone models: Xiaomi note 5 pro, Xiaomi MI A1, and Realme 2, were used to study the inter-and intra-phone repeatability. Images were transformed into RGB, HSV and LAB color spaces and best features were selected according to the response of illumination and

smartphone variation. Here, mean color channel values (RGB, HSV, and Lab) have been analysed statistically with the change in 2 parameters 1) illumination condition and 2) smartphone model/make. It was observed that out of nine, four features were predominantly invariant to both, and therefore, can be used for concentration estimation. Further classification of different concentrations was done using machine learning algorithms: Logistic regression, Support vector machine (SVM) model with different kernels and Random forest (RF). The performance of these machine learning algorithms has been compared to arrive at the suitable algorithm for concentration estimation.

II. MATERIALS AND METHODS

A. *Methods*

Fig.1 illustrates the method to quantify albumin concentration using smartphone. At first, experiments were performed on standard solutions of albumin. Albumin solution with 10 different concentrations: 4000 mg/dL, 1000 mg/dL, 500 mg/dl, 250 mg/dL, 125 mg/dL, 62.5 mg/dL, 31.25 mg/dL, 15.625 mg/dL and 7.8125 mg/dL, were prepared using standard serial dilution method. Image acquisition using smartphone was carried out in different lighting conditions, using incandescent (3500K) and fluorescent(6500K) light sources at different incident angles. Each smartphone was placed normal to the strip at a fixed distance of 10 cm on a platform. When image is acquired using smartphone, the shadow casted by the smartphone, makes visible or invisible intensity change on the strip, and largely influences the intensity values over whole strip. Although, the effect of shadow on the strip has not been considered in most of the quantification procedures [19][6][22][27]. As shown in Fig.2 (a, c, e), the shadows were different for each smartphone, and there is a huge influence of ambient light conditions on the strip and the background, affecting the color values. To compensate the change in the shadow region, which, in turn, changes the intensity on strip, an illumination source could be placed in the proximity of the strip. The smartphone flash light (LED) could be a viable alternative to the additional light source, to compensate ambient light condition and shadow effect of smartphone The shadow-effect can be minimized by placing the LED light of smartphone near the strips as shown in Fig.2 (b,d,f).Color features of the acquired image; mean values of RGB, HSV and Lab values, were computed using machine learning algorithms. Based upon the response of color features with illumination condition and smartphone variability, features that are least affected by the ambient lighting and smartphone variability, were selected. These features were used by the machine learning classifiers to classify color values corresponding to 10 different albumin concentrations.

B. *Solution preparation*

In clinical settings, protein concentration in urine is measured semi-quantitatively (+, ++, +++, ++++) corresponding to the trace concentration of 30 mg/dL,100 mg/dL,300 mg/dL and 2000 mg/dL respectively, provided on the reference chart. Commercially available urine dipstick (Uristix, Siemens), which is impregnated with chemical reagent (Bromophenol blue), was used for the initial screening of proteinuria. In this method, the acidic groups of the dye react with basic groups of any protein to give blue color. These strips come with a chart which acts as a reference as shown in Fig. 3(a). In the present study, 10 different concentrations were chosen to arrive at more informed decision. An albumin solution with concentration of 4g/dL was used for the sample preparation. Standard protocols of serial dilution were followed to prepare the samples at different concentrations, which were validated using the standard clinical chemistry analyzer. The 10 concentration values are 4000 mg/dL, 1000 mg/dL, 500 mg/dL, 250 mg/dL, 125 mg/dL, 62.5 mg/dL, 31.25 mg/dL, 15.625 mg/dL and 7.8125 mg/dL. Test strips were dipped into the solutions of different concentration and respective test strip images were captured after 60 seconds. This allows the chemical reaction to complete, resulting in the colorimetric variations in the strips as shown in Fig. 3(b). Experiments were performed in two different settings: Constant illumination (Experiment I) and variable lighting condition (Experiment II). These experiments were performed multiple times on multiple days, and, solutions were prepared each day, for each experiment. In experiment I, solution was prepared once, as it was not a lengthy experiment. In this study, images were captured 5 times in a time interval of 1 minute. For experiment II, Solution was prepared thrice. Images were captured two times in a interval of 1 minute. Also, Experiment I was replicated in experiment II (300 lux and 3500K color temperature condition). Extracted mean colour values in Experiment II were found to be similar to Experiment I under the given experimental condition.

C. *Experimental Design and Image acquisition*

Images of strips were acquired under constant illumination (Experiment I) and variable lighting condition (Experiment II). Color changes over the strip were quantified using machine learning algorithms to estimate the protein (albumin) concentration.

1) Experiment I: In this study, all the lighting conditions were kept constant and images were acquired using three smartphones; Realme 2, Xiaomi note 5 pro and, Xiaomi MI A1, hereafter referred as S1, S2 and S3 respectively. It was the controlled illumination (300 lux) in a dark room environment, prepared using an incandescent Halogen light source having a constant color temperature (3500K). The illumination was regularly monitored using a lux meter. Smartphone was placed at a distance of 10 cm normal to the

urine strip over a platform and light source was kept at 45° for uniform illumination. Here, two factors majorly influence the colorimetric values at the strip; (1) Shadow of smartphone, and (2) Smartphone camera settings. In most of the studies reported in literature [22][27], camera settings were kept constant. In this study, camera settings were kept in automatic mode instead of keeping constant. Also, different smartphone have different CMOS sensor, and hence different spectral sensitivity. Therefore. Experiment I was planned in two steps. In the first step, all the lighting conditions (Color temperature, illumination and angle of light source) were kept constant and the effect of flash of the 1st smartphone camera (S1) was studied, thus keeping the spectral sensitivity constant. Images of the strips at all 10 concentrations were captured using S1 in two modes; Flash light "Without Flash" and "Flash ON" to observe the effect of flash on colorimetric values. The aim of the 1st experiment was to infer that keeping all the conditions constant, whether machine learning models are able to classify all concentration values as additional five concentrations have been introduced instead of five standard concentration values of albumin. Furthermore, although light source is placed at constant position, color values can change with slight change in camera settings, and shadow on the strip. Hence, flash light was employed in 1st experiment to analyze its effect on individual intensity values and whether it can reduce the variance among intensity values for one concentration.

In the second step of experiment I, keeping all lighting conditions constant, all three smartphones were used to observe the difference in color values obtained using respective smartphones**.** As shown in Fig.2, each smartphone draws distinct shadows on the testing paper and strip. Hence, color values will be different for each smartphone in "Without Flash" condition. Also, individual smartphone possesses distinct LEDs, hence, their Lambertian profile would be different when Flash light will be used. As shown in Fig.2 (b, d, e), it can be observed that each smartphone draws different Lambertian profile with "Flash ON". Hence, in this experiment, we analyzed the effect of different smartphones on colorimetric values, and how Flash can lessen the effect of shadow on colorimetric values among different smartphones.

***2) Experiment II*:** In this study, variable lighting condition was introduced using another light source having different color temperature (6500K). Here, 120 images were acquired sequentially at six different incident angles, without flash, corresponding to different concentrations of albumin ranging from 4000 mg/dL to 7 mg/dL.

Machine learning models were employed to estimate the albumin concentration in both the experiments. Additional experiments were performed to observe the effect of variation in smartphone and illumination on colorimetric values, and how combination of flash and machine learning classifiers can better perform in variable lighting conditions.

D. Dataset

Effect of different camera sensors, optics and camera settings in different illumination conditions were studied using three smartphones. In the first experiment, 296 images were acquired in 2 modes ("Flash ON" and "Without Flash") at 10 different concentrations using 3 different smartphones. In the second experiment, images were acquired in variable lighting conditions and at different incident angles with respect to strip. Here, 6 images were acquired sequentially for 6 different angles for a given concentration of the albumin. The repeatability was ensured by imaging each strip twice for each incident angle. Thus, a total 1400 images were acquired in both the modes.

E. Pre-processing and Feature extraction

Images acquired using different smartphones were preprocessed in MATLAB version19. Images acquired using smartphone were in JPEG format in RGB color space. RGB color space, have some shortcomings as it has negative values for some spectral range. Therefore, RGB color values were converted into HSV and LAB color space. They are more robust to variations in external lighting conditions. Region of interest (ROI), which is the albumin sensor pad on the urine strip was cropped out from the image as shown in Fig.4. Mean values of all individual channels were calculated for the images acquired in both modes; "Flash ON" and "Without Flash".

F. Machine learning Classifier

With multiple independent input variables, ambient light conditions, inter and intra-smartphone model camera settings; machine learning algorithms are the viable alternative to classify and estimate the changes in concentration. Logistic regression (LR), SVM and Random RF were applied to classify 10 different albumin concentrations by extracting color features. Description of each classifier algorithm is given in the following sub-sections.

1) Logistic regression

Logistic is a statistical model, wherein, Logistic function is used to model the binary dependent variable. Here, logistic function is an S-shaped curve which takes real values and map them into a value within the range between 0 and 1. The major difference between logistic and linear regression is that, logistic regression gives discrete value outputs, whereas linear regression gives a continuous output[30].

2) Support Vector Machine (SVM)

It is a supervised learning algorithm used for classification and regression task. SVM is based on finding the optimal hyper plane between various classes [31]. Intuitively, a good hyperplane is the one that has maximum

margin (separation) between two classes. When the data is not linearly separable in original feature space, kernalized SVMs are used. The idea behind kernel based learning is to map input data into high dimensional feature space defined by a kernel function. SVM is a constrained optimization problem and to get the optimal hyper plane, one need to approach this problem through Lagrange multiplier. The solution of primal and dual problem gives the optimal parameters and support vectors. As a result, the solution of SVM has only a significant subset of the training data called support vectors. Our solution depends on these support vectors and thus the complexity of the model is also reduced. Further details can be found in [32].

3) Random forest

RF is also a supervised learning algorithm, however, it is based on the ensemble of decision trees. It is a bagging (Bootstrap Aggregation) i.e. random input sampling with replacement model. Decision trees are very sensitive to training data and to remedify this ensembling is used along with decision trees. Being an ensemble model, RF is robust to noise and irrelevant features. Similar to SVM, RF does not over-fit and it has robustness to noise and irrelevant features and almost no fine-tuning of parameters is needed to produce good predictions [32].

G. *Evaluation of the Classification Models*

Classification accuracy was estimated using weighted average accuracy for calculation of accurate model prediction. In addition, Estimation of confusion matrix depicts more precise results of the classifier. Therefore, precision, recall and F-number metrics were calculated to measure the accuracy of the model. Precision is the ratio of true positive cases to that of total number of positive cases which is named as **positive predictive value** (PPV) and is given as

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

Recall is the ratio of true positive values to that of total number of true positive and false negative values. It is also called **sensitivity** or **true positive rate** and is estimated using eq.2. Similarly, F-number metrics is calculated using eq.3.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$F = 2\frac{Precisionx\ Recall}{Precision+Recall} \quad (3)$$

III. RESULTS AND DISCUSSION

A. *Results*

Fig.5 shows the intensity heatmaps of the images captured using smartphones S1, S2 and S3 in "Without Flash" and "Flash ON" modes. As shown in Fig.5 (a, c, e), shadows vary with each smartphone model as presented in the heatmap depicting the spatial distribution of intensity on the strip. The respective heatmaps of images taken using "Flash ON" are shown in Fig.5 (b, d, f). Here, the effect of shadow and ambient light variations on the testing paper and strip are reduced. However, as depicted in the Fig. 5(b, d, f), the spatial distribution of light intensity and hence the intensity profile on the imaging area differs with smartphone models due to different Lambertian profiles of the smartphone LED. The experimental observations are summarized below. As mentioned in section 2 C, Experimental Design and Image acquisition, two experiments were performed to study the effect of flashlight and other illumination conditions.

1) Experiment I

a) *Effect of flash*

In the first step of Experiment I, the effect of Flash on color values using 1st smartphone (S1) was studied by plotting 'hue' feature of HSV color space at different concentration values as shown in Fig.6. Concentration values ranging from 4000 mg/dl to 7.1825 mg/dl were referred as C1 to C10. In Fig. 6, under, the "Without Flash" condition, it was observed that the concentration value pairs (C2,C3) , (C4,C5) and, (C8,C9and C10) were overlapping with each other and can be misclassified by classifier. However, on the application of flash on the strip images, it can be observed that individual classes were more distinct and can give better classification accuracy among different classes.

In the second step of experiment I, keeping all lighting conditions constant, all three smartphones were used to observe the difference in color values obtained using respective smartphones. Under, the "Without Flash" condition, significant difference in intensity values for each color space were observed from one smartphone to other smartphone as shown in Fig.7(a). When the Flash Light of the smartphone was switched ON, difference in color values got reduced. From Fig. 7(b), a significant overlapping in l, a, b and h color values can be observed in all three smartphones. Although the camera spectral sensitivity and Lambertian profile was different for each smartphone, color features (l, a, b, h) were observed to be invariant to the smartphone variation. Hence, inter-phone repeatability can be achieved using these features along with "Flash ON" to quantify albumin concentrations. To validate the hypothesis, machine learning algorithms were employed on the data extracted from two smartphones and tested on third smartphone.

b) Classification

For classification, 10 concentrations were treated as ten different classes and total 296 mages were captured using smartphone camera in "Without Flash" and "Flash ON" modes. Average values of individual channels for RGB, HSV and Lab values were extracted in the region of interest (ROI). For the classification task, three machine learning models were employed. Logistic regression model was selected as baseline model for classification approach. Training and testing of classifier on the entire dataset were performed by splitting the data into 70-30 training and

testing sets respectively. Grid search method was used to find out the best parameters to train classifiers. Classification accuracy was calculated on both datasets ("Flash ON" and "Without Flash") with different set of features; 'All Features (RGB, HSV and Lab)'; 'h, l, a, b'; 'h, a, b'; 'h, a'; and 'h '.

Initially, logistic regression was used to classify different concentrations and an accuracy of 33% on training set and 31% on testing set was achieved on the dataset recorded "Without Flash". Same model was also employed on the images of the "Flash ON" dataset. Here, the accuracy improved to 35% in training dataset and 33% on testing dataset. SVM classifier was employed on the dataset with different kernels and SVM with 'poly' kernel was found to yield the best classification accuracy among all the SVM models employed. The classification accuracy improved to 81%, when 'poly' kernel was applied on the dataset with "Flash ON". In addition, SVM with other kernels ('Linear', 'Radial basis function (rbf)') were also employed for classification. To further improve the classification accuracy, RF was employed on training data, where an accuracy of 92% with "Flash ON" was achieved on testing data. Comparison of the accuracy achieved using RF and SVM classifiers is given in Fig.8. The confusion matrix for "Without Flash" dataset using RF is shown in Fig.9(a), where five classes; '2','3','5','7'and '9'; were predicted correctly with maximum class support size. In addition, three elements of class '6' and '9'; two elements of class'0'; and one element of class '4', were predicted correctly. With "Flash ON" is shown in Fig.9(b), seven out of ten classes; '2','3','4','5','7','8' and '9'; were predicted correctly with maximum class support size. Also, two elements of each class '0' and '1' and one element of class '6' were predicted correctly. Hence, RF works with the maximum efficiency even for "Without Flash" dataset.

In addition, other evaluation metrics; precision, recall and F-number, were also calculated to study the performance of different classifiers as given in Tables IV and V. Table IV shows the evaluation results for all three classifiers for "Without Flash" data. Logistic regression performs poorly and its precision, Recall and F1- score was very poor with a score of 0.06, 0.25 and 0.10 respectively. Table V shows the comparison of all classifiers for the "Flash ON" dataset. The evaluation metrics shows improvement for individual classifier for the "Flash ON" dataset in comparison to "Without Flash" data. Also, RF yielded the highest value of evaluation metrics. Precision, recall and F-measure were calculated for individual class in case of RF. Tables VI and VII showed that RF yield good classification accuracy for overall classes, results a weighted average of 0.94 precision, 0.92 recall and 0.90 F-number. F-number for individual class vary in the range of 0.25 to 1. Whereas precision lies between 0.67 to 1 and in general higher in comparison to recall, except in the case of 31.25 mg/dl.

In the above experiment, as mentioned earlier, dataset was split into 70-30 ratio, where, 70% data was used in training the model and the models were tested on 30% data. Further experiments were conducted, where the data of only two smartphones (S1 and S2) were used to train the model and the data acquired using third smartphone (S3) was used for testing. It was observed that 'l, a, b' features of Lab color space and 'h' feature of HSV color space were least affected using different smartphone models as shown in Fig. 7. This hypothesis was examined by implementing the machine learning model on different set of features, one set being all the 9 features, to train the model. The other set of features used for training included 'l, a, b, h', 'h, a, b', 'h, a' and 'h' features. As shown in Fig.10, RF obtained the highest accuracy in 'l, a, b, h' feature set with 80% accuracy using S3 on "Flash ON" dataset. Therefore, it can be concluded that combination of 'l, a, b, h' feature set using RF in "Flash ON" mode, will yield accurate classification and compensate inter-phone variability.

2) Experiment II

a) Effect of flash

In experiment II, variability in illumination condition was introduced by using another light source having different color temperature (6500K) at six different incident angles. As depicted in Fig. 11, under, "Without Flash "condition, intensity values at different concentration were overlapping with each other. However, under, the "Flash ON" mode, intensity values were differentiable. Fig. 12 shows the effect of color temperature on individual color channel values at both color temperatures. In the "Flash ON" mode, color values were invariable for l, a, b and h, which was later validated using machine learning classifiers.

b)Classification

Total of 1400 images, corresponding to 10 different concentrations, were tested in experiment II. It was found that the Logistic regression did not yield desired accuracy, even under constant lighting conditions, and, therefore, only SVM and RF classification models were employed for the classification. SVM classifier was employed on the dataset with different kernels wherein, SVM with 'poly' kernel was found to yield better classification accuracy amongst all the SVM models employed. The classification accuracy improved to 71%, when 'poly' kernel was applied on the dataset with "Flash ON" mode, as shown in Fig 13. Later, application of RF further improves the classification accuracy to 82% in "Flash ON" mode on the testing data. Fig.14(a) shows the confusion matrix for "Without Flash" dataset using RF, where three classes; '3', '5', and '7'; were predicted correctly with class support size greater than 20. With "Flash ON", as shown in Fig. 14(b); 4 out of ten classes; '3', '4', '5' and '8'; were predicted correctly with class support size greater than 20. Similar to 1st experiment, RF works with the maximum efficiency even for "Without Flash" dataset. Comparison of the accuracy achieved using

RF and SVM classifiers is given in Fig.13. In addition, other evaluation metrics calculated to study the performance of different classifiers are provided in Tables V and VI. Evaluation results using RF on "Without Flash" data are shown in Table V. Precision, recall and F-measure were calculated for individual class for RF yielding the weighted average of 0.90 precision, 0.90 recall and 0.90 F-number.

Further, to evaluate the performance of our approach, RF model was first trained on the dataset corresponding to the lighting condition at 3500K color temperature and its testing accuracy was calculated. It yielded good classification accuracy when dataset was split into 70-30 ratio as shown in Fig.13. Further, this trained model was fed with the data of selected feature sets, corresponding to the lighting condition at 6500K color temperature, totally unknown to the model. This yielded the highest accuracy for 'l, a, b, h' feature set in the "Flash ON" mode as shown in Fig 15.

B. Discussion

In most of the studies related to health conditions, urinalysis are paid less attention in comparison to blood analysis. However, protein in urine, which is dominated by the presence of albumin, carries a high-risk factor. Presence of albumin in urine in a very small quantity (microalbuminuria condition) can be an early sign of CKD. It can also be associated with the high risk of cardiovascular diseases and certain type of cancer [33][34]. The dipstick method can play a major role in early screening of albumin in urine to avoid CKD complications. In general practice, dipsticks are not sensitive enough for microalbuminuria (<300 mg/dL) which is most significant for early disease diagnosis of CKD. Hence, we used ten different concentrations which belongs to normal (7.1825 mg/dL, 15 mg/dL and 30 mg/dL), microalbuminuria (62 mg/dL, 150 mg/dL and 250 mg/dL), and macroalbuminuria (500 mg/dL, 1000 mg/dL, 2000 mg/dL and 4000 mg/dL) to have more significant diagnosis of CKD and its stages. In this study, images of commercially available urine strips were captured using smartphone under different illumination conditions. The strips were illuminated at different incident angles using incandescent and Fluorescent light sources with color temperatures of 3500K and 6500 K respectively. This helps to understand the behavior of different color spaces with change in albumin concentration under different lighting conditions. The distance between strip and the smartphones were kept constant (10 cm), normal to the urine strip using a platform to avoid any variation in the illumination at strip. Also, the shadow casted by the smartphone on the strip and background (white paper) is different as shown in Fig.1 (a, c, e). The variation in shadow is due to the different position of the camera on the smartphone. T. Kong *et al* had proposed the use of flash light to mitigate ambient light condition for microfluidic assays [28]. Although, the Flash light of smartphone decreases the effect of ambient light, applicability of Flash light has its own drawback as Flash light follows a Lambertian profile which will have maximum intensity at the center and decreases in gradient fashion thereafter. Here, Lambertian effect was avoided by placing the microfluidic assay device at the place with maximum intensity. However, it is difficult for the user to locate it on a white background. This constraint has been done away with in the present study as selected features were least affected by the position of the strip. Here, the strip was placed randomly and imaged using different smartphones kept at a fixed distance of 10 cm.

The results obtained in the "Flash ON" mode, highlight the fact that the usage of Flashlight noticeably affects the classification accuracy. Effect of Flash was clearly observed in Experiment I, where use of Flashlight decreased the intra-class variability and increased inter-class variation. In the second study (Experiment II), it nearly nullified the effect of angle of incidence onto the strip. In similar study, M.E. Solmaz, *et al* had used SVM and RF for the classification of hydrogen peroxide test strips using RGB, HSV and LAB color values as features. This yielded 100% cross validation accuracy for SVM and 87% for RF [27]. However, when the same SVM model was used as an application software by the user, the accuracy decreased to 87%. Two primary reasons for relatively poor performance of the classification models were given as i) change in the ambient light condition, and ii) variations in the crop size chosen by the user with respect to the laboratory settings. It should be noted that the crop size was chosen uniformly for the training set. In our study, these factors, which drastically change the classification accuracy, were taken care of, by using the Flash light of the smartphone. This mitigates the effect of ambient light sources. In general, the effective illumination of Flash of smartphone is around 700-800 lux, which can effectively decrease the effect of ambient light illumination as illumination in general vary between 50 lux (family living room) to 500 lux (day light). Also, as shown in Fig.2 (b, d, f), it effectively compensates the effect of shadow on paper due to different smartphones, which was not considered in previous studies [17][22]. In addition, the crop size of albumin sensor part was randomly chosen, as user can crop the sensor part in different pixel sizes.

Further, the study of different features suggests that 'l, a, b' of Lab color space and 'h' feature of HSV color space were least affected by the variation of smartphone and variable lighting conditions. Further, this hypothesis was experimentally verified by classifying all the concentration values by choosing different feature sets; "All features", 'l,a,b,h' , 'h,a,b' , 'h,b', 'h', which shows that classification model yielded the highest accuracy for 'l,a,b,h' feature sets. In previous studies, A. Mathaweesansurn *et al* used 'h' feature for the classification of albumin [17]. Application of 'l, a, b, h' feature set in this study further improved the accuracy in general and for inter-phone repeatability in particular.

RF classified most of the classes in both the modes; "Flash ON" and "Without Flash" (92% and 82% respectively for constant and for variable lighting condition), whereas for

SVM, the accuracy was lower than RF (81% for "Flash ON" data and 72% for "Without Flash" data) as given in Fig. 9 and 15. In particular, logistic regression performs very poor for most of the classes. Furthermore, as can be seen in the confusion matrix of RF classifier, prediction of correct class was lesser for higher concentrations (>500 mg/dl). The poor behavior of classification models for these classes can be explained by the fact that there is high variability in color values for these classes and they do not change uniformly with concentration. Specifically, the distribution of data points in these classes are nonlinear, hence linear models; such as logistic regression which act as linear model in multiclass classification for certain range of values, exhibited poor performance and was able to classify the concentration values in only two classes; class '6' and '8'. However, nonlinear models such as SVM with 'Rbf' and 'Poly' kernels improved the accuracy upto 60%. The optimization of two parameters using grid search further increased the accuracy as it provided the best parameters for optimization. On the contrary, RF is a bagging decision tree model, which splits the whole data into small subsets according to best features. At each split, instead of taking whole set of features, it took only small random subset of features. Thus, some strong features are selected from the entire pool of datasets. This was taken care of by RF as it outperforms other employed classifiers.

From clinical point of view, detection of lower concentrations with high precision and accuracy, belonging to normal and microalbuminuria range, is more significant. Thus, higher concentration values (>500 mg/dl) can be neglected. Our method was tested on standard albumin concentration which is major part of protein in urine. However, further investigation is needed to investigate its efficacy on urine samples. The proposed approach was able to classify different concentration of albumin by using smartphone in variable illumination conditions.

The prime motivation of the present study is to understand the feasibility of an automatic smartphone based, PoC system to quantify different concentrations of albumin in urine sample. This will enable more accurate detection and quantification of albumin at lower concentrations (Normal and micro albuminuria), clinically very significant in the early diagnosis of kidney disease. In future, the effect of positional and angular variation of the smartphone with respect to the strip, may also be analyzed and optimized. Furthermore, a 3D printed fixture to hold the smartphone, can be used to perform the smooth measurements without changing any positional and angular parameters. Moreover, as machine learning models become more robust and generalized with large dataset, its practical applicability in clinical settings can be enhanced by dealing with a larger dataset. Also, to make this system more universally acceptable, the spectral sensitivity of smartphone cameras needs to be addressed with more smartphone models.

IV. CONCLUSION

In this work, a smartphone-based quantification of urinary albumin using commercially available dipsticks, was performed at lower concentration. This is clinically very important for early detection of CKD and cardiovascular disease. Images of chemically impregnated urine strip were acquired in ambient conditions without the use of any accessory to mitigate the effect of ambient light conditions. However, even at constant illumination and color temperature, ambient light conditions and shadow of smartphone significantly change the color values. With change in smartphone model, shadow region of the smartphone on surface changes so thus the intensity value of strip. Also, the camera settings of smartphones vary with respective models and Manufacturers. To account for shadow effect, light source of smartphone was used with "Flash ON". The variation of results in color is due to different camera settings and different spectral sensitivities among different smart phone cameras under constant/similar illumination condition. The intensity distribution is different for all three smart phones. It shows high variability in average intensity values of smart phones in different color spaces. Quantification of urine protein was done by extracting color features in RGB, HSV and LAB color spaces. Best feature set; 'l, a, b, h' was selected by analyzing the effect of smartphone and lighting variation. Further, variable lighting conditions were introduced to observe the robustness of our approach. Machine learning algorithms were employed to classify different concentrations. RF classification algorithm in conjunction with camera mode "Flash ON", yields an accuracy of 92% and 82% in constant and variable lighting conditions respectively. The Confusion Matrix/Error Matrix allows visualization of performance of algorithm. The proposed method can be effectively utilized for quantification of urinary albumin concentrations in low resource PoC settings.

REFERENCES

[1] T. K. Chen, D. H. Knicely, and M. E. Grams, "Chronic Kidney Disease Diagnosis and Management: A Review," *JAMA - Journal of the American Medical Association*, vol. 322, no. 13. American Medical Association, pp. 1294–1304, 01-Oct-2019.

[2] S. Varughese and G. Abraham, "Chronic kidney disease in India: A clarion call for change," *Clinical Journal of the American Society of Nephrology*, vol. 13, no. 5. American Society of Nephrology, pp. 802–804, 07-May-2018.

[3] P. Cockwell and L. A. Fisher, "The global burden of chronic kidney disease," *The Lancet*, vol. 395, no. 10225. Lancet Publishing Group, pp. 662–664, 29-Feb-2020.

[4] P. E. De Jong and G. C. Curhan, "Screening, Monitoring, and Treatment of Albuminuria: Public Health Perspectives," *J Am Soc Nephrol*, vol. 17,

pp. 2120–2126, 2006.

[5] D. Kumar and D. Banerjee, "Methods of albumin estimation in clinical biochemistry: Past, present, and future," *Clin. Chim. Acta*, vol. 469, no. August 2016, pp. 150–160, 2017.

[6] J. Il Hong and B.-Y. Chang, "Lab on a Chip Development of the smartphone-based colorimetry for multi-analyte sensing arrays †," vol. 14, p. 1725, 2014.

[7] S. C. Kim, U. M. Jalal, S. B. Im, S. Ko, and J. S. Shim, "A smartphone-based optical platform for colorimetric analysis of microfluidic device," *Sensors Actuators, B Chem.*, vol. 239, pp. 52–59, 2017.

[8] S. Lee, D. O'Dell, J. Hohenstein, S. Colt, S. Mehta, and D. Erickson, "NutriPhone: A mobile platform for low-cost point-of-care quantification of Vitamin B 12 concentrations," *Sci. Rep.*, vol. 6, no. March, pp. 1–8, 2016.

[9] D. J. You, T. S. Park, and J. Y. Yoon, "Cell-phone-based measurement of TSH using Mie scatter optimized lateral flow assays," *Biosens. Bioelectron.*, vol. 40, no. 1, pp. 180–185, Feb. 2013.

[10] V. Oncescu, M. Mancuso, and D. Erickson, "Cholesterol testing on a smartphone," *Lab Chip*, vol. 14, no. 4, pp. 759–763, 2014.

[11] A. Carrio, C. Sampedro, J. L. Sanchez-Lopez, M. Pimienta, and P. Campoy, "Automated low-cost smartphone-based lateral flow saliva test reader for drugs-of-abuse detection," *Sensors (Switzerland)*, vol. 15, no. 11, pp. 29569–29593, 2015.

[12] R. Yang *et al.*, "Color Space Transformation-Based Smartphone Algorithm for Colorimetric Urinalysis," *ACS Omega*, vol. 3, no. 9, pp. 12141–12146, Sep. 2018.

[13] R. Thakur *et al.*, "Development of smartphone based lateral flow device for the quantification of LH and E3G hormones," *IEEE Sens. J.*, pp. 1–1, Jul. 2020.

[14] F. Giavazzi *et al.*, "A fast and simple label-free immunoassay based on a smartphone," *Biosens. Bioelectron.*, vol. 58, pp. 395–402, 2014.

[15] B. Berg *et al.*, "Cellphone-Based Hand-Held Microplate Reader for Point-of-Care Testing of Enzyme-Linked Immunosorbent Assays," *ACS Nano*, vol. 9, no. 8, pp. 7857–7866, 2015.

[16] A. F. Coskun, R. Nagi, K. Sadeghi, S. Phillips, and A. Ozcan, "Albumin testing in urine using a smart-phone."

[17] A. Mathaweesansurn, N. Maneerat, and N. Choengchan, "Sensors and Actuators B : Chemical A mobile phone-based analyzer for quantitative determination of urinary albumin using self-calibration approach," *Sensors Actuators B. Chem.*, vol. 242, pp. 476–483, 2017.

[18] S. Akraa *et al.*, "AC," *J. Netw. Comput. Appl.*, 2018.

[19] L. Shen, J. A. Hagen, and I. Papautsky, "Point-of-care colorimetric detection with a smartphone," *Lab Chip*, vol. 12, no. 21, pp. 4240–4243, Nov. 2012.

[20] H. Karlsen and T. Dong, "Smartphone-Based Rapid Screening of Urinary Biomarkers," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 2, pp. 455–463, 2017.

[21] K. R. Konnaiyan, S. Cheemalapati, M. Gubanov, and A. Pyayt, "MHealth Dipstick Analyzer for Monitoring of Pregnancy Complications," *IEEE Sens. J.*, vol. 17, no. 22, pp. 7311–7316, Nov. 2017.

[22] A. Y. Mutlu, V. Kiliç, G. K. Özdemir, A. Bayram, N. Horzum, and M. E. Solmaz, "Smartphone-based colorimetric detection: Via machine learning," *Analyst*, vol. 142, no. 13, pp. 2434–2441, 2017.

[23] M. Ra, M. S. Muhammad, C. Lim, S. Han, C. Jung, and W. Kim, "Smartphone-Based Point-of-Care Urinalysis Under Variable Illumination," vol. 6, no. October 2017, 2018.

[24] M. E. Solmaz, A. Y. Mutlu, G. Alankus, V. Kılıc, A. Bayram, and N. Horzum, "Sensors and Actuators B : Chemical Quantifying colorimetric tests using a smartphone app based on machine learning classifiers," vol. 255, pp. 1967–1973, 2018.

[25] E. V. Woodburn, K. D. Long, and B. T. Cunningham, "Analysis of Paper-Based Colorimetric Assays with a Smartphone Spectrometer," *IEEE Sens. J.*, vol. 19, no. 2, pp. 508–514, Jan. 2019.

[26] B. Coleman, C. Coarsey, and W. Asghar, "Cell phone based colorimetric analysis for point-of-care settings," *Analyst*, vol. 144, no. 6, pp. 1935–1947, 2019.

[27] T. Kong *et al.*, "Accessory-free quantitative smartphone imaging of colorimetric paper-based assays," *Lab Chip*, vol. 19, no. 11, pp. 1991–1999, 2019.

[28] C. P. Wen, Y. C. Yang, M. K. Tsai, and S. F. Wen, "Urine dipstick to detect trace proteinuria: An underused tool for an underappreciated risk marker," *American Journal of Kidney Diseases*, vol. 58, no. 1. W.B. Saunders, pp. 1–3, 2011.

[29] S. L. White, R. Yu, J. C. Craig, K. R. Polkinghorne, R. C. Atkins, and S. J. Chadban, "Diagnostic accuracy of urine dipsticks for detection of albuminuria in the general community," *Am. J. Kidney Dis.*, vol. 58, no. 1, pp. 19–28, Jul. 2011.

[30] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, 1974.

[31] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A Practical Guide to Support Vector Classification," 2003.

[32] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[33] L. Jørgensen, I. Heuch, T. Jenssen, and B. K. Jacobsen, "Association of albuminuria and cancer incidence," *J. Am. Soc. Nephrol.*, vol. 19, no. 5, pp. 992–998, May 2008.

[34] L. Jørgensen, T. Jenssen, I. Heuch, and B. K. Jacobsen, "The combined effect of albuminuria and inflammation on all-cause and cardiovascular mortality in nondiabetic persons.," *J. Intern. Med.*, vol. 264, no. 5, pp. 493–501, Nov. 2008.