Reprinted from
Journal of Imaging Science and Technology® 65(6): 060407-1–060407-15, 2021.

Smartphones' Skin Colour Reproduction Analysis for Neonatal Jaundice Detection

Mekides Assefa Abebe▲, Jon Yngve Hardeberg▲, and Gunnar Vartdal
Colour and Visual Computing Laboratory, Picterus AS; Gjøvik, Norway
E-mail: mekides.abebe@ntnu.no

Abstract. *In recent years, smartphone-based colour imaging systems are being increasingly used for Neonatal jaundice detection applications. These systems are based on the estimation of bilirubin concentration levels that correlates with newborns' skin colour images corresponding to total serum bilirubin (TSB) and transcutaneous bilirubinometry (TcB) measurements. However, the colour reproduction capacity of smartphone cameras are known to be influenced by various factors including the technological and acquisition process variabilities. To make an accurate bilirubin estimation, irrespective of the type of smartphone and illumination conditions used to capture the newborns' skin images, an inclusive and complete model, or data set, which can represent all the possible real world acquisitions scenarios needs to be utilized. Due to various challenges in generating such a model or a data set, some solutions tend towards the application of reduced data set (designed for reference conditions and devices only) and colour correction systems (for the transformation of other smartphone skin images to the reference space). Such approaches will make the bilirubin estimation methods highly dependent on the accuracy of their employed colour correction systems, and the capability of reducing device-to-device colour reproduction variability. However, the state-of-the-art methods with similar methodologies were only evaluated and validated on a single smartphone camera. The vulnerability of the systems in making an incorrect jaundice diagnosis can only be shown with a thorough investigation of the colour reproduction variability for extended number of smartphones and illumination conditions. Accordingly, this work presents and discuss the results of such broad investigation, including the evaluation of seven smartphone cameras, ten light sources, and three different colour correction approaches. The overall results show statistically significant colour differences among devices, even after colour correction applications, and that further analysis on clinically significance of such differences is required for skin colour based jaundice diagnosis.* © *2021 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2021.65.6.060407]

1. INTRODUCTION

Advances in smartphone camera technologies has led to smartphone-based imaging being utilized for many skin related applications. The need for low-cost, portable and remote medical diagnostics solutions has made smartphone based imaging a potential alternative for skin related medical applications [1]. For example, early detection of skin based cancers such as melanoma can be greatly facilitated by smartphone cameras [2–5]. Remote screening and monitoring of skin lesions such as urticaria, aczema, psoriasis, infantile hemangioma and other conditions with photographs were also shown to provide additional information for physicians and further reinforce their diagnosis [6].

The colour accuracy of smartphone-based imaging for teledermoscopy systems is usually limited by illumination condition, camera resolution, magnification or zoom, camera dynamic range, and related others. To help resolve such issues, most teledermoscopy systems additionally utilize Dermatoscopes: microscopic device with magnification power, integrated LED light sources and polarizers. Dermatoscopes, attached on top of the camera lens, have proven to enhance the colour accuracy and visualization of skin lesions [4, 5, 7]. Computer vision techniques such as camera calibration as well as machine learning based lesions detection and classification are then applied to help enhance and perform preliminary/screening diagnosis.

In addition to the skin related diseases, smartphone-based colour imaging systems are being introduced for neonatal jaundice detection applications. Neonatal jaundice, with higher levels of bilirubin accumulation, is known to cause permanent brain damage. It is also indicated as the reason for more than 100,000 infants annual worldwide deaths, mostly in sub-Saharan Africa and south Asia [8]. Diagnosis of neonatal jaundice is professionally done by measuring total serum bilirubin (TSB) and transcutaneous bilirubinometry (TcB) with specialized devices and laboratory equipment, which are usually expensive and require invasive tests. To help avoid painful medical tests for the newborns and greatly reduce the medical instrument costs, an increasing interest towards smartphones-based solutions is being observed.

One such jaundice detection systems was launched by the Norwegian company, Picterus AS. Anders et al. [8] proposed bilirubin levels estimation method for new born babies from colour-calibrated smartphone images. Their approach uses a mathematical skin reflectance model, according to the bio-optics of newborns, for simulation of various newborns' skin reflectance. The resulted reflectance is then used to create the system's database of *RGB* colour versus bilirubin concentration value pairs. With the database, the Picterus smartphone-based application is applied to capture the newborn babies' skin colour, apply colour correction, and estimate the corresponding bilirubin level. More explanation on the method [8] is given in a separate section, Bilirubin Estimation Method, below.

▲ IS&T Members.

Received Aug. 9, 2021; accepted for publication Nov. 9, 2021; published online Dec. 7, 2021. Associate Editor: Chang-Hwan Son.

1062-3701/2021/65(6)/060407/15/$25.00

As known in most colour imaging applications, smartphone-based colour reproduction quality suffers from many challenges that are inherent to the applied camera technologies and acquisition processes [9]. Colour inconstancy is one of these common challenges [9, 10]. Colour of stimuli captured with a particular camera and illumination condition will not remain the same if reacquired with another illumination condition. Smartphone cameras also have their own distinct device spaces (with a certain dynamic range and colour gamut), which is determined by the sensor, colour filter array and optical component technologies used to create the camera [11].

Therefore, to perform accurate skin colour and machine learning based bilirubin level predictions, all illumination conditions and smartphone devices (which are going to possibly be used in real application scenarios) must be represented in the data set. Generating such data set will be very challenging, considering time and storage constraints, the amount of smartphone types and their variants, as well as the diversity of possible acquisition conditions. The common work around to the colour inconsistency and device-to-device variability challenge, in most colour imaging state-of-the-art works as well as the Picterus system, is utilizing colour constancy and colour correction algorithms. The Picterus system, for example, only setup the database for one reference camera and illumination condition, and it applies real time colour correction algorithm to calibrate the skin colours (captured by an unknown end users' smartphone) to the reference space. Over the years, various colour correction algorithms have been introduced for various colour reproduction applications. The proposed approaches can be classified as model based (which utilize perceptual appearance models [12–14]) and empirical (which are purely data oriented and predictive methods [8]) approaches.

The accuracy of such jaundice level prediction systems strongly rely on their colour correction algorithms to generate similar colours as the smartphone cameras and illumination conditions, represented in their data set. Despite the smartphone camera and illumination setup used for the acquisition, the colour difference of a newborn skin colour should remain negligible compared to its captured colours with the reference camera and illumination condition. A slight discrepancy of colour, in this regard, may lead to wrong jaundice diagnosis. Consequently, a thorough investigation of the colour inconsistency of individual smartphone cameras (under varied illumination conditions) as well as the device-to-device variabilities (even after colour correction) is much needed for such applications.

In this work, we present an in-depth analysis of smartphone-based skin colour reproduction differences for the Picterus jaundice monitoring application. The variability of skin colour values among various smartphone types as well as many light sources is presented. For the colour correction purposes, three different colour correction methods (selected from both model based and empirical approaches) are also evaluated. The results of the analysis showed the dominant performances of empirical and machine learning based approaches for such applications. The presence of a statistically significant colour differences among devices, even after the application of the best performing colour correction algorithm, is also observed. We believe that the overall discussions of this work can provide additional and important insights towards the current challenges corresponding to the process of smartphone-based skin colour reproduction applications (in particular bilirubin level estimation systems) and will be a valuable resource for researchers in the field.

In the following section, the state-of-the-art colour correction and bilirubin estimation methods will be briefly discussed. Afterwards, the description of the analyzed Picterus system, the evaluation methodology as well as the evaluation results and discussions will be presented. Finally, the limitations of our study and future work ideas will be given in the conclusion section.

2. RELATED WORK

As stated before, the skin colour of a newborn differs when captured by different smartphone cameras or under different illumination conditions. For accurate and efficient modeling as well as accurate prediction of bilirubin concentration from the *RGB* values of newborns' skin images, the colour consistency of the smartphone-based imaging process should be greatly enhanced, and the device-to-device differences need to be reduced. In this regard, many calibrations and colour correction techniques for accurate colour constancy and colour reproduction applications have been introduced over the years. In this section, a summary of the state-of-the-art techniques, which have been utilized in colour imaging (in general) and with respect to jaundice monitoring systems, is presented.

2.1 *Calibration and Colour Correction Methods*

The main purpose of camera calibration is to determine the relationship between input scene radiance and camera response [9, 15, 16]. As the ISO standards for the calibration and characterizations of Digital Still Cameras (DSCs) [15, 16] recommend, the characterization process can be model-based or empirical. Model based characterization requires knowledge of the colour matching functions, correlation statistics of scene data, device spectral sensitivities, and related others. Due to the wide ranges of scenes which could be encountered in real world application, different assumptions such as the maximum ignorance assumption are used to simplify the calibration process and the derivation of transformation matrices. The device spectral sensitivities are also not usually provided by the manufacturers and have to be estimated from reference patch measurements (as in empirical approaches) or measured with proper monochromatic light-based laboratory setups (for model-based cases) [9, 16].

The model-based approaches, in most practical applications, utilize a matrix-based characterization to linearly transform the device dependent values into CIE tristimulus or other device independent spaces. This matrix based

calibration is based on the Luther-Ives condition which assumes that the spectral sensitivities of all colorimetric cameras are linearly related to the human visual system's Colour matching functions [14, 17, 18]. However, such assumption may not hold for most current cameras and the matrix-based calibration is just an approximation [14]. Also, the resulting transformation matrix is usually optimized for a given reference viewing condition, which further makes the matrix to be dependent on conditions such as the illumination and reference white.

Hence, to elevate the influences of illuminations and reference whites on colour reproduction, most characterization applications apply additional white balancing methods. The white balancing process further transforms the colours, taking the destination reference white point and illumination into account. In most colour reproduction solutions such as Colour appearance models, the simulation of the human visual system's chromatic adaptation process is used to achieve better Colour constancy. Human visual models such as von Kries, Bradford, Sharp - based on sharpened sensors, CAT02 - optimized for minimizing CIELAB differences, and related others have been applied for such chromatic adaptation transform applications [13].

The availability of well-equipped and properly established laboratories to measure and design accurate model-based systems is not always guaranteed, due to financial and infrastructural constraints. Hence, many real time applications perform colour correction and calibration with the absence of measured camera spectral sensitivities, following various empirical approaches. The empirical approaches usually estimate the corresponding colours (for the desired destination viewing conditions) based on mathematical or machine learning models, which are trained on a data set containing many sample colour transformation pairs [9, 19]. Regression methods such as linear least-squares and weighted least-squares regression, generalized polynomial transform, Thin-Plate Spline interpolation, and related others are mostly utilized in this regard. An increasing color constancy work is also, recently, observed using deep learning approached for better accuracy [20, 21]. According to past studies [8, 19], model-based techniques were often outperformed by such empirical techniques as long as the application is known and only intended for limited medium and illuminants. Enough training data should also be available for better predictive accuracy. However, the model-based strategy, being oblivious to scene statistics, was generally robust across different illumination conditions.

2.2 *Bilirubin Estimation Methods*

Smartphone-based colour imaging techniques, along with the discussed colour correction methods, have been adopted for creating affordable and non-invasive jaundice monitoring system for newborns. One recent study on smartphone based device independent colorimetric measurements [12] proposed general colour constancy method, which utilizes illumination subtraction technique using image captured with flash and no-flash settings. To achieve device invariability, the method follows the common colour management workflow, using the CIE tristimulus *XYZ* colour space as their device connection space (DCS) [22]. Even if the method follows a model-based approach (since most camera and viewing information remains unknown) proper and more accurate chromatic adaptation models were not utilized. When it comes to methods proposed for Neonatal jaundice detection application, however, almost all of the few smartphone-based solutions are designed with empirical solutions [8, 23–27].

Mustafa Aydin et al. [23] used kNN (k-Nearest Neighbor) and SVR (Support Vector Regression) algorithms on the data set, containing 40 image and bilirubin data of newborns with jaundice 40 without, to fit and predict bilirubin levels. For generating the image data set, images of 80 newborn babies were taken with printed 8 colour calibration card and colour (average *RGB*, YC_bC_r and *Lab* values) and colour gradient (average Sobel filter results in *RGB* space) features were extracted. However, the normalization of the images with the average *RGB* value will not represent a proper white-balancing solution, as claimed by the authors. Their results shown in the referenced paper [23] signaled the failures of the simple normalization approach from removing the effects of illumination, specular reflections, and shadows. The BiliCam application [25], detailed later, enhanced both the white balancing and regression limitations of the Mustafa Aydin et al.'s method. The white balancing approach was replaced by using the white calibration patch-based normalization and the machine learning regression is improved by increasing the number of features (by extracting from two flash and without flash images of the same sample) and reducing the dimensionality (with principal component analysis - PCA). Their bilirubin level estimation process was also made to be performed with 5 different (both from linear and non-linear) regression algorithms and averaging the results in such a way that their bilirubin level predictions could avoiding false negatives and potential undetected risks.

Recent studies validate the correlation of skin colour with the bilirubin levels of the newborn. A study on healthy Caucasian newborns reported that only the green and blue channel of the dermatoscope assisted smartphone images were showing a significant correlation with the bilirubin measurements [24]. According to similar study by Swarna et al. [26] on evaluating the reliability of jaundice estimation using the smartphone application BiliCam on 35 neonates, good correlation between the BiliCam estimation and the serum bilirubin levels were presented, particularly for the chest area images. Correlation of skin colour with bilirubin levels has also been corroborated by similar studies performed by Picterus [8, 28]. An in-depth description of the Picterus system is provided in a later section.

Most of the studies conducted for designing efficient jaundice monitoring system were only using and validating their proposed applications on one or two smartphone. The number of light sources used in each study were also very limited. However, as stated before, the smartphones' technological differences as well as the acquisition condition

variations create colour inconstancy and device-to-device variability. Such variations, unless fully represented in the data set, are likely to lead to misdiagnosis. Therefore, to perform accurate predictions at all times (despite the device, illumination, and human skin colour variations), more complex modeling or generating an inclusive data set becomes necessary. Some solutions, such as the one proposed by Picterus system, introduced alternative approach that allows prediction systems to use reduced data set representing only reference camera and viewing conditions (to which all the other unknown camera captures can be transformed to). Such methods require additional colour correction transformation which further strengthen the need for verification on various smartphones and illumination conditions. Accordingly, this work performs a thorough evaluation of colour reproduction differences for seven smartphone camera types together with ten different light sources. The goal of the study is to find out if the colour differences of the various skin images of the evaluated cameras (captured under the evaluated illuminations and colour corrected by various colour correction algorithms) will be negligible as anticipated.

3. PICTERUS BILIRUBIN ESTIMATION METHOD

Like previously discussed smartphone-based solutions [25], Anders et al. proposed smartphone images and colour correction solutions based, bilirubin levels estimation method for newborns. Their approach uses a mathematical skin reflectance model and calculates a level of bilirubin concentration based on an optical diffusion model of skin, or Monte Carlo simulations of skin optics [8, 28]. Therefore, to use *RGB* sensor values of smartphone cameras for the estimation of the bilirubin levels, one must find the appropriate mapping to the corresponding numerical skin parameters which will result in similar skin colour. Such mapping can be achieved through regression or lookup table-based approaches [28]. Currently, for computational efficiency purposes and reduced data set, a lookup table is only created for Nikon camera and a halogen-based illuminant data. Hence, the Nikon camera spectral sensitivities (Nikon Ref., shown in Figure 5) and spectral power distribution of the halogen illuminant (Picterus Ref. A, shown in Figure 2) will be used as a reference for the calibration of other smartphone data. For stronger colour correction accuracy, the authors recommend to use their own spectrally printed calibration card (which includes around 24 reflectances of the numerically simulated skins and grey patches) to calibrate skin colours captured with other smartphone cameras. The measured reflectance of the calibration card patches are later used with reference camera sensitivities (Nikon Ref.) and illumination (Picterus Ref. A) to empirically correct the captured skin colour using a Gaussian process regression method and the corresponding bilirubin concentration is retrieved from the numerical model lookup table. The overall workflow of the Picterus system is illustrated in Figure 1.

The Picterus bilirubin level estimation system is previously evaluated through cross-sectional prospective study,

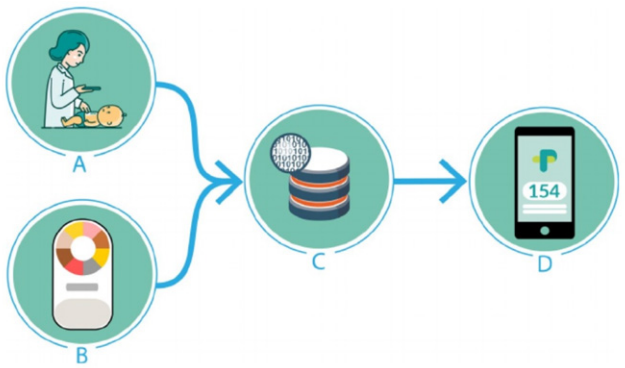

Figure 1. General workflow of Picterus bilirubin estimation system (Taken from [8]). a) Raw image of the newborn skin or sclera with the calibration card is taken. b) The colour of the calibration card with the reference conditions will be used to calibrate the skin or sclera colour. c) The corresponding/closer colour to the calibrated colour is selected from the simulation data base. d) The bilirubin concentration of the selected pair is display as the final estimation.

conducted at two different Norwegian hospitals [8]. The study was performed on 302 newborn infants (from which 76 of them had severe jaundice) and the correlation between image estimations and the results of TSB and TcB was investigated. According to the results, the image-based system had a good correlation ($Pearson's r > 0.83$) to the TSB and TcB for all Caucasian cases and significantly lower correlation ($Pearson's r < 0.75$) for non-Caucasian cases. However, the only smartphone used in the study was Samsung Galaxy S7. Similarly, the images were acquired under illumination conditions with only 46-watt halogen lamps and under the room ambient lighting conditions.

As discussed in the earlier sections, the colour reproduction quality of smartphone cameras differs and depends on various factors. Therefore, the accuracy of the Picterus bilirubin estimation system depends on the colour constancy as well as device-to-device variability of the calibration system used to adapt smartphones' data to the references. As a result, the analysis and evaluation of different smartphone cameras' performances under wider types of illumination conditions becomes highly relevant. Accordingly, in this work, we intend to evaluate the colour correction (colour consistency) variability among different smartphone cameras (including Samsung Galaxy S7); with a consideration that lower variability will lead to similar bilirubin estimation accuracy. In addition to the Gaussian process regression method used by the Picterus system, the performance of other selected calibration methods (both empirical and model based) will also be assessed in terms of increasing colour constancy and reducing device-to-device variability of skin colour reproduction.

4. EVALUATION METHODOLOGY

For the intended study of device-to-device skin colour reproduction variability, a total of six smartphone cameras (listed in Table I) two studio light sources, which are available at our colour laboratory (colourlab - CL studio *A* and *D*50)

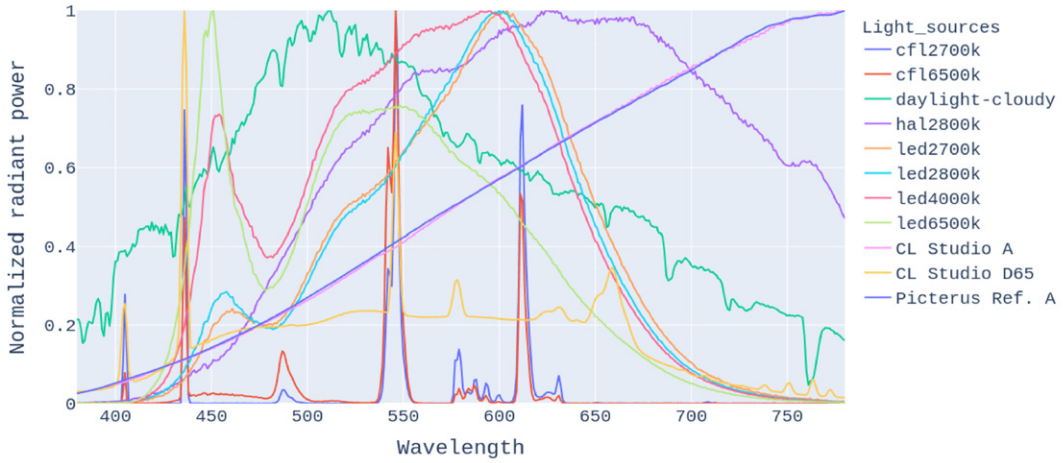

Figure 2. Spectral power distributions of the evaluated light sources. ''CL studio A'' and ''CL studio D50'' are the light sources used to capture the colour targets in colourlab's studio, and hence used in our captured sensor values based evaluation. ''Picterus Ref. A'' is the reference light source used to generate the Picterus's *RGB* versus bilirubin concentration pairs data set. All the other light sources are used for our simulation-based evaluations.

Table I. Evaluated Smartphone Cameras and the corresponding ISO settings used for camera raw captures. Various ranges of shutter speeds were used, for each camera, to capture different exposure *RGB* values. However, all the camera sensor values used to generate the results of this work are acquired using auto shutter speed setting. Note: the ID provided in this table will be the ones used when referring to the corresponding phones in the rest of this document.

ID	Phone	Lens	ISO
H1	Huawei P30 Pro	5.6 mm	100
H2	Huawei P20 Pro	5.6 mm	100
S1	Samsung Galaxy S7	4.2 mm	100
S2	Samsung Galaxy S10+	4.3 mm	50
R1	Redmi Note 9 Pro	Micro 1.9 mm	100
R2	Redmi Note 9 Pro	Wide 5.4 mm	100
A1	Apple iPhone SE	4 mm	100

and around eight other measured light sources (given in Fig. 2) are utilized. Also, for calibration, colour correction, and evaluation purposes, the *RGB* captured values as well as spectral reflectance of all the patches of the three different colour targets (shown in Figure 6) were collected. In order to illustrate the workflow of our evaluation process, a flow chart is generated and shown in Figure 3.

As shown at the right end side of the work flow, Fig. 3, the proposed study will compare colour corrected *RGB* values of the captured skin colour test patches (with the seven smartphone cameras) as well as their simulated *RGB* values (using the patches' measured spectral reflectance and the cameras' calibrated spectral sensitivities) with respect to their corresponding *RGB* values computed with the reference condition and reference camera. The process of model-based camera sensor *RGB* simulation, the studio acquisition process of the raw *RGB* values, as well as the generation of the Picterus reference *RGB* values are briefly explained in the following subsections. As depicted in the flow chart, Fig. 3, the simulated and measured *RGB* values of the test skin patches will be first colour corrected to the Picterus reference device space, prior to their comparison to the Picterus reference *RGB* values. Accordingly, a total of three colour correction methods (discussed in the Evaluated colour Correction Methods section) were investigated for this purpose.

4.1 *Studio Acquisition and Preprocessing*

The captured raw *RGB* values of all the seven smartphone cameras used in the evaluation are acquired by capturing three colour targets, shown in Fig. 6, in camera raw mode. All smartphones (listed in Table I) provide camera raw *RGB* values in Digital Negative (DNG) file format [29]. Therefore, from the DNG files of all our captures, the demosaiced linear and 16-bits *RGB* values were extracted (no white balancing applied) with the help of the Libraw image decoder library and the python wraper API, rawpy [30, 31]. As shown in the Table I, the acquisition were performed with fixed ISO setting (mostly at 100) and using auto shutter speed settings. The linearity of the cameras, in raw mode, is also verified by examining the opto-electronic conversion functions (OECFs) of the cameras according to the ISO standard guidelines [15].

However, during our acquisition processes, we have noticed illumination non-uniformity from our studio light sources (used from the Judge II Gretagmacbeth colour booth). Such non-uniformity is very common in most light sources if an appropriate diffuser is not used [32]. Therefore, after 16-bits images are extracted, we have applied a spatial non-uniformity correction based on uniform white colour target and dark image captures of all evaluated phones under the same exposure settings. Based on other related works [33, 34], given the raw *RGB* values of the white D_W and black D_B images' pixels and their averages over the whole images' area ($\overline{D_W}$ and $\overline{D_B}$), the *RGB* values of the colour target images (D_I) can be corrected according to Eq. (1). Sample results of

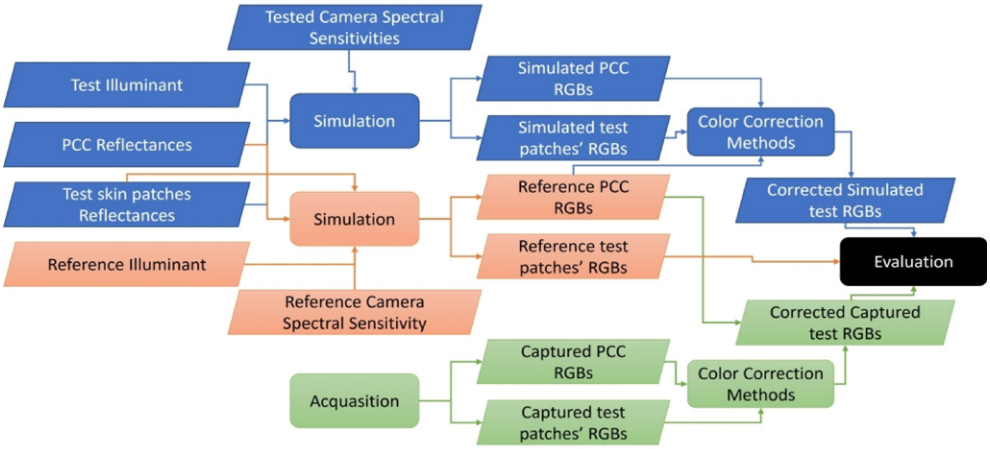

Figure 3. The work flow illustration of the process that we have followed to generate our evaluation data. The *RGB* values of the captured as well as simulated colour target patches will be first extracted or generated. The extracted values are then calibrated to the reference conditions and evaluated against the reference *RGB* values.

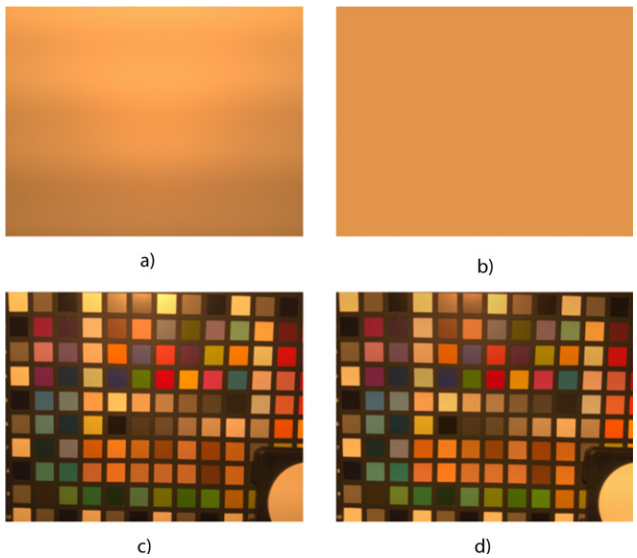

Figure 4. X-Rite white balance uniform grey (a) and colour checker digital SG (c) semi-gloss colour targets. The images in the left column are the raw images captured with the Samsung S7 (S1) camera on auto-exposure setting and under illuminant ''CL studio A''. The corresponding right column images are the results of the spatial illumination uniformity correction. The raw images were read in 16-bits linear DNG format but a normalization to 8-bit and gamma of 2.2 is applied here, for visualization purposes. It should be noted that no white balancing is applied to these images.

this spatial uniformity correction are presented in Figure 4. The corrected 16-bits images ($D_I^{'}$) will then be used to extract the average *RGB* values of individual colour patches of each captured colour targets:

$$D_I^{'} = \frac{(\overline{D_W} - \overline{D_B})(D_I - D_B)}{D_W - D_B}. \quad (1)$$

4.2 *Camera Characterization and Capturing Simulation*

In addition to real sensor value acquisitions, our study also assessed simulated camera sensor values of colour targets, for all smartphone cameras and ten light sources described earlier. The camera raw response, D_{S_i}, model for linear DSCs, in the i^{th} spectral channel, is mathematically expressed with Eq. (2), where C_i, P_r and L represent the spectral sensor sensitivity of the DSC, the spectral reflectance of the colour target patches and the spectral incident radiance power measurements of the light sources in the sensitive spectral region of the device($[\lambda_{\min}, \lambda_{\max}]$) [33, 35–37]. In our application, for smartphone cameras, i ranges from 1 to 3 representing the *RGB* colour channels. The constant K can be computed according to the exposure time used during the acquisition or can be empirically computed from estimated and real captured *RGB* sensor values of some grey patches. In our prior experimentation with multiple exposure data, we were able to observe that the K value for accurate simulation of real sensor values varies greatly with exposure time and it should be computed accordingly:

$$D_{S_i} = K \int_{\lambda_{\min}}^{\lambda_{\max}} C_i(\lambda) P_r(\lambda) L(\lambda) \Delta\lambda. \quad (2)$$

Since none of the manufacturers of the evaluated smartphones provided their corresponding camera response functions and spectral sensitivities, we needed to characterize each phone camera in our laboratory. To measure the spectral sensor sensitivities of the smartphone cameras, we have performed colour characterization following the ISO standard, ISO-17321 [16]. Following the recommendations from the standard, we used a Bentam TCM300 monochromator (capable of producing monochromatic lights in wider spectral range: 200 nm–3 μm) as a source of narrow band illuminations. To achieve narrower bandwidth and stronger radiance power, we replaced the Bentham QTH light source with COB LED based light source. This light source is assembled by colour lab colleague using 30W LED with 5000 K colour temperature and wavelength range from 380 nm to 800 nm (of which majority radiance is concentrated in the visible range: 400 nm–720 nm). The

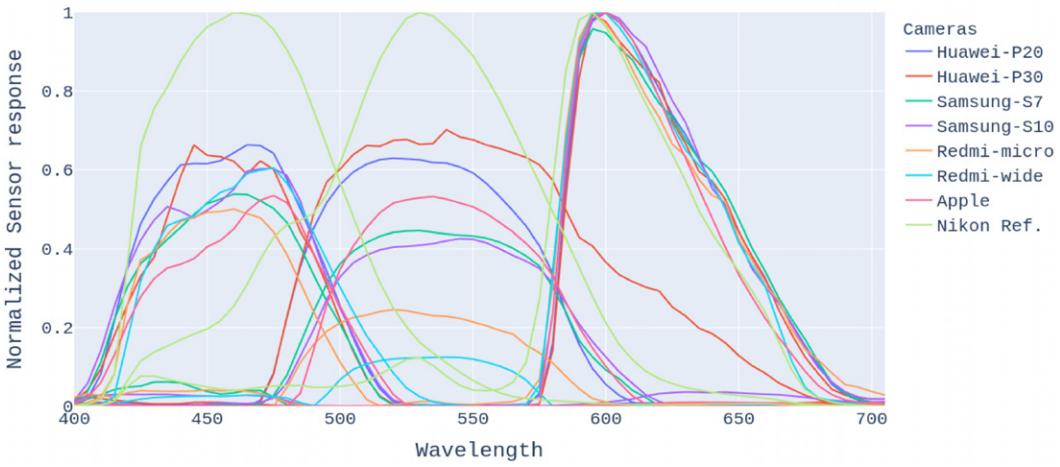

Figure 5. Final camera spectral sensitivities of all characterized smartphone cameras.

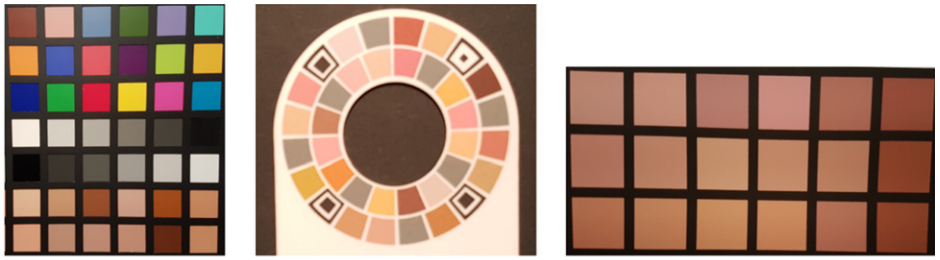

Figure 6. Colour targets used for calibration, colour correction and test skin patches. Fig. (a) is a section of the Colour checker digital SG target containing only the patches with in raw $2-8$ and column E to J. Fig. (b) is a calibration card prepared by Pictures with reflectance selected from their simulated skin data set. Note that the black background is not the part of the calibration card. Fig. (c), on the other hand, is a collection of test skin colour patches generated with simulation of various levels of bilirubin concentrations.

resulted spectral sensor sensitivities of all characterized cameras are shown in Fig. 5.

Conversely, the reflectance of the colour target patches, the studio light source spectral radiance powers and the monochromator spectral radiances are all measured using Konika Minolta Spectroradiometer, model CS-2000. The measured reflectance of the patches of the three colour targets are shown in Figure 7.

According to Andrew et al. [14], the actual raw values obtained in practice are quantized values, modeled by taking the integer part of Eq. (2), and it is useful to subsequently normalize the simulated D_S and captured D_I values to the range $[0, 1]$ by dividing them with the raw clipping point (2^16 in 16-bits). With this normalization it can be considered that the reference white of the camera raw space is unity vector, Eq. (3). As it can be seen from the plots, Fig. 2, the spectral power of the light sources is also normalized to 1 for more comparable results:

$$\begin{bmatrix} D_{S_R} \\ D_{S_G} \\ D_{S_B} \end{bmatrix} = \begin{bmatrix} D_{I_R} \\ D_{I_G} \\ D_{I_B} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{3}$$

4.3 *Evaluated Colour Correction Methods*

As explained previously, the bilirubin estimation of the Picterus system is performed based on their data set of *RGB* versus bilirubin concentration pairs. These paired *RGB* values are created based on their skin simulation spectra, reference camera spectral sensor sensitivity (given as "Nikon Ref." in Fig. 5), and illuminant A (given as "Picterus Ref. A" in Fig. 2), according to Eq. (2). Therefore, for the accurate bilirubin estimation of skin colours captured by other smartphone cameras, the captured *RGB* values must be appropriately colour corrected and transformed to the reference device space. In this regard, we have decided to try and assess the colour reproduction differences among three selected and state-of-the-art colour correction methods. The selected methods include a perceptual model based chromatic adaptation transform, CAT02, and two other empirical (machine learning based) colour correction approaches, Root Polynomial and Gaussian Process Regressions. A brief description of the methods are provided as follows.

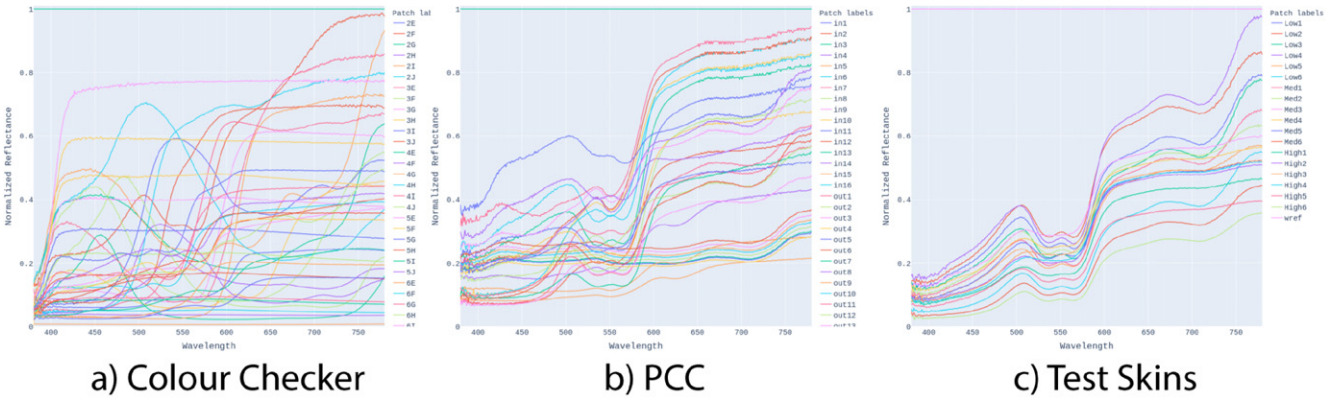

Figure 7. The measured spectral reflectance of the colour target patches, shown in Fig. 6. All the measurements are performed using Konika Minolta Spectroradiometer, model CS-2000.

4.3.1 *Chromatic Adaptation Transforms*

Chromatic adaptation refers to the capability of the human visual system to adjust the responsivity of its colour mechanisms and preserve the appearance of a given stimuli, which is presented under varying illumination conditions. Chromatic adaptation models are perceptual models which are designed to predict corresponding colour (for a given viewing condition) of a particular stimuli colour, which is observed under another viewing condition [38]. As given in Eq. (4), to accurately model the physiological mechanisms of chromatic adaptation, most chromatic adaptation models works in the cone excitation *LMS* space. The models predict the adjusted cone signals $L_aM_aS_a$ from a given initial cone signal $L_iM_iS_i$ and that of the adapting white stimulus $L_{\text{white}}M_{\text{white}}S_{\text{white}}$, according to some function f:

$$L_aM_aS_a = f(L_iM_iS_i, L_{\text{white}}M_{\text{white}}S_{\text{white}}, \ldots). \tag{4}$$

Cone excitation *LMS* values are usually approximated by a 3×3 matrix linear transformations of the corresponding CIE tristimulus *XYZ* values. This chromatic adaptation can then be inverted to compute the corresponding colour in another adapting viewing condition. The combined process of such forward and inverse transformation is generally referred as chromatic adaptation transform (CAT) [38]. Over the years, several types of chromatic adaptation transforms have been introduced [39, 40]. Among them, we have chosen the most commonly used CAT02 as a representative for the perceptual model-based colour correction approaches.

CAT02: After extensive experimental evaluation of various linear CATs , the CIE TC 8-01 recommended CAT02, which is based on a matrix optimized to a wide variety of corresponding colours [38]. To transform the *RGB* test colours to that of the reference *RGB* space, CAT02 is utilized as in Eq. (5):

$$\begin{bmatrix} R_{\text{ref}} \\ G_{\text{ref}} \\ B_{\text{ref}} \end{bmatrix} = M^{-1}_{xyz_{\text{ref}}} M^{-1}_{\text{CAT02}} A_W M_{\text{CAT02}} M_{xyz_{\text{test}}} \begin{bmatrix} R_{\text{test}} \\ G_{\text{test}} \\ B_{\text{test}} \end{bmatrix}, \tag{5}$$

where,

$$M_{\text{CAT02}} = \begin{bmatrix} 0.4002 & 0.7076 & 0.0808 \\ -0.2263 & 1.1653 & 0.0457 \\ 0.0 & 0.0 & 0.9182 \end{bmatrix} \tag{6}$$

and

$$A_W = \begin{bmatrix} \frac{L_{\text{white}_{\text{ref}}}}{L_{\text{white}_{\text{test}}}} & 0.0 & 0.0 \\ 0.0 & \frac{M_{\text{white}_{\text{ref}}}}{M_{\text{white}_{\text{test}}}} & 0.0 \\ 0.0 & 0.0 & \frac{S_{\text{white}_{\text{ref}}}}{S_{\text{white}_{\text{test}}}} \end{bmatrix}. \tag{7}$$

The matrix used to convert a device *RGB* values to the device independent CIE *XYZ* space $M_{xyz_{\text{test}}}$ is generated, for all the tested and reference smartphone cameras, through individual camera characterization process (outlined in ISO-17321, [16]). The resulted matrices are then optimized (using least squares regression) for all tested illuminants (Fig. 2). For the calibration of the cameras with the two studio light sources, the measured data of the colour Checker SG colour target and the CIE1931 2° observer colour matching functions [9, 14] were used. The adapting cone excitation values, on the other hand, are computed from the adapting whites of the test as well as reference scenes, according to Eq. (8) and Eq. (9):

$$\begin{bmatrix} L_{\text{white}_{\text{ref}}} \\ M_{\text{white}_{\text{ref}}} \\ S_{\text{white}_{\text{ref}}} \end{bmatrix} = M_{\text{CAT02}} M_{xyz_{\text{ref}}} \begin{bmatrix} R_{\text{white}_{\text{ref}}} \\ G_{\text{white}_{\text{ref}}} \\ B_{\text{white}_{\text{ref}}} \end{bmatrix} \tag{8}$$

Table II. The mean and standard deviation of RMSE and ΔE of the colour correction results for captured sensor values of the Skin test target patches. The results are given for the three calibration approaches, and they are averaged over the light source and acquisition modes.

Camera	Correction	RMSE		ΔE	
		Mean	σ	Mean	σ
A1	CAT	0.123	0.098	8.942	2.962
	GP	0.045	0.038	5.701	2.048
	RP	0.045	0.041	5.584	2.072
	Org	0.115	0.093	6.848	2.377
H2	CAT	0.215	0.105	7.487	2.540
	GP	0.090	0.063	6.721	2.608
	RP	0.093	0.066	4.796	1.697
	Org	0.205	0.107	7.329	2.167
H1	CAT	0.244	0.112	20.507	24.237
	GP	0.092	0.057	7.992	3.126
	RP	0.103	0.061	6.034	2.841
	Org	0.237	0.113	9.019	2.896
R1	CAT	0.172	0.103	18.781	5.723
	GP	0.109	0.110	12.856	8.020
	RP	0.118	0.108	11.029	4.721
	Org	0.230	0.093	11.075	3.708
R2	CAT	0.117	0.073	18.945	7.498
	GP	0.206	0.181	18.868	13.873
	RP	0.199	0.177	16.857	8.197
	Org	0.219	0.063	9.378	3.640
S2	CAT	0.091	0.069	11.142	3.068
	GP	0.052	0.030	8.722	4.812
	RP	0.058	0.031	6.720	3.550
	Org	0.142	0.077	8.735	3.010
S1	CAT	0.085	0.062	11.856	3.535
	GP	0.051	0.037	7.084	3.606
	RP	0.051	0.033	5.416	1.999
	Org	0.139	0.083	8.761	2.762

$$\begin{bmatrix} L_{\text{white}_{\text{test}}} \\ M_{\text{white}_{\text{test}}} \\ S_{\text{white}_{\text{test}}} \end{bmatrix} = M_{\text{CAT02}} M_{xyz_{\text{test}}} \begin{bmatrix} R_{\text{white}_{\text{test}}} \\ G_{\text{white}_{\text{test}}} \\ B_{\text{white}_{\text{test}}} \end{bmatrix}. \quad (9)$$

4.3.2 *Root-Polynomial Regression*

In addition to using the underlying physical model for defining the forward and inverse transformations of colours from different domains, empirical approaches can also be used to perform data fitting or interpolation. In colour imaging, several mathematical and machine learning techniques have been used for colour correction applications [9, 41]. In this study, we have chosen the Root-Polynomial Regression method, proposed by Finlayson et al. [42], to represent the empirical approaches in our evaluation. In their study, the method showed better performance in terms of computational efficiency and preserving colour appearance at different exposure levels, compared to the polynomial regression (Which in turn is known to have superior colour constancy performances than other linear interpolation techniques [28]).

Root-Polynomial regression (RP) is another form of least-squares fitting wherein the characterization function (usually 3×3 matrix) is approximated by a polynomial, as shown in Eq. (10). The root polynomial terms of the training colours P_{train} contains $1xm$ polynomial terms (example is shown in Eq. (11) for 3^{rd} degree polynomial) and the corresponding weight matrix $\boldsymbol{A}$ contains $mx3$ polynomial weights to be optimized, Eq. (12):

$$RGB_{\text{ref}} = P_{\text{train}} A, \quad (10)$$

$$P_{\text{train}} = \begin{bmatrix} r & g & b & \sqrt{rg} & \dots & \sqrt[3]{rg^2} & \dots & \sqrt[3]{rgb} \end{bmatrix}, \quad (11)$$

$$P_{\text{train}} = \begin{bmatrix} w_{R_{\text{ref}},0} & w_{G_{\text{ref}},0} & w_{B_{\text{ref}},0} \\ w_{R_{\text{ref}},1} & w_{G_{\text{ref}},1} & w_{B_{\text{ref}},1} \\ \vdots & \vdots & \vdots \\ w_{R_{\text{ref}},m} & w_{G_{\text{ref}},m} & w_{B_{\text{ref}},m} \end{bmatrix}. \quad (12)$$

The *RGB* components of the training data set are replaced by their polynomial expansion in order to cast the polynomial regression problem to that of least-squares problem. If we collect all N training samples and form the $Nx3$ captured RGB_{ref} and their Nxm polynomial terms P_{train}, then the polynomial weight parameters can be optimized according to Moore-Penrose inverse [9], as given in Eq. (13). For more detailed explanation of the method, the readers are recommended to refer to the original paper [42]:

$$A = (P_{\text{train}}^t P_{\text{train}})^{-1} P_{\text{train}}^t RGB_{\text{ref}}. \quad (13)$$

4.3.3 *Gaussian Process Regression*

The picterus system, described in the earlier sections, also employs empirical approach to calibrate and colour correct the *RGB* captures of smartphone cameras to that of the reference device (Nikon Ref.) space, which is used to generate their bilirubin data set. After pre-processing of the captured raw colour correction target, Fig. 4, along with the skin colours for illumination uniformity correction and other degradation enhancements, the system applies Gaussian Process Regression to colour correct and transform the skin colour values to the reference device space [28].

Gaussian Process Regression (GP), is a nonlinear regression using gaussian processes and Bayesian approach [43]. The method has a capability of capturing a wide variety of relations between the input (test *RGB* in our case) values and outputs (reference *RGB*) by utilizing a theoretically infinite number of parameters (using gaussian processes) and letting the data determine the level of complexity through the means of Bayesian inference. The mathematical details of gaussian process regression requires quite lengthy explanations of the gaussian processes as well as Bayesian inferences, which an interested reader can find in the provided references [43, 44]. However, for generating the proposed evaluation results of the gaussian process regression-based colour

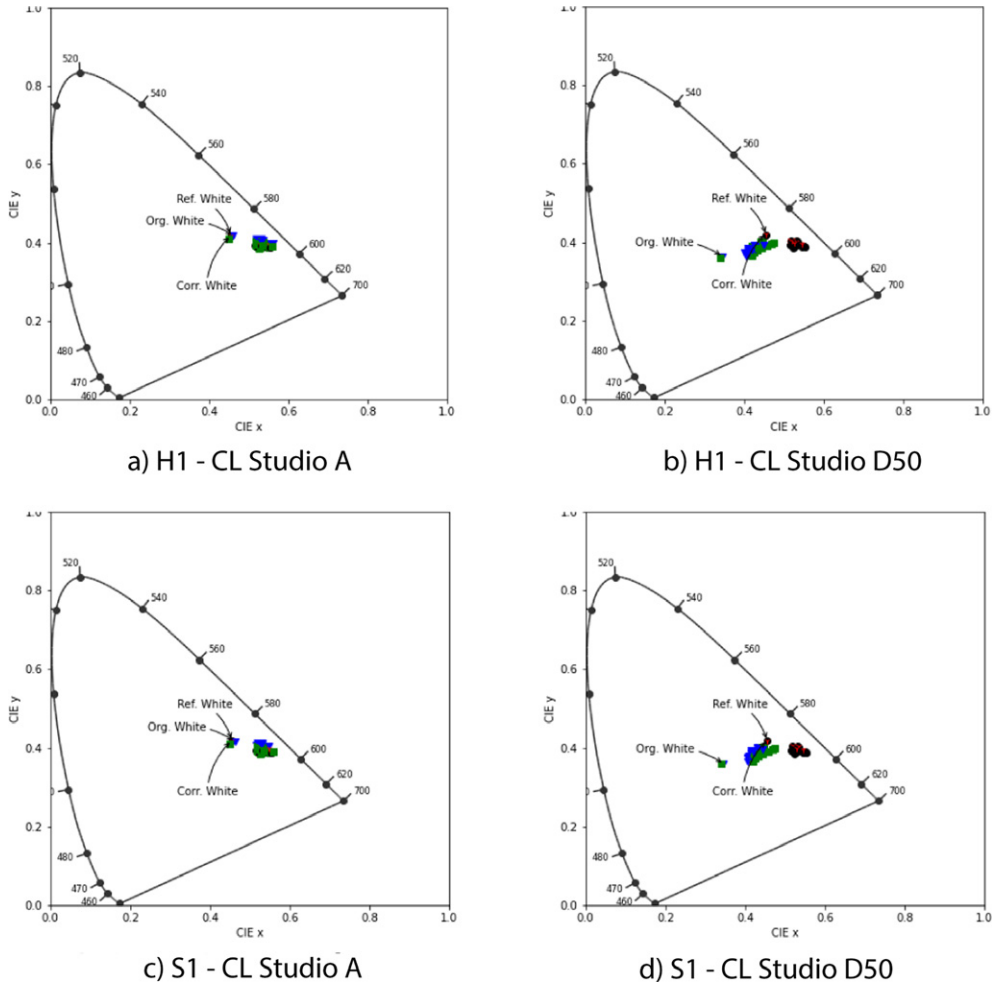

Figure 8. Chromaticity values of the skin test patches before and after Gaussian processes transformation. The *xy* values of the patches computed from the reference camera (in black circles), test cameras (H1 and S1 in blue triangle), and the corrected results of the Gaussian processes (in red) are provided in the CIE1931 chromaticity diagram. The corresponding measured *xy* values of the patches with the spectroradiometer are also shown with green squares.

correction results, we have applied the scikit-learn python implementation of the algorithm proposed by Rasmussen et al. [44]. The smartphone sensor values of the Picterus calibration card, Fig. 6, were used to train the regression model with the default initial kernel and alpha parameter settings.

5. COLOUR CORRECTION EVALUATION RESULTS AND DISCUSSION

As described in the previous section, Fig. 3, we have generated both captured and simulated *RGB* sensor values for all the seven smartphone cameras under different light sources. In this section, the colour correction results of all collected data with three evaluated colour correction methods (discussed in the previous section) will be presented and evaluated in both captured and simulated scenarios. The colour reproduction variability analysis among evaluated smartphone devices will also be given and the acceptability of using one reference camera for bilirubin estimation (as used in the Picterus system) will be discussed accordingly.

5.1 *Colour Correction Evaluation with Captured Sensor Values*

All the raw *RGB* sensor values of the Skin test colour target, shown in Fig. 6, patches (resulted from the acquisition process described in the Studio acquisition and pre-processing section) for all evaluated cameras and under two studio light sources (shown as "CL Studio A" and "and CL Studio D65" in Fig. 2) are colour corrected to the reference camera space (Nikon with A illuminant). The three colour correction approaches together with the captured and measured Picterus calibration target reference data (for training the two regression models) were employed. The reference raw *RGB* sensor values of the same Skin test patches are also computed with the reference Nikon spectral sensor sensitivities and measured Picterus reference A illuminant. Finally, the colour difference between the colour correction results of all the smartphone sensor values and reference sensor values are computed, both in the linear *RGB* space (as *RMSE* values) and the perceptual *CIELab* colour space (as ΔE values).

To transform the device dependent sensor values of all the cameras to the device independent CIE chromaticity *XYZ* values, all the smartphone cameras were characterized and their corresponding 3×3 transformation matrix is generated. Prior to computing colour differences, each matrix is optimized for each evaluated light source based on the colour Checker DSG colour target patches' (shown in Fig. 6) measured data, as explained in the ISO-17321 standard [9, 16]. The *CIELab* values used to compute the CIEΔE differences are computed using the Konika Minolta reference white patch reflectance (shown as "wref" in Figure 7.) as a reference white [38].

The final colour difference results are summarized and given in Table II and III. The values presented in Table II show the average colour differences of the evaluated colour correction methods, averaged over the evaluated studio light sources and acquisition modes (with flash or without flash). The results show that the empirical approaches lead to closer colours to the references rather than the chromatic adaptation transform method. This inference holds for all the evaluated smartphone cameras and colour difference matrices. As described in the related sections, machine learning based approaches performs better when the application domain and conditions are fully known and properly represented in the training data set, as it is the case in our evaluation. The skin test colour target patches, used for our testing, and the Picterus calibration card patches, used for training the models, have a similar skin colour distribution. However, the colour differences of the regression models' results may go higher if tested with colour patches which are completely out of the training data distribution.

Therefore, due to their better performances in our application, the rest of our analysis will only be using the results of the two regression algorithms (RP and GP). Accordingly, the average colour differences of the smartphones for the two studio light sources as well as acquisition modes (averaged over the two regression based colour correction methods) is provided in Table III. Additionally, Figure 8 also shows the CIE tristimulus visualizations of sample results (for Huawei P30 and Samsung Galaxy S7 smartphones). As it can be seen from the figure, the tristimulus values of both phones before and after corrections (for the studio light CL A) are similar to that of the reference and measured tristimulus values. This is mainly due to the similarity of the studio light CL A and the reference light source, Picterus Ref. A. However, as it can be seen from the results of the two smartphones under the CL D50 light sources, the colours become closer to the reference values only after the application of the colour correction methods. Also, as mentioned previously, for a camera to be colorimetric and satisfy the Luther condition, the 3×3 calibration matrix is optimized with the CIE1931 2° observer colour matching function [16]. As the result, all the calibration matrices resulted from our characterizations of all the cameras can also be transforming their corresponding *RGB* values to approximate locations in the CIE *Yxy* space. On the other hand, according to the results shown in Table III, the effects of the flash on the enhancement of colour reproduction various depending on the type of smartphone and used light sources. As shown in one of the state-of-the-art bilirubin estimation methods [25], the flash and no-flash acquisition modes can be applied to expand the training image features and enhance prediction accuracy.

Table III. The mean and standard deviation of RMSE and ΔE of the colour correction results for captured sensor values of the Skin test target patches. Only, the results of the best performing calibration methods (the Gaussian process and Root polynomial, as shown in Table II) are used for computing averages for the two studio light source measurements. Note: the phrase "MA" and "MD50" represent the CL studio A and CL studio D50 light sources, respectively.

Camera	Light	Mode	RMSE		ΔE	
			Mean	σ	Mean	σ
A1	MA	Flash	0.0289	0.016	5.1	1.764
		No_flash	0.0812	0.05	6.269	2.3782
	MD50	Flash	0.0526	0.026	6.394	1.9804
		No_flash	0.0182	0.024	4.806	1.5893
H2	MA	Flash	0.026	0.021	5.748	2.1373
		No_flash	0.0898	0.057	4.787	1.8111
	MD50	Flash	0.1207	0.038	6.789	2.9739
		No_flash	0.1292	0.071	5.71	2.1626
H1	MA	Flash	0.1088	0.035	8.711	3.1471
		No_flash	0.096	0.058	6.635	3.3323
	MD50	Flash	0.0549	0.026	5.141	1.8196
		No_flash	0.1302	0.076	7.564	2.9593
R1	MA	Flash	0.1918	0.108	14.3	4.8224
		No_flash	0.19	0.104	14.77	5.5936
	MD50	Flash	0.0455	0.031	12.12	8.1951
		No_flash	0.0262	0.014	6.584	3.7188
R2	MA	Flash	0.357	0.096	25.17	9.2487
		No_flash	0.3712	0.105	26.22	10.569
	MD50	Flash	0.0252	0.02	8.53	5.2656
		No_flash	0.0565	0.035	11.53	6.9494
S2	MA	Flash	0.0511	0.023	7.101	2.2015
		No_flash	0.0776	0.033	13.12	4.5386
	MD50	Flash	0.0639	0.024	4.671	2.0752
		No_flash	0.0274	0.018	5.992	1.9728
S1	MA	Flash	0.0186	0.018	5.306	1.5408
		No_flash	0.0872	0.031	8.546	4.5033
	MD50	Flash	0.0433	0.023	5.393	1.721
		No_flash	0.0539	0.026	5.755	2.0905

5.2 *Colour Correction Evaluations with Simulated Sensor Values*

At times, the available studio light sources can be limited in number. Even with high number of available light sources, measuring and capturing colour targets for all the possible smartphone and light source combinations

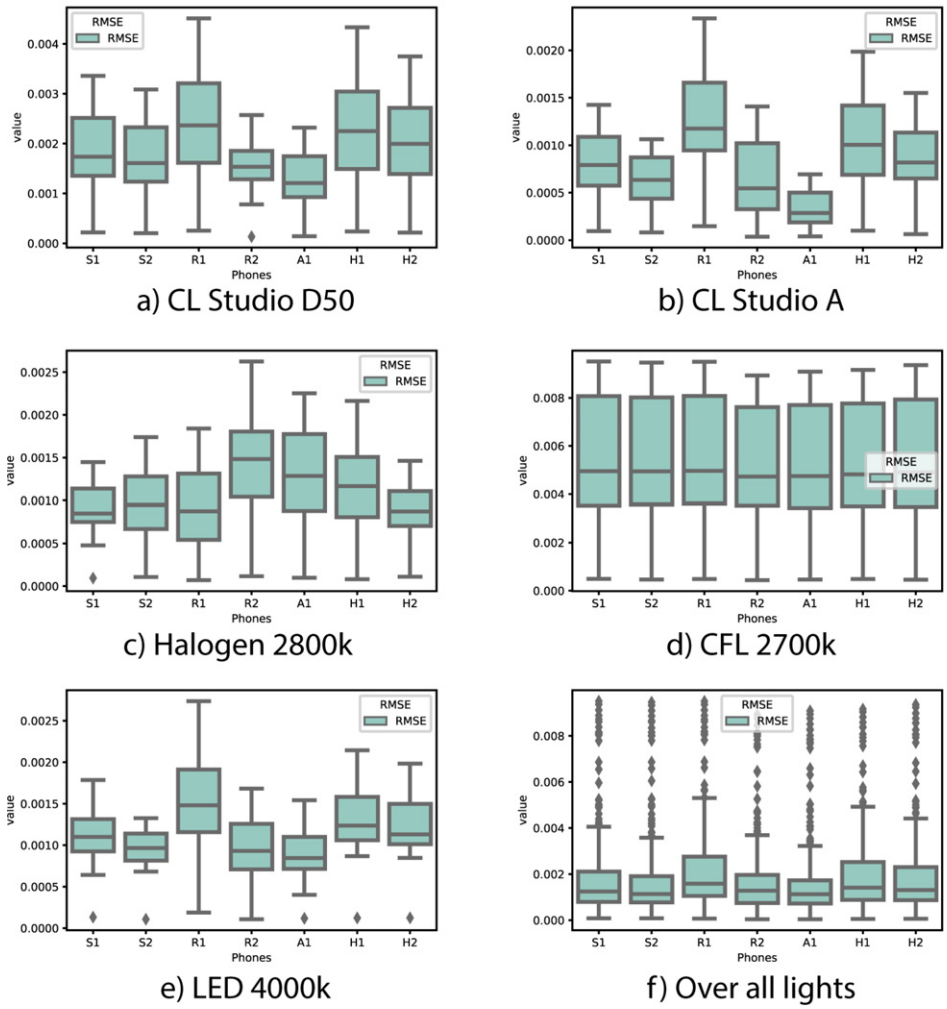

Figure 9. RMSE difference distribution for different camera capture simulations under different light sources.

requires great effort and may take a long time. Therefore, for effective evaluation and study purposes, the application of the camera sensor value simulation approaches (described in the previous Camera Characterization and Capturing Simulation section) can be a good idea. In this case, to be able to assess the colour reproduction variability among the evaluated smartphones for a wider range of illumination conditions, we have measured the sensor spectral sensitivities of the smartphone cameras and simulated their sensor *RGB* results, accordingly.

After the simulation for all the reference, training, as well as test skin colour data, the results of the simulated smartphones sensor values using the ten illuminations were colour corrected, using the two regression approaches. The average colour differences between the corrected smartphones' sensor simulation results and the reference sensor values are computed for all evaluated scenarios, similarly with the captured sensor values evaluation. The distributions of the colour difference for some selected light sources as well as over the whole ten light sources is provided as box plots, given in Figure 9, and 10. Just by looking at the distributions of the *RMSE* and ΔE of all the evaluated smartphones, an evident device-to-device difference is visible.

Consequently, in order to verify if the noticed device-to-device difference is significant, a one-way ANOVA (ANalysis Of VAriance) analysis is performed. ANOVA uses variance-based F test to check the group mean equality, by testing the non-specific null hypothesis i.e. all group means are equal. The resulting F and p values of the ANOVA analysis for the selected light sources and over the entire sets of light sources are given in Table IV. The p values obtained from both the *RMSE* and ΔE analysis are significant ($p < 0.05$), except for the results of compact fluorescent lamp (CFL). To have further insight onto which pairs of the smartphones have significantly different colour reproduction results, we additionally performed multiple pairwise comparison (post hoc comparison) analysis using Tukey's honestly significant difference (HSD) tests. The results of the HSD tests for both colour difference matrices are visualized in Figure 11.

According to the various device-to-device colour reproduction difference results, shown so far, it is obvious

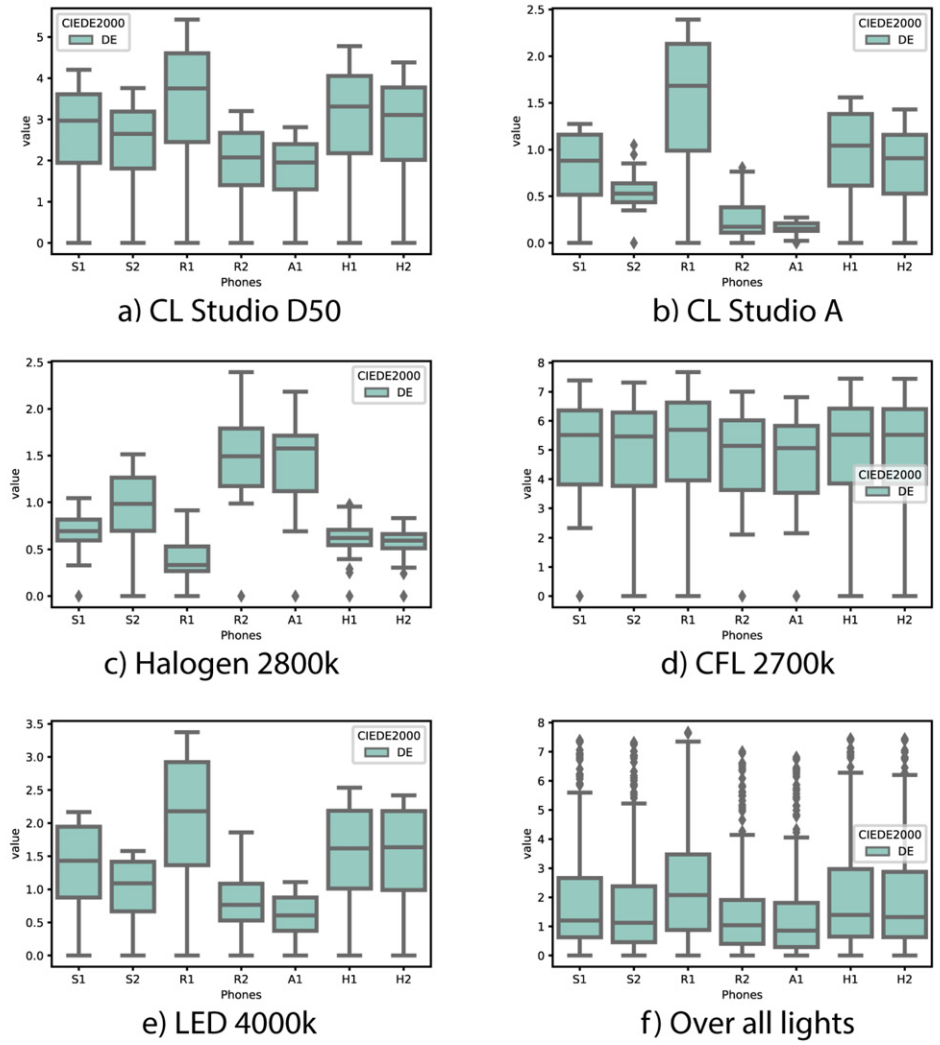

Figure 10. ΔE difference distribution for the different camera capture simulations for some selected light sources.

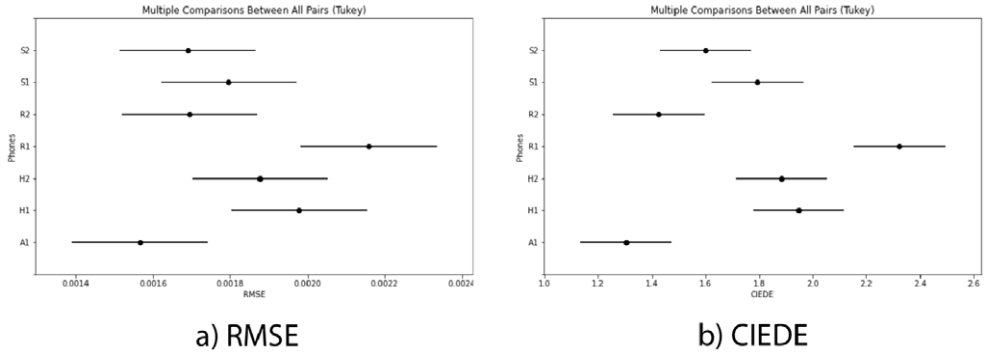

Figure 11. POST-HOC testing results with Tukey Honestly Significant Difference(HSD) method.

that there is significant difference of corrected colour results of different smartphone cameras. Among all the evaluated phone, Apple and Redmi-micro phones resulted in the most significantly smaller and larger colour differences, respectively. Such results shows that, since the difference between the two phones is higher, the bilirubin estimation results can also differ greatly. Therefore, unless such device variabilities are handled with proper colour correction systems in general, jaundice detection system based on colour calibration could always lead to dangerous misdiagnosis.

Table IV. One-way ANOVA analysis results for evaluating the significance of the difference among the corrected values of the evaluated smartphone cameras.

	RMSE		DE	
Light_source	F	*p*_value	F	*p*_value
CL studuio A	24.12763	1.21E-22	44.765	2.18E-37
CL studuio D50	8.889243	7.93E-09	8.5862	1.60E-08
CFL 2700k	0.070653	0.9986184	0.4488	0.845528
Hal 2800k	10.43215	2.32E-10	54.129	5.29E-43
LED 4000k	14.53775	2.70E-14	17.44	6.07E-17
LED 2800k	11.47066	2.24E-11	37.832	7.66E-33
Over all lights	5.627539	0.000008	17.871	2.06E-20

6. CONCLUSION AND FUTURE WORK

Of late, the number of smartphone-based solutions for monitoring jaundice levels in newborns are increasing. Since most of the proposed solutions depend on the colour reproduction qualities of smartphone cameras, the proposed work presents a thorough analysis of device-to-device colour reproduction variability. The study includes the results of the investigations for many smartphone cameras, light sources as well as colour correction algorithms. The results showed a significant variability among devices even after colour correction with one of the best regression algorithms.

In addition to device variability, it was observed that many of the smartphone-based bilirubin solutions are highly prone to image degradation during image acquisition process, inaccuracies in simulations of skin colour reflectance and generation of representative calibration reference targets. Moreover, the amount as well as intersectionality limitations of existing skin color measurement and simulation data sets further hinders the accuracy of the diagnosis. Most datasets introduce bias towards Caucasian skin tones and higher misdiagnoses rate could be occur in dark skinned newborns.

Therefore, as a future work, increasing use of real and inclusive measured skin data sets, with various illumination condition as well as smartphone devices, can lead to increased performances. Skin reflectance data set, bilirubin measurements and mathematical models should also be more inclusive to different shades of skin colours, including the ones with higher melanin concentration. With more representative data set, advanced machine learning methods, such as deep learning, can be applied to alleviate most of the limitations of the existing jaundice monitoring solutions. Convolutional autoencoders as well as other appropriate networks can be utilized for more accurate colour corrections, preprocessing and quality enhancement, as well as bilirubin level estimation purposes.

REFERENCES

[1] S. Majumder and M. J. Deen, "Smartphone sensors for health monitoring and diagnosis," Sensors **19** (2019).

[2] R. J. Friedman, D. S. Rigel, and A. W. Kopf, "Early detection of malignant melanoma: The role of physician examination and self-examination of the skin," CA: A Cancer J. Clinicians **35**, 130–151 (1985).

[3] T. Wadhawan, N. Situ, K. Lancaster, X. Yuan, and G. Zouridakis, "Skinscan©: A portable library for melanoma detection on handheld devices," *2011 IEEE Int'l. Symposium on Biomedical Imaging: From Nano to Macro* (IEEE, Piscataway, NJ, 2011), pp. 133–136.

[4] B. Cugmas and E. Štruc, "Accuracy of an affordable smartphone-based teledermoscopy system for color measurements in canine skin," Sensors **20** (2020) (Basel, Switzerland).

[5] S. M. H. Rizvi, T. Schopf, A. Sangha, K. Ulvin, and P. Gjersvik, "Teledermatology in norway using a mobile phone app," PLoS One **15**, 1–5 (2020).

[6] T. Hubiche, L. Valério, F. Boralevi, E. Mahe, C. B. Skandalis, and A. Phan, "P. del Giudice for the Research Group of the French Society of Pediatric Dermatology (Groupe de Recherche de la Société Française de Dermatologie Pédiatrique), "Visualization of Patients' Skin Lesions on Their Smartphones: A New Step During Dermatology Visits"," JAMA Dermatol. **152**, 95–97 (2016).

[7] J. Kato, K. Horimoto, S. Sato, T. Minowa,, and H. Uhara, "Dermoscopy of melanoma and non-melanoma skin cancers," Frontiers Med. **6**, 180 (2019).

[8] A. Aune, G. Vartdal, H. Bergseng, L. L. Randeberg, and E. Darj, "Bilirubin estimates from smartphone images of newborn infants' skin correlated highly to serum bilirubin levels," Acta Paediatrica **109**, 2532–2538 (2020).

[9] G. Sharma and R. Bala, *Digital Color Imaging Handbook* (CRC Press, Boca Raton, FL, 2017).

[10] M. Ebner, "Color Constancy," *The Wiley-IS&T Series in Imaging Science and Technology* (John Wiley, Hoboken, NJ, 2007).

[11] M. A. Abebe, *"Perceptual content and tone adaptation for HDR display technologies"*, Ph.D. thesis (Poitiers, 2016).

[12] M. Nixon, F. Outlaw, and T. S. Leung, "Accurate device-independent colorimetric measurements using smartphones," PLoS One **15**, 1–19 (2020).

[13] S. Süsstrunk and G. Finlayson, "Evaluating chromatic adaptation transform performance," *Proc. IS&T/SID CIC13: Thirteenth Color and Imaging Conf.* (IS&T, Springfield, VA, 2005), pp. 75–78.

[14] D. A. Rowlands, "Color conversion matrices in digital cameras: a tutorial," Opt. Eng. **59**, 110801 (2020).

[15] I. 42, "Photography–electronic still-picture cameras–methods for measuring opto-electronic conversion functions (oecfs)," *ISO 14524 :2009*, 2009.

[16] I. 42, "Graphic technology and photography—colour characterisation of digital still cameras (dscs)–part 1: Stimuli, metrology and test procedures," *ISO 17321–1:2012*, 2012.

[17] R. V. Luther, "Aus dem gebiet der farbreizmetrik," Z. Technishe Phys. **12**, 540–558 (1927).

[18] P. Hung, "Color theory and its application to digital still cameras," in *Image Sensors and Signal Processing for Digital Still Cameras*, edited by J. Nakamura (Informa, London, 2006), pp. 205–221.

[19] P. Hubel, J. M. Holm, G. Finlayson, and M. Drew, "Matrix calculations for digital photography," *Proc. IS&T/SID CIC5: Fifth Color and Imaging Conf.* (IS&T, Springfield, VA, 1997), pp. 105–111.

[20] S. Bianco, C. Cusano, and R. Schettini, "Color constancy using cnns," *2015 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)* (IEEE, Piscataway, NJ, 2015), pp. 81–89.

[21] H.-H. Choi and B.-J. Yun, "Deep learning-based computational color constancy with convoluted mixture of deep experts (cmode) fusion technique," IEEE Access **8**, 188309–188320 (2020).

[22] P. Green, "Color Management: understanding and using ICC Profiles," *The Wiley-IS&T Series in Imaging Systems and Technology* (John Wiley, Hoboken, NJ, 2010).

[23] M. Aydın, F. Hardalaç, B. Ural, and S. Karap, "Neonatal jaundice detection system," J. Med. Syst. **40**, 1–11 (2016).

[24] S. B. Munkholm, T. Krøgholt, F. Ebbesen, P. B. Szecsi, and S. R. Kristensen, "The smartphone camera as a potential method for transcutaneous bilirubin measurement," PLoS One **13**, e0197938 (2018).

[25] J. A. Taylor, J. W. Stout, L. de Greef, M. Goel, S. N. Patel, E. K. Chung, A. Koduri, S. R. McMahon, J. A. Dickerson, E. A. Simpson, and E. C. Larson, "Use of a smartphone app to assess neonatal jaundice," Pediatrics **140** (2017).

[26] S. Swarna, S. Pasupathy, B. Chinnasami, D. Manasa, and B. Ramraj, "The smart phone study: assessing the reliability and accuracy of neonatal jaundice measurement using smart phone application," Int. J. Contemp. Pediatrics **5**, 285–289 (2018).

[27] T. S. Leung, F. Outlaw, L. W. MacDonald, and J. Meek, "Jaundice eye color index (jeci): quantifying the yellowness of the sclera in jaundiced neonates with digital photography," Biomed. Opt. Express **10**, 1250–1256 (2019).
[28] V. Gunnar, R. Lise Lyngsnes, A. Anders, and K. Aleksander, "Image based bilirubin determination," U.S. Patent 10,799,150 (2020).
[29] A. INC., "Digital negative (DNG) specification." Accessed: 2021-07-13.
[30] L. LLC, "LibRaw: RAW image decoder library." Accessed: 2021-07-13.
[31] M. Riechert, "rawpy's documentation" Accessed: 2021-07-13.
[32] T. W. Sawyer, A. S. Luthman, and S. E. Bohndiek, "Evaluation of illumination systems for wide-field hyperspectral imaging in biomedical applications," Proc. SPIE **10068**, 1006818 (2017).
[33] J. Y. Hardeberg, H. Brettel, and F. J. M. Schmitt, "Spectral characterization of electronic cameras," Proc. SPIE **3409** (1998).
[34] S. Westland and C. Ripamonti, *Computational Colour Science using MATLAB* (John Wiley, Hoboken, NJ, 2004).
[35] O. Burggraaff, N. Schmidt, J. Zamorano, K. Pauly, S. Pascual, C. Tapia, E. Spyrakos, and F. Snik, "Standardized spectral and radiometric calibration of consumer cameras," Opt. Express **27**, 19075–19101 (2019).
[36] J. Jiang, D. Liu, J. Gu, and S. Süsstrunk, "What is the space of spectral sensitivity functions for digital color cameras?," *2013 IEEE Workshop on Applications of Computer Vision (WACV)* (IEEE, Piscataway, NJ, 2013), pp. 168–179.
[37] S. Quan, N. Ohta, and X. Jiang, "Comparative study on sensor spectral sensitivity estimation," Proc. SPIE **5008** (2003).
[38] *Chromatic Adaptation Models* ch. 9, pp. 181–198. John Wiley and Sons, Ltd, 2013.
[39] P. Green and T. Habib, "Chromatic adaptation in colour management," in *Computational Color Imaging*, edited by S. Tominaga, R. Schettini, A. Trémeau, and T. Horiuchi (Springer International Publishing, Cham, 2019), pp. 134–144.
[40] C. Li, Z. Li, Z. Wang, Y. Xu, M. R. Luo, G. Cui, M. Melgosa, M. H. Brill, and M. Pointer, "Comprehensive color solutions: Cam16, cat16, and cam16-ucs," Color Res. Appl. **42**, 703–718 (2017).
[41] C. F. Andersen and J. Hardeberg, "Colorimetric characterization of digital cameras preserving hue planes," *Proc. IS&D/SID CIC13: Thirteenth Color and Imaging Conf.* (IS&T, Springfield, VA, 2005), pp. 141–146.
[42] G. D. Finlayson, M. Mackiewicz, and A. Hurlbert, "Color correction using root-polynomial regression," IEEE Trans. Image Process. **24**, 1460–1470 (2015).
[43] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions," *J. Math. Psychol.* (Elsevier, Amsterdam, 2018), Vol. 85, pp. 1–16.
[44] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)* (The MIT Press, Cambridge, MA, 2005).