Screening neonatal jaundice based on the sclera color of the eye using digital photography

Terence S. Leung,[1,*] Karan Kapur,[1] Ashley Guilliam,[1] Jade Okell,[2] Bee Lim,[2] Lindsay W. MacDonald,[3] and Judith Meek[2]

[1]Department of Medical Physics & Biomedical Engineering, University College London, UK
[2]The Neonatal Care Unit, Elizabeth Garrett Anderson Wing, University College London Hospitals Trust, UK
[3]Department of Civil, Environmental & Geomatic Engineering, University College London, UK
[*]t.leung@ucl.ac.uk

Abstract: A new screening technique for neonatal jaundice is proposed exploiting the yellow discoloration in the sclera. It involves taking digital photographs of newborn infants' eyes (n = 110) and processing the pixel colour values of the sclera to predict the total serum bilirubin (TSB) levels. This technique has linear and rank correlation coefficients of 0.75 and 0.72 (both $p<0.01$) with the measured TSB. The mean difference ($\pm$ SD) is 0.00 $\pm$ 41.60 µmol/l. The receiver operating characteristic curve shows that this technique can identify subjects with TSB above 205 µmol/l with sensitivity of 1.00 and specificity of 0.50, showing its potential as a screening device.

OCIS codes: (300.6550) Spectroscopy, visible; (330.1710) Color, measurement.

References and Links

1. "NICE clinical guideline 98: Neonatal jaundice," (National Institute for Health and Clinical Excellence, London, 2010).
2. J. M. Rennie, *Rennie & Roberton's Textbook of Neonatology* (Elsevier Health Sciences, 2012).
3. R. Keren, K. Tremont, X. Luan, and A. Cnaan, "Visual assessment of jaundice in term and late preterm infants," Arch. Dis. Child. Fetal Neonatal Ed. **94**(5), F317–F322 (2009).
4. L. I. Kramer, "Advancement of dermal icterus in the jaundiced newborn," Am. J. Dis. Child. **118**(3), 454–458 (1969).
5. P. Szabo, M. Wolf, H. U. Bucher, J. C. Fauchère, D. Haensse, and R. Arlettaz, "Detection of hyperbilirubinaemia in jaundiced full-term neonates by eye or by bilirubinometer?" Eur. J. Pediatr. **163**(12), 722–727 (2004).
6. G. Nagar, B. Vandermeer, S. Campbell, and M. Kumar, "Reliability of transcutaneous bilirubin devices in preterm infants: a systematic review," Pediatrics **132**(5), 871–881 (2013).
7. S. N. El-Beshbishi, K. E. Shattuck, A. A. Mohammad, and J. R. Petersen, "Hyperbilirubinemia and transcutaneous bilirubinometry," Clin. Chem. **55**(7), 1280–1287 (2009).
8. M. Ahmed, S. Mostafa, G. Fisher, and T. M. Reynolds, "Comparison between transcutaneous bilirubinometry and total serum bilirubin measurements in preterm infants <35 weeks gestation," Ann. Clin. Biochem. **47**(1), 72–77 (2010).
9. D. De Luca, E. Zecca, P. de Turris, G. Barbato, M. Marras, and C. Romagnoli, "Using BiliCheck™ for preterm neonates in a sub-intensive unit: Diagnostic usefulness and suitability," Early Hum. Dev. **83**(5), 313–317 (2007).
10. K. Jangaard, H. Curtis, and R. Goldbloom, "Estimation of bilirubin using BiliChek, a transcutaneous bilirubin measurement device: Effects of gestational age and use of phototherapy," Paediatr. Child Health **11**(2), 79–83 (2006).
11. T. Karen, H. U. Bucher, and J. C. Fauchère, "Comparison of a new transcutaneous bilirubinometer (Bilimed) with serum bilirubin measurements in preterm and full-term infants," BMC Pediatr. **9**(1), 70 (2009).
12. S. Sanpavat and I. Nuchprayoon, "Transcutaneous bilirubin in the pre-term infants," J. Med. Assoc. Thai. **90**(9), 1803–1808 (2007).
13. E. T. Schmidt, C. A. Wheeler, G. L. Jackson, and W. D. Engle, "Evaluation of transcutaneous bilirubinometry in preterm neonates," J. Perinatol. **29**(8), 564–569 (2009).
14. L. Y. Siu, L. W. Siu, S. K. Au, K. W. Li, T. K. Tsui, Y. Y. Chang, G. P. Lee, and N. S. Kwong, "Evaluation of a transcutaneous bilirubinometer with two optical paths in Chinese preterm infants," Hong Kong J. Paediatr. **15**, 132–140 (2010).
15. W. A. Willems, L. M. van den Berg, H. de Wit, and A. Molendijk, "Transcutaneous bilirubinometry with the Bilicheck in very premature newborns," J. Matern. Fetal Neonatal Med. **16**(4), 209–214 (2004).

16. S. Yasuda, S. Itoh, K. Isobe, M. Yonetani, H. Nakamura, M. Nakamura, Y. Yamauchi, and A. Yamanishi, "New transcutaneous jaundice device with two optical paths," J. Perinat. Med. **31**(1), 81–88 (2003).
17. C. Romagnoli, E. Zecca, P. Catenazzi, G. Barone, and A. A. Zuppa, "Transcutaneous bilirubin measurement: Comparison of Respironics BiliCheck and JM-103 in a normal newborn population," Clin. Biochem. **45**(9), 659–662 (2012).
18. L. d. Greef, M. Goel, M. J. Seo, E. C. Larson, J. W. Stout, J. A. Taylor, and S. N. Patel, "BiliCam: Using Mobile Phones to Monitor Newborn Jaundice," in *The 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* 2014)
19. T. S. Leung, A. Guilliam, L. MacDonald, and J. Meek, "Investigation of the relationship between the sclera colour of the eye and the serum bilirubin level in newborn infants," in *42nd Annual Meeting of the International Society on Oxygen Transport to Tissue,* (London, 2014).
20. P. G. Watson and R. D. Young, "Scleral structure, organisation and disease. A review," Exp. Eye Res. **78**(3), 609–623 (2004).
21. A. N. Bashkatov, E. A. Genina, V. I. Kochubey, and V. V. Tuchin, "Optical properties of human sclera in spectral range 370–2500 nm," Opt. Spectrosc. **109**(2), 197–204 (2010).
22. J. J. Kuiper, "Conjunctival Icterus," Ann. Intern. Med. **134**(4), 345–346 (2001).
23. N. J. Talley and S. O'Connor, *Clinical Examination: A Systematic Guide to Physical Diagnosis* (Elsevier Health Sciences APAC, 2014).
24. N. Cohen, "A Color Balancing Algorithm for Cameras," Standford Project Report. (2011).
25. J. F. Hair, *Multivariate Data Analysis* (Prentice Hall, 2010).
26. D. J. Sheskin, *Handbook of parametric and nonparametric statistical procedures* (crc Press, 2003).
27. J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," Lancet **327**(8476), 307–310 (1986).
28. M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine," Clin. Chem. **39**(4), 561–577 (1993).
29. A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," Pattern Recognit. **30**(7), 1145–1159 (1997).
30. W. J. Youden, "Index for rating diagnostic tests," Cancer **3**(1), 32–35 (1950).
31. S. Leartveravat, "Transcutaneous bilirubin measurement in full term neonate by digital camera," Medical Journal of Srisaket Surin Buriram Hospitals **24**, 105–118 (2009).
32. Z. Liu and J. Zerubia, "Skin image illumination modeling and chromophore identification for melanoma diagnosis," Phys. Med. Biol. **60**(9), 3415–3431 (2015).

1. Introduction

Neonatal jaundice is a common condition affecting newborn babies - 60% of term infants and 80% of preterm infants in the first week of life, and 10% of breastfed babies are still jaundiced at 4 weeks old [1]. It is caused by increased red blood cell breakdown or decreased liver function, usually a physiological process, nonetheless causing an accumulation of bilirubin. In most cases no treatment is required. However, if the total serum bilirubin (TSB) level exceeds an age dependent treatment threshold [1], phototherapy has been shown to reduce the TSB level to avoid kernicterus, a condition whereby a high level of bilirubin has crossed the blood brain barrier causing short and long term neurological dysfunction.

Jaundice can also indicate other life threatening conditions. For instance, haemolytic disease of the newborn (a condition in which antibodies from the mother attack the newborn's red blood cells), neonatal infection and liver diseases can all cause bilirubin levels to rise [2]. Early diagnosis of these pathological conditions would allow urgent treatments to be administered in a timely fashion, improving the chance of recovery significantly.

In the UK newborns who have had a low risk perinatal course are discharged from hospital within 1-2 days of birth. They and those born at home are routinely visited by a community midwife who assesses their wellbeing with respect to feeding, weight gain and jaundice. If the baby appears jaundiced the midwife should measure the bilirubin level either with a transcutaneous bilirubinometer (TcB) or with a blood sample. If the baby appears unwell or the bilirubin measurement exceeds 250 µmol/l, the baby should be referred to the paediatric team in the nearest hospital. In a worldwide setting there is a wide variation in the availability of community midwives and local paediatric teams. A screening tool using digital photography (e.g. a mobile phone camera) could have wide applications in settings where TcBs and blood tests are too expensive and difficult to obtain.

One noticeable feature of jaundiced patients is the appearance of yellow discoloration on their skin and sclera caused by a higher level of TSB (yellow in colour) in the blood. It is visually detectable when TSB levels exceed 85 μmol/l. The degree of the yellow discoloration is dependent on the TSB level. An experienced healthcare worker can often identify jaundiced patients by visual inspection. The extent of neonatal jaundice can also be quantified by a 5-point scale first proposed by Kramer [3–5]. However, colour perception can be severely influenced by the ambient lighting and the skin pigment [5]. Such screening practice is therefore unable to provide an objective, reliable and accurate result [3].

For a more quantitative measurement, customised spectrometers have been developed which illuminate the skin with control lighting in direct contact and evaluate the reflected light at multiple wavelengths. Currently, the two most popular commercial TcBs include BiliChek (Philips Healthcare) and Jaundice Meter JM102/JM103 (Draeger Medical Inc., previously marketed by Konica Minolta Inc.) [5–16]. The skin pigment is the most influential confounding factor for the skin colour approach. To minimise its effect, the two TcBs adopt two very different approaches: BiliChek exploits 137 wavelengths in the visible spectrum between 380 and 760 nm to resolve the concentrations of melanin, haemoglobin, bilirubin and dermis based on their known absorption spectra [7, 17]; JM102/JM103, on the other hand, only uses two colours of light in blue and green (with peak wavelengths of 450 nm and 550 nm), as well as a special probe design which allows the reflected light to be measured by two detectors at two different distances. The two optical density signals measured in this way account for light with two different penetration depths beneath the skin, one deeper than the other. The difference between the two signals is therefore more sensitive to the deeper subcutaneous tissue where the effect of skin pigment is minimised [7, 17].

A recent study also shows that a smartphone's camera (iPhone 4S) can be used to predict TSB level based on the skin colour [18]. This approach essentially makes use of only three broadband colours (red, green and blue) that can be detected by a conventional smartphone's camera. This technique converts the three base colours into 2 colour spaces and uses 5 different machine learning algorithms to predict the TSB level.

This current study reports an alternative approach based on the sclera colour to predict bilirubin level and quantify jaundice. The main advantage is that the sclera is less influenced by pigments such as melanin. Healthy sclera is white, independent of the ethnicity of the subject. Without the influence of a major confounding factor, the quantification of the yellowness and therefore the prediction of bilirubin level become much easier.

Our study involves a digital camera which is used to take photos of newborn infants' eyes and does not require any contact with the infants. The digital image is then analysed by a computer program written in Matlab, a high-level technical computing language. Preliminary results were presented at an international conference in July 2014 [19]. The sample size has since been expanded and the methodology improved, as reported in this paper. The ultimate aim of this work is to develop a screening tool which allows more effective referral of suspected patients for blood tests, rather than replacing them completely.

2. Methods

2.1 Study protocol and data collection

The study has been approved by the National Research Ethics Service Committee (London – City Road and Hampstead). The data were collected in the Advanced Neonatal Nurse Practitioner Clinic at University College London Hospital. The majority of these babies were referred by midwives conducting a routine health check after discharge from the hospital. Some appointments were made for babies with pre-existing prolonged or physiological jaundice. As the TSB blood test formed part of the routine clinical procedure during the consultation, no further blood sampling was required specifically for this study. The photographs were taken before the blood test as babies tended to be more compliant at this

time. The results for the blood test were processed within twenty minutes thereafter to prevent any error from short-term fluctuations in bilirubin measurements.

The camera used was a Nikon D3200, which had a 24.6 megapixel CMOS sensor, with a prime macro lens of 60 mm focal length. Manual focus was used to ensure that the sclera was in focus and programmed auto mode was selected. This allowed the shutter speed and aperture size to vary depending on light entering the sensor to prevent over/under exposure. The ISO value (sensitivity of the digital sensor to light) was kept at ISO 1600 as a compromise between maintaining fast shutter speed and allowing enough light to expose the picture correctly. White balance was also maintained at the '4' setting for fluorescent lighting and kept constant to ensure that colour temperature did not influence colour analysis on a picture-to-picture basis. A customised colour chart was utilised to match the colour of the sclera to a reference value on the chart so that any yellow difference in sclera colour can be correctly associated with a physiological effect rather than a change in spectrum of ambient lighting. The images were captured in RAW format to preserve as much tonal information as possible for each image taken.

2.2 Subjects

A total of 133 newborn infants were photographed for this study. Since the ambient lighting can seriously affect the result of this study, only the 110 samples taken in a specific clinic room have been included in the subsequent analysis. The reasons for exclusion are: 18 samples were collected from other locations, 2 samples had erroneous colour channel responses, 1 baby would not fully open his eyes, and 2 TSB blood samples could not be recorded. Figures 1(a) and 1(b) show the distribution of TSBs in relation to the gestational and postnatal ages, respectively. As expected, Fig. 1(b) shows that the TSB tends to drop with the postnatal age. The database includes 67 white and 43 non-white infants.
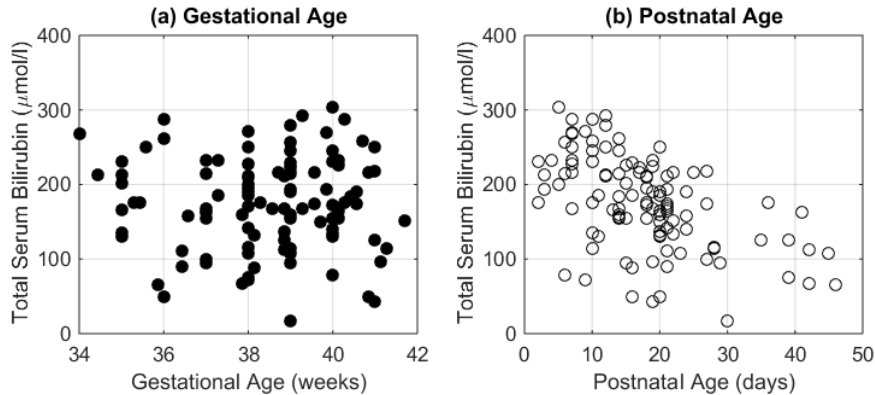

Fig. 1. The distribution of total serum bilirubin levels in relation to (a) gestational, and (b) postnatal ages (n = 110).

2.3 Optical properties of the sclera

The sclera is often known as "the white of the eye" which is the opaque, outer layer of the eye, surrounding the iris. The sclera is mainly composed of collagen, elastin, proteoglycans and glycoproteins [20]. Since the collagen has a slow turnover, the sclera has a low metabolic rate and requires only a low blood supply. In comparison to other parts of the body, e.g., skin, the sclera is less vascularised which also means that the colour of a healthy sclera is less influenced by the colour of the blood. In newborn infants, the sclera is thin and often appears blue, revealing the colour of the underlying uveal layer [20].

The optical properties of human sclera have been studied previously [21], showing that the optical transport scattering coefficient decreases monotonically from 370 nm to 1300 nm, similar to the general spectral characteristics of biological tissues. The study also shows that

between 400 and 900 nm, the absorption coefficient of the human sclera has a simple spectral trend and does not have the spectral characteristics found in rabbit and pig sclera which are influenced by the scleral melanin unique only to animals [21]. Since the human sclera is free of melanin, it appears white in all ethnicities.

The sclera is overlaid by the clear, lubricating layer conjunctiva, which some have argued is the main origin of the yellow discolouration associated with a higher level of bilirubin in jaundiced patients [22, 23]. It has been pointed out that the bilirubin has a strong affinity for elastin and it is the innermost layer of conjunctiva and the most superficial part of the sclera that are rich in elastic fibres, and therefore account for the yellow discoloration [22]. Some have simply noted that the bilirubin is mainly deposited in the vascular conjunctiva rather than the avascular sclerae [23]. While acknowledging the subtlety in the interpretation, we have adopted the more common term of "sclera colour" to describe the "combined colour" of conjunctiva and sclera in this paper.

2.4 Signal analysis

The raw images were originally saved in Nikon's proprietary NEF format, and then converted into TIF format using Nikon's ViewNX 2 software. These images were analysed using the mathematical software Matlab. The pixel coordinates of the sclera, skin (for comparison purpose) and the reference colour were manually identified in an image viewer. The reference colour was taken as the white patch in the colour chart. Figure 2 shows an example of pixels identified in this way. Three colour indexes for red, green and blue, represented by *R, G* and *B* here, were calculated by averaging 900 neighbouring pixels (30 × 30 region) centred on the pre-defined pixel coordinates. This process resulted in colour indexes: R_{eye}, G_{eye} and B_{eye} for the sclera, R_{skin}, G_{skin} and B_{skin} for the skin, and R_{ref}, G_{ref} and B_{ref} for the reference colour. Normalised versions of the colour indexes were also calculated for both the sclera and skin, e.g., $R_{eye,nor} = R_{eye} / R_{ref}$ is the normalised red colour index of the sclera. The normalisation can be considered as a colour balancing process in digital photography [24].

The next step was to develop a multivariate model based on the measured TSB levels and colour indexes in the database [25]. The multiple linear regression was performed using Matlab and its Statistics and Machine Learning Toolbox. Using colour indexes as the independent variables and the measured TSB values as dependent variables (*y*), a quadratic model was developed:

$$y_{eye} = C_0 + C_1 R_{eye} + C_2 G_{eye} + C_3 B_{eye} + C_4 R_{eye} G_{eye} + C_5 R_{eye} B_{eye}$$
$$+ C_6 B_{eye} G_{eye} + C_7 R_{eye}^2 + C_8 G_{eye}^2 + C_9 B_{eye}^2 \tag{1}$$

where y_{eye} is the predicted TSB value, coefficients $C_0 = 373$, $C_1 = -3385$, $C_2 = 5137$, $C_3 = -2391$, $C_4 = -14945$, $C_5 = 1987$, $C_6 = 7163$, $C_7 = 8680$, $C_8 = 1295$ and $C_9 = -3906$. This quadratic model was chosen to provide a reasonably high R-squared value while avoiding overfitting the data, as in the case of some higher order polynomial models. By performing multiple linear regression on four sets of colour indexes ($R_{skin}/G_{skin}/B_{skin}$, $R_{skin,nor}/G_{skin,nor}/B_{skin,nor}$, $R_{eye}/G_{eye}/B_{eye}$ and $R_{eye,nor}/G_{eye,nor}/B_{eye,nor}$), four models were produced which could then be used to predict four sets of TSB levels: y_{skin}, $y_{skin,nor}$, y_{eye} and $y_{eye,nor}$.
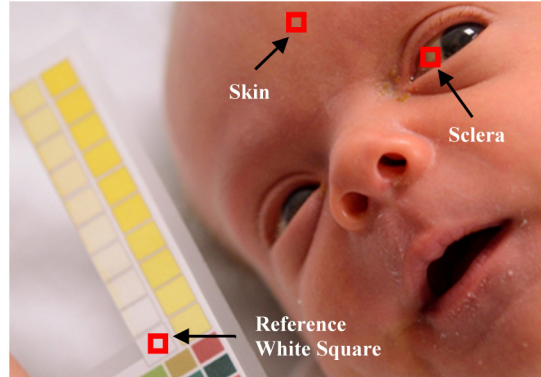
Fig. 2. An example of the image collected in the UCL Hospitals. The colour chart is shown on the left. The pixel values for the sclera, skin and white reference have been used in the processing. (This image has been media consented by the parents).

2.5. Statistical analyses

The statistical analyses focused on comparing the measured and predicted TSB levels in terms of their correlation, bias, sensitivity and specificity as a screening tool. The linear correlation between the measured and predicted TSB levels was assessed by calculating Pearson's linear correlation coefficient r [26]. For a non-parametric correlation with less influence from outliers, Spearman's rank correlation coefficient ρ was calculated [26]. To assess the agreement between the conventional and new techniques, Bland-Altman analysis was carried out [27].

Since the ultimate aim of this work is to develop a screening tool which can make a binary decision (referral or non-referral), the accuracy of detecting subjects above a certain TSB screening level (y_{sc}) is also assessed using the receiver operating characteristic (ROC) curve [28]. The ROC curves have been calculated based on the predicted TSB levels, with reference to the measured TSB levels as the true answers. Suppose the new technique aims to identify subjects with a TSB level above a screening threshold, e.g., > y_{sc} = 205 µmol/l. If one uses a low cut-off threshold (y_{cut}) of 50 µmol/l (i.e. all subjects with predicted TSB levels above 50 µmol/l would be considered to have TSB levels above 205 µmol/l), the true positive rate (TPR) would be very high, albeit at the expense of a high false positive rate (FPR). As y_{cut} increases, both the TPR and FPR would drop in a non-linear fashion. The ROC curve summarises the TPR and FPR for a range of y_{cut}, allowing the selection of a particular y_{cut} with certain sensitivity (TPR) and specificity (1-FPR). The "area under the ROC curve" (AUC) was also calculated which indicated the general accuracy of the screening technique [29].

3. Results

The results are presented for four sets of predicted TSB levels including y_{skin}, $y_{skin,nor}$, y_{eye} and $y_{eye,nor}$ as described in section 2.4. Although the main focus of this study was on using the sclera colour to predict TSB levels, the skin colour was included here for comparison purpose.

3.1 Correlations between the measured and predicted serum bilirubin levels

Table 1 summarises the values of Pearson's linear correlation coefficient r and Spearman's rank correlation coefficient ρ between the measured and predicted TSB levels. It can be seen that both the Pearson and Spearman correlations for using the sclera colour (normalised:0.75/0.72; un-normalised: 0.75/0.72) are much higher than those for the skin colour (normalised:0.54/0.50; un-normalised: 0.56/0.54).

Figure 3 shows the scatter plots of the measured and predicted TSB levels using the sclera colour, i.e., Fig. 3(a), and skin colour, i.e., Fig. 3(b) as the predictors, respectively. An identity line is also shown in each figure which indicates that the predicted TSBs tend to overestimate in the lower TSB range (around 100 µmol/l) and underestimate in the higher TSB range (around 250 µmol/l). It can be seen that the data points are generally closer to the identity line using the sclera colour as the predictor (see Fig. 3(a) for y_{eye}) in comparison to using the skin colour (Fig. 3(b) for y_{skin}).

Table 1. Correlation between the measured and predicted TSB levels

Predicted Serum Bilirubin level	Region of interest	RGB values normalised by white reference colour ?	Pearson's linear correlation (r), CI: 95% Confidence Intervals	Spearman's rank correlation (ρ)
$y_{skin,nor}$	Skin	Yes	0.54 (CI: 0.39 –0.66, p<0.01)	0.50 (p<0.01)
y_{skin}	Skin	No	0.56 (CI: 0.42 – 0.68, p<0.01)	0.54 (p<0.01)
$y_{eye,nor}$	Sclera	Yes	0.75 (CI: 0.65 – 0.82, p<0.01)	0.72 (p<0.01)
y_{eye}	Sclera	No	0.75 (CI: 0.65 – 0.82, p<0.01)	0.72 (p<0.01)

3.2 Bland-Altman plot

In the previous section, it has been established that the TSB predicted by the RGB values of the sclera without normalisation, i.e., y_{eye}, provides a good performance, which is the focus in the following analysis. Figure 4 shows the Bland-Altman plot which is often used to assess the agreement between two measurement methods [27]. The mean difference is close to zero because of the condition imposed by the least squares multiple linear regression algorithm. The standard deviation is 41.60 µmol/l. Although the vast majority of data points (all except eight) lie within the 95% confidence intervals, i.e., 0 ± 81.54 µmol/l (mean ± 1.96 standard deviations), the TSB differences are generally large, indicating that the predicted TSBs based on the sclera colour are not accurate enough to replace the TSBs by blood sampling on an individual basis. Similar to the scatter plots in Fig. 3, the Bland-Altman plot also shows the tendency of predicted TSBs overestimating in the lower TSB range and underestimating in the higher TSB range. Similar trends have also been found in a commercial TcB (BiliChek) [5].
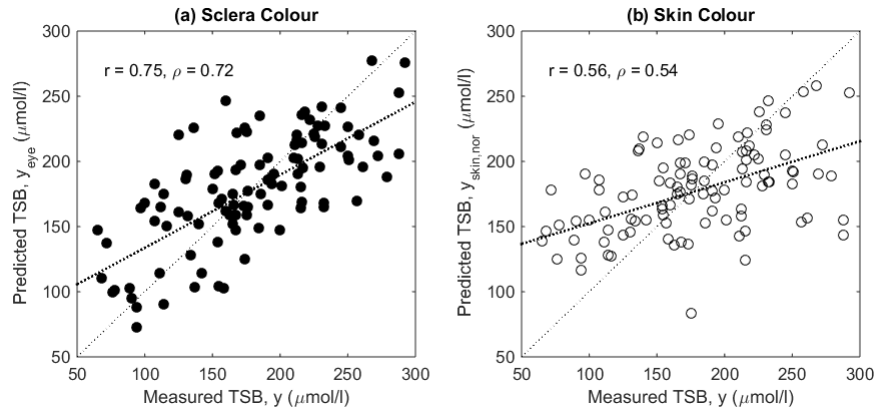

Fig. 3. Scatter plots for the measured and predicted TSB levels based on (a) the sclera colour (un-normalised) and (b) the skin colour (un-normalised). The Pearson's linear correlation (r), the Spearman's rank correlation (ρ) and linear regressed lines are also shown (n = 110).
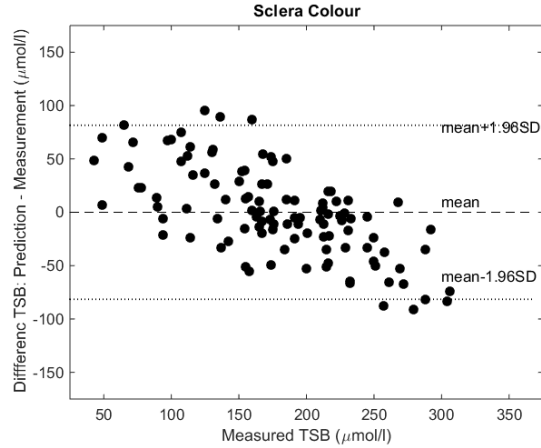
Fig. 4. The Bland Altman plot showing the agreement between the conventional blood sampling method (measurement) and the photographic method (prediction) based on the sclera colour: the mean difference is zero and most of the differences lie between ± 1.96 standard deviations, i.e., ± 81.54 µmol/l.

3.3 Receiver operating characteristic curves

Figure 5 depicts the ROC curve for a screening threshold (y_{sc}) at 205 µmol/l. The optimal cut-off threshold ($y_{cut}*$) can be identified using the Youden's J statistic which results in the maximum J = sensitivity + specificity - 1 = TPR – FPR [30]. This optimal point represents the best compromise between a high TPR and a low FPR, and is, in this case, found to be $y_{cut}*$ = 194 µmol/l (hollow circle), with sensitivity = 0.80 and specificity = 0.86. However, one can also choose a particular y_{cut} to raise the sensitivity which is important in a screening test. A y_{cut} at 162 µmol/l (filled circle) has been chosen to raise the sensitivity to 1.00 at the expense of a reduced specificity = 0.50.
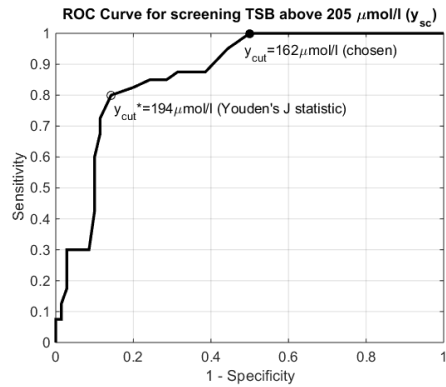

Fig. 5. The ROC curve for screening TSB above 205 umol/l: the optimal cut-off threshold (open circle) based on the Youden's J statistic, $y_{cut}*$ = 194 µmol/l, resulting in sensitivity = 0.80 and specificity = 0.86. To increase the sensitivity, the cut-off threshold is reduced to 162 µmol/l (filled circle) which raises the sensitivity to 1.00 at the expense of a reduced specificity = 0.50.

4. Discussions

4.1 Ambient lighting

The images taken in this study relied on the ambient lighting for illuminating the sclera, which inevitably also affected the prediction performance. When images taken under various

ambient lightings were used for TSB prediction, the correlations between sclera/skin colours and the measured TSB were poor (result not shown here). Using the sclera samples collected from the same location for prediction, the correlations become much stronger as shown in Table 1. The results from Table 1 may first seem surprising that the normalised RGB values, which can also be considered to have undergone a white balancing process, do not improve the correlations, in comparison to the un-normalised RGB values. However, the white balancing is normally used to reduce the effect of colour differences under different illuminations. In our case, the same illumination (fluorescent light in the clinic room) has been used for the whole data set and therefore the need for white balancing is less important. In fact, the introduction of white balancing can be counterproductive because the white reference colour can be an additional source of noise when it receives a different illumination due to multipath reflections. In general, the region of interest is also affected by reflected light from the surrounding areas. For example, a bright colour jumper worn by the mother holding the baby could affect the RGB pixel values of the image. The variability of the lighting environment causes errors in the predicted TSB, leading to a high standard deviation of 41.60 µmol/l in the Bland Altman plot in Fig. 4.

4.2 Comparisons with other techniques

Numerous studies have compared the agreement between TcB and TSB [5–16]. One review in 2013 summarised the results from 22 studies conducted between 1982 and 2012 ([5–16]), mainly on BiliChek and JM-103 (and its earlier versions) [6]. Two other studies adopted digital photography to predict TSB based on the skin colour in the sternum and forehead [18, 31]. The results of these studies have been summarised in Table 2. It can be seen that the linear correlation coefficient r for other studies ranges between 0.83 and 0.86 while for this study it is lower at 0.75. The standard deviations of the mean differences reported in [6] are 24.06 µmol/l for the sternum site and 29.46 µmol/l for the forehead site [6], in comparison to ours at 41.60 µmol/l for the sclera site. Our database lacks subjects with TSB values in both the high (>250 µmol/l) and low (<100 µmol/l) ends of the range, which may in turn affect the r value.

Table 2. Comparison of result obtained by various studies (conversion for bilirubin: 1 mg/dL = 17.1 µmol/l)

Study	Method	Body site	Pearson's linear correlation, r (n)	Bland-Altman test: mean difference [a] ± SD (µmol/l)
Nagar 2013 [6]	TcB: BiliChek, JM-102/103	Forehead	0.83 (n N/A; 16 pooled studies)	−0.06 ± 29.46 (n = 912; 11 pooled studies)
		Sternum	0.83 (n N/A; 10 pooled studies)	3.80 ± 24.06 (n = 265; 5 pooled studies)
Leartveravat 2009 [31]	Digital photography	Sternum	0.86 (n = 61)	N/A
de Greef 2014 [18]	Digital photography	Sternum & Forehead	0.84 (n = 100)	N/A
Leung 2015 (This study)	Digital photography	Sclera	0.75 (n = 110)	0.00 ± 41.60

[a]mean difference defined as: predicted TSB – measured TSB.

For screening purpose, the ROC curve results may be more relevant. Table 3 shows the ROC curve results from three studies [5, 17] including ours. A previous study found that JM-102 performed better than BiliChek for screening subjects with a y_{sc} = 250 µmol/l using a y_{cut} = 230 µmol/l, achieving a sensitivity of 0.97 and a specificity of 0.83 [5]. Another study also found that JM-103 was more accurate than BiliChek in screening for subjects with a y_{sc} = 205 µmol/l using a y_{cut} = 144 µmol/l, achieving a sensitivity of 1.00 and a specificity of 0.42 [17], which are comparable to our results: sensitivity = 1.00 and specificity = 0.50 in screening for subjects with a y_{sc} = 205 µmol/l using a y_{cut} = 162 µmol/l. Although the UK based NICE guidelines require the screening technique to identify patients with y_{sc} > 250 µmol/l for referrals [1], we are unable to report this result because our database lacks samples above 250

μmol/l. The AUC of our technique at 0.87 is also comparable to others' which range between 0.89 and 0.98.

Table 3. Comparison of results from ROC curves

Study	Method	Body site	n	Screening Threshold, y_{sc} (μmol/l)	Cut-off Threshold, y_{cut} (μmol/l)	Sensitivity	Specificity	AUC
Szabo 2004 [5]	TcB: BiliChek	Forehead	140	250	190	0.94	0.72	0.92
Szabo 2004 [5]	TcB: JM-102	Sternum	140	250	230	0.97	0.83	0.98
Romagnoli 2012 [17]	TcB: BiliChek	Forehead	630	205	157	0.99	0.30	0.89
Romagnoli 2012 [17]	TcB: JM-103	Forehead	630	205	144	1.00	0.42	0.94
Leung 2015 (This study)	Digital photo-graphy	Sclera	110	205	162	1.00	0.50	0.87

5. Conclusions

This work has shown that by using the sclera colour to predict the TSB level, reasonably high correlations ($r = 0.75$ and $\rho = 0.72$) with the measured TSB can be achieved. Although the correlations may still not be high enough for this technique to be used to predict the absolute TSB level on an individual basis, it is reasonably robust as a screening technique in identifying subjects with TSBs above 205 μmol/l, achieving a sensitivity of 1.00 and specificity of 0.50, comparable to those achieved by commercial TcBs such as JM-103 and BiliChek [17]. In comparison to skin, sclera is less affected by other confounding factors such as melanin (ethnicity), allowing a relatively simple analysis technique, i.e., multiple linear regression, to be used in this work and still producing encouraging results. This technique can also be readily implemented as a low cost smartphone app, making it more accessible to the wider community and developing countries. Unlike the contact based TcBs which require disposables, this technique is non-contact, making it easier and more economical to use.

In order to further improve the performance, in our next phase of work, certain improvements will be made. First, to avoid the influence of ambient light, the room light will be turned off and a diffuse LED flashlight will be used to gently illuminate the face of the baby. In this way, the photos can be taken in different locations while ensuring the same illumination is used. Second, the lighting environment can be better controlled by putting the baby in a specially designed "cot" with non-reflective surfaces, similar to the light cabinet (a.k.a viewing booth or light box) used in the textile industry for visual inspection of fabrics and garments. This environment would reduce the multipath problem and therefore minimise the influence of the colour objects in the surrounding on the captured images. Third, the algorithm can also be improved by adopting techniques such as the independent component analysis, which has been shown to provide a better identification of pigment colour [32].