

# Screening for Neonatal Jaundice by Smartphone Sclera Imaging

*Felix Outlaw*

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
**Doctor of Philosophy**  
of  
**University College London.**

Department of Medical Physics and Biomedical Engineering  
University College London

September 17, 2021

I, Felix Outlaw, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

# Abstract

Jaundice is observed in over 60% of neonates and must be carefully monitored. If severe cases go unnoticed, death or permanent disability can result. Neonatal jaundice causes 100,000 deaths yearly, with low-income countries in Africa and South Asia particularly affected. There is an unmet need for an accessible and objective screening method. This thesis proposes a smartphone camera-based method for screening based on quantification of yellow discolouration in the sclera.

The primary aim is to develop and test an app to screen for neonatal jaundice that requires only the smartphone itself. To this end, a novel ambient subtraction method is proposed and validated, with less dependence on external hardware or colour cards than previous app-based methods. Another aim is to investigate the benefits of screening via the sclera. An existing dataset of newborn sclera images (n=87) is used to show that sclera chromaticity can predict jaundice severity.

The neoSCB app is developed to predict total serum bilirubin (TSB) from ambient-subtracted sclera chromaticity via a flash/ no-flash image pair. A study is conducted in Accra, Ghana to evaluate the app. With 847 capture sessions, this is the largest study on image-based jaundice detection to date. A model trained on sclera chromaticity is found to be more accurate than one based on skin. The model is validated on an independent dataset collected at UCLH (n=38).

The neoSCB app has a sensitivity of 100% and a specificity of 76% in identifying neonates with  $TSB \geq 250 \mu\text{mol/L}$  (n=179). This is equivalent to the TcB (JM-105) data collected concurrently, and as good as the best-performing app in the literature (BiliCam). Following a one-time calibration, neoSCB works without specialist equipment, which could help widen access to effective jaundice screening.

# Impact Statement

The primary contribution of this work is the development of the neonatal scleral-conjunctival bilirubinometer (neoSCB) app. Oftentimes, visual inspection is the only available means to screen for jaundice. This is subjective and difficult for the untrained individual, leading to missed cases. This can result in death and disability. neoSCB may offer an objective means of screening for jaundice, leading to fewer untreated severe cases of jaundice. The screening performance of the neoSCB app is superior to visual inspection and as good as purpose-made screening devices (transcutaneous bilirubinometers – TcBs).

Apart from reduction in quality of life, permanent disabilities impose a greater socio-economic burden on families in less economically developed countries, where there are higher rates of jaundice-related death and disability. In a global health context, the ubiquity of smartphones means the app can be used in parts of the world with fewer healthcare provisions, where there is limited access to advice from medical professionals or TcBs.

The neoSCB app was developed so as not to rely on phone attachments or reference cards. This may help with the adoption of the tool, as it does not depend on supply chains or assembly or maintenance steps. This may increase access to jaundice screening in low-resource settings.

The neoSCB app is cheaper than a TcB, even if a suitable smartphone must be purchased. This is an advantage, as reducing costs is important in both high and low-income countries.

Telemedicine is evolving beyond the remote consultation, with more diagnostic services available via smartphone. neoSCB can help remote communities catch

jaundice early, as well as reducing the number of false alarms. In remote areas, false alarms can involve long, costly, and sometimes dangerous trips to far away hospitals. In the context of the COVID-19 pandemic, new parents may be wary of trips to hospitals. A tool like neoSCB can provide reassurance if the baby is not at risk, and a definitive signal to seek medical attention if the baby is at risk.

Apart from the impact of a future deployment of the neoSCB screening app, this work makes several academic contributions. It confirms that sclera chromaticity is more indicative of systemic bilirubin levels than skin chromaticity, and it confirms that phototherapy leads to subsequent underestimation by TcB devices. These confirmations can help inform further research into jaundice screening.

This work showcases a new application of ambient subtraction to recover ambient-independent colour. Further, it shows that this can be done on a smartphone, and that a screen-as-illumination modality is also feasible. These insights may help other researchers explore new ways to monitor disease via smartphone colour imaging. For example, some literature has considered using smartphone camera images to diagnose anaemia by quantifying the pallor of the palpebral conjunctiva.

Finally, the Detecting Jaundice in African Neonates (DJAN) study collected 847 separate image sets. This is the largest study of its kind, and the first in a majority Black African population. This database can be used to further refine the neoSCB algorithm or develop entirely new methods for image-based jaundice detection.

# Publications Resulting

## Journal Articles

- Terence S Leung, Felix Outlaw, Lindsay W MacDonald, and Judith Meek. Jaundice Eye Color Index (JECI): quantifying the yellowness of the sclera in jaundiced neonates with digital photography. *Biomedical optics express*, 10(3):1250–1256, 2019
- Felix Outlaw, Miranda Nixon, Oluwatobiloba Odeyemi, Lindsay W MacDonald, Judith Meek, and Terence S Leung. Smartphone screening for neonatal jaundice via ambient-subtracted sclera chromaticity. *Plos one*, 15(3):e0216970, 2020
- Miranda Nixon, Felix Outlaw, and Terence S Leung. Accurate device-independent colorimetric measurements using smartphones. *Plos one*, 15(3):e0230561, 2020

## Conference Proceedings

- Felix Outlaw, Judith Meek, Lindsay W MacDonald, and Terence S Leung. Screening for neonatal jaundice with a smartphone. In *Proceedings of the 2017 International Conference on Digital Health*, pages 241–242, 2017
- Felix Outlaw, Miranda Nixon, Nana Okai Brako, Lindsay W MacDonald, Judith Meek, Christabel Enweronu-Laryea, and Terence S Leung. Smartphone colorimetry using ambient subtraction: application to neonatal jaundice screening in Ghana. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 172–175, 2019
- Miranda Nixon, Felix Outlaw, Lindsay W MacDonald, and Terence S Leung. The importance of a device specific calibration for smartphone colorimetry. In *27th Color and Imaging Conference Final Program and Proceedings*, pages 49–54, 2019

# Conferences Attended

## Presentations

- 7th International Conference on Digital Health, London, July 2017
- The EGA Institute for Women's Health 15 Year Celebration Conference, London, June 2019
- 60th Anniversary Celebration and Summer Meeting of The Neonatal Society, Cambridge, June 2019
- BioMedEng Conference, London, September 2019

## Posters

- Progress in Colour Studies (PICS), London, September 2016
- Institute for Women's Health 12th Annual Conference, London, July 2017
- Institute for Women's Health 13th Annual Conference, London, April 2018
- Saving Lives at Birth DevelopmentXChange, Washington DC, July 2018
- i-sense Digital Health Conference, London, July 2019
- ACM International Joint Conference on Pervasive and Ubiquitous Computing (UBICOMP), London, September 2019

## Academic Meetings Attended

- Colour Vision Meeting & Palmer Lecture and CRS Sponsored Lecture, London, January 2017
- Royal Society Scientific Meeting on Understanding Images in Biological and Computer Vision, London, February 2018
- UCL Institute for Healthcare Engineering Global Healthcare Engineering Symposium, London, July 2019

# Contributions

Important contributions to the work laid out in this thesis were made by Alister Lam (AL), Christabel Enweronu-Laryea (CEL), Dorothy Owusu Mensah (DOM), Felix Outlaw (FO), Fiona Young (FY), Judith Meek (JM), Josephine Windsor-Lewis (JWL), Lindsay MacDonald (LM), Miranda Nixon-Hill (MNH), Oluwatobiloba Odeyemi (OO), Patrick Laske (PL), Rhodaline Odoi (RO), Terence Leung (TL), and Yifan Zhang (YZ). Here follows a summary of their contributions:

## **Development of Ambient-subtracted Sclera Chromaticity Approach**

JM and TL proposed imaging the sclera; FO and TL proposed ambient subtraction; LM helped to refine idea; FO proposed a chromaticity-based jaundice metric.

## **Development of Subtracted Signal-to-Noise Metric**

FO, MNH, and TL identified the need; FO and MNH devised the metric; FO and MNH collected flash/ no-flash lab data; MNH analysed the images and wrote the software; MNH determined appropriate SSNR threshold.

## **Development of One-time Calibration Solution**

FO and MNH suggested the solution, with input from LM; MNH captured the image data, wrote the software, and developed the mappings.

## **Chromaticity (Snooker Ball) and Ambient Subtraction Demonstrations**

LM and TL proposed the experiments; FO prepared the samples, collected the image data, and performed the analysis.

## **UCLH Nikon Study**

JM and TL provided the dataset; LM contributed raw image processing script; TL contributed analysis code; FO re-analysed data and developed analysis code



further; FO and TL proposed the Jaundice Eye Colour Index, with input from LM.

### **UCLH Smartphone Study**

FO, JM, and TL devised study; FO and TL determined app features; FO developed app; JM supervised collection of data; OO collected the data; FO developed analysis code and analysed the data.

### **Ghana Smartphone Study**

CEL, FO, JM, and TL determined study aims and data collection protocol; TL managed the project; TL oversaw equipment procurement, with help from FO and JM; FO and TL decided on improvements to app interface and features, with feedback from DOM and RO; AL improved and refactored app; FO, FY, and JWL further improved app; MNH developed and produced the diffuser for the back-facing flash; FY wrote the app user guide, with help from FO and MNH; FO created the data collection sheets, with input from CEL, JM, and TL; CEL oversaw collection of data; DOM and RO collected the data; FO managed secure data transfer and storage; FO managed database; FO segmented image data, with help from FY, JWL, PL, and YZ; MNH contributed software tool to identify colour card patches; FO wrote analysis software and performed all analyses; FO trained and tested app prediction model.

# Acknowledgements

Many people were essential to the completion of this thesis. Dr Terence Leung was a fantastic supervisor throughout. He always put my development as a researcher first, affording me opportunities to try new things, learn, and travel. He taught me the ins and outs of the world of research. His enthusiasm and passion for research made for extremely generative and enjoyable meetings. His broad experience in biomedical engineering, signal processing, and optics helped me understand the context of my own investigations. His thoughtful and patient mentorship paved the way for everything I did at UCL.

The expertise and practical experience of neonatologist Dr Judith Meek was invaluable for this project. Her deep familiarity with the clinical context of neonatal jaundice informed all aspects of this research. Judith always asked the right question at the right time to focus my research in the direction it needed to go. She always kept the goal of producing an output that would improve outcomes for newborns front and centre. She provided me with timely encouragement and helped me take perspective. This thesis would not have been possible without her.

Professor Lindsay MacDonald introduced me to colour science. His passion for the subject was infectious, as was his curiosity and cheerfulness. His great expertise in colour science helped me contextualise my project in the field of colour measurement, and his feedback was always incisive. Lindsay was immensely generous with his time, always willing to explain a concept, offer a piece of equipment, lend a book, or introduce a connection.

Without Professor Chrisabel Enweronu-Laryea, our study in Ghana would have been impossible. Her dedication ensured things got done at all stages of the

project. As well as being an inspirational doctor and researcher, Christabel was a wonderful host to us on our visits to Ghana.

Miranda Nixon-Hill was a fantastic colleague and friend through my PhD experience. Working with Miranda was both immensely productive and immensely fun. Our collaboration – bouncing new ideas off each other, performing regular sanity checks, and making lab time less of a lonesome endeavour – certainly benefited me greatly. Miranda is also a brilliant scientist whose diligent work is essential to this thesis, as reflected in the number of times she is cited. I hope the secrets of secret specular unravel themselves for you.

I am also extremely grateful for the contributions to the project from UCL researchers Patrick Laske, Josephine Windsor-Lewis, Fiona Young, and Yifan Zhang. This thesis owes more to their work than they probably realise. They made great efforts improving the app, doing the painstaking task of manual segmentation, and representing the project and the group both within and outside UCL.

The Ghana project relied on the essential contribution of many team members: Dr Nana Brako, Dr Ayegua Hagan Seneadza, Dr Genevieve Insaidoo, Dr Mary Ani-Amponsah, Eugene Gyakari, Rhodaline Odoi, and Dorothy Owusu Mensah. Without these people the study would have been impossible. I am sorry that I did not get to meet them all in person. The project has a particular debt of gratitude to the incredible hard work of Rhodaline and Dorothy, who collected the vast majority of the data in often challenging circumstances. Perhaps even more challenging, they made sure we did not get lost in Ghana's biggest market during our visit.

I thank all of my wonderful friends for putting up with me for the last four years. Thanks for keeping me laughing.

I thank my partner Sophia for being infinitely patient, supportive, and understanding. Every challenge I have overcome since the beginning of the PhD has been overcome with her backing.

I thank my parents, Denise and Chris, for their unconditional support. They really are my role models in life. Any success of mine is thanks to the opportunities and encouragement they have given me.

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>   | <b>23</b> |
| 1.1      | Motivation and Aims . . . . .                               | 23        |
| 1.2      | Thesis Objectives and Overview . . . . .                    | 25        |
| 1.3      | Jaundice Background . . . . .                               | 27        |
| 1.3.1    | What is Jaundice? . . . . .                                 | 27        |
| 1.3.2    | Disease Burden . . . . .                                    | 29        |
| 1.3.3    | Screening and Diagnosis Methods . . . . .                   | 30        |
| 1.3.4    | Limitations of Existing Methods . . . . .                   | 32        |
| 1.3.5    | Advantages of Smartphone-based Screening . . . . .          | 33        |
| 1.4      | Digital Colour Imaging . . . . .                            | 34        |
| 1.4.1    | Digital Image Formation Pipeline . . . . .                  | 34        |
| 1.4.2    | Consistent Colour . . . . .                                 | 42        |
| 1.5      | Review of Image-based Jaundice Detection Research . . . . . | 46        |
| <b>2</b> | <b>Clinical Imaging Studies</b>                             | <b>51</b> |
| 2.1      | Overview . . . . .  | 51        |
| 2.2      | Clinical Datasets Used In This Thesis . . . . .             | 51        |
| 2.2.1    | Study 1: UCLH Nikon Study . . . . .                         | 51        |
| 2.2.2    | Study 2: UCLH Smartphone Study . . . . .                    | 53        |
| 2.2.3    | Study 3: DJAN Smartphone Study . . . . .                    | 56        |
| 2.2.4    | Comparison of Datasets . . . . .                            | 62        |
| 2.2.5    | Objectives of Clinical Dataset Analyses . . . . .           | 62        |

|          |  |           |
|----------|--|-----------|
| <b>3</b> | <b>Development of the neoSCB App</b>   | <b>64</b> |
| 3.1      | Overview . . . . .   | 64        |
| 3.2      | Challenges To Be Addressed . . . . .   | 64        |
| 3.2.1    | Image Capture . . . . .  | 65        |
| 3.2.2    | Image Processing: ROI Selection . . . . .  | 66        |
| 3.2.3    | Image Processing: Extracting Colour Information . . . . .                          | 67        |
| 3.2.4    | Jaundice Level Estimation . . . . .  | 68        |
| 3.3      | Ambient-subtracted Sclera Chromaticity . . . . .                                   | 69        |
| 3.3.1    | Ambient Subtraction . . . . .  | 70        |
| 3.3.2    | Subtracted Signal-To-Noise Ratio . . . . .   | 72        |
| 3.3.3    | Chromaticity Space . . . . .   | 73        |
| 3.3.4    | Imaging The Sclera . . . . .   | 75        |
| 3.3.5    | Device Independence via One-Time Calibration . . . . .                             | 76        |
| 3.3.6    | Pipeline Overview . . . . .  | 78        |
| 3.4      | Proof Of Principle Experiments . . . . .   | 79        |
| 3.4.1    | Snooker Ball Demonstration . . . . .   | 79        |
| 3.4.2    | Jaundice Eye Colour Index . . . . .  | 81        |
| 3.4.3    | Ambient Subtraction Demonstration . . . . .  | 83        |
| 3.5      | App Development . . . . .  | 85        |
| 3.5.1    | Hardware Requirements . . . . .  | 85        |
| 3.5.2    | Development Process and Features . . . . .   | 87        |
| 3.5.3    | App Walk-Through . . . . .   | 88        |
| 3.5.4    | Safety Measurements . . . . .  | 90        |
| 3.6      | Summary . . . . .  | 95        |
| <b>4</b> | <b>Development of SCB Model to Predict Total Serum Bilirubin from Chromaticity</b> | <b>96</b> |
| 4.1      | Overview . . . . .   | 96        |
| 4.2      | Real-world Subtraction Evaluation . . . . .  | 96        |
| 4.2.1    | Methods . . . . .  | 98        |
| 4.2.2    | Implementation . . . . .   | 99        |

|          |   |            |
|----------|---|------------|
| 4.2.3    | Results . . . . .   | 100        |
| 4.2.4    | Discussion . . . . .  | 102        |
| 4.3      | Sclera and Skin Chromaticity . . . . .                            | 104        |
| 4.3.1    | TSB-chromaticity Correlations . . . . .                           | 106        |
| 4.4      | Modelling the Relationship between Chromaticity and TSB . . . . . | 107        |
| 4.4.1    | Simple Linear Regression . . . . .                                | 108        |
| 4.4.2    | Effect of Phototherapy . . . . .                                  | 109        |
| 4.4.3    | Multiple Linear Regression . . . . .                              | 112        |
| 4.5      | SCB Model Evaluation and Validation . . . . .                     | 113        |
| 4.5.1    | Effect of Segmentation . . . . .                                  | 113        |
| 4.5.2    | Effect of Repeats . . . . .                                       | 116        |
| 4.5.3    | Performance Consistency . . . . .                                 | 117        |
| 4.5.4    | Validation on UCLH Data . . . . .                                 | 119        |
| 4.5.5    | Comparison with Literature . . . . .                              | 121        |
| 4.6      | Summary . . . . .   | 123        |
| <b>5</b> | <b>Screening Performance of the neoSCB App</b>                    | <b>124</b> |
| 5.1      | Overview . . . . .  | 124        |
| 5.2      | Screening Neonates at Risk . . . . .                              | 125        |
| 5.2.1    | Screening and Decision Thresholds . . . . .                       | 125        |
| 5.2.2    | ROC Curves . . . . .  | 126        |
| 5.2.3    | Discriminant Analysis Approach . . . . .                          | 127        |
| 5.2.4    | Comparison with Literature . . . . .                              | 129        |
| 5.3      | Comparison to TcB . . . . .                                       | 130        |
| 5.3.1    | Effect of Phototherapy . . . . .                                  | 133        |
| 5.4      | Comparison to Visual Inspection . . . . .                         | 134        |
| 5.5      | Applicability and Usability . . . . .                             | 137        |
| 5.5.1    | Real-time ROI Selection and Result . . . . .                      | 139        |
| 5.6      | Summary . . . . .   | 141        |
| <b>6</b> | <b>General Conclusions</b>  | <b>142</b> |

|   |            |
|---|------------|
| <b>Appendices</b>                         | <b>150</b> |
| <b>A Data Collection Sheets</b>           | <b>150</b> |
| <b>B neoSCB User Guide</b>                | <b>153</b> |
| <b>C UCLH Smartphone Study Statistics</b> | <b>158</b> |
| <b>Bibliography</b>                       | <b>160</b> |

# List of Figures

|     |  |    |
|-----|--|----|
| 1.1 | Newborn undergoing phototherapy. . . . .   | 28 |
| 1.2 | NICE treatment thresholds for jaundice. . . . .  | 29 |
| 1.3 | Flowchart for the screening and diagnosis of jaundice. . . . .                           | 31 |
| 1.4 | Example of a RGB camera spectral sensitivity profile. . . . .                            | 35 |
| 1.5 | Molar absorption coefficients of skin chromophores. . . . .                              | 37 |
| 1.6 | Essential steps of raw camera data processing pipeline. . . . .                          | 39 |
|     |  |    |
| 2.1 | Example image captured in UCLH Nikon Study. . . . .                                      | 52 |
| 2.2 | Example image pair captured by app in UCLH Smartphone Study. . .                         | 54 |
| 2.3 | Typical image capture scenario for Phase 1a of DJAN study. . . . .                       | 57 |
| 2.4 | Phases of the DJAN study. . . . .  | 59 |
| 2.5 | Flowchart showing number of included capture sessions in DJAN<br>study. . . . .          | 59 |
| 2.6 | Histogram of TSB values for DJAN data. . . . .   | 60 |
| 2.7 | Histograms of gestational age, postnatal age, and birth weight for<br>DJAN data. . . . . | 61 |
|     |  |    |
| 3.1 | Diagram of the human eye. . . . .  | 76 |
| 3.2 | Pipeline for calculation of ambient-subtracted sclera chromaticity. . .                  | 79 |
| 3.3 | RGB channel images and relative contrast for five painted balls. . . .                   | 80 |
| 3.4 | Snooker ball line profiles in blue channel and blue chromaticity. . . .                  | 81 |
| 3.5 | LG Nexus 5X with screen illumination on and off. . . . .                                 | 83 |
| 3.6 | Rendering of printed JECI patch colours. . . . .   | 84 |
| 3.7 | Ambient subtraction proof of principle for a range of JECI values. . .                   | 85 |



|      |  |     |
|------|--|-----|
| 3.8  | Picture of flash diffuser fitted to Samsung S8. . . . .  | 86  |
| 3.9  | Screenshots of splash, welcome, and camera preview screens of neoSCB app. . . . .                  | 89  |
| 3.10 | Screenshots of review screen of neoSCB app. . . . .  | 90  |
| 3.11 | Action spectra for retinal hazard. . . . .   | 92  |
| 3.12 | Samsung S8 LED flash relative effective irradiance with and without diffuser. . . . .              | 93  |
| 4.1  | Picture of X-Rite ColorChecker Passport. . . . .   | 98  |
| 4.2  | White balance comparison via mean chromaticity error per patch. . . . .                            | 101 |
| 4.3  | Neonatal sclera and sternum chromaticity in xy chromaticity space via colour card. . . . .         | 104 |
| 4.4  | Neonatal sclera and sternum chromaticity in xy chromaticity space via ambient subtraction. . . . . | 105 |
| 4.5  | Sclera xy chromaticity for 595 neonates. . . . .   | 108 |
| 4.6  | y chromaticity against TSB for phototherapy and non-phototherapy treated neonates. . . . .         | 110 |
| 4.7  | Illustration of sclera ROI segmentation methods. . . . .   | 114 |
| 4.8  | Bland-Altman plot for DJAN study. . . . .  | 118 |
| 4.9  | SCB correlation with TSB for UCLH data. . . . .  | 120 |
| 4.10 | Bland-Altman agreement plot for SCB and TSB for UCLH data. . . . .                                 | 121 |
| 4.11 | SCB correlation with TSB for three-repeat DJAN data. . . . .                                       | 123 |
| 5.1  | SCB and TSB for DJAN data. . . . .   | 126 |
| 5.2  | ROC curve for neoSCB from DJAN data. . . . .   | 127 |
| 5.3  | Discriminant analysis in xy chromaticity space. . . . .  | 128 |
| 5.4  | ROC curve for three-repeat neoSCB from DJAN data. . . . .  | 130 |
| 5.5  | Correlation between TSB and JM-105 TcB values. . . . .   | 131 |
| 5.6  | Bland-Altman plots for neoSCB and JM-105 data. . . . .   | 134 |
| 5.7  | Box plots of TSB for responses to visual assessment question. . . . .                              | 136 |

5.8 Histogram of time taken for three repeats and real-time result generation. . . . . 138

5.9 Real-time SCB vs TSB. . . . . 140

5.10 Real-time SCB vs offline segmentation SCB. . . . . 141

# List of Tables

|     |  |     |
|-----|--|-----|
| 1.1 | Comparison of visual inspection and transcutaneous bilirubinometer as screening methods. . . . . | 33  |
| 2.1 | Comparison of phases of DJAN study. . . . .  | 58  |
| 2.2 | Comparison of clinical study datasets. . . . .   | 62  |
| 4.1 | Performance metrics for six white balance and ambient subtraction methods. . . . .               | 101 |
| 4.2 | Sclera chromaticity correlation with TSB. . . . .  | 106 |
| 4.3 | Skin chromaticity correlation with TSB. . . . .  | 107 |
| 4.4 | Simple linear models for TSB prediction. . . . .   | 109 |
| 4.5 | Single explanatory variable correlation for phototherapy and non-phototherapy groups. . . . .    | 111 |
| 4.6 | Multiple linear models for TSB prediction. . . . .   | 112 |
| 4.7 | TSB prediction performance for jaundice detection apps compared. .                               | 122 |
| 5.1 | Screening performance for jaundice detection apps compared. . . .                                | 129 |
| 5.2 | neoSCB and JM-105 confusion matrices. . . . .  | 132 |
| 5.3 | Screening performance for jaundice detection apps and TcBs compared. . . . .                     | 132 |
| 6.1 | Summary of future work. . . . .  | 148 |

# List of Abbreviations

|        |   |
|--------|---|
| AIB    | Automated Image-based Bilirubin                               |
| ANNP   | Advanced Neonatal Nurse Practitioner                          |
| AUC    | Area Under Curve  |
| CCD    | Charge-Coupled Device   |
| CCT    | Correlated Colour Temperature                                 |
| CI     | Confidence Interval   |
| CIE    | Commission Internationale de l'Eclairage                      |
| CMOS   | Complementary Metal Oxide Semiconductor                       |
| CSS    | Camera Spectral Sensitivity                                   |
| DJAN   | Detecting Jaundice in African Newborns                        |
| DNG    | Digital Negative  |
| DSLR   | Digital Single-Lens Reflex                                    |
| G6PD   | Glucose-6-Phosphate Dehydrogenase                             |
| GPS    | Global Positioning System                                     |
| HSL    | Hue-Saturation-Lightness                                      |
| ICNIRP | International Commission on Non-Ionizing Radiation Protection |

|        |   |
|--------|---|
| IFE    | Image Formation Equation                          |
| IR     | Infrared  |
| ISO    | International Organization for Standardization    |
| JECI   | Jaundice Eye Colour Index                         |
| JPEG   | Joint Photographic Experts Group                  |
| LED    | Light-Emitting Diode                              |
| LEDC   | Less Economically Developed Country               |
| MAE    | Mean Absolute Error                               |
| NEF    | Nikon Electronic Format                           |
| neoSCB | neonatal Scleral-Conjunctival Bilirubinometer     |
| NHS    | National Health Service                           |
| NICE   | National Institute for Health and Care Excellence |
| NICU   | Neonatal Intensive Care Unit                      |
| RBC    | Red Blood Cell                                    |
| RGB    | Red-Green-Blue                                    |
| ROC    | Receiver Operating Characteristic                 |
| ROI    | Region Of Interest                                |
| SCB    | Scleral-Conjunctival Bilirubin                    |
| SE     | Standard Error                                    |
| SNR    | Signal-to-Noise Ratio                             |
| SPD    | Spectral Power Distribution                       |

|      |  |
|------|--|
| SSNR | Subtracted Signal-to-Noise Ratio                         |
| TcB  | Transcutaneous Bilirubinometer/ Transcutaneous Bilirubin |
| TIFF | Tagged Image File Format                                 |
| TLS  | Total Least Squares                                      |
| TSB  | Total Serum Bilirubin                                    |
| UCL  | University College London                                |
| UCLH | University College London Hospitals                      |
| UCS  | Uniform Chromaticity Space                               |
| UV   | Ultraviolet  |

## **Chapter 1**

# **Introduction**

### **1.1 Motivation and Aims**

Smartphones are now used in a variety of healthcare contexts. At the time of writing, some national governments are using smartphones to suppress the spread of the COVID-19 pandemic by tracking exposed individuals via GPS. Altogether less topical and controversial, many smartphones count steps and report the daily totals as a default behaviour. Yet smartphones were originally conceived neither as tools to monitor the spread of disease nor as pedometers. These innovations made use of existing smartphone functionalities – the GPS unit and the accelerometer – to deliver solutions where medical devices and healthcare infrastructure were not.

Digital colour imaging is now used for various medical purposes such as managing wound healing [7], burn severity assessment [8], melanoma assessment [9], anaemia detection [10], and a variety of eye examinations [11] [12]. There has been a trend towards quantitative image-based diagnostics, where once only qualitative assessment was used. Until recently only a dedicated digital camera would have the quality required to capture useful clinical images. Now smartphone image quality has improved significantly, providing opportunities for point-of-care imaging solutions. For example, urine-based strip tests can be read and interpreted in the home setting using a smartphone camera application [13].

The aim of this thesis is to develop and test an app to measure neonatal jaundice severity so it can serve as a screening tool. The approach is based on the

quantification of the yellow discolouration in the sclera and skin that characterises jaundice. The risk of death and disability from jaundice rises the longer the condition goes untreated, so early identification is essential [14]. In this thesis it is argued that a smartphone-based method could provide an alternative to both visual inspection and expensive, purpose-built devices, going some way towards addressing the unmet need for accessible and objective jaundice screening.

The scalability of smartphone-based healthcare solutions gives them a great potential for impact. Smartphones are now ubiquitous: Almost half of the world's population own one, including a substantial fraction in the least economically developed countries [15]. This means people who have limited access to healthcare resources in less economically developed countries (LEDCs) can benefit from smartphone-based innovations. Jaundice affects LEDCs disproportionately, with a greater incidence of long-term impairment and death [16]. Throughout this thesis, the proposed solutions are considered in the light of how scalable they would be in a low-resource setting, with preference given to approaches that require minimal add-ons to the smartphone itself.

Smartphones are not a panacea. Smartphone hardware and software can be difficult to use due to heterogeneity between different makes and models [17]. For this reason, some of the most well-established smartphone healthcare applications use the device as a datalogger and interface for a purpose-built external sensor, such as a continuous glucose monitor. The question of how to achieve device independent results while using data from the inbuilt camera is therefore central to this thesis.

Sclera-based screening apps have been studied in adult populations, but little attention has been directed towards imaging the newborn sclera. In this work, the feasibility and predictive utility of sclera colour quantification is compared to skin colour quantification.



## 1.2 Thesis Objectives and Overview

The objectives of this thesis are:

1. Identify the desirable features of an image-based jaundice detection app and the key challenges to be overcome
2. Propose an imaging solution and jaundice metric that satisfies these criteria
3. Build an app that implements this solution
4. Establish the app is feasible to use in a clinical settings by a pilot study
5. Test the colorimetric accuracy of the app
6. Determine whether the sclera or skin is a better site from which to predict total serum bilirubin (TSB), and whether the sclera-based approach is affected by the phototherapy history of the newborn
7. Collect data in a larger population and use the images and measured TSB values to train and test a model to predict TSB from the jaundice metric
8. Investigate the factors affecting TSB prediction performance
9. Calculate the screening performance of the app in identifying newborns with TSB greater than a clinically relevant threshold
10. Compare the app screening performance to that of the transcutaneous bilirubinometer

In the rest of this chapter, the background to this work is presented. Firstly, the problem of neonatal jaundice is described. Current methods of jaundice screening are introduced and their limitations are discussed, motivating the need for a smartphone-based solution. Following this, there is an overview of the process by which a digital image is produced and how colour information is derived from images. Finally, current research efforts towards image-based jaundice detection are reviewed.

In Chapter 2, the three clinical studies used in this thesis are described. Details are provided of the scopes, enrolment processes, imaging technologies, and clinical data collected. The datasets are compared to each other. Lastly, there is a summary of how each study is applied to address the above objectives.

In Chapter 3, the thought process behind the development of the neonatal scleral-conjunctival bilirubinometer (neoSCB) screening app is explained. The challenges related to screening for jaundice using a smartphone camera are outlined (Objective 1). Based on these considerations, a novel method based on ambient-subtracted sclera chromaticity is suggested (Objective 2). The principles of the method are demonstrated on clinical and non-clinical image data. Finally, the smartphone hardware requirements are outlined and the neoSCB app is described (Objective 3).

In Chapter 4, a model relating total serum bilirubin (TSB) to ambient-subtracted sclera chromaticity is sought using data from a pilot study at UCLH (Objective 4) and a larger study conducted in Ghana (Objective 7). First, the ambient subtraction method is compared to some white balance methods on colour card data collected in a clinical environment (Objective 5). Skin chromaticity on the sternum is compared to sclera chromaticity to determine which is a better predictor of neonatal TSB (Objective 6). The variability due to segmentation and between repeat measurements is quantified. The effects of gestational age, phototherapy history, and other clinical factors on prediction accuracy are investigated (Objective 8). Once the model is selected, it is trained on the Ghana dataset and tested on the UCLH dataset. The performance is compared to other studies that have sought image-based TSB prediction.

In Chapter 5: Screening performance, the neoSCB app screening performance is quantified (Objective 9). It is compared to transcutaneous bilirubinometry (Objective 10). The real-world applicability of the app and possible next steps are discussed.

## 1.3 Jaundice Background

### 1.3.1 What is Jaundice?

Jaundice is caused by a build-up of bilirubin in the body, which manifests as a yellow discolouration of the skin and whites of the eyes (sclera). Bilirubin is a yellow pigment that occurs naturally in the body as a breakdown product of heme, which is present in red blood cells (RBCs). When a RBC is first broken down, the bilirubin is in an unconjugated, fat-soluble form and is found bound to albumin. In the liver, the bilirubin becomes conjugated and thus water-soluble. In this form it can be excreted. Anything that disrupts or overloads this pathway can cause jaundice, as bilirubin fails to be cleared from the body quickly enough. For example, adult patients with liver problems, such as alcoholic liver disease or viral hepatitis, may present with jaundice because the conjugation process is impaired.

The focus of this work is neonatal jaundice. Jaundice is very common in neonates, with around 60% of term and up to 80% of premature neonates affected in the first week of life [14]. There are several contributing factors: newborns naturally have a higher rate of hemolysis (RBC breakdown), as fetal RBCs have shorter lifespans; the enzyme that conjugates bilirubin in the liver takes some time to gain function after birth; bilirubin is re-absorbed into circulation by the gut; bruising during childbirth is common, creating an greater load of hemolysis [18]. In the majority of cases, this normal, so-called physiological jaundice resolves without any need for intervention [14]. Pathological jaundice refers to jaundice that results from an underlying health condition, which can lead to a more rapidly rising level of jaundice and a greater risk of adverse outcomes. Examples of such conditions include ABO blood group incompatibility between mother and newborn, Rhesus disease, and inherited deficiency in the Glucose-6-Phosphate Dehydrogenase (G6PD) enzyme, all of which increase the rate of hemolysis. Unconjugated bilirubin is neurotoxic, and it can cross the blood-brain barrier if levels are elevated for a prolonged period. In the short term, this can lead to acute bilirubin encephalopathy. Symptoms of bilirubin encephalopathy include arching of the back, listlessness, and high-pitched crying. If untreated, permanent damage to the brain can ensue, which is called kernicterus.

Death or lifelong disability such as cerebral palsy or deafness can result [19].

In the most severe cases, exchange transfusion is used to urgently reduce the concentration of bilirubin in the blood. In moderate cases, the safer, non-invasive option of phototherapy is preferred. Figure 1.1 shows a baby undergoing phototherapy. The eyes are shielded from the light.

Phototherapy involves shining blue light in the wavelength range 460-490nm onto the exposed skin of the baby. While bilirubin absorbs most strongly around 460nm, the longer wavelengths penetrate deeper into the tissue. These wavelengths affect bilirubin by the processes of photoisomerisation and photodegradation, with the former being more significant [20]. The isomers created can be excreted by the body in bile and urine, bypassing the need for conjugation in the liver [21]. These isomers appear in the blood soon after phototherapy is applied, but it may be some time before all isomers are cleared from the superficial tissues. While isomers are cleared from the skin, they are simultaneously reabsorbed from the blood, along with unconjugated bilirubin, until a new equilibrium is reached [21].

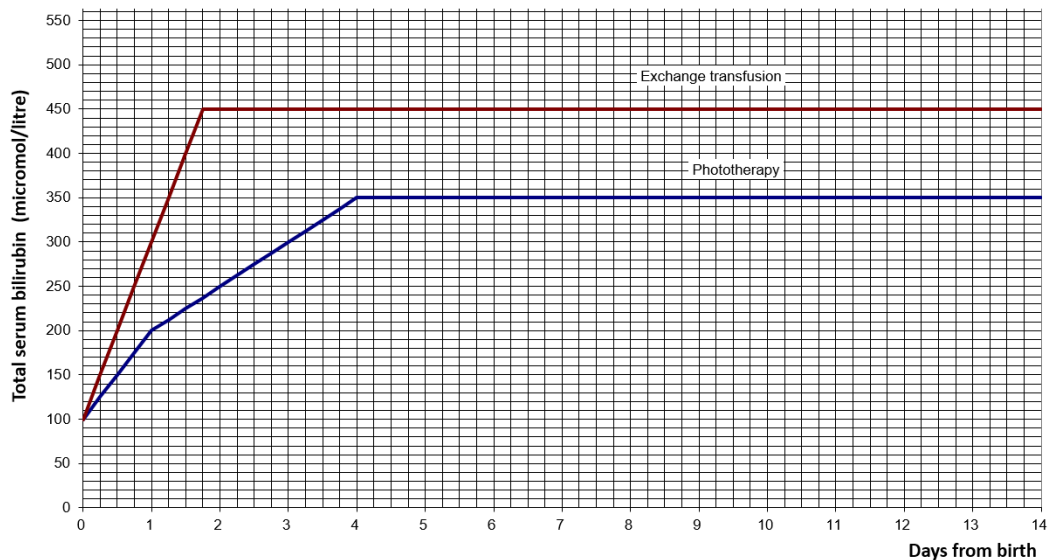
The isomers of bilirubin created during phototherapy also have a yellow colour. As they are cleared from the superficial tissue, the yellowness fades, a process some-



**Figure 1.1:** Newborn undergoing phototherapy. “phototherapy” by Jennifer Pack is licensed with CC BY-NC 2.0 (<https://creativecommons.org/licenses/by-nc/2.0/>).

times referred to as “bleaching” [22]. This means that contact-based estimations by optical devices such as the transcutaneous bilirubinometer develop a discrepancy with total serum bilirubin (TSB) in the hours after phototherapy [22] [23] [24]. Studies showed this discrepancy is absent or less pronounced in skin areas covered during phototherapy, which provides more evidence for a bleaching effect on exposed skin [25] [26] [27] [28] [29].

Treatment threshold graphs such as the one in Figure 1.2 are produced by the National Institute of Health and Care Excellence (NICE). These indicate the recommended treatment based on the neonate age and bilirubin level. Premature babies are vulnerable to excess bilirubin (also known as hyperbilirubinaemia) at lower concentrations. Therefore, different thresholds are applied depending on the gestational age of the neonate [14].



**Figure 1.2:** Treatment thresholds for a gestational age of 38 weeks. For babies in the first 4 days of life, thresholds for treatment are lower. Graph from NICE Clinical Guideline 98 (2010) [14].

### 1.3.2 Disease Burden

According to Bhutani et al. [16], worldwide in 2010, hyperbilirubinaemia was responsible for 114,000 deaths, 62,000 cases of developmental delay, 64,000 cases of hearing loss, 35,000 cases of athetoid cerebral palsy, and 18,000 individuals affected with other neurological disabilities. In less economically-developed coun-

tries (LEDCs), the disease burden due to jaundice is higher. Travel to healthcare facilities may be expensive or arduous, fewer community healthcare professionals such as midwives are available, and home births are more common [30] [31]. These three factors mean more cases of jaundice go undiagnosed, which greatly increases the risk of permanent disability. Furthermore, jaundice-related disability imposes a higher socio-economic cost on families in LEDCs than in developed nations [32] [33]. One of the most affected regions in the world is sub-Saharan Africa. In addition to the reasons mentioned above, jaundice can be particularly difficult to notice in Black African neonates as their skin pigmentation is darker. G6PD enzyme deficiency is also more common in this region. This inborn metabolic condition causes red blood cells to be broken down at an increased rate, which increases the chance of jaundice [34]. In Ghana, Adei-Atiemo et al. identified hyperbilirubinaemia as “the most significant and preventable risk factor for cerebral palsy” [35].

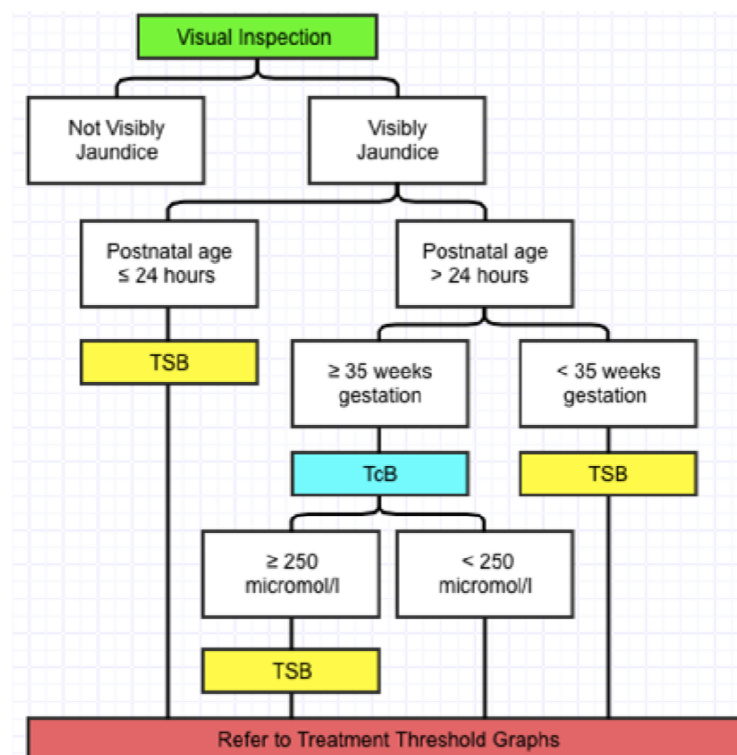
### 1.3.3 Screening and Diagnosis Methods

The gold standard test for diagnosis of jaundice is a blood test to determine the total serum bilirubin (TSB), a concentration measured in micromoles per litre ( $\mu\text{mol/L}$ ) or milligrams per decilitre (mg/dL). A heel lance is used to extract a blood sample. This can be painful for babies and distressing for parents. The need to travel to a hospital to carry out such a blood test also makes them impractical for routine jaundice monitoring, especially in remote areas. Furthermore, the result is not immediately available, and lab measurement turnaround times can vary considerably. As it is neither possible nor desirable to monitor bilirubin levels via blood tests alone, neonates must be regularly screened for jaundice in other ways. Most commonly this is done via a visual inspection, ideally by a midwife or doctor. The NICE clinical guidelines advise new parents to check for any yellow discolouration under a bright, preferably natural source of light by pressing lightly on the skin. In babies with darker skin tones, it is recommended to inspect the gums and whites of the eyes for signs of jaundice [36].

A transcutaneous bilirubinometer (TcB) is a hand-held optical device that, when pressed to the skin of the forehead or chest, can provide an estimate for TSB.

The reflectances of specific wavelengths are measured to infer the optical density of bilirubin in the skin. In the UK, the JM-103/5 (Draeger UK) and the Bilichek (Philips Healthcare) are the two TcBs available for clinical use. The JM-103/5 measures reflectance of 450nm and 550nm light. It mitigates the effect of other skin chromophores (primarily melanin and hemoglobin) using a dual optical pathway system. This is possible because the longer path length provides more information about the deeper, subcutaneous layer (where bilirubin accumulates), while the shorter path length provides more information about the dermal and epidermal layers (where the other skin chromophores predominate). The Philips Bilichek uses reflectance data from 137 different wavelengths in the visible range to resolve an estimate of concentrations of all skin chromophores [37].

TcBs are useful screening tools because they are non-invasive and give an on-the-spot, objective indication of jaundice severity. The TcB reading is in the same units as the TSB ( $\mu\text{mol/L}$  or  $\text{mg/dL}$ ). It is compared to a clinically relevant TSB



**Figure 1.3:** Flowchart for the screening and diagnosis of jaundice. Reproduced with permission from Ref. [38].

threshold to determine whether or not the baby requires a blood test. Therefore, even though a numerical reading is provided by the TcB, the outcome is a binary screening decision.

Figure 1.3 summarises the decision tree currently used in the UK for screening and diagnosis of jaundice. If jaundice is visible, babies at home are referred to a hospital. Particularly at-risk groups (premature or less than one day old) are given blood tests at once.

A positive TcB screening result is always followed by a blood test before treatment is recommended. The accuracy of the TcB is inferior to measurement of TSB; TSB is the gold standard to which all screening techniques must be compared.

### **1.3.4 Limitations of Existing Methods**

Visual inspection is subjective and unreliable, and thus should not be relied upon as the sole method for identifying jaundiced babies [36] [39] [40] [41]. Visual identification of jaundice is especially difficult for inexperienced individuals such as parents, and it is even harder to identify jaundice in babies with darker skin complexions (for both professionals and parents).

TcBs have a few important disadvantages. They are expensive, at approximately £3,400 per device (NHS costing report, 2010 [36]). As a result, they are not always available to midwives visiting in the community. In resource-poor settings TcBs are prohibitively expensive. Furthermore, it is not well-established that TcBs are equally effective for all ethnicities, with the literature split [34] [36] [42] [43]. Olusanya et al. reported an overestimation of TSB in Black African neonates for both the JM-103 and Bilichek TcBs, which could lead to unnecessary treatment [34].

Table 6.1 summarises the accuracy and accessibility of visual inspection and the TcB. As a standalone screening approach, visual inspection lacks accuracy, while the TcB is not sufficiently accessible. There is thus a need for an inexpensive, accessible, objective screening method, especially in a global context.



**Table 1.1:** Comparison of visual inspection and transcutaneous bilirubinometer.

| Method                               | Accuracy vs TSB  | Accessibility  |
|--------------------------------------|--|--|
| Transcutaneous Bilirubinometer (TcB) | Validated accuracy in term and near-term neonates up to 250 $\mu$ mol/L – TcBs are effective as an objective screening tool at this threshold [36]                   | Low availability, especially outside of clinical setting; High unit cost       |
| Visual Inspection                    | Not easily quantified due to subjectivity; High negative predictive value when performed correctly [39] [44]; Poor in identifying clinical severity of jaundice [44] | No equipment needed; Training required in best practices for visual inspection |

### 1.3.5 Advantages of Smartphone-based Screening

In the last few years, smartphone manufacturers have allowed developers to access raw camera data, opening the door to point-of-care diagnosis based on the measurement of colour. At the same time, hardware and software advances have increased the quality of images captured with smartphones.

If smartphone cameras can be leveraged to provide a new screening method for jaundice, the potential impact is large. As discussed in Section 1.3.2, sub-Saharan African populations suffer disproportionately from jaundice-related mortality and disability. Current screening tools are not available or unfit, and there is less contact with trained healthcare workers.

A smartphone-based screening method would offer several advantages over existing methods. Like TcBs, they are objective, portable, non-invasive, and offer a result at the point-of-care. However, they are an order of magnitude cheaper than TcBs. They are also more accessible, as smartphones are becoming ubiquitous in even the most resource-poor settings. Finally, TcBs are contact-based, which increases the risk of spreading infection or the need for disposables, while digital photography can be completely contact-free.

The benefits of such a technique would be felt more keenly in LEDCs. In LEDCs, lab results can be unreliable or slow to arrive, which may delay time-critical treatment. In this situation, clinicians may err on the side of caution and

recommend phototherapy. Unnecessary phototherapy can cause dehydration and, when phototherapy units are in short supply, block access to treatment for other neonates [45].

## 1.4 Digital Colour Imaging

### 1.4.1 Digital Image Formation Pipeline

A digital image is an image discretised in both spatial coordinate and intensity [46]. It is stored as an array of integer values, which encode brightness level. In the case of a 2D greyscale image, the digital image can be stored as a 2D array. In the case of a 2D colour image, each colour channel has its own 2D array. RGB colour images have red, green, and blue colour channels, and can be stored as an  $m \times n \times 3$  array, where  $m$  and  $n$  are the dimensions of the image in pixels.

The process of digital image formation depends not only on the contents of the scene being imaged, but also on the nature of the light illuminating the scene, as well as the characteristics of the imaging device. Once an image is captured, further post-processing steps can be employed to improve the appearance of an image or emphasise certain features.

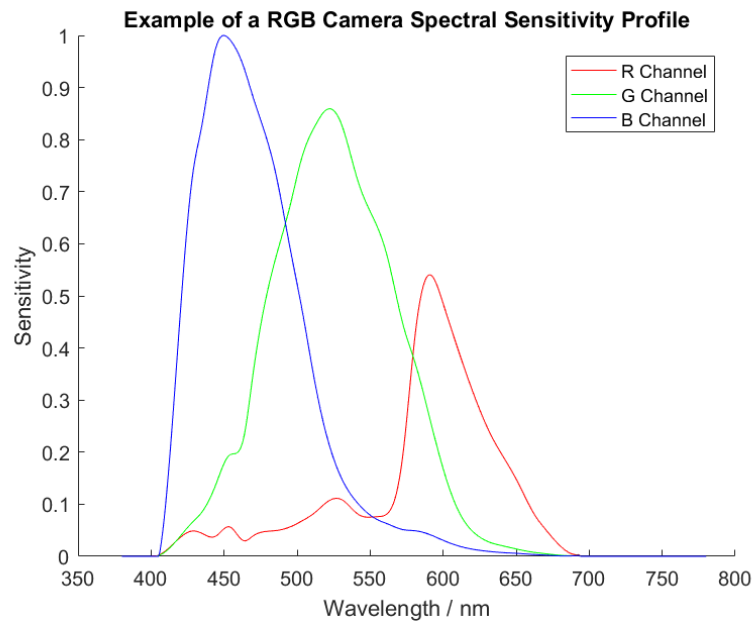
In this section, the key ingredients of digital image formation - camera, scene, lighting, and post-processing - will be explored in turn, followed by equations that can be used to model this process.

#### 1.4.1.1 The Camera

One of the most important characteristics of a RGB camera is the camera spectral sensitivity (CSS). The CSS relates scene radiance to measured RGB triplets. It can be thought of as the relative probability of a photon of a given wavelength triggering a R, G, or B count. The shape of the sensitivity curve depends on the sensitivity of the photosensor<sup>1</sup> and the transmittance of the optical components before it (most importantly the RGB colour filter array and the infrared (IR) cut off filter), which

---

<sup>1</sup>The great majority of smartphones employ complementary metal-oxide semiconductor (CMOS) image sensors for their cameras. CMOS sensors are chosen over charge-coupled device (CCD) sensors because they are cheaper, draw less power, achieve faster data-throughput rates, and allow on-chip processing such as noise reduction and analogue-to-digital conversion [47].



**Figure 1.4:** Example of a RGB camera spectral sensitivity profile (Nikon D3200). R, G and B channels are sensitive to different regions of the visible spectrum. CSS measured using LED colour target method of MacDonald et al. [48].

are all functions of the incident light's wavelength. Figure 1.4 shows an example CSS.

It is useful to know the CSS of a camera for applications including multispectral imaging, colour constancy, and camera simulation. Characterising the camera's response to light in this way is often a necessary step towards using it as a scientific or medical instrument.

The *radiometric response function* of a camera relates irradiance at the sensor to image pixel value. In the ideal case, this is a linear function up to some maximum value. This maximum value is the *saturation point*.

The *dynamic range* of a camera is the range of intensities the camera is capable of recording, from the dimmest (one standard deviation greater than the noise floor), to the brightest (the intensity that will saturate a sensor) [49].

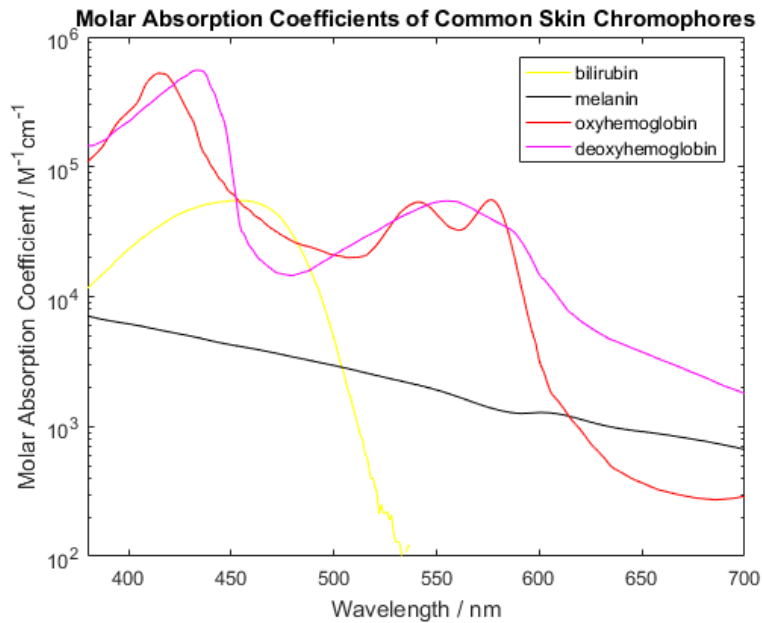
The *bit depth* of a camera is the number of distinct values that a pixel can take. At 12-bit, this number is  $2^{12} = 4096$ , at 14-bit, this number is  $2^{14} = 16384$ .

### 1.4.1.2 The Scene

The colour of an object in a digital image depends on the nature of the prevailing illumination. However, even under consistent illumination an object may not produce the same sensor response when imaged by two cameras if the CSS of one is not the same as the other. The stable intrinsic property of an object in a scene that affects its colour is its reflectance spectrum.

The reflectance spectrum of a material is the fraction of incident light reflected by that material for every wavelength of visible light. This depends on how photons are scattered and absorbed within the body of the material. Scattering is caused by inhomogeneities that cause variation in the refractive index at a microscopic scale. In biological tissue, visible light is scattered by organelles found in cells, such as mitochondria, and inhomogeneities in the extracellular matrix [50]. Absorption can occur when the energy of a photon matches the energy of an electronic transition from one molecular energy configuration to another. Molecules that absorb light in the visible region of the spectrum are called chromophores. Photons of different wavelengths carry different amounts of energy, so will be absorbed to greater or lesser degree by a given chromophore. In Figure 1.5, the molar absorption coefficients of melanin, oxyhaemoglobin, and deoxyhaemoglobin are plotted. These four biological chromophores are the most important in determining the colour of skin. Haemoglobin absorbs in the blue and green regions of the spectrum, so white light that passes through a haemoglobin-rich tissue becomes dominant in red. Similarly, bilirubin (responsible for jaundice) has a yellow colour because it absorbs strongly in the blue region of the spectrum.

An important distinction must be made between two types of reflected light: diffuse and specular. Diffusely reflected light enters the body of a material and is scattered multiple times before being re-emitted from the surface. This is the component of reflected light that is of diagnostic utility, because it contains information about the absorption and scattering characteristics of the tissue [52]. Specular reflection, on the other hand, describes light that is directly reflected from the surface of the material. It has approximately the same spectral power distribution as the



**Figure 1.5:** Molar absorption coefficients of skin chromophores: oxyhaemoglobin, deoxyhaemoglobin, melanin and bilirubin. Data from Ref. [51].

light source itself. It is highly directional, and more pronounced on glossy, wet, or metallic surfaces. By contrast, diffuse reflection occurs in all directions; a perfectly diffusive surface has the same luminance when viewed from any angle. This idealised surface is described as a Lambertian surface.

### 1.4.1.3 The Illumination

Common sources of light include daylight, incandescent lighting, fluorescent lighting, and light-emitting diodes (LEDs). Fluorescent lamps pass a current through a mercury vapour. The excited mercury produces ultraviolet (UV) light that causes the phosphor tube coating to glow [53]. Fluorescent lamps are commonly used in commercial and government buildings as they are more energy efficient than incandescent bulbs, emitting less power as heat.

LEDs exploit *electroluminescence*, the emission of light from a material through which a current is passed. Semi-conductors with energy band gaps in the visible spectrum are used for LEDs. White LEDs rely on a combination of this and a phosphor coating, like fluorescent lamps. To create a white light, a blue light LED and a phosphor that fluoresces yellow are combined.

Incandescent bulbs pass current through a wire filament to heat it to high enough temperatures to emit light. The sun is also an incandescent light source: it emits light due to its temperature. The spectrum emitted by an incandescent light source depends principally on its temperature, as well as on the material properties.

A *black body* is a conceptual material whose spectral power output depends only on its temperature. Plank's formula gives this relationship, which is derived by considering a body that absorbs all wavelengths of incident radiation in thermal and radiative equilibrium. The peak of the emission spectrum of a black body moves toward shorter wavelengths at higher temperatures.

The relationship between colour and temperature is not only a theoretical consideration. The *correlated colour temperature (CCT)* of a light source is the temperature of a black body whose spectrum appears most similar to the light source [53]. This can be used to describe the spectrum of any light source, even when that spectrum is very dissimilar to that of a black body (such as a fluorescent tube). Incandescent lights are relatively redder than most other sources (CCT approximately 2400-2600K), while direct sunlight is more white (CCT approximately 5000-6000K). Fluorescent lights are intermediate in CCT (2700-5000K).

Although CCT can be a useful description of light sources, it does not fully characterise the source. One object may appear to have the same colour under two different light sources, leading one to assume they are equivalent. However, an object with a different spectral reflectance may appear quite different under those same two light sources. This phenomenon is called *metamerism*.

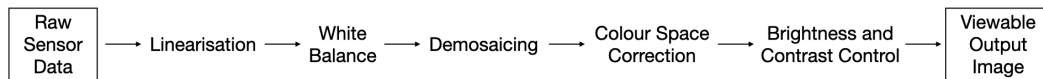
#### 1.4.1.4 File Formats and Post-processing

Various post-processing operations can be performed on raw image data, which is the data initially recorded by the sensor array. These frequently include compression (to reduce the file size for storage and transfer), gamma encoding (to make the range of available tones more perceptually uniform), and white balancing (in order to render neutral objects as neutral, rather than as the prevailing scene illumination colour). JPEG images are created using all of these post-processing steps.

A raw format file contains all the information captured by the sensor. For scien-

tific applications, the goal is to exploit this information rather than create a visually appealing image. Therefore, a lossless rather than lossy file format is preferred. JPEG is an example of a lossy file format because compression takes place.

Figure 1.6 summarises the essential steps for processing raw image data in a lossless manner without adding non-linearity.



**Figure 1.6:** Essential steps of raw camera data processing pipeline [54].

**Linearisation** Some digital cameras may store image data using a non-linear transform. The inverse transform must first be applied to recover linear data.

**White Balance** R and B values are scaled relative to the G value such that objects of neutral colour (grey or white) appear neutral in the image.

**Demosaicing** The raw data is originally greyscale because each sensor element has either a red, green, or blue colour filter. Pixel values for the other two channels must be interpolated from neighbouring pixels.

**Colour Space Correction** Camera spectral sensitivity functions vary so RGB values are not comparable between devices until a transform is applied to bring the triplets into a common colour space. This transform can be summarised by a 3x3 matrix or lookup table.

**Brightness and Contrast Control** This step is only necessary for viewing purposes. A global scaling can make the image brighter. A non-linear transform (gamma) can be used to make detail in dark areas more visible but should not be used if a linear output is needed.

#### 1.4.1.5 Image Formation Equation

The image formation process can be summarised in an equation. In the simplest case: the scene imaged is composed of flat, coplanar objects (the “Mondrian World”

assumption, after the artist); the scene is uniformly illuminated by a single illumination; surfaces are Lambertian (perfectly diffusive). In this case, the pixel value for channel  $k$  at position  $x$ ,  $I_{k,x}$ , is given by Equation 1.1.  $E(\lambda)$  refers to the uniform illumination spectrum,  $S(\lambda, x)$  refers to the object reflectance spectrum at position  $x$ , and  $Q_k(\lambda)$  refers to the CSS of channel  $k$ , where  $k=R,G,B$ . The integral is over the entire spectrum  $\omega$ .

$$I_{k,x} = \int_{\omega} E(\lambda)S(\lambda, x)Q_k(\lambda)d\lambda \quad (1.1)$$

Equation 1.1 can be developed to describe more general situations. In general, illumination is not uniform; the term for illumination may be position-dependent,  $E(\lambda, x)$ . An extra term can be added to represent the effect of specular reflection. In this so-called *dichromatic model* of image formation, reflection is modelled as a linear combination of diffuse and specular components [55]. For the specular component, the surface reflectance term  $S$  can be assumed to be wavelength-independent for most dielectric inhomogeneous objects (the neutral interface reflection assumption [56]). Therefore, this term can be incorporated into a position-dependent factor outside the second integral. This factor,  $w_s(x)$ , also encapsulates the geometrical effect of viewing angle, surface orientation, and incident light direction on the intensity of the specular term. A similar scaling,  $w_d(x)$ , for the diffusive term also depends on surface orientation and incident light direction, but not on viewing angle because this component is Lambertian by construction [57]. Equation 1.2 includes these extensions, and will be referred to as the Image Formation Equation (IFE) hereafter.

$$I_{k,x} = w_d(x) \int_{\omega} E(\lambda, x)S(\lambda, x)Q_k(\lambda)d\lambda + w_s(x) \int_{\omega} E(\lambda, x)Q_k(\lambda)d\lambda \quad (1.2)$$

It is important to note that this equation still falls short of encapsulating all the effects that lead to the creation of an image. For example, it does not countenance more than one distinct source of light, reflections of light from one surface to



another (interreflections), or the presence of noise.

#### 1.4.1.6 Noise

The *signal-to-noise ratio* (SNR) is an important consideration in any measurement.

*Photon noise*, also known as *shot noise*, is inherent to any measurement of light, regardless of the equipment used. This is because it is a feature of the light itself: light is quantised in packets called photons, and the arrival of photons is a random process. The Poisson distribution describes the variation in arrival time. This distribution is used for any process where events occur at a fixed rate per unit time whilst also being independent of each other. In the case of light, the expected rate of arrival of photons is proportional to the irradiance. If the average rate of arrival is  $\lambda$ , then over time interval  $t$  the expected number of photons incident on a sensor element is  $\lambda t$ , and the discrete probability distribution for the number of incident photons  $N$  is given by Equation 1.3.  $k$  is any non-negative integer.

$$Pr(N = k) = \frac{e^{-\lambda.t}(\lambda.t)^k}{k!} \quad (1.3)$$

One feature of a Poisson distribution is that the expected value is equal to the variance. Therefore, the standard deviation is proportional to the square root of the expected value: when measuring light, the photon noise is signal-dependent. This means that even in the absence of any other noise contributions, the photon noise gives us a fundamental upper limit for the signal to noise ratio, which is proportional to the square root of the intensity of the signal. Hence, when photon noise is the dominant source of noise, longer exposure times result in images with better SNR [58].

In the case of large photon counts, the central limit theorem allows photon (Poisson) noise to be accurately approximated by a Gaussian distribution with mean and variance each given by  $\lambda t$  [59] [60]. For captures involving smaller photon counts, other sources of image noise become relatively more important [58].

*Thermal noise*, or *dark current*, refers to the signal generated within the bulk of the sensor that is not caused by incident light. Dark current is produced as a result

of thermal fluctuations, which cause electron-hole pairs to form and thus add noise to the output voltage [61]. Like photon noise, these thermal electrons are generated according to Poisson statistics. Longer exposure times and higher temperatures are associated with increased dark current noise.

One other important source of noise is *read noise*. This encompasses noise due to converting the photon-generated charge to voltage, amplifying, and digitising the signal. For charge-coupled device (CCD) sensors, read noise is the same for every pixel, as charges are read out in series through the same output. For complementary metal-oxide semiconductor (CMOS) sensors, each column - or sometimes even each pixel - has its own readout structure, so charges can be converted in parallel. This means frame rates for CMOS sensors are faster, and that read noise is not the same for every pixel. Instead, read noise is best described by a distribution, normally characterised by its root-mean-square noise value.

## 1.4.2 Consistent Colour

Thus far the process of image formation has been explored through the factors that affect the recorded digital image. In this section, the concepts used to define and compare different colours are introduced, as well as some approaches to achieving repeatable colour in practice by accounting for device- and illumination-specific effects.

### 1.4.2.1 Colour Spaces and Transforms

When an image is first recorded by the camera sensor, it is said to be in the camera's internal or raw space. The values cannot be interpreted unless something is known about the way the camera records light. They are device dependent. Device independent colour spaces are required to compare between device outputs and display colours.

The CIE (International Commission on Illumination) 1931 XYZ colour space is both device independent and a "reference" colour space, meaning it can represent all colours visible to the typical human eye. It relies on measurements of the average human response to colour stimuli taken by Wright and Guild in the 1920's [62]. The

CIE used these data to define a standard observer, from which the XYZ colour space results. Many other colour spaces exist, but most are defined with direct reference to the CIE 1931 model. The XYZ space can be used as an intermediate step when transforming between camera raw values and display colour spaces such as sRGB.

To convert between camera raw space and XYZ space it is common to use a mapping. Such a mapping can be developed using a set of RGB values and corresponding XYZ values. These can either be simulated, if the camera CSS is known, or measured by imaging a colour card with known XYZ values. All mappings introduce some degree of error due to the fact CSS's cannot be described as linear combinations of the human cone sensitivities.

Importantly, such a mapping is only optimal for the illumination under which the colour card was imaged: Using a mapping developed under one illumination on image data capture under a different illumination will give sub-optimal results. For this reason, mappings under two or more illuminants are often provided in image metadata for use by colour space conversion algorithms.

#### 1.4.2.2 Standard Illuminants and White Points

In addition to the standard observer, the CIE also define standard illuminants. These are spectra used to represent common and useful lighting conditions. Illuminant E has equal energy in all parts of the visible spectrum. Illuminant series D represent natural daylight, with D65 most similar to noon light (CCT of approximately 6500K) and D50 described as horizon light (CCT of approximately 5000K).

The white point of an illuminant is given by the colour values assigned to an object that reflects equally in all parts of the visible spectrum. It is necessary to define white in this way because human vision exhibits *colour constancy*. This is the observation that we adapt to our lighting environment so that objects appear to have a stable colour under a wide range of illuminations (under the warm glow of a candle or the cool illumination of an overcast day, a sheet of white paper always appears white). Cameras do not have such a facility, which is why white balancing is required.

Chromatic adaptation transforms are used to emulate the chromatic adaptation

that gives us colour constancy. These are used to estimate the colour that would be recorded under a different illumination than the one used to capture the image. For example, if an object's colour under a D50 illuminant is specified by one triplet XYZ(D50), a chromatic adaptation transform could provide an estimate for XYZ(D65), the colour under a D65 illuminant.

### 1.4.2.3 Achieving Consistent Colour

We have now seen why it is necessary to account for both CSS and illumination to reliably describe the colour of an object. The problem of correcting for CSS is made easier by the fact that the camera can be characterised before the image is captured. However, the nature of the scene illumination cannot be known ahead of the capture, and affects not only the required white balance but also the optimal choice of mapping, as discussed. Therefore, the question of how to estimate scene illumination from image data is central to achieving consistent colour.

Broadly, two approaches are used to attain reliable colour descriptors for objects in digital images: computational colour constancy and invariance methods. Computational colour constancy aims to render a scene as it would appear under some canonical illuminant by first estimating the illumination under which the image was captured. Invariant techniques attempt to find image features that do not change when certain aspects of the imaging conditions change (e.g. illumination or imaging device) [57].

Colour constancy algorithms cannot work for images with arbitrary scene contents and illumination, because the degrees of freedom are too high to be recovered from an array of RGB triplets [63]. Therefore, assumptions are needed to make the problem tractable. Some techniques make assumptions on the contents of the scene. One simple method assumes that the average contents of the scene would be neutral when properly colour balanced. As such, any deviation must be due to colour cast from the illumination. This is known as the “grey world” assumption [64]. The “white point” method assumes that the brightest part of an image is a perfectly reflective white object, and thus its colour reflects the colour of the prevailing illumination [65]. A third method assumes that the scene contains a gamut

of colours that is representative of that illumination and calculates the transform that would take those colours to a gamut representative of the desired canonical illuminant [66]. Often, assumptions about the possible illumination spectra are used to constrain the space of possible solutions. For example, to estimate illumination spectra Finlayson et al. suggest comparing the image gamut to a range of candidate illumination gamuts and assuming the best match is the true illumination [67]. Some methods assume that the illumination is approximately Plankian, reducing the problem of estimating illumination spectral power distribution (SPD) to estimating the correlated colour temperature of the illumination [68].

Invariant techniques, unlike colour constancy algorithms, do not output RGB images. They involve an algebraic manipulation of the terms of an image formation model such as Equation 1.2 to find quantities that are independent of some terms (under certain assumptions). For example, assuming an object without any specular reflection, normalised RGB (chromaticity) is invariant to image intensity, object orientation, and incident light direction [69].

Of particular relevance to this thesis, some literature exists on the use of flash/no-flash image pairs of the same scene to mitigate the effect of the illumination. DiCarlo et al. estimates the scene as it would appear under flash only using ambient subtraction [70]. Given that the flash SPD and CSS are known, surface reflectance can be estimated using a linear model of surface reflectance. This information is then used along with the ambient-only image to classify the ambient illumination as one of a set of candidates. One advantage of this technique over other colour constancy methods is that different parts of the scene are treated separately. Thus, in images with two or more areas of ambient light this approach can correctly classify each. Two limitations are important. Firstly, this technique cannot be applied to distant objects as the flash will not reach them. Secondly, the surface reflectance recovery is based on only three knowns, so only a three-dimensional surface model can be used. Real objects have surface reflectance functions that are better described by five to seven basis functions [71]. Drew and Lu also use flash/no-flash pairs to estimate scene illumination, but without first estimating surface reflectance [72].

By assuming narrowband camera sensors, Plankian lights, and Lambertian surfaces, they show that ambient illumination can be identified without any *a priori* knowledge about the flash SPD, CSS, or scene. Under these assumptions, the log of the ratio between the ambient-only and estimated flash-only images is invariant to camera and surface properties.

The suitability of each of the above approaches depends on what image data is available, what the desired output is, what information is known *a priori*, and what assumptions it is reasonable to make.

## 1.5 Review of Image-based Jaundice Detection Research

Some research exists on the use of digital photography to detect jaundice. The earliest example came from Leartveravat et al. in 2009 [73]. Using a Sony Cybershot digital camera and a custom reference colour card, images of the sternum were captured and manually processed in Photoshop. After white balance, region selection and conversion to CYMK colour space, a correlation of 0.86 was achieved between (Y - M) and TSB across 61 subjects.

Leung et al. (2015) used a Nikon D3200 DSLR (Digital Single-Lens Reflex) camera to photograph the eye of 110 babies immediately before their blood test [74]. All images were captured in the same room to minimise ambient light variation. RGB values from square regions of interest (ROIs) on the skin (forehead) and sclera were fitted to a quadratic model for predicting TSB values. Linear correlation coefficients of 0.75 ( $p < 0.01$ ) and 0.56 ( $p < 0.01$ ) between estimated and measured TSB were found, respectively, for the sclera ROI and the skin ROI. This confirmed the hypothesis of the authors that the sclera is a better region from which to estimate TSB, as it is free from melanin and hemoglobin. They conclude that the technique is feasible as a screening method, and show that for a screening threshold of  $205\mu\text{M}$  it has performance comparable to that of the latest generation of TcBs [74]. Limitations of this study include the relative lack of data at high TSB values (only a few subjects had TSB greater than  $250\mu\text{mol/L}$ ), and that it was not

shown to be effective in more than one lighting environment or for more than one camera.

Rizvi et al. (2019) also used a digital camera to image the sclera for jaundice estimation [75]. 50 healthy and 50 jaundiced neonates were recruited. The images captured by a Samsung camera were transferred to a Samsung tablet. An app called BiliCapture on the tablet aided the user in identifying a region of interest to be analysed. A correlation of 0.86 was achieved in 50 jaundiced neonates.

de Greef et al. (2014) introduced the first smartphone-based method for jaundice detection called BiliCam [76]. The technique involves a colour card that is placed above the navel of the newborn. This serves two purposes. Firstly, it allows for a degree of quality assurance in the capture image: The mobile application guides the user in the positioning of the phone using a viewfinder that should match the shape of the colour card. Also, by examining the standard deviation of the colour card pixel values the application is able to warn the user if the view is occluded or the lighting is inconsistent. Secondly, the colour card is used for colour calibration.

The BiliCam app captures images with the phone flash on and off, and extracts features from the skin colour coordinates in various colour spaces, as well as colour gradients. These features are fed into multiple machine learning algorithms which are then combined to produce an estimate for the TSB. In a study of 100 newborns this estimate correlated with the gold-standard blood test with a linear correlation 0.84. Although the colour card-based approach allows for effective white balancing and image quality control, it may also prove a limitation of the technique. As the author notes, it is not clear how reproducible their custom card is, and whether variation due to different printers or papers would affect the result. As in the study by Leung et al. (2015), only one model of camera was used (in this case, the iPhone 4S back-facing camera). Taylor et al. evaluated the performance of the BiliCam app installed on iPhone 5S smartphones in a multi-ethnic sample of 530 newborns in seven different sites in the US [77]. The colour accuracy of the printed colour card was tested at the time of printing. A correlation of 0.9 with TSB values was observed, with similar performance observed in all ethnic groups. The authors con-

clude that BiliCam is as effective as a TcB in screening for neonatal jaundice.

Aune et al. (2020) produced an app that estimates TSB using a colour card and skin image [78]. An optical model of newborn skin was used that has parameters for skin thickness, melanin and haemoglobin levels. Some patches on the colour card were designed to have similar reflectance properties as newborn skin. This requires spectral printing, but in this study a consumer printer was used and the resulting colours were measured with a spectrometer. They estimate the cost of production of such a card at less than one US dollar. In a study across two hospitals, the app-estimated TSB correlated with TSB with a coefficient of 0.84 across 302 newborns. The correlation among Caucasian newborns was found to be significantly higher than the correlation among non-Caucasians.

BiliScan (BeiShen Healthcare Technology Co., Shenzhen, China) is the most extensively studied jaundice detection app. Like the later version of BiliCam and the system by Aune et al., it uses a colour card with a hollow centre to guide the image capture and colour correct the image. However, in this case the user of the app must print the colour card, so it cannot require specially coated-paper, advanced printing methods, or a measurement of its spectral reflectance. Images are captured under ambient illumination and uploaded to a server which uses a machine learning algorithm to calculate an “Automated Image-based Bilirubin” (AIB) value. Rong et al. (2016) used the BiliScan app to image 215 neonates and found a correlation of 0.77 with TSB [79]. The largest study to date (Dong et al., 2020, n=369) found a similar correlation of 0.76, but only within a range of bilirubin values 5 to 15 mg/dL (86 to 257  $\mu\text{mol/L}$ ), finding the method unreliable for values less than 5 mg/dL and greater than 18mg/dL (308 $\mu\text{mol/L}$ ) [80]. Yang et al. (2018, n=296 from 194 subjects), also found that BiliScan worked best in a central range of 10 to 20 mg/dL, but was inferior to the TcB outside this range [81]. Ren et al. (2020) found a correlation between AIB and TSB of 0.78, and no difference in performance based on postnatal or gestational age [82]. However, they found a significantly lower performance for those images captured during the night hours as compared to daylight hours. Swarna et al. (2018) conducted the only evaluation of BiliScan



on a non-Chinese neonatal population [83]. A lower correlation of 0.6 was reported across 35 newborns.

Aydin et al. (2016) used a smartphone and a colour card printed by a consumer printer, which was placed on the chest [84]. A combination of Gaussian filtering, thresholding, and pixel similarity methods were used to identify skin in the images of 80 neonates, of which 40 were healthy. A correlation of 0.85 with TSB values was achieved using an ensemble of two machine learning approaches, and the system had 85% accuracy in identifying which babies belonged to the jaundiced group.

Some researchers have investigated contact-based methods for imaging the skin using a smartphone. Sufian et al. (2018), used a housing for the smartphone camera containing LEDs and a reference colour card [85]. A plastic window allowed imaging of the skin. The housing allows ambient light to be removed as a confounding factor, and the contact blanches the skin, lowering the influence of haemoglobin on the skin colour. After calibrating six different phones using reference objects, they were brought into agreement. A non-linear model based on the ratio of the Y and Z value in XYZ colour space gave a correlation of 0.9 in a sample of 48 neonates. Munkholm et al. (2018) imaged the skin using a smartphone in direct contact with the newborn forehead, using a dermatoscope, and using a dermatoscope with a green filter [86]. The aim of the filter was to minimise the influence of haemoglobin on the recorded colour. The dermatoscope standardises the imaging distance and lighting, and a significant correlation with recorded TSB in 64 neonates was found. The filtered and direct method results did not reach significance. Padidar et al. (2019) used a similar approach, imaging the forehead using a 100x clip-on microscope attachment on a smartphone [87]. They used a three-colour, printed colour card for reference captures before the measurement capture. With a sample size of 113, a correlation of 0.48 with TSB was found.

The group from the University of Washington that created BiliCam later produced BiliScreen (Mariakakis et al. [88]). This is a smartphone-based method for screening for jaundice in adults, with a view to catching cases of pancreatic cancer

earlier. Here, the sclera colour is analysed rather than the skin. The study author argues that because the healthy range of bilirubin for adults is much lower (less than  $22\mu\text{M}$  rather than less than  $257\mu\text{M}$ ), it is necessary to look to the sclera. Two methods are employed to ensure consistent colour measurements in different ambient lighting environments. One applies a colour correction algorithm based on a colour card like BiliCam, but cut into the shape of glasses that can be worn. This also makes automatic localisation of the eyes easier. The other method uses a 3D-printed goggles that block out ambient light, into which the mobile phone is slotted. Here, the flash is used in torch mode to provide the light. Automatic segmentation algorithms identify the sclera, and colour features from various colour spaces are used with a random forest regression to produce TSB prediction models. In a study with 70 volunteers, the linear correlation with measured TSB was 0.78 for the glasses method and 0.89 for the 3D-printed goggles method. Two solutions to the problem of variation between smartphone cameras are proposed: either to train the model on each available device separately, or to incorporate a calibration step before data collection using a colour target array of known reflectance properties.

Three other papers have been published that aim to detect jaundice in adults using digital imaging [89] [90] [91]. Like Mariakakis, they also image the sclera, and rely on either custom made glasses, light-blocking goggles or housing for the camera. These accessories are used in the automatic detection of sclera regions and also mitigate the effect of ambient light. Such methods rely on cooperation from the subject, and so would not be practical for neonates.

## **Chapter 2**

# **Clinical Imaging Studies**

## **2.1 Overview**

In this chapter, there follows an overview and comparison of the clinical datasets analysed in this thesis. All three studies involve the collection of paired TSB readings and digital images from jaundiced newborns. The three studies are described in the order in which they were conducted:

1. UCLH Nikon Study
2. UCLH Smartphone Study
3. Ghana Smartphone Study

The data collected in these studies is analysed in subsequent chapters to address the thesis objectives.

## **2.2 Clinical Datasets Used In This Thesis**

### **2.2.1 Study 1: UCLH Nikon Study**

The first clinical dataset used in this investigation was collected at University College London Hospital (UCLH). A digital camera was used to capture images of the sclera and forehead of 133 newborns. The aim was to investigate the potential of digital imaging for estimating the severity of jaundice, and to compare the sclera and skin as sites of measurement. Ethical Approval was obtained from the London - City Road and Hampstead NHS Research Ethics Committee.

Subjects that required a blood test were recruited from the outpatient clinic. Many of the subjects were referred by a visiting midwife during a routine check-up, while others were attending appointments to monitor ongoing physiological jaundice. There were no exclusion criteria based of postnatal or gestational age. Babies who had received phototherapy within the preceding 24 hours were ineligible. The images were captured in the Advanced Neonatal Nurse Practitioner (ANNP) clinic of UCLH, and the blood test followed within twenty minutes to limit the possibility of a change in bilirubin level in the intervening period. TSB was determined from a blood sample obtained via heel prick. The sample was spun in a centrifuge before insertion into a point-of-care bilirubinometer.

Images were captured using a Nikon D3200 24.2-megapixel DSLR camera with a 60mm macro lens. ISO was fixed at 1600, while exposure time and aperture control were set automatically by the camera. Images were saved in Nikon's raw file format, NEF (Nikon Electronic Format). The sclera was brought into focus manually. The lighting in the ANNP clinic was not controlled beyond ensuring that the fluorescent ceiling lighting was switched on. For subjects one to 82, a custom printed, L-shaped yellows scale and colour patch array was included in the capture. Figure 2.1 shows a typical capture for this phase of the study. For subjects 83 to 133, a laminated CameraTrax three-by-two-inch 24-patch colour card was included.



**Figure 2.1:** Example image captured in UCLH Nikon Study, including custom colour reference card. Media consent for this image was obtained from all parties.

Analysing this dataset, Leung et al. (2015) showed the sclera-region RGB

values gave a stronger correlation with TSB than did the forehead-region RGB values [74]. This study used a quadratic polynomial model. Coauthored by this author, and based on the same dataset, Leung et al. (2019) later showed that a simple linear model based on chromaticity could achieve similar correlation with TSB [1]. However, this analysis used a smaller sample ( $n=87$ ) because it relied on raw image files, which were only available for 87 of the 133 subjects.

### **2.2.2 Study 2: UCLH Smartphone Study**

A clinical study was carried out in the UCLH Neonatal Care Unit and Postnatal Ward between January 2017 and February 2018. The aim was to investigate the accuracy of image-based jaundice estimation via the sclera using a smartphone. Ethical Approval was obtained from the London - City Road and Hampstead NHS Research Ethics Committee. Parents were informed of their baby's eligibility for the study and what would be involved in the data collection process, including potential benefits and downsides. After allowing them to consider, verbal and written consent was obtained. Parents were reminded that giving consent did not oblige them to take part in the study, and that they could drop out at any point without it affecting the care of their baby.

This study used the front-facing camera of an LG Nexus 5X smartphone, with illumination provided by the screen of the smartphone itself.

The pictures were taken at whatever time was most convenient during the appointment so as not to disrupt the assessment or treatment of the baby. Often, the data was collected before the baby underwent a blood test, while he or she was lying supine on the assessment table. Babies were found to be more compliant before the blood test, as the heel prick to draw blood is a painful procedure. In other cases, the images were captured with the baby in the arms of a parent. Around the time of feeding babies open their eyes, making data collection more straightforward. The pictures were always taken within 20 minutes of the blood draw so the TSB reading would be accurate at the time of image capture.

Image capture involved holding the phone approximately 10-20cm from the face. The volume keys were used to initiate the capture sequence, which took ap-

proximately 1.5 seconds and captured one image with illumination from the screen and one without. Multiple image pairs were captured for each subject to ensure at least one pair was usable. The user would aim to ensure that the image was not blurred and that the sclera was visible. However, due to the clinical setting this was not always possible.

Figure 2.2 shows an example of an image pair captured by the app in the flash-on and flash-off conditions, with the corresponding phone images adjacent.



**Figure 2.2:** Example image pair captured by app in UCLH Smartphone Study. (A) Smartphone screen illuminating subject. (B) Subject illuminated by smartphone. (C) Smartphone in flash-off condition. (D) Subject capture under ambient illumination only. Media consent for these images was obtained from all parties. Figure originally published in Ref [2].

The exposure time and ISO value were fixed for all captures to 1/25 seconds and 250, respectively. These values were chosen as a compromise between minimizing motion blur and using as much of the dynamic range of the sensor as possible,

after extensive testing in hospital and office-like lighting environments (predominantly fluorescent strip lighting). However, the fixed capture settings did not always result in a well exposed capture, which is a limitation of this dataset.

The TSB measurement began with a blood draw using a heel lance. A few drops were collected using a capillary tube, which was promptly sealed with resin. A centrifuge spun the tube to separate the serum from the hematocrit, and the serum was analysed using a bilirubinometer to give a value for the total serum bilirubin (TSB). This method allowed a result at the point of care. Sometimes this method was not available, and a sample was instead sent for laboratory testing, which is a longer process. Only measurements taken using the centrifuge and bilirubinometer method were used for this study.

A data collection sheet was filled in for each subject. For confidentiality reasons, only the time and date of image capture was used to link the image to the datasheet. The time of the first image and last image capture was noted on the datasheet, and each baby received an integer subject number. Other information collected included the baby's gestational age, postnatal age, whether he or she had received phototherapy within the previous 24 hours, where the images were captured (including description of the lighting present), the TcB measurement (if one was made), and the point-of-care TSB measurement. A copy of the data collection sheet used can be found in Appendix A.

A total of 51 babies were imaged using the application (16 female, 35 male, median postnatal age 11 days). All but four of the 51 image sets were captured in the same room (the Advanced Neonatal Nurse Practitioner (ANNP) Clinic of the UCLH Neonatal Ward). For eight of the subjects, not enough sclera was visible to identify a usable region of pixels in the sclera. For another two subjects, results from the point-of-care bilirubinometer were not available. 41 subjects remained for analysis. Histograms representing the distributions of TSB, gestational age, and postnatal age for these subjects can be found in Appendix C.

Based on the UCLH dataset, this author published a journal article investigating ambient-subtracted sclera chromaticity for smartphone jaundice screening [2].

### 2.2.3 Study 3: DJAN Smartphone Study

The Detecting Jaundice in African Newborns (DJAN) study was conducted in Ghana. Funded by Saving Lives at Birth<sup>1</sup>, this study aimed to test the smartphone-based screening technique in the setting in which it stands to have the most impact. The study also tested the practicality of the app in both hospital and rural community healthcare settings. Images were captured in the Greater Accra Regional Hospital between March 2019 and February 2020, and in community health centres and the Holy Family Hospital in Kwewu West District, a rural area of Ghana, between February and March 2020. A total of 847 capture sessions took place, with 765 individuals enrolled, making this the largest study on image-based bilirubin detection to date. Ethical approval was obtained from the Ghana Health Service Ethical Review Committee and UCL Research Ethics Committee.

In the DJAN study, the imaging app used the smartphone back-facing camera and the smartphone LED flash, which was fitted with an acrylic diffuser as a precaution. Four different Samsung S8 (Samsung Electronics Co. Ltd., South Korea) smartphones were used.

Only babies in need of a blood draw for TSB estimation were included in the study. Immediately prior to the heel prick blood draw, the baby was imaged with the smartphone app and had a measurement taken with the Draeger JM-105 transcutaneous bilirubinometer (Draeger UK). Serum bilirubin concentration measurements were made using a centrifuge (SciSpin Haematocrit Micro Centrifuge, SciQuip Ltd., UK) and bilirubinometer (Bilimeter 3D, Pfaff Medical GmbH, Germany).

In this study, some babies were undergoing treatment for jaundice, others were attending postnatal clinic for scheduled appointments. Images were taken in a variety of locations, and this was recorded along with the nature of the prevailing

---

<sup>1</sup>Saving Lives at Birth: A Grand Challenge for Development funds “groundbreaking prevention and treatment approaches for pregnant women and newborns in poor, hard-to-reach communities around the time of childbirth” [92]. It is jointly supported by Grand Challenges Canada, the U.S. Agency for International Development (USAID), the Norwegian Agency for Development Cooperation (Norad), the Bill & Melinda Gates Foundation, the UK’s Department for International Development (DFID) and the Korea International Cooperation Agency (KOICA) [92].



light (indoor/ natural/ both). No measurement was made of the ambient light environment. If the baby was undergoing phototherapy, or had received phototherapy within the last 24 hours, this was recorded. Whether the parents and staff thought the baby appeared jaundiced was also noted. The data collection sheet used can be found in Appendix A.



**Figure 2.3:** Typical image capture scenario for Phase 1a of DJAN study. Media consent for this image was obtained from all parties.

The app was used to capture three flash/ no-flash image pairs of the sclera and one of the sternum. For the first 167 studies, an X-Rite ColorChecker Passport was included in the shot for these captures. Figure 2.3 shows a typical image capture scenario for this phase of the study. It was time-consuming to simultaneously position the full colour card inside the shot and obtain a good quality image of the sclera with sufficient flash illumination. Thereafter, the protocol was changed, and the colour card was captured separately, after the other captures. This was an easier capture process and resulted in sclera images with larger regions of interest and higher subtracted signal-to-noise ratios. The colour card was imaged over the ster-

| DJAN Study Phase | Principle Location                                 | Capture Protocol   | Real-time Feedback/ Result from App |
|------------------|--|--|-------------------------------------|
| 1a               | Greater Accra Regional Hospital (Procedure Room)   | x3 sclera image pairs, x1 chest image pair (all with colour card in frame) | No                                  |
| 1b               | Greater Accra Regional Hospital (Procedure Room)   | x3 sclera image pairs, x1 chest image pair, x1 colour card image pair      | No (first 80%)<br>Yes (final 20%)   |
| 2                | Greater Accra Regional Hospital (Postnatal Clinic) | x3 sclera image pairs (no colour card)                                     | Yes                                 |
| 3                | Holy Family Hospital (Playroom)                    | x3 sclera image pairs (no colour card)                                     | Yes                                 |

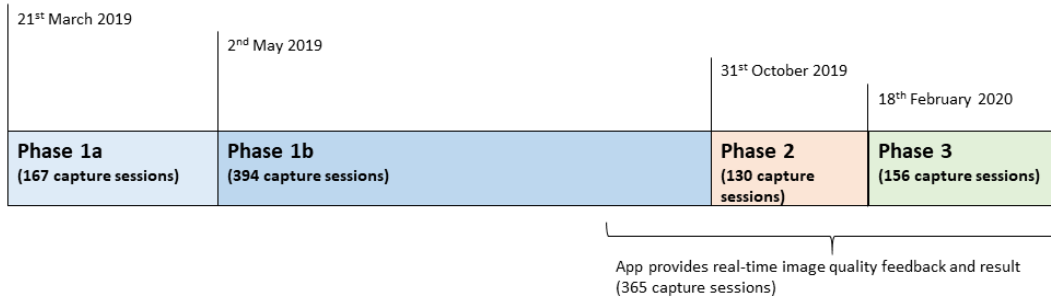
**Table 2.1:** Comparison of phases of DJAN study.

num of the baby immediately after the other captures to minimise the risk of any changes in incident ambient light between baby images and colour card images.

Images were captured with approximately 10-20 centimetres between the camera and the sclera, sternum, or colour card. When possible, the camera preview was used to angle the phone to avoid specular highlights on the sclera. The app was updated after six months to provide real-time feedback to the user on the signal-to-noise in the captured image. The image capture user guide for the app can be found in Appendix B. This was provided to the Ghana team in August 2019 and explains the protocol for capturing good images using feedback from the real-time image review screen included in the latest app version.

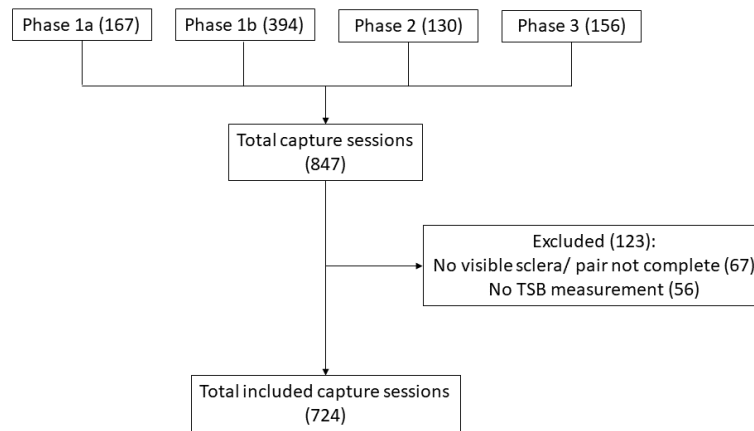
Table 2.1 shows the different phases of the DJAN study and how they differed from one another. Figure 2.4 shows the different phases of the study by date and number of capture sessions. In Phase 1, all capture sessions took place in the Procedure Room. In Phase 2, 45% of capture sessions took place in the Postnatal Clinic, 41% in the Procedure Room, and the remaining 14% in various other locations including the maternity ward and the neonatal intensive care unit (NICU). In Phase 3, at Holy Family Hospital, 94% of capture sessions took place in the Playroom, with the rest mostly in the NICU.

Figure 2.5 shows the inclusion pipeline for capture sessions. After removing



**Figure 2.4:** Phases of the DJAN study. Relative width of segments corresponds to the number of capture sessions recorded in each phase.

capture sessions where either no TSB was recorded or there was no image pair with visible sclera, 724 capture sessions remained. 76% of sessions excluded due to a lack of TSB reading were from the Postnatal Clinic in Phase 2. This is because the neonate was attending a scheduled check up appointment and was not deemed at risk of jaundice after visual inspection and a TcB reading.



**Figure 2.5:** Flowchart showing number of capture sessions and reasons for exclusion in DJAN study.

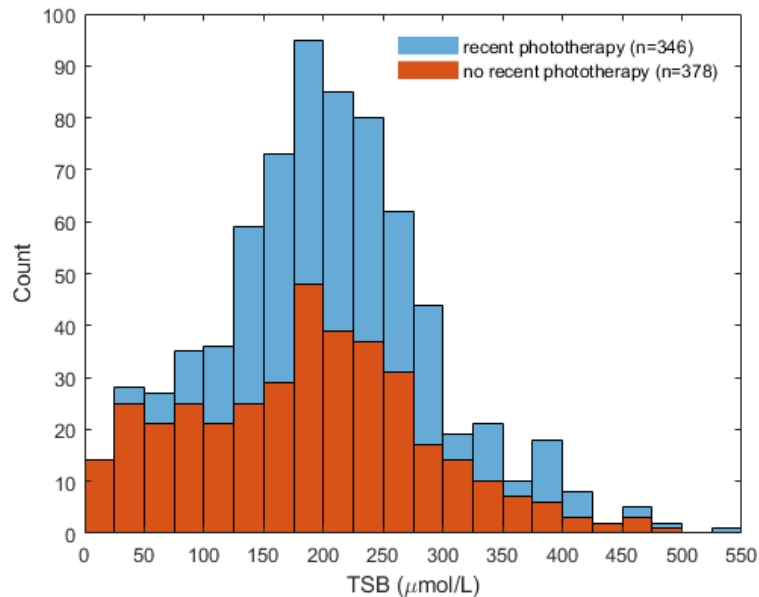
The included 724 capture sessions form the basis of the analysis in subsequent chapters. 403 of them involved male subjects, and 321 involved female subjects. Subjects were sometimes imaged more than once. From the 724 sessions, 651 image sets were from the first capture session for a given subject, 59 were from a

second-time session, 13 were from a third-time session, and one was from a fourth-time session.

### 2.2.3.1 DJAN Subject Statistics

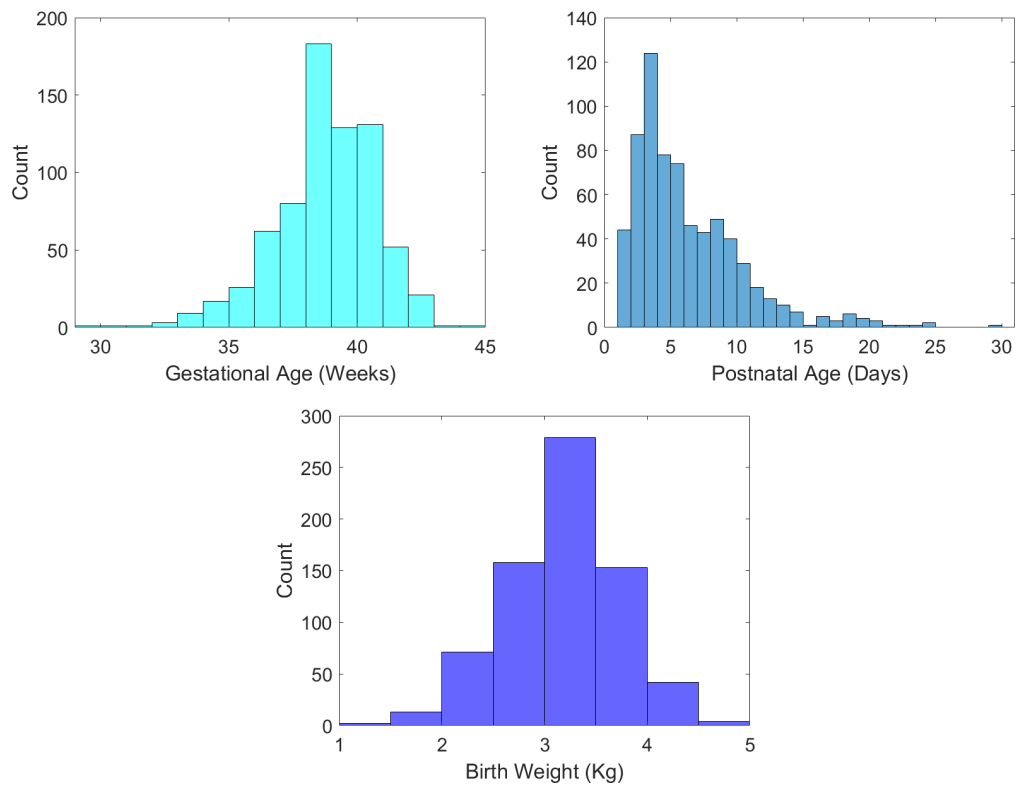
In this section, statistics for the DJAN subject data are displayed in a series of plots. Similar plot for the UCLH Smartphone Study can be found in Appendix C.

Figure 2.6 shows a histogram of TSB values for all 724 capture sessions. The legend shows the numbers of capture sessions involving subjects who had either no recent exposure to phototherapy (378) or who had had phototherapy within the last 24 hours (346). The relative number of no-phototherapy babies is higher at low TSB values. Overall, the median TSB is  $198\mu\text{mol/L}$  and the maximum is  $540\mu\text{mol/L}$ .



**Figure 2.6:** Histogram of TSB values for DJAN study (n=724).

Figure 2.7 shows the gestational age, postnatal age, and birth weight histograms for the DJAN dataset. The median gestational age was 28 weeks 4 days. 32 neonates were preterm (less than 35 weeks gestational age), while 88 were near-term (35-37 weeks gestational age). The median postnatal age was 4 days 22 hours and the median birth weight was 3.1kg.



**Figure 2.7:** Histograms of gestational age, postnatal age, and birth weight for DJAN study (n=724).

## 2.2.4 Comparison of Datasets

Table 2.2 summarises the three datasets that are used in the subsequent chapters.

|                                    | UCLH Nikon Study               | UCLH Smartphone Study          | DJAN Smartphone Study                    |
|------------------------------------|--------------------------------|--------------------------------|--|
| Data Collection Dates              | July 2014 – August 2014        | January 2017 – February 2018   | March 2019 – March 2020                  |
| Number Capture Sessions Included   | 87                             | 41                             | 724                                      |
| Phototherapy within preceding 24h? | No                             | No                             | 346 Yes, 378 No                          |
| Principle Location                 | ANNP Clinic, UCLH, UK          | ANNP Clinic, UCLH, UK          | Multiple, Ghana                          |
| Majority Ethnicity                 | Caucasian (59%)                | Caucasian (53%)                | Black African (100%)                     |
| Device Used                        | Nikon D3200 DSLR               | LG Nexus 5X                    | Samsung S8 (x4)                          |
| Accessories                        | Custom colour card             | None                           | LED diffuser, X-Rite colour card         |
| Ambient Light                      | Indoor - fluorescent (100%)    | Indoor - fluorescent (100%)    | Indoor - fluorescent (98%), Natural (2%) |
| Active Illumination                | None                           | Phone screen                   | Phone back-facing flash                  |
| Site Imaged                        | Sclera, Forehead               | Sclera                         | Sclera (x3), Sternum (x1)                |
| TcB Used                           | None                           | None                           | Draeger JM-105                           |
| TSB Method                         | Centrifuge and bilirubinometer | Centrifuge and bilirubinometer | Centrifuge and bilirubinometer           |

**Table 2.2:** Comparison of clinical study datasets.

## 2.2.5 Objectives of Clinical Dataset Analyses

Study 1: UCLH Nikon Study data was analysed by this author, with the conclusion that a simple chromaticity-based sclera colour metric could be predictive of the level of jaundice as long as the data was treated in a scene-referred space [1]. This

implied the need for raw image data, informing the development of the app and on which smartphones it would be able to run. Therefore, although these data were not collected using a smartphone, they were used to inform the proposal of the sclera chromaticity as a metric of jaundice (Objective 2) and the requirements of the neoSCB app, which are described in Chapter 3.

Study 2: UCLH Smartphone Study was the first to pilot the neoSCB app developed by this author (Objective 3: Build App). It showed that it was feasible to image the newborn eye with a flash/ no-flash image pair in a clinical setting using a smartphone (Objective 4) [2]. In Chapter 4, it serves as a separate, unseen dataset on which to validate the SCB model (Objective 7).

Study 3: Ghana Smartphone Study is analysed in Chapter 4 and Chapter 5. In Chapter 4, it is used to test the colorimetric accuracy of the app (Objective 5), compare the efficacy of skin-based and sclera-based chromaticity measurements in predicting TSB (Objective 6), train the final SCB model (Objective 7), and investigate the factors that affect the predictive performance of the SCB model (Objective 8). In Chapter 5, it is used to quantify the screening performance of the neoSCB app (Objective 9), and compare it to the screening performance of the JM-105 TcB (Objective 10).

## Chapter 3

# Development of the neoSCB App

### 3.1 Overview

In this chapter, the development of the neoSCB (neonatal scleral-conjunctival bilirubinometer) app and the principles behind it are outlined. It begins with an overview of the key challenges that need to be addressed by any image-based jaundice screening technique. Next, the concepts of ambient-subtracted chromaticity and subtracted signal-to-noise ratio are explained. Ambient subtraction is essential to avoid the need for a colour card or light-blocking housing, which is a key aim of this work. These ideas are then demonstrated on both clinical and experimental data as a proof of principle. Finally, the specification and development of the app itself is described.

### 3.2 Challenges To Be Addressed

The goal of this work is to create a system that can detect jaundice by colour imaging. It is helpful to consider the key sources of error in such a system. This means that any proposed solution can be considered in light of how well it addresses these challenges.

Let us consider the most general example of a system that uses an RGB image to derive information about the bilirubin level of a newborn. It can be broken down into three steps:

1. **Image Capture** How the image is captured: the hardware, camera settings, and protocol for capture including who takes the image and in what environ-



ment.

2. **Image Processing** The operations applied to the image. This comprises of two parts: A selection of the relevant region of interest and a derivation of a colour descriptor or set of features.
3. **Jaundice Level Estimation** The colour information derived from the image must be related to the jaundice level of the baby in some way. This may result in an estimate for the serum bilirubin concentration or directly in a screening decision or recommendation.

Despite the apparent simplicity of such a pipeline, there are a number of challenges to be addressed.

### 3.2.1 Image Capture

#### Light environment

A jaundice detection app must be effective in a range of illumination environments. Among the studies in camera-based jaundice detection, the methods to correct for the influence of ambient light either rely on a colour card or aim to replace the ambient light with another light source of known properties (or some combination of the two). Each solution has its own challenges concerning the light environment required for image capture. If using a colour card, then the light must be spatially consistent and bright enough to provide an image of sufficient signal-to-noise ratio. On the other hand, if the technique relies on a light source of known colour, then the contribution from the ambient background must be fully mitigated.

Some studies that rely on a colour card use the smartphone flash as a means of active illumination [76] [78]. Recognising that the use of a flash alone does not in itself guarantee sufficient illumination of the subject, these studies also attempt to standardise the distance between the smartphone and the subject.

Studies that aim to replace ambient illumination with a known illumination must discount the ambient contribution to the signal. These studies have all used contact-based approaches using dermatoscopes or dermatoscope-like attachments to the smartphone camera so that ambient light is blocked out [85] [86].

### **Quality assurance**

Apart from signal-to-noise ratio, the images captured must meet several other criteria to be usable. The region of interest, be it the skin or the sclera, must be in the image, large enough, and in focus. Motion blur due to the subject or camera moving may disqualify the image. Some assurance of image suitability is therefore necessary if the final result is to be trusted.

Almost all studies to date have manually discarded inadequate images during their offline analysis. However, an app should be usable by a midwife or parent after no or limited instruction. This requires an automatic method of discarding images that are not usable. Some proposals involving colour cards have a viewfinder on the app camera preview, which allows the user to align the smartphone in the right plane and at the right distance with respect to the baby [83] [76] [78]. This helps ensure the entire colour card is within the capture. Some also use a “hollow” colour card, with a window in the centre to frame a region of interest of the chest [78] [77]. BiliCam uses the colour card to check for inconsistent lighting and occlusion by checking that the standard deviation of the pixel values within a certain colour patch is below a certain threshold [76].

### **3.2.2 Image Processing: ROI Selection**

#### **Identifying relevant region of interest**

The skin or sclera region must be identified. Whether done automatically, semi-automatically, or manually, this step can introduce error. For example, pixels belonging to the iris may be included in the sclera segmentation.

In previous work, this step is often done manually. Automatically identifying the region of interest can be aided by introducing accessories that can be easily detected by algorithms. For example, the systems that involve glasses worn by adults can use fiducial markers printed on the glasses or the colour of the glasses to identify the eyes [89] [88]. In the same way, hollow colour cards also simplify the identification of a relevant region of interest. Contact-based methods further simplify the task of ROI identification: the area of interest will always occupy the same part of the image.

### **Rejecting non-representative pixels**

Within the region of interest, some pixels may need to be excluded from the final analysis if they are not representative of the tissue colour. In the sclera, filtering eyelashes, blood vessels, and specular highlights may be required. Similarly, in the skin, bruises or specular regions would need to be identified and excluded, either manually or automatically, to avoid a spurious result.

BiliScreen empirically determined thresholds in HSL (Hue-Saturation-Lightness) colour space to exclude glare ( $L > 220$ ), eyelashes ( $L < 5$ ), and vessels ( $H < 15$ ) [88].

### **3.2.3 Image Processing: Extracting Colour Information**

#### **Device- and ambient-independent colour**

The colour estimate derived from the image should be independent of the ambient light environment and the hardware used to capture the image. Variation between camera spectral sensitivities, even between cameras of the same make and model, must be accounted for [6]. Similarly, to the extent that the light environment is not controlled by the method of capture, the colour recorded must be corrected so that the bilirubin level is not overestimated or underestimated. This is essential as any model that relates colour to bilirubin level depends on the accuracy of the input colour values.

The majority of image-based techniques described to date use some variation of a reference standard or colour card. A mapping developed from a colour card has the advantage of simultaneously correcting for variations between camera spectral sensitivities and lighting environments. Methods that remove the influence of ambient light, such as the adult goggle instantiation of BiliScreen, or the contact-based method of Sufian et al. still must address the variation between camera spectral sensitivities. They each suggest a one-time calibration for each device before it is used [85] [88].

#### **Confounding chromophores**

Tissue colour can vary due to chromophores besides bilirubin, such as melanin and haemoglobin. Variation in the concentration of each of these can influence the

bilirubin concentration estimate, and this should be accounted for in some way.

Aune et al. use a physics-based model of light transport in skin to model the effect of variation in haemoglobin and melanin [78]. Most approaches rely on a purely machine learning-based approach to understanding the relationship between bilirubin and colour in the presence of confounding chromophores. Contact-based imaging techniques discuss the benefit of blanching the skin to remove the influence of haemoglobin [85] [86]. Munkholm et al. investigated the use of a coloured green filter to minimise the effect of haemoglobin on the measurement, but found no performance improvement [86].

Leung et al. compared sclera and skin regions and found a better correlation between colour and bilirubin concentrations in the sclera [74]. This is due to the absence of haemoglobin and melanin in the sclera. Apart from Leung et al. and the work presented in this thesis, only Rizvi et al. have attempted to image the newborn sclera [75]. However, all four studies on image-based jaundice detection in adults have opted to image the sclera [88] [89] [91] [90]. The sclera colour is considered a more sensitive measure of jaundice levels, but this must be traded off against the added difficulty in image acquisition when the subject is a newborn who cannot cooperate.

### 3.2.4 Jaundice Level Estimation

#### Generalisability of the model used

If a prediction model is trained on a set of paired colour and bilirubin values, care must be taken to ensure that the model holds on unseen data. As well as the danger of overfitting, there is the possibility that the sample used for training was not representative. For example, if the model was trained on a dataset wherein all subjects were of the same ethnicity, it should be confirmed that it performs equally well in others before being deployed. This also applies to other factors that may plausibly influence the relationship between colour and bilirubin level, including, for example, gestational age. Even a physics-based model that is completely training-free must be deployed within a certain scope of validity.

The two most extensively tested apps are BiliCam and BiliScan. BiliCam was

validated in a multi-ethnic sample and across multiple sites [77]. It also used multiple imaging devices, although they were all the same make and model (iPhone 5S). Many studies have been published on BiliScan, but only one has tested the app on a non-Chinese population, in which a relatively low performance was observed [83].

### **Communicating the result**

Once the estimate has been arrived at, it should be communicated to the user in a manner that is intelligible and actionable. This depends in part on whether the intended user is a healthcare professional or a untrained individual. A TSB concentration may not be easily interpreted by a parent, so an instruction may be more appropriate. Finally, the user should be made aware of the degree of uncertainty in the estimate. This may take the form a published sensitivity and specificity, or a measurement-specific uncertainty value with which to qualify the result.

## **3.3 Ambient-subtracted Sclera Chromaticity**

The key advantages of using a smartphone for jaundice screening are accessibility, and the possibility for a contact-free, objective measurement. The ubiquity of smartphones enables a solution that can be scaled and help identify cases of jaundice in a variety of settings, including low- and middle-income countries particularly affected by jaundice.

The approach proposed in this work is intended to fully exploit the comparative advantages offered by smartphones (accessibility and scalability) while still addressing the challenges discussed in the preceding section.

To date, work on image-based jaundice detection can be categorised into one of two groups based on the method for achieving device- and ambient-independence:

1. Rely on a colour card to account for ambient and device variation: Additional illumination is sometimes used to guarantee sufficient illumination, but no attempt is made to reduce the ambient contribution to the recorded signal.
2. Block ambient light: Another source is used to illuminate the subject. A calibration step is needed beforehand to account for the combined light and camera properties.

The disadvantage of a colour card approach is the need to distribute and maintain a colour card printed to high standards. Aune et al. estimated the cost of printing a colour card using a calibrated printer at as little as one US dollar per card, and noted that the card could be laminated so as to be preserved, sanitised and reused [78]. However, reliance on a supply chain in the case of loss or damage undermines the potential for easy access in less-economically developed parts of the world. Lamination can cause specular reflections from the card. If the user is instructed to print a card on a consumer printer the colour accuracy cannot be guaranteed.

Systems that block ambient light involve an attachment to the smartphone. This poses distribution and affordability challenges, although maintenance may be less difficult than for colour cards. However, these systems are contact-based and so may risk spreading infection.

In this work, an approach based on ambient subtraction is proposed. By imaging the subject twice, once under the smartphone illumination and once under only ambient, the ambient contribution is explicitly subtracted. This approach is most similar to the second approach above, but does not require a contact-based measurement to remove the ambient contribution. In this way it minimises the number of accessories needed, which may reduce the barriers to adoption.

While BiliCam and the method proposed by Aune et al. both involve taking flash and no-flash measurements, they do not attempt to explicitly subtract the ambient contribution. Instead, the recorded values under each condition are treated as separate features to be input into a machine learning model.

### 3.3.1 Ambient Subtraction

The ambient subtraction method can be summarised as follows. Suppose there is a scene illuminated by an illumination  $A$ , which we identify as the ambient illumination. A second illumination  $F$  (for flash) can be turned on and off at will. Two images of the scene are captured: the first with the illumination  $F$ , the second without illumination  $F$ . The illumination  $A$  is present for both captures. These images are labelled  $I^{A+F}$  and  $I^A$ , respectively. By subtracting, pixel-wise,  $I^A$  from  $I^{A+F}$ , an

estimate for the image of the scene under flash illumination only,  $\hat{I}^F$ , is calculated.

Let us consider a single pixel in an image,  $I_x$ , where  $x$  serves to index the pixels.  $I_x$  corresponds to a particular point on a surface of an object in the scene being imaged. The camera outputs an RGB triplet,  $[I_{x,R}^{A+F}, I_{x,G}^{A+F}, I_{x,B}^{A+F}]$ , for the flash capture, and another RGB triplet for the ambient-only capture,  $[I_{x,R}^A, I_{x,G}^A, I_{x,B}^A]$ .

In the ideal case, the RGB triplet measured with both light sources is simply the addition of the two triplets under the two separate light sources. The flash-only values can be estimated by a vector subtraction:  $\hat{I}_k^F = I_k^{A+F} - I_k^A$ .

Clearly, several assumptions must hold true for this method to return an accurate estimate of  $I^F$ . First of all, the camera must respond to light in a linear fashion. If the pixel count is not proportional to the light intensity incident on the camera sensor, then the contributions to the image from illuminations  $A$  and  $F$  cannot be considered additive and independent. The subtraction method will thus fail. It is reasonable to assume that manufacturers aim to build sensors with linear responses over their expected working ranges. How well typical camera sensors conform to this assumption of linearity is ultimately an empirical question, and one of great importance here. Percentage of linearity is a metric that can be used to indicate how linear pixel output is with respect to photovoltage [93]. If a given pixel is saturated in one or both of the input images, ambient subtraction is impossible. This depends on the dynamic range of the sensor, which depends on the electron well capacity. Failure due to pixel saturation can be thought of as a particular case of failure due to non-linearity. A final way that a camera's response may be non-linear is due to post-processing of the image. Gamma compression is one commonly-used post-processing step that maps linear pixel values to non-linear values, increasing contrast in shadows and decreasing contrast in highlights. To avoid this, the ambient-subtracted image should be calculated from linear input images.

Pixelwise ambient subtraction also requires that there is no movement between the first and second capture. The pixel at location  $x$  in the ambient-only image must correspond to the same point on the same object as the pixel at location  $x$  in the flash with ambient image. This means the camera must remain stationary with respect to

the object. By averaging over a larger region of interest, small movements can be safely ignored.

The ambient illumination must be the same for both captures. Not only must the intensity and spectral power distribution remain constant, the direction of incidence on the object must also stay the same. If there is movement of the source of ambient light with respect to the object, the distribution of ambient light reflected from the object will be different and subtraction will not remove the ambient contribution. Areas in shadow and areas with specular reflection will also change, which will again cause subtraction to give spurious results.

### 3.3.2 Subtracted Signal-To-Noise Ratio

Even if both flash and no-flash images have good signal-to-noise ratio, it is possible that the subtracted result has a very low signal-to-noise ratio. This can happen when the ambient light dominates the smartphone flash, either because the phone is too far from the subject or because the ambient light is particularly bright. There is a need to quantify the signal-to-noise ratio of the post-subtraction signal, as this is the part of the signal that originates from the smartphone flash.

The signal is assumed to be subject to only shot noise as a simplifying assumption. The signal recorded at each pixel thus has a Poisson distribution. By the Central Limit Theorem, this image signal can be modelled as being Gaussian distributed with mean and variance equal to the expected value of the signal. For a particular pixel, we consider an additive noise model as in Equation 3.1, where the recorded signal  $I$  is the sum of the “true” signal  $\langle I \rangle$  and a noise term  $n$ , given by a Gaussian with zero mean and variance given by the expected signal value,  $n \sim \mathcal{N}(0, \langle I \rangle)$ .

$$I = \langle I \rangle + n \quad (3.1)$$

Subtracting two such signals,  $I^{A+F}$  and  $I^A$ , results in a signal with noise greater than the noise of either of the input signals,  $n^{A+F}$  and  $n^A$ . The variance of each must be added to give an estimate for the variance of the subtracted signal, which



is then given by  $I^{A+F} + I^A$ . We define the subtracted signal-to-noise ratio (SSNR) as in Equation 3.2, as the ratio of the estimate for the flash-only pixel value and the estimate for its standard deviation. The same definition is proposed by Hui et al. (2016) [94].

$$SSNR = \frac{I^{A+F} - I^A}{\sqrt{I^{A+F} + I^A}} \quad (3.2)$$

Equation 3.2 implies that the best SSNR is achieved when 100% of the signal is from the flash illumination (when there is no ambient contribution at all). For low flash signal and high ambient signal, it is more advantageous to increase the flash contribution by one unit rather than to decrease the ambient contribution by one unit. On the other hand, if the flash signal contribution is high and the ambient contribution is low, the SSNR is best improved by further decreasing the ambient contribution.

In work coauthored by this author, Nixon et al. (2020) experimentally demonstrated an SSNR threshold of 3.4 guaranteed adequate ambient-subtracted chromaticity accuracy for the same two smartphone models used in this work (Samsung S8 and LG Nexus 5X) [3]. This was determined using 172 patches from the Macbeth ColorChecker DC colour card as test targets. The limit was chosen such that chromaticity error after subtraction was within one standard deviation of the error that was observed when no ambient light was present at all. Beyond an SSNR of 3.4, increasing SSNR ceased to improve the recovered chromaticity accuracy. The same threshold is used in this work to discard flash/ no-flash image pairs with insufficient SSNR.

### 3.3.3 Chromaticity Space

One consequence of using the subtraction method is an indeterminacy in the intensity of the subtracted signal. The signal after subtraction could be increased by reducing the flash-target distance, increasing the flash power, or imaging a more reflective object. Without additional information – or a reference standard in the scene – it is impossible to distinguish between these factors. Therefore, only the

relative raw RGB values are meaningful after ambient subtraction.

Chromaticity values, denoted by lowercase letters  $rgb$ , are given by normalising the  $RGB$  values by the summation of the three. Chromaticities are invariant to changes in light intensity and exposure time, as these affect all three channels by the same proportion. Chromaticities are also invariant to shading due to object geometry for the same reason. By using chromaticity the affect of varying the distance between object and smartphone can also be ruled out.

The efficacy of ambient-subtracted chromaticity as a method of quantifying the degree of yellow discolouration can be motivated on theoretical grounds by examining the image formation equation (IFE). This also reveals what implicit assumptions exist in our approach.

The first important assumption is that we are able to avoid areas of specular reflection in choosing a representative region of pixels to analyse. This means the specular term of the IFE can be ignored, resulting in Equation 3.3 for the ambient-only image. Note the superscript on the illumination variable  $E^A(\lambda, x)$  to differentiate ambient from flash. The geometry and visibility scaling term  $w_d^A(x)$  also has this superscript, as it depends on the illumination direction of incidence (see Section 1.4.1.5).

$$I_{k,x}^A = w_d^A(x) \int_{\omega} E^A(\lambda, x) S(\lambda, x) Q_k(\lambda) d\lambda \quad (3.3)$$

Assuming a linear camera response, the flash and ambient contributions to pixel count should be independent, meaning they can be written as separate terms in the ambient-plus-flash IFE (Equation 3.4).

$$I_{k,x}^{A+F} = w_d^A(x) \int_{\omega} E^A(\lambda, x) S(\lambda, x) Q_k(\lambda) d\lambda + w_d^F(x) \int_{\omega} E^F(\lambda, x) S(\lambda, x) Q_k(\lambda) d\lambda \quad (3.4)$$

Then the subtraction of Equation 3.3 from Equation 3.4 gives the flash-only image estimate (Equation 3.5), as long as there has been no movement so that all geometry and position-dependent terms are the same in each image (and the ambient

spectral power distribution has not changed between captures).

$$\hat{I}_{k,x}^F = I_{k,x}^{A+F} - I_{k,x}^A = w_d^F(x) \int_{\omega} E^F(\lambda, x) S(\lambda, x) Q_k(\lambda) d\lambda \quad (3.5)$$

Invariant features are often calculated under the equal-energy illuminant assumption [57]. This implies a white light which is consistent across the visible spectrum. Under this assumption of white flash, we can remove the illumination term from the integral, giving Equation 3.7.  $c_k$  is the integral of surface reflectance and CSS for channel  $k$ .

$$\hat{I}_{k,x}^F = w_d^F(x) E^F(x) \int_{\omega} S(\lambda, x) Q_k(\lambda) d\lambda \quad (3.6)$$

$$= w_d^F(x) E^F(x) c_k \quad (3.7)$$

Finally, taking the blue channel as an example, normalising  $\hat{I}_B^F$  gives the blue channel chromaticity.  $F$  and  $x$  notation is dropped for clarity.

$$\hat{b} = \frac{E w_d c_B}{E w_d c_R + E w_d c_G + E w_d c_B} = \frac{c_B}{c_R + c_G + c_B} \quad (3.8)$$

Equation 3.8 shows that the ambient-subtracted chromaticity is invariant to surface orientation, viewing angle, illumination direction, and illumination intensity. It depends only on CSS and object reflectance properties.

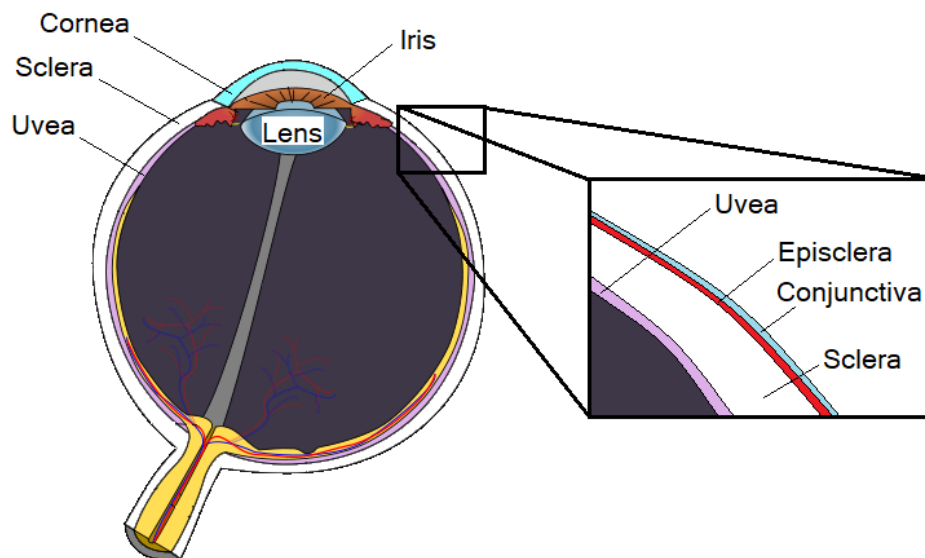
### 3.3.4 Imaging The Sclera

In this work, the sclera is imaged because it is free from the confounding influence of melanin and hemoglobin. The yellow discolouration observed here is therefore independent of the ethnicity of the subject. Although imaging the newborn eye presents its own challenges, we seek to show that the improved sensitivity and generalisability compared to skin imaging makes it worthwhile.

From histological studies it has been noted that the majority of the yellowing is due to the accumulation of biliurbin in the conjunctiva rather than the sclera proper [95], and that the correct description of this symptom should therefore be

conjunctival icterus rather than scleral icterus [96]. Figure 3.1 shows a cross-sectional diagram of the eye, including an inset showing the outermost layers of the eye. In contrast with the sclera proper, the conjunctiva and episclera are vascularised and have a high content of the protein elastin, which has a high affinity for bilirubin [97]. Therefore, it is in these layers that bilirubin accumulates [96].

In this work, the term “sclera” rather than “conjunctiva” is used when discussing sclera colour. This is consistent with similar literature on the identification of jaundice through digital imaging. The term “sclera colour” is used with the understanding that this means the apparent colour of the sclera region of the eye (as opposed to the iris or pupil), and that the cause for the discolouration is the bilirubin in the translucent episclera and conjunctival layers above it.



**Figure 3.1:** Cross-sectional diagram of the human eye. Inset is a cross-section of the outermost part of the eye, including the sclera, episclera, and conjunctiva.

### 3.3.5 Device Independence via One-Time Calibration

After the ambient subtracted chromaticity has been calculated, the result is supposed to be independent of the ambient light, imaging geometry and flash intensity. However, the result from one smartphone could not be compared to that from another because it depends on the smartphone CSS and flash.

To standardise measurements across a range of devices we must introduce a

calibration step. This involves transforming values measured under smartphone flash that are in the camera raw space to values under a canonical illuminant in a reference colour space. Assuming the flash and camera properties are stable over time, this calibration only needs to be done once per device.

There are several ways to develop the required transform. In this work we propose using an X-Rite ColorChecker Classic colour card. The 24 colour patches have known XYZ values, measured using a X-Rite ColorMunki Photo spectrophotometer. By imaging the card with the smartphone we record the corresponding RGB values in the raw space under the flash illumination only. Given the 24x3 matrix of raw recorded values  $\mathbf{R}$  and the 24x3 matrix of XYZ values  $\mathbf{H}$ , a least-squares solution for the mapping  $\mathbf{M}$  between them can be calculated using Equation 3.9.

$$\mathbf{M} = (\mathbf{R}^T \mathbf{R})^{-1} \mathbf{R}^T \mathbf{H} \quad (3.9)$$

Other methods to calibrate the system require expensive specialist equipment to characterise the spectral output of the flash and the CSS. Using a colour card the calibration can be done with less equipment: only the smartphone and card are needed.

In Equation 3.9,  $\mathbf{M}$  is a 3x3 matrix. It describes a linear relationship between the raw RGB values and the target XYZ values. This mapping will not be completely free from error. Higher order polynomial relationships can also be developed, and may offer greater accuracy.

To achieve device-independent ambient-subtracted chromaticity values, the post-subtraction RGB values are converted to XYZ values using a 3x3 calibration matrix. From here, the xy chromaticity values are derived. Using this approach, polynomial mappings are not appropriate because of the indeterminacy in the scaling of the post-subtraction signal. The chromaticity calculation only mitigates for flash intensity and smartphone-subject distance if the linear relationship between RGB and XYZ is maintained.

For the same reason, colour spaces that are non-linearly related to XYZ space are not suitable for this application. Another suitable space in which to compare

ambient-subtracted chromaticities is the 1976 CIE UCS (Uniform Chromaticity Space). This chromaticity space is more perceptually uniform than the  $xy$  chromaticity space. The UCS chromaticity coordinates ( $u', v'$ ) are calculated as shown in Equation 3.10.

$$\begin{aligned} u' &= \frac{4X}{X + 15Y + 3Z} = \frac{4x}{-2x + 12y + 3} \\ v' &= \frac{9Y}{X + 15Y + 3Z} = \frac{9y}{-2x + 12y + 3} \end{aligned} \quad (3.10)$$

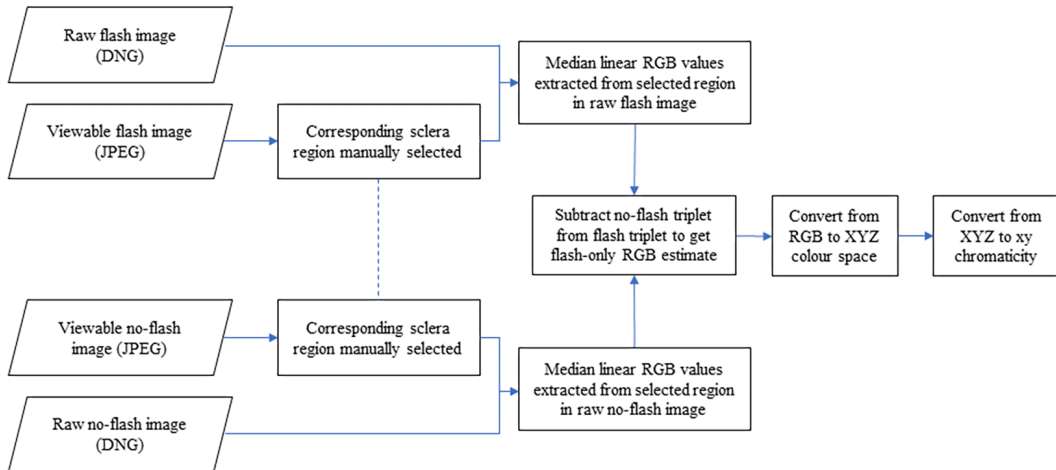
The one-time device calibration method was performed for each device used in the DJAN study following the method described by Nixon et al. (2019) [6]. A tripod was used to stabilise the smartphone and the 24 colour patches of the ColorChecker were captured using the app in an otherwise darkened room. A second image of the ColorChecker's grey card was taken with minimal movement in order to correct for any non-uniformity in the intensity of the flash across the colour card [6].

### 3.3.6 Pipeline Overview

Figure 3.2 shows the steps required to calculate the ambient-subtracted chromaticity of the sclera. Note that the subtraction is performed on the median value of a manually selected region of interest, not pixelwise. This is to allow for small movements between the flash and no-flash captures - the two images are not perfectly coregistered, as would be required for pixelwise subtraction. By identifying the region of interest manually, the user can account for small movements, and ensure the same sclera region is selected in each image. In Chapter 4, different manual segmentation approaches are compared on clinical datasets.

The median is chosen over the mean as a summary statistic so that the result is more robust to non-representative colour regions, such as specular reflections and blood vessels in the sclera. The median value for a set of RGB values in the ROI can be calculated in two ways. By taking the median value of each channel separately, it is possible that the resulting median RGB triplet does not exist anywhere as a recorded RGB triplet in the ROI. The geometric median generalises the median in

higher dimensions. The geometric median in a three-dimensional space is the point with the minimum sum of distances to all other points. Although this ensures the resulting RGB value corresponds to a single specific pixel in the ROI, it is more computationally expensive to calculate.



**Figure 3.2:** Pipeline for calculation of ambient-subtracted sclera chromaticity. Adapted from figure in Ref. [2].

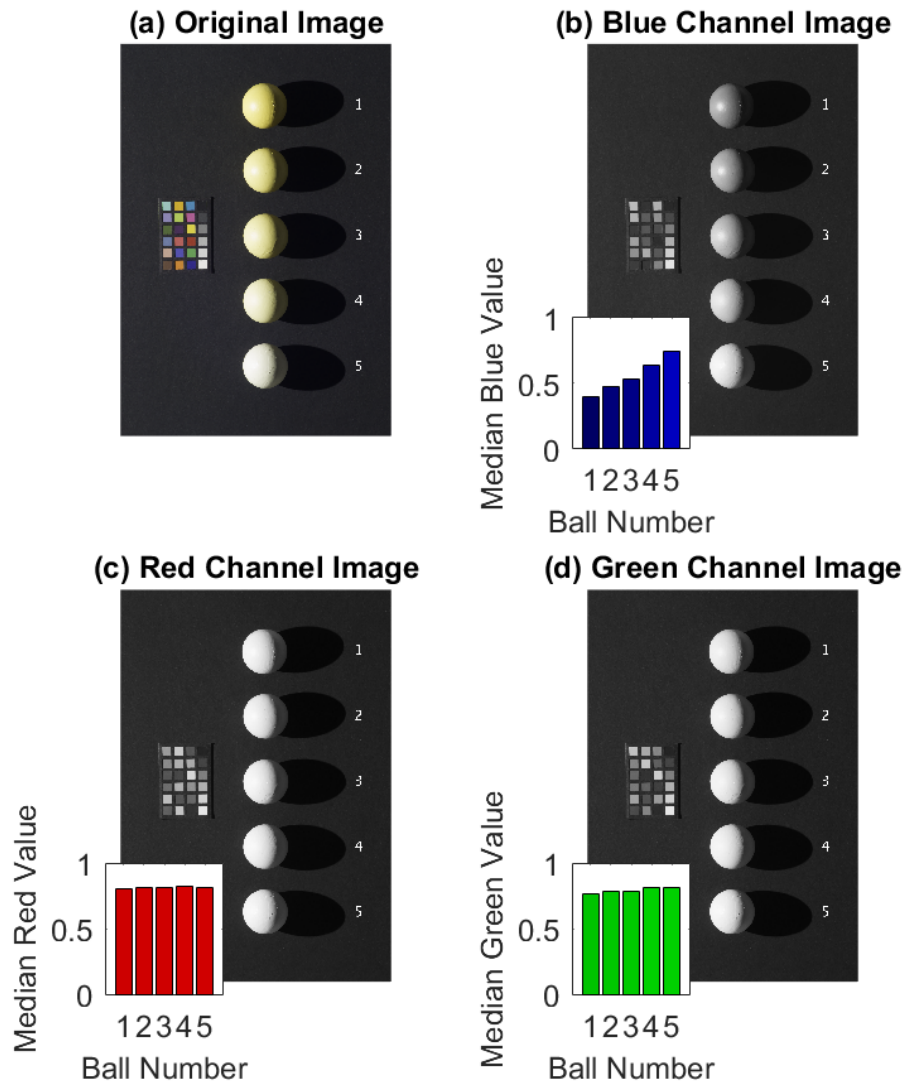
## 3.4 Proof Of Principle Experiments

In the previous section, the theory behind the ambient-subtracted chromaticity approach was laid out. In what follows, two proof of principle experiments are used to demonstrate chromaticity and ambient subtraction.

### 3.4.1 Snooker Ball Demonstration

A set of five snooker balls were painted with a mix of white and yellow paint in varying proportions. As shown in Figure 3.3, the blue channel shows the greatest relative contrast between the different degrees of yellowness. This is because yellow objects absorb blue light.

As discussed in the previous section, the absolute blue channel value cannot be used to quantify the degree of yellowness. A more intense light or longer exposure time can both increase the absolute blue channel count. Furthermore, different points on a curved object receive light at a different angle, which causes shading.

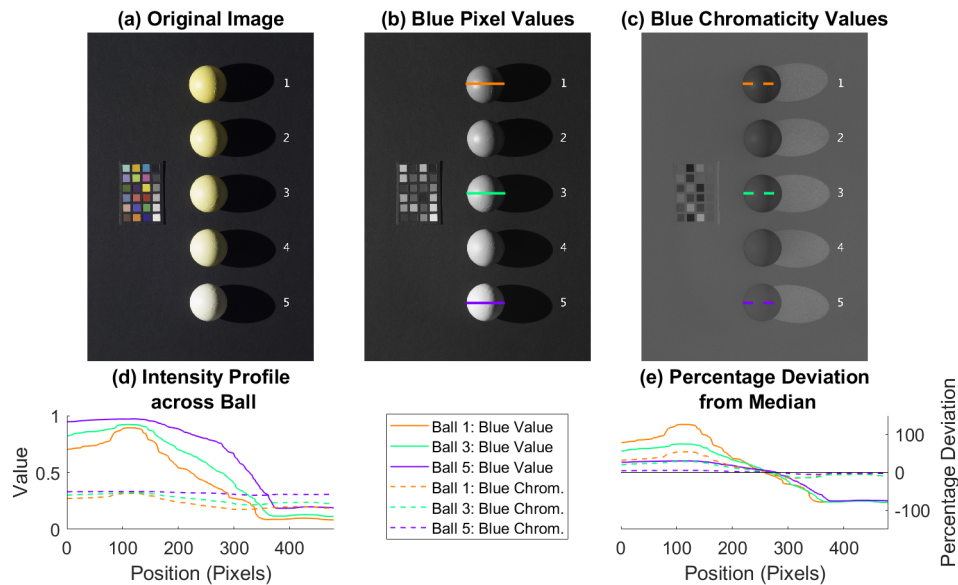


**Figure 3.3:** RGB channel images and relative contrast for five painted balls.

Using chromaticity mitigates these problems. Figure 3.4 compares the blue channel value to the blue chromaticity value for the same snooker ball image. The line profiles shown in Figure 3.4 (d) demonstrate a much greater variability in the absolute channel value than in the chromaticity value. The left side of the balls show some degree of specular reflection, while the right is in shade. As discussed in the theory section, chromaticity is invariant with respect to shading, but areas of specular reflection have a different chromaticity compared to the bulk of the object. Figure 3.4 (e) shows that the chromaticity of the shaded area is almost the same as the



median for the whole, but the specular region shows some deviation. The overall percentage deviation is much less in chromaticity than in the raw channel value. In our application, where the relative degree of yellowness must be quantified and the absolute intensity of the recorded signal is indeterminate following subtraction, a metric based on blue chromaticity offers good contrast.



**Figure 3.4:** Snooker balls: (a) original image, (b) blue channel value image, and (c) blue chromaticity image. (d) shows how the intensity value varies in the blue channel compared to how it varies in blue chromaticity for a horizontal line profile. (e) shows the same as (d) but as a percentage deviation from the median value for each ball. Figure originally published in Ref. [1].

### 3.4.2 Jaundice Eye Colour Index

The dataset first analysed by Leung et al. (2015) (UCLH Nikon Study) was reanalysed to evaluate blue chromaticity as a metric for jaundice in the sclera [74] [1]. 87 subjects with raw images available were included in the analysis, with TSB ranging from  $17\mu\text{M}$  to  $304\mu\text{M}$ . As described in Chapter 2, these data were captured using a Nikon D3200 camera. All data were collected in the Advanced Neonatal Nurse Practitioner Clinic at University College London Hospital.

The raw NEF format images were converted to raw 16-bit linear TIFF images using *dcrw* (version 9.27 by Dave Coffin 2016 [98]) and processed in MATLAB (MathWorks, Inc.). Sclera regions were manually segmented and the median blue

chromaticity was calculated in the camera raw space. Using a simple linear regression between TSB and raw space blue chromaticity, a correlation of -0.73 was observed.

In the original study (n=110), a regression was trained using the sclera RGB values produced using white balance from the camera fluorescent preset and the proprietary software of the Nikon ViewNX 2 software to convert into TIFF images. A regression involving cross terms and quadratic terms gave a correlation of 0.75 (95% confidence interval (CI): 0.65 - 0.82).

These results showed that by using a chromaticity-based metric motivated by first principles consideration of the image formation process, a similar predictive power may be possible with a much simpler predictive model. This is important as models involving fewer terms are more parsimonious and more likely to maintain their validity across different datasets.

In the work on the adult jaundice detection system BiliScreen, Mariakakis et al. generated features based on five colour spaces and the six ratios between RGB channels for a total of 21 colour features [88]. Using automatic feature selection methods they found that the green-to-blue channel ratio feature has the most explanatory power. Like blue chromaticity, this involves the blue channel, which captures the variation in yellow, and a normalisation by a channel not affected by bilirubin absorption: the green channel. In this way, Mariakakis et al. arrive at a similar conclusion via purely statistical methods as we have via the first-principles considerations above.

Although blue chromaticity has been shown to effectively capture the colour variation due to jaundice in the sclera, a definition based on a raw colour space cannot be used as a basis for comparison. Therefore, Jaundice Eye Colour Index (JECI), defined in Equation 3.11, is proposed as a metric by which sclera yellowness can be objectively compared [1]. JECI is based on z chromaticity from the XYZ colour space. The negative sign is included so that larger positive JECI values correspond to more jaundiced individuals. An offset  $z_0$  is included such that when z chromaticity is the same as under the D65 illuminant white point ( $z = z_0$ ), JECI is

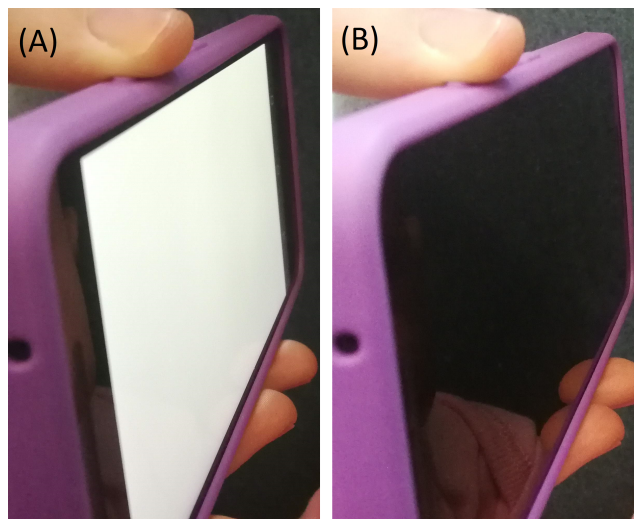
zero. To three significant figures,  $z_0$  is 0.358.

$$JECI = z_0 - z \quad (3.11)$$

### 3.4.3 Ambient Subtraction Demonstration

In the UCLH Nikon dataset all images were captured in approximately the same light conditions and it was not necessary to mitigate the effect of ambient light. We have proposed ambient subtraction as a means to discount the effect of ambient illumination. Here we show that ambient subtraction can improve the accuracy of a chromaticity estimate of a set of printed test patches of different JECI levels.

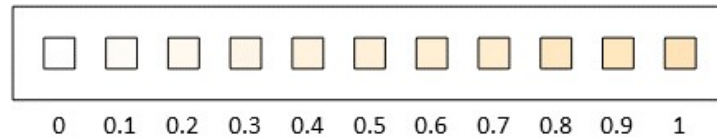
An LG Nexus 5X smartphone is used to capture images. The screen is used as a source of diffuse illumination and the front-facing camera is used to capture two consecutive images, one with the screen “flash” on, one with it off. Figure 3.5 shows the Nexus 5X smartphone with screen illumination on (Figure 3.5 (A)) and off (Figure 3.5 (B)). The ISO and exposure are set according to the first, flash image. The ambient light is provided in a controlled manner by a TaoTronics TT-DL09 LED desk lamp, which has a variable colour temperature. Two colour temperatures were used: 2700K (warm) and 6500K (cool). The smartphone was fixed in place relative to the test target centre at distance of 15cm and an angle of 45°.



**Figure 3.5:** LG Nexus 5X with screen illumination on (A) and off (B).

The test target was a set of 11 square patches of increasing JECI value, from 0

to 1.0 in increments of 0.1. 0 is white and 1 the most yellow. A given JECI value corresponds to a chromaticity, so the Y value is indeterminate. The Y value for each patch was chosen so that it would be within the printer gamut. The accuracy of the printed chromaticities was checked using an X-Rite ColorMunki Photo, and Figure 3.6 shows a rendering of these colours as printed.



**Figure 3.6:** Rendering of printed JECI patch colours as measured by X-Rite ColorMunki Photo. JECI values range from 0 to 1 in steps of 0.1.

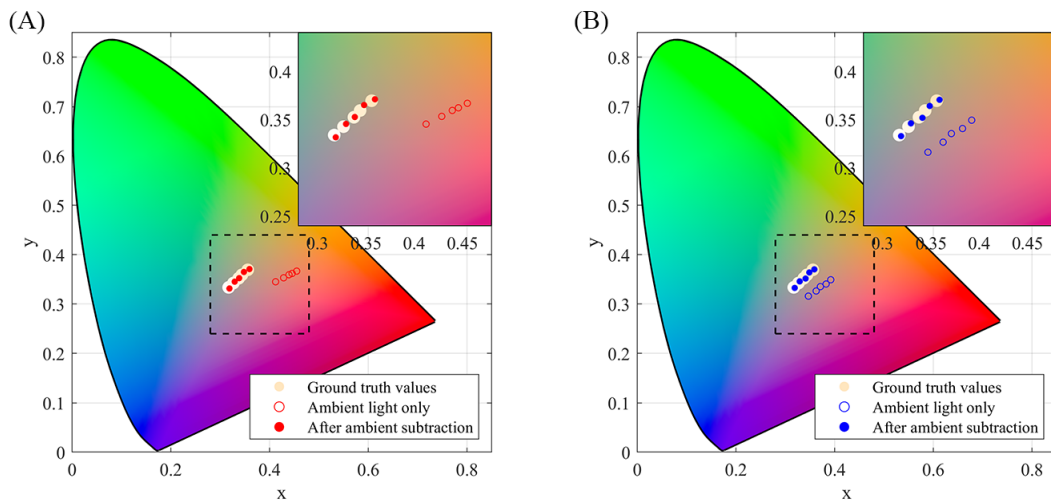
The pipeline shown in Figure 3.2 was used to calculate the ambient-subtracted chromaticity for each patch under each colour temperature of ambient illumination.

Figure 3.7 shows the results in the xy chromaticity space. The coloured gamut represents the human visual space. The results for every second patch are plotted, with the ground truth values represented by the solid circles of their respective patch colour. The filled circles are the results after ambient subtraction, and the empty circles are the recorded values under the ambient light.

The accuracy of the ambient-subtracted chromaticity estimation is better than the ambient-only chromaticity estimate for both warm (Figure 3.7 (A)) and cool (Figure 3.7 (B)) ambient illuminations.

Although this proof of principle experiment is restricted to one smartphone and two ambient illuminations, it shows that ambient subtraction using a smartphone is possible and can improve the accuracy of the chromaticity estimate for a range of colours typical of the jaundiced sclera.

Chapter 4 starts with an evaluation of the colorimetric accuracy of ambient subtraction with the neoSCB app. The results are compared to several other methods using colour card data collected in clinical conditions in the Ghana Smartphone Study.



**Figure 3.7:** Ambient subtraction demonstration in  $xy$  chromaticity space for a range of JECI values. Ground truth values are represented by yellow circles. Accuracy of chromaticity estimation following ambient subtraction (filled circles) is better than ambient-only condition (empty circles). (A) Warm colour temperature (2700K). (B) Cool colour temperature (6500K). Figure originally published in Ref. [2]. CIE chromaticity diagram rendered using Ref. [99].

## 3.5 App Development

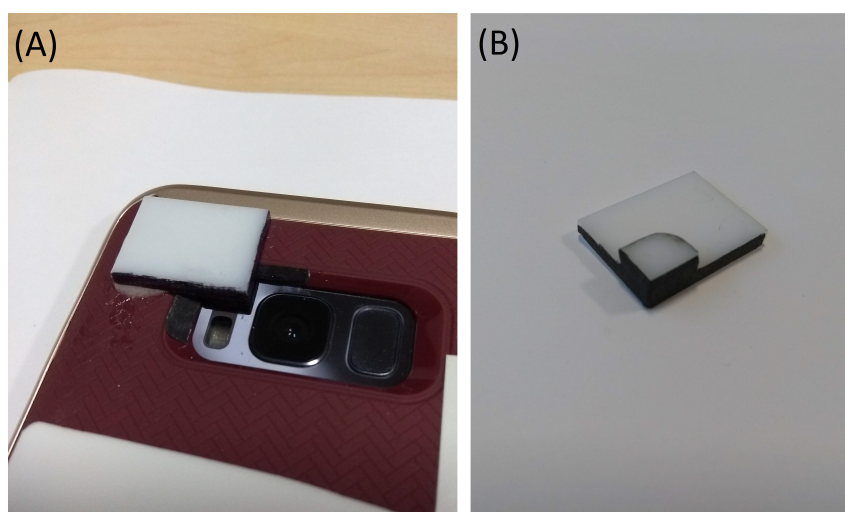
### 3.5.1 Hardware Requirements

In the preceding sections, some important requirements of the smartphone were assumed:

1. Linearity: the smartphone camera's response to light should be linear. This is needed for ambient subtraction.
2. Access to raw image data: without this facility, it is impossible to do a calibration and unknown proprietary algorithms will influence the output colour values.
3. Consistency and stability of flash: whether the screen or LED flash is used as an illumination, its colour must be consistent between captures and also stable over time.
4. Safety of flash: it must be confirmed that it is safe to image the newborn eye with the flash.

Ultimately, two models of smartphone were used in the clinical studies in this thesis. For the UCLH Smartphone Study, a single LG Nexus 5X smartphone was used. The flash was provided by the screen illumination and the front-facing camera was used. The screen based illumination was advantageous because it provided a diffuse and gentle illumination to the subject. On the other hand, it became clear that it was difficult to image the subject's eye without being able to see a preview of the capture. The number of smartphones offering access to the raw pixel data from the front-facing camera is also limited. At the time of writing, Google Nexus and Google Pixel smartphone models offer front-facing camera raw support, but the majority of Android device models do not.

An alternative approach was selected for the DJAN study in Ghana. By using the back-facing camera a better quality image could be obtained by an inexperienced user, and a greater majority of smartphones would be able to support raw capture in this mode. A set of four devices of the same make and model were used (Samsung S8). Although it is confirmed in this study (see Section 3.5.4) and elsewhere that a smartphone flash does not pose a risk to the newborn eye, a diffuser was added to the flash for the sake of the study to avoid any discomfort for the subjects or concern from their parents. This diffuser was designed and made by Miranda Nixon-Hill (UCL Centre for Doctoral Training in Medical Imaging). The diffuser is made from



**Figure 3.8:** The flash diffuser for the Samsung S8. (A) Diffuser fitted in place. (B) Diffuser before attachment.

acrylic and fitted to the shape of the phone case. The edge is painted black to prevent light entering the camera directly from the edge of the diffuser. Figure 3.8 shows the diffuser design and how it is attached to the smartphone.

### 3.5.2 Development Process and Features

The first app version was developed by this author for use in the UCLH Smartphone Study. The app was coded in Java in the Android Studio Integrated Development Environment (version 3.0.1). Two images were captured in quick succession using a predefined exposure setting. The screen was entirely white and at full brightness for the first capture and entirely off for the second. When not in use, the screen is on. The screen turns off momentarily for the second capture. This is to allow the subject to acclimatise to the brightness and to avoid a short bright flash that could be overstimulating. The volume key was used to initiate the capture sequence so that the user would not have to touch the screen to take an image, as this could block the light. Images were saved in both raw (Adobe DNG) and JPEG formats.

The second version of the app was developed for the Samsung S8 back-facing camera for use in the DJAN study. Contributions to the development process were made by Alister Lam (Frugal Spark Ltd), Josephine Windsor-Lewis (UCL Biomedical Engineering), and Fiona Young (UCL Natural Sciences). As in the previous version, the flash is on continuously until the second capture. This has an added benefit when using the back-facing camera, as the capture preview screen can be used to position the smartphone so that the flash specular reflection is not on the sclera. This version has an auto-exposure routine which sets the exposure time and ISO setting according to the flash image conditions. This means the flash image is well exposed in a variety of ambient conditions, and the no-flash image is not overexposed. A welcome screen on starting the app and a review screen after capture ensure that the battery is not drained by the flash remaining in torch mode. As in the previous version, both DNG and JPEG formats are saved for both captures.

After capture, the user is able to see both images to assess their suitability. By panning and zooming on the flash image, a green box is positioned in the sclera ROI on both flash and no-flash images. This means that a real time calculation can be

performed by the app. The SSNR is calculated using the ROI in each image. If it is sufficient, and an RGB-XYZ mapping is found for the specific device being used, a chromaticity value or bilirubin estimate can be calculated for the user immediately.

Appendix B includes more details and the usage guidance for this version of the neoSCB app.

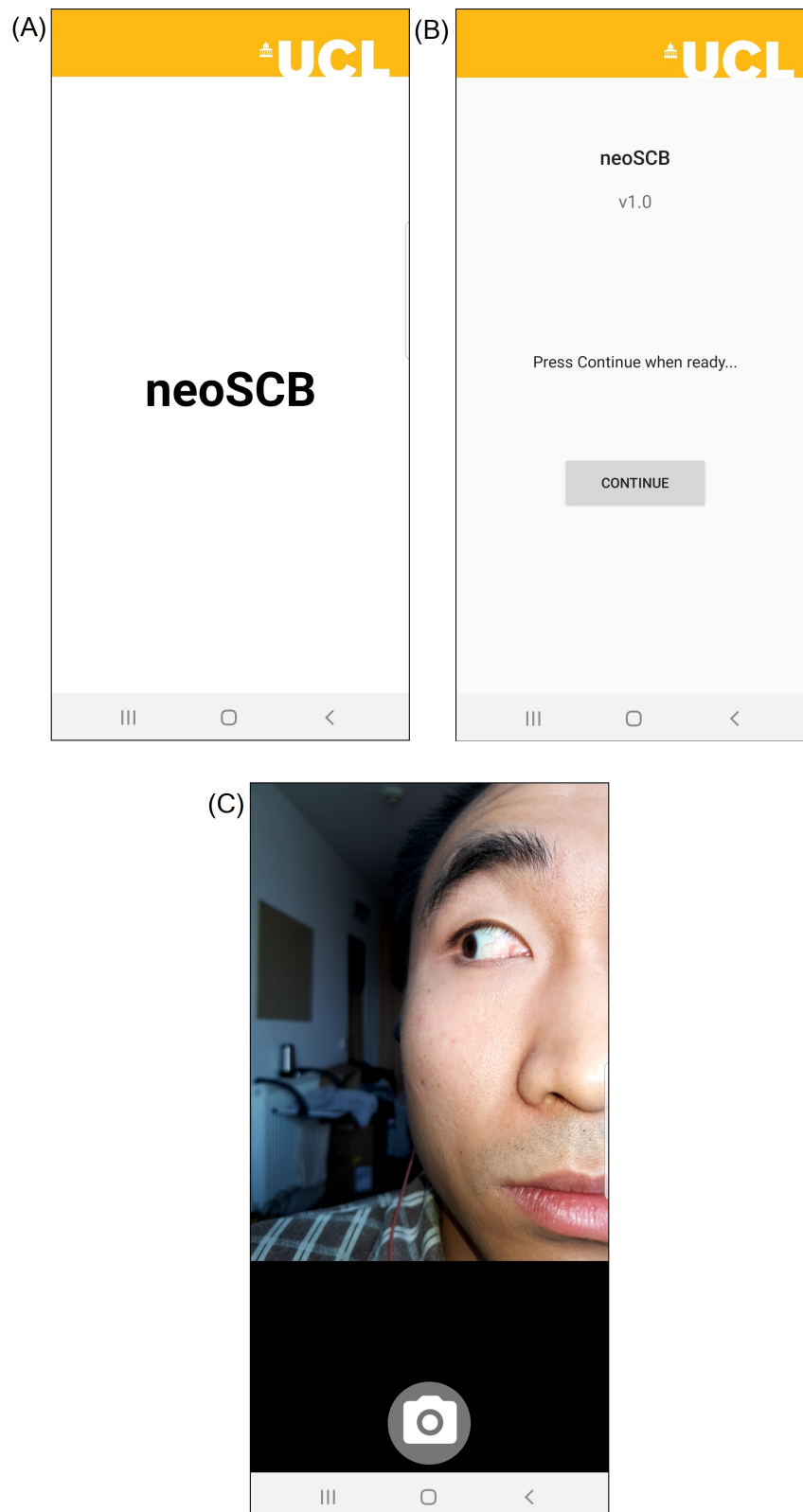
### 3.5.3 App Walk-Through

Screenshots from the latest app version, neoSCB v1.0, running on a Samsung S8, are shown in Figure 3.9 and Figure 3.10 below.

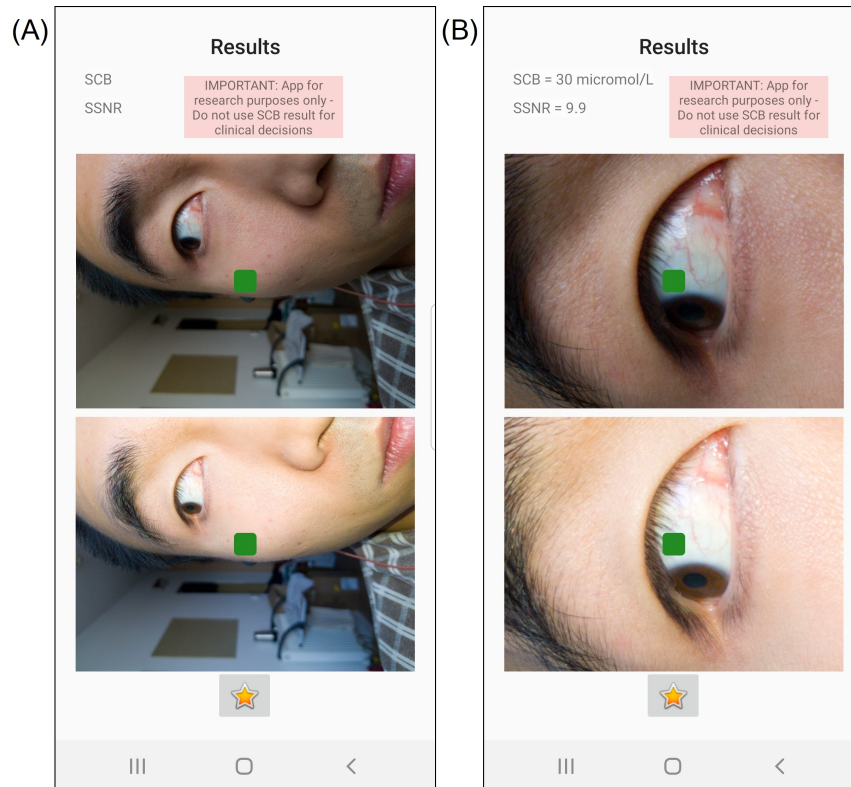
On starting the app, the user sees a splash screen (Figure 3.9 (A)), followed by a welcome screen (Figure 3.9 (B)). Once the user is ready to take a measurement, they press the **Continue** button. From this point until the capture of the flash image the flash is on continuously in torch mode. Figure 3.9 (C) shows the camera preview. When the sclera is in view and the subject is still, the photo capture button at the bottom is pressed. This automatically captures a flash/ no-flash image pair.

Once the capture sequence is concluded, the app shows the review screen (Figure 3.10 (A)). The lower, flash image can be swiped and pinched in order to pan and zoom the image relative to the green square. These view changes affect the no-flash image in the same way. Once the green square is satisfactorily positioned in the sclera in both images, the user presses the star-shaped calculate button. The SSNR and Scleral-Conjunctival Bilirubin (SCB) values are calculated according to the process laid out in this chapter. The result is displayed as in Figure 3.10 (B). The red box shows a disclaimer. This warns the user that the app is still under development and should therefore not be used to make consequential clinical decisions.





**Figure 3.9:** Screenshots from a Samsung S8 running the neoSCB app. (A) Splash screen. (B) Welcome screen. (C) Camera preview screen.



**Figure 3.10:** Screenshots from a Samsung S8 running the neoSCB app. (A) Review screen before positioning region of interest (green box). (B) Review screen after green box has been positioned and SCB and SSNR have been calculated.

### 3.5.4 Safety Measurements

It is important to verify that the flash illumination is safe for the eye. The internationally accepted committee which defines exposure limits for laser radiation and broadband optical radiation is the International Commission for Non-Ionizing Radiation Protection (ICNIRP).

Smartphone back-facing flash illumination is LED-based. LEDs are treated as incoherent optical sources by all safety guidelines. Unlike lasers, they have broadband spectra and are not collimated [100]. The 2013 ICNIRP guidelines for incoherent optical radiation are therefore appropriate for LEDs [101].

There are two types of damage to the retina that can arise from incoherent visible light. The first is photothermal damage. This arises due to the thermal energy deposited in the retina by wavelengths between 380nm and 1400nm. This energy dissipates as heat over some timescale, so it is the rate of deposition of thermal en-

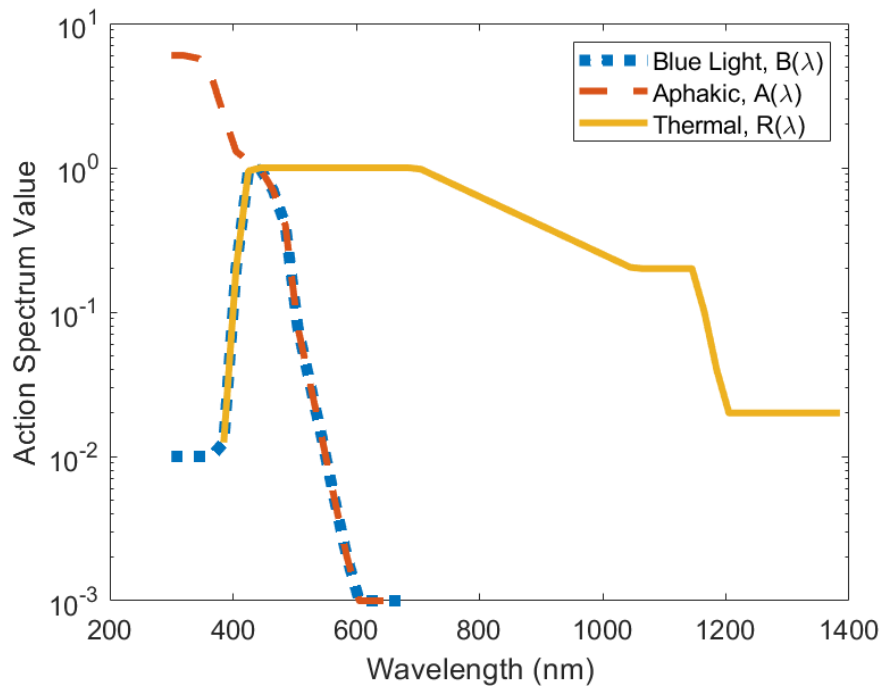
ergy that is the crucial factor in determining source safety. The second mechanism is photochemical damage. This is sometimes referred to as “blue light” damage because the wavelengths that cause it are between 300nm and 550nm. Photons of these wavelengths cause the damage when absorbed by the pigmented layer of the retina [102]. This form of damage is purely dose dependent, meaning that either a longer exposure or a higher irradiance can increase the overall photochemical damage [101].

Some wavelengths are more efficient in causing either photochemical or photothermal damage than others. Action spectra are defined for each mechanism of damage. The purpose of these is to weight the source radiance at each wavelength by the relative potential to do harm at that wavelength. This spectrally weighted radiance is known as the effective radiance. The effect of light of different wavelengths is cumulative, so an integral or summation across the spectrum is needed to estimate the potential danger. Equation 3.12 shows such a summation. The effective thermal radiance ( $L_R$ , in watts per steradian per square meter) is the sum across all wavelengths ( $\lambda$ , in nm) of the product of the thermal action spectrum ( $R(\lambda)$ ) and the spectral radiance ( $L_\lambda$ , in watts per steradian per square meter per nanometer).

$$L_R = \sum_{\lambda} L_{\lambda} \times R(\lambda) \times \Delta\lambda \quad (3.12)$$

A similar equation to Equation 3.12 is used to calculate the effective radiance for photochemical damage, with the blue light action spectrum  $B(\lambda)$ . A modified blue light action spectrum is used when the subject has no crystalline lens. This is called the aphakic condition, and may occur during, for example, an eye surgery. The lens absorbs a large fraction of incident ultraviolet light, so without it the retina becomes more vulnerable to photochemical damage. The aphakic action spectrum,  $A(\lambda)$ , is also used for newborns, as the lens of children under the age of two has a much higher transmittance in violet and ultraviolet [101]. Figure 3.11 shows all three action spectra for retinal hazard on a logarithmic scale.

To determine if a source is safe, the effective radiance is compared to an exposure limit defined by the ICNIRP. These limits depend on the length of exposure



**Figure 3.11:** Action spectra for retinal hazard.  $R(\lambda)$  is the photothermal action spectrum,  $B(\lambda)$  is the photochemical action spectrum, and  $A(\lambda)$  is the aphakic photochemical action spectrum. Data from Ref. [101].

and the image size on the retina. For the reasons mentioned above, the photothermal limit is expressed as a radiance, while the photochemical limit is expressed as radiant dose (a time-integrated radiance).

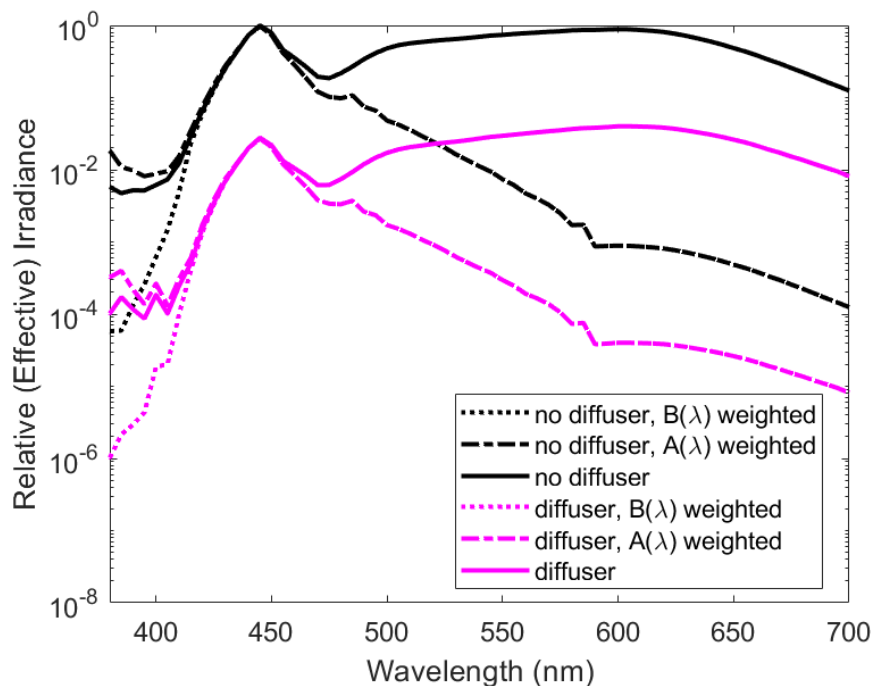
It is well established that LEDs do not have sufficient radiant power to cause retinal photothermal damage [100]. Photochemical damage is more of a concern, especially if there is significant power in the UV region of the spectrum. White LEDs use a UV emission at 390nm or blue emission at 450nm in combination with a phosphor. A peak in the 460-480nm region coincides with the peak of the  $B(\lambda)$  action spectrum, and in the aphakic case, UV radiation is particularly harmful (see Figure 3.11).

Point (2018) considered both the aphakic adjustment and the geometry of the newborn eye to investigate whether the blue light exposure limits were safe for newborns [103]. Point concludes that limits should be lower by a factor of 2.8 to account for the newborn pupil diameter and focal length. A similar factor of 3 was arrived at by Mactier et al. (2008) by considering the biometry of the newborn eye [104].

Point found little difference between effective radiance of white LEDs under aphakic and phakic conditions due to the low output in the violet and ultraviolet. In contrast, he found the aphakic adjustment became important for the calculation of the effective radiance of halogen and fluorescent lights.

Irradiance measurements were performed using a Spectrometer (HR2000+, Ocean Optics, Inc., Florida, USA) of the Samsung LED flash at a separation of 10cm. The fibre probe was mounted in line with the LED and translated perpendicular to this axis until a peak irradiance was observed in the SpectraSuite software user interface, whereupon the three measurement repeats were made with and without the acrylic diffuser. Figure 3.12 shows the resulting average relative irradiance and the average relative effective irradiance after weighting by the action spectra  $B(\lambda)$  and  $A(\lambda)$ .

The peak irradiance output of the Samsung S8 back-facing flash is at 445nm. The introduction of the diffuser attenuates the LED by a factor between 40 (at 380nm) and 15 (at 700nm): there is a greater attenuation at (more dangerous)



**Figure 3.12:** Spectral measurements of Samsung S8 LED flash relative irradiance and relative effective irradiance under blue light phakic and aphakic assumptions with and without diffuser.

shorter wavelengths. Figure 3.12 shows minimal difference between the relative effective irradiance after weighting by phakic and aphakic blue light action spectrum. The difference is only significant below 400nm, where, as noted by Point, power output for (blue-phosphor) white LEDs is low.

Several studies have investigated fundus photography using a smartphone as an indirect ophthalmoscope, and most use the back-facing LED torch for illumination without employing a diffuser [105] [106] [107]. Goyal et al. (2019) used this approach to investigate retinopathy of prematurity among premature newborns [108]. Even with a medically induced dilation of the pupil such applications are deemed safe.

Kim et al. (2012) evaluated the safety of the iPhone 4 (Apple Inc., Cupertino, USA) flash for use as an indirect ophthalmoscope [109]. They calculated retinal irradiances would be two orders of magnitude below the relevant safety standard for ophthalmic instruments produced by the International Organisation for Standardisation (ISO 15004-2.2) for both photothermal and photochemical mechanisms. Hong et al. (2017) concluded that the iPhone 6 and 6S were safe by a similar margin for a one-minute exposure [110]. The ISO standard is itself one order of magnitude below the damage threshold [111]. Kim et al. also note that the retinal irradiance from the smartphone is an order of magnitude less than standard indirect ophthalmoscopes [109].

In conclusion, the safety of LED smartphone flashes is well documented. Indirect ophthalmoscopes provide a widely used point of comparison for the safe use of an LED on the neonatal eye. The spectrometer measurement geometry was not sufficiently well specified to calculate the absolute effective radiance of the Samsung S8 flash. However, the measurements showed that the flash has a typical LED profile with very little UV component. They also showed that the diffuser attenuated the irradiance by two orders of magnitude, guaranteeing a high margin of safety.

## 3.6 Summary

In this chapter the key challenges in smartphone camera-based jaundice detection were explored, leading to a proposal for a method based on ambient subtracted sclera chromaticity. This method is motivated by the need to create a system that leverages the affordability and accessibility of smartphones to be usable in LMICs by avoiding the use of accessories or add-ons. In the literature, one of two approaches were adopted for achieving device- and ambient-independent colour readings: colour cards and contact-based, ambient light blocking housings. Ambient subtraction after a one-time calibration combines the advantages of both to give a contact-free solution that does not require a colour card to be distributed to all potential users and maintained in complex clinical and home environments.

The theory of ambient-subtracted chromaticity was outlined, giving insight into the advantages as well as the assumptions behind the approach. Two proof of principle experiments demonstrated the advantages of a chromaticity metric and the feasibility of ambient subtraction. A simple sclera chromaticity-based metric was shown to correlate with bilirubin levels in a clinical dataset.

Several questions posed by this exploration remain unanswered: Is ambient subtraction still feasible in a real-world clinical context with moving subjects and varied lighting? How else could ambient illumination be mitigated without a colour card? Does thresholding based on subtracted signal-to-noise (SSNR) guarantee sufficient accuracy? How much more sensitive is the sclera measurement site compared to the skin? How should the sclera be segmented? Does specular reflection need to be avoided entirely due to its influence on chromaticity?

In Chapter 4 these questions will be explored using the datasets from UCLH Smartphone Study (LG Nexus 5X front-facing camera) and the DJAN study in Ghana (Samsung S8 back-facing camera with LED diffuser).

## **Chapter 4**

# **Development of SCB Model to Predict Total Serum Bilirubin from Chromaticity**

### **4.1 Overview**

In this chapter the relationship between ambient-subtracted sclera chromaticity and TSB is modelled using clinical data from Ghana. First, the ambient-subtraction approach is compared to some other approaches to correct for the influence of ambient light, one of which also makes use of a flash/ no-flash image pair. Next, sclera chromaticities for DJAN subjects are plotted, as measured via ambient subtraction and via colour card mapping. Various models for TSB prediction from chromaticity are compared. The finalised model is validated on the UCLH dataset.

### **4.2 Real-world Subtraction Evaluation**

One of the aims of this thesis is to explore low-cost methods for accurate colorimetry that do not involve the use of a colour card. Ambient subtraction was introduced in Chapter 3 as a candidate solution. In this section, ambient subtraction is compared to other methods that do not require a colour card. The methods are compared using images from the DJAN study in Ghana. Colour cards were included in the first phase of data collection as a reference. Given the known colour values of the card,



it is possible to quantify the accuracy of the methods against a ground truth.

The methods are listed below. They are all calculated from the image data of one or both of the flash/ no-flash pair.

1. **MaxRGB** Division of each channel by the maximum recorded value in that channel is used to white balance the image. This assumes that there exists at least one object that is maximally reflective for each channel.
2. **Grey World** Division of each channel by the average of that channel across all pixels is used to white balance the image. This assumes the average colour of the objects in the scene is neutral.
3. **White Patch Reference** A neutral reference standard is included in the capture. Division of each channel by the patch mean value is used to white balance the image. This assumes that the same light source is illuminating the reference patch as is illuminating the rest of the scene.
4. **Petschnigg White Balance [112]** This method involves a flash/ no-flash image pair. If the flash colour is known *a priori* and corrected for, then the difference image is proportional to surface albedo<sup>1</sup>. The ratio between ambient pixel value and albedo isolates the influence of ambient light, and is calculated per pixel. The average of the ambient light estimate across the image is used calculate global white balance multipliers. This approach assumes that objects are diffuse and that the scene is illuminated by a single colour illumination.
5. **Ambient Subtraction** The method described in Chapter 3 also involves a flash/ no-flash image pair. If the flash colour is known *a priori* and corrected for, then the difference image is proportional to surface albedo. This is similar to a local application of the method of Petschnigg et al.. No assumption needs

---

<sup>1</sup>Albedo is defined as the fraction of incident electromagnetic energy reflected by a surface. It is often quoted integrated over a wavelength range. Here, the albedo is three-dimensional, with values across the red, green, and blue channels of the CSS. It is an illumination-independent surface property.

to be made about the number of illuminations in the scene as a whole, but the illumination and albedo are assumed to be constant within the ROI.

### 4.2.1 Methods

The colour card was an X-Rite ColorChecker Passport. All colour card images associated with a DJAN study subject that were fully visible in the capture and without saturated regions were included in this analysis. The colour card was measured with a ColorMunki Photo to determine ground-truth XYZ values. Median RGB values for each patch were extracted from a linear 16-bit TIFF developed by ddraw, with no white balance applied. Figure 4.1 shows a picture of the colour card used and the number assigned to each patch.



**Figure 4.1:** Picture of X-Rite ColorChecker Passport with patch numbers.

In each case, following the white balance step, the 3x3 linear mapping found in the DNG metadata was used to convert RGB values to XYZ values. Following this, xy chromaticity values were calculated and compared to ground truth chromaticity values by quantifying the distance in xy space. The metadata mapping (rather than the one-time calibration mapping developed under flash illumination) was used in order to compare these methods strictly by their white balance accuracy.

## 4.2.2 Implementation

In the general case, when performing white balance, we seek a kernel  $\mathbf{W}(\mathbf{x})$  of multipliers to apply to an ambient-only image  $I^A$  to give a white balanced image  $I^{wb}$ . This is expressed in Equation 4.1 for colour channel  $k$ .

$$\hat{I}_k^{wb}(\mathbf{x}) = \mathbf{W}_k(\mathbf{x})I_k^A(\mathbf{x}) \quad (4.1)$$

**MaxRGB White Balance** A triplet of white balance multipliers is calculated from the maximum value of each channel and applied globally, as shown in Equation 4.2.

$$W_k^{mv} = \frac{1}{\max_x I_k(\mathbf{x})} \quad (4.2)$$

In the implementation of this algorithm, the image is first resized to have a height of 64 pixels using bicubic interpolation, while keeping the aspect ratio constant. MATLAB's *imresize* function is used. This makes it less likely that the maximum value is a small area of specular reflection or a defective pixel, and has been shown to improve performance [65].

**Grey World White Balance** A triplet of white balance multipliers is calculated from the average channel value and applied globally, as shown in Equation 4.3.

$$W_k^{gw} = \frac{\sum_k \sum_x I_k(\mathbf{x})}{3 \sum_x I_k(\mathbf{x})} \quad (4.3)$$

**White Patch White Balance** A triplet of white balance multipliers is calculated from the average value of each channel in the reference neutral and applied globally, as shown in Equation 4.4.

$$W_k^{wp} = \frac{\sum_k \sum_{x \in wp} I_k(\mathbf{x})}{3 \sum_{x \in wp} I_k(\mathbf{x})} \quad (4.4)$$

### Petschnigg Albedo Ratio White Balance

Petschnigg et al. (2004) note that the difference image, which is the flash-only estimate  $\hat{I}^F$  given by ambient subtraction, is proportional to the albedo of the surface

$\alpha_k(\mathbf{x})$  if the flash colour  $e^F$  is accounted for [112].

$$\alpha_k(\mathbf{x}) = \frac{I_k^{A+F}(\mathbf{x}) - I_k^A(\mathbf{x})}{e_k^F} = \frac{\hat{I}_k^F(\mathbf{x})}{e_k^F} \quad (4.5)$$

The ambient illumination at  $\mathbf{x}$ ,  $e^A(\mathbf{x})$ , is then given by the recorded value in the no-flash image divided by the albedo at each channel  $k$ :

$$e_k^A(\mathbf{x}) = \frac{I_k^A(\mathbf{x})}{\alpha_k(\mathbf{x})} \quad (4.6)$$

$e^A(\mathbf{x})$  has unknown scale due to the geometric shading factor and flash-object distance being unknown. To calculate white balance multipliers, the ambient colour is averaged over the set of pixels  $\zeta$  for which the no-flash image  $I_k^A(\mathbf{x}) > \tau_1$  and the object albedo  $\alpha_k(\mathbf{x}) > \tau_2$ .  $\tau_1$  and  $\tau_2$  are thresholds chosen to increase the robustness of the estimate. In this implementation we follow Petschnigg and set both thresholds to 0.02 [112]. Equation 4.7 gives the multipliers for the albedo ratio method of white balance proposed by Petschnigg.

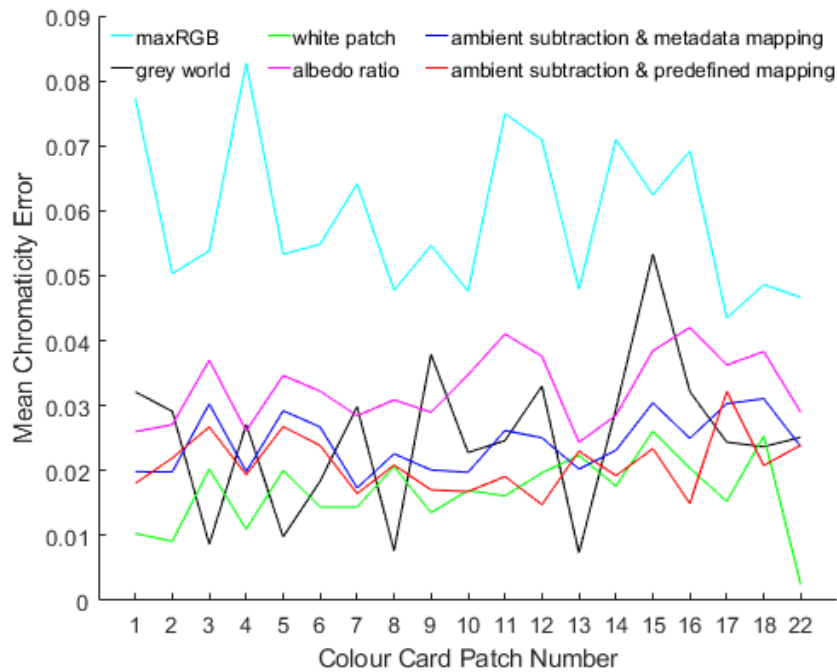
$$W_k^{ar} = \frac{\sum_k \sum_{x \in \zeta} e_k^A(\mathbf{x})}{3 \sum_{x \in \zeta} e_k^A(\mathbf{x})} \quad (4.7)$$

**Ambient Subtraction** The ambient subtraction method introduced in Chapter 3 is not strictly a method for white balance. A white balance kernel is not sought. Instead ambient subtraction is applied only to a specific region of interest to recover the local surface albedo as defined in Equation 4.5.

### 4.2.3 Results

Figure 4.2 shows the mean chromaticity error for each of the white balance methods outlined above for each colour card patch. Patches for which the SSNR is less than the 3.4 threshold are not included. Ambient subtraction using the mapping to XYZ provided in the image metadata is compared to ambient subtraction using a custom mapping from the one-time calibration step.

Table 4.1 compares the mean, median, and 95<sup>th</sup> percentile chromaticity error across all patches. Only one of the neutral patches, patch 22, is included so as



**Figure 4.2:** White balance comparison via mean chromaticity error per patch.

**Table 4.1:** Performance metrics for six white balance and ambient subtraction methods.

|   | Chromaticity Error Across All Images and Patches |              |                 |
|---|--|--------------|-----------------|
|   | Median   | Mean         | 95th Percentile |
| maxRGB                                    | 0.044  | 0.060        | 0.163           |
| Grey World                                | 0.026  | 0.026        | 0.048           |
| White Patch                               | 0.016  | <b>0.016</b> | <b>0.031</b>    |
| Albedo Ratio                              | 0.027  | 0.034        | 0.081           |
| Ambient Subtraction (metadata mapping)    | 0.022  | 0.025        | 0.054           |
| Ambient Subtraction (calibration mapping) | <b>0.015</b>                                     | 0.021        | 0.052           |

not to overweight the error in recovering neutral chromaticity. As we can see in Figure 4.2, the white patch method, which is based on patch 21, is most accurate in recovering neutrals. maxRGB white balance performs the worst overall. Using a predefined mapping is found to be more accurate than a metadata mapping when using ambient subtraction. Ambient subtraction with a predefined mapping has the smallest median chromaticity error overall. White patch white balance has a lower mean and 95<sup>th</sup> percentile chromaticity error.

#### 4.2.4 Discussion

The results suggest that the ambient subtraction approach outlined in Chapter 3 gives more accurate chromaticity recovery than simple grey world or maxRGB white balance algorithms. This is true not in general but for our specific application of imaging a newborn at close range. For example, flash/ no-flash methods would not be effective for correcting for ambient illumination in large scenes because the flash would not illuminate far-away objects. By testing the algorithms on colour card image data collected concurrently with the clinical data, we can be confident that these results will also apply to the human subject data.

Grey world and maxRGB approaches both assume a single ambient light source throughout the image. This assumption is likely to be violated in a situation with a distinct foreground and background such as an image of a face. It is fairly common in indoor settings for both natural and artificial light to be present to varying degrees in different parts of an image. Furthermore, grey world assumes a broad gamut of colours will be present in the image. The images collected are dominated by either the face of the subject or fabric that is often colourful.

The flash/ no-flash method of Petschnigg et al., referred to as the albedo ratio method, also assumes a single source of ambient illumination. On the other hand, by ignoring pixels with a small subtracted value, the foreground will have more influence on the white balance multipliers than the background. This makes it more likely that the ambient light estimation will be relevant to the object of interest. However, for our application, there is no need to calculate white balance multipliers for the entire image and using data from the entire image. Instead we are able to identify a single region of interest and calculate an ambient-independent albedo chromaticity for that region. Across this region it is more likely that there is only a single ambient illumination colour.

Hui et al. demonstrate that it is possible to achieve a per-pixel white balance kernel  $\mathbf{W}(\mathbf{x})$  using a flash/ no-flash image pair [94]. They note that the albedo chromaticity is the same as the image chromaticity in an ideally white-balanced image. Following this observation, and assuming that the channel intensity of the

no-flash image is unchanged after white balance, the ratio of albedo chromaticity to no-flash image chromaticity gives a per-pixel triplet of white balance multipliers. This approach has the advantage of making no assumption about the number of independent sources of ambient illumination in the scene; the multipliers may be different for each pixel. The main drawback of this approach is the need for registered images.

The methods of Hui and Petschnigg, as well as grey world and maxRGB, are all proper white balance methods in that the result is a kernel that can be applied to one image (say, the no-flash image) as a whole. The relative intensity is preserved in different regions of the image so the white-balanced result has an output that can be viewed. Importantly, the image intensities in the white-balanced images are not absolute but relative, so only the chromaticity can be related to the object of interest. In our application, there is no need for a viewable image, so it is sufficient to directly calculate the ambient-subtracted chromaticity. Thus, returning to the grouping of illumination mitigation methods introduced in Section 1.4.2.3, our approach is better described as an invariant technique rather than a colour constancy technique.

The white patch white balance method allows for knowledge of the absolute intensity of the scene because it has a known reflectance. As discussed, this advantage must be traded off against the need to have a well-maintained reference neutral available in all circumstances in which the screening app is used.

The methods of Hui and Petschnigg and the ambient subtraction method all require knowledge of the flash colour, and so a reference neutral is required for a one-time calibration. A trained individual would be able to maintain a neutral card and calibrate as many smartphones as needed. There would be no need to position the card in the image shot or disinfect it between uses.

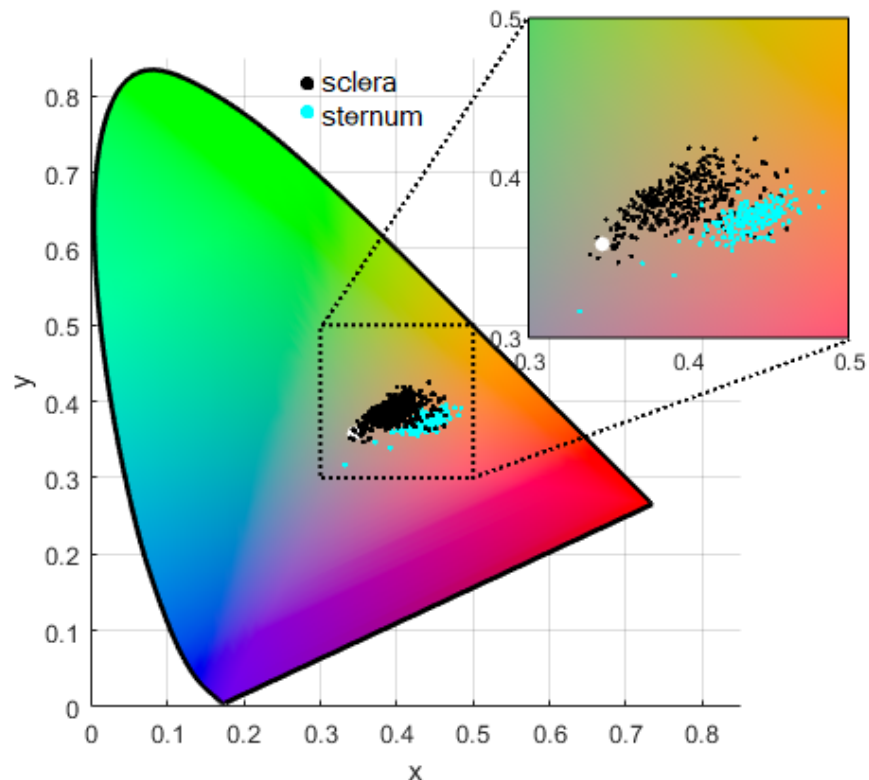
The results also show the benefit of using a device-level mapping developed from a colour card under the flash illumination. The performance is notably better compared to application of flash multipliers followed by a model-level, metadata-derived mapping. As both approaches require a calibration step, the full colour card mapping is preferable.

### 4.3 Sclera and Skin Chromaticity

In what follows, the sclera and skin chromaticities recovered using ambient subtraction are compared to those recovered using a mapping developed from a colour card.

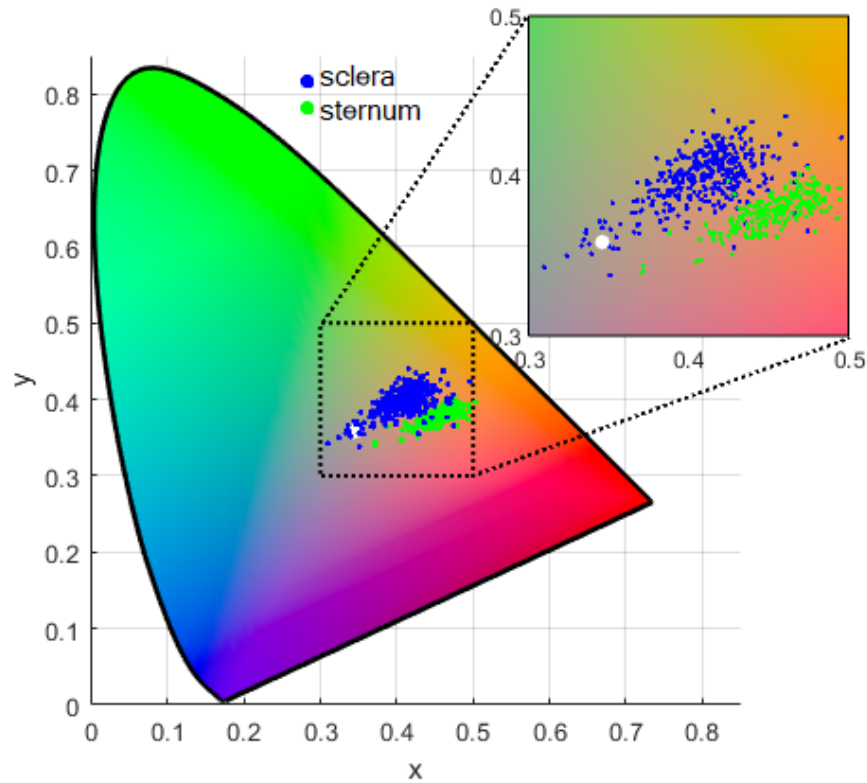
Chest images were not collected after Phase 1 of the DJAN study. Likewise, colour cards were only employed for Phase 1. In Phase 1a, the colour card was included in each shot. In Phase 1b, the colour card was imaged separately after the chest and sclera images.

For the sclera chromaticity data, 361 subjects are included. This represents the subset of the DJAN database with a complete no-flash colour card image without saturation in any of the patches and with a suitable sclera region of sufficient SSNR for subtraction. In cases with multiple suitable repeats for a given capture session, the chromaticity values are averaged. The equivalent number for chest images is



**Figure 4.3:** xy chromaticity for sclera (361 neonates) and sternum (192 neonates) as measured by app with colour card method. CIE chromaticity diagram rendered using Ref. [99].





**Figure 4.4:** xy chromaticity for sclera (361 neonates) and sternum (192 neonates) as measured by app with ambient subtraction method.

192. This is relatively small because many images pairs fail to meet the SSNR threshold.

Figure 4.3 shows the xy chromaticity values obtained using the colour card method and Figure 4.4 shows the xy chromaticity values obtained using the ambient subtraction method. In both cases, the sclera chromaticity values form a cluster ranging from the white point, D50, towards more yellow hues of increasing saturation. Meanwhile, skin chromaticity values are more red in hue and fewer points are near the white point. The subtraction-derived data are less tightly clustered for both sclera and skin. Some sclera chromaticity values exist on the blue side of the white point.

For sclera, the colour card median chromaticity is (0.396,0.386), with an x chromaticity range of [0.339,0.464] and a y chromaticity range of [0.347,0.425], while the ambient subtraction median chromaticity is (0.412,0.400), with an x chromaticity range of [0.310,0.496] and a y chromaticity range of [0.338,0.441].

For skin, the colour card median chromaticity is (0.439,0.374), with an x chromaticity range of [0.332,0.484] and a y chromaticity range of [0.317,0.395], while the ambient subtraction median chromaticity is (0.455,0.379), with an x chromaticity range of [0.371,0.503] and a y chromaticity range of [0.340,0.405].

### 4.3.1 TSB-chromaticity Correlations

In this section, the correlations between TSB and chromaticity are reported in order to complete the descriptive account of the data. The task of developing a predictive model is undertaken in the next section, Section 4.4.

Table 4.2 shows the correlations with TSB for sclera chromaticity recovered via ambient subtraction and colour card methods for 361 neonates. Table 4.3 displays the same information for the chest skin data from 192 neonates.

For both methods and measurement sites, y and x chromaticities are positively correlated with TSB and z chromaticity is negatively correlated with TSB. Equally consistent is the relative strength of these correlations: y shows the strongest correlation, followed by z and finally x.

Ambient subtraction results give consistently weaker correlations than colour card results. For skin data the discrepancy is particularly wide, and x and z correlations do not reach significance. For sclera data, the choice of method affects the y correlation most, with coefficients of 0.55 and 0.63 for ambient subtraction and colour card methods, respectively.

The skin and sclera data subsets are different, which means a strict comparison is not possible. However, it is obvious that skin chromaticity provides a weaker correlation with TSB than sclera chromaticity for both methods.

These results confirm the hypothesis that the sclera is a better measurement

**Table 4.2:** Sclera chromaticity correlation with TSB for the two methods, n = 361.

| Sclera Chromaticity Variable | Correlation with TSB: Ambient Subtraction | Correlation with TSB: Colour Card |
|------------------------------|---|-----------------------------------|
| x                            | 0.45 (p<0.01)                             | 0.44 (p<0.01)                     |
| y                            | 0.55 (p<0.01)                             | 0.63 (p<0.01)                     |
| z                            | -0.55 (p<0.01)                            | -0.58 (p<0.01)                    |

**Table 4.3:** Skin chromaticity correlation with TSB for the two methods,  $n = 192$ .

| Skin Chromaticity Variable | Correlation with TSB: Ambient Subtraction | Correlation with TSB: Colour Card |
|----------------------------|---|-----------------------------------|
| x                          | 0.06 ( $p=0.42$ )                         | 0.23 ( $p<0.01$ )                 |
| y                          | 0.22 ( $p<0.01$ )                         | 0.32 ( $p<0.01$ )                 |
| z                          | -0.12 ( $p=0.09$ )                        | -0.28 ( $p<0.01$ )                |

site than the skin. Also as expected, z chromaticity is negatively correlated with TSB because bilirubin absorbs in the blue region of the spectrum.

In the absence of ground truth reflectance spectrum measurements of the sclera, we cannot conclude that colour card chromaticity results are more accurate than ambient subtraction chromaticity results. On the other hand, we can conclude that they correlate better with TSB.

The two methods agree on important features of the chromaticity distribution, including the relative strength of the x, y, and z correlations with TSB. Given that the aim of this work is to produce a accessory-free technique, the ambient subtraction method is preferred.

Based on these considerations, the focus of the subsequent analysis will be on ambient-subtracted sclera chromaticity data.

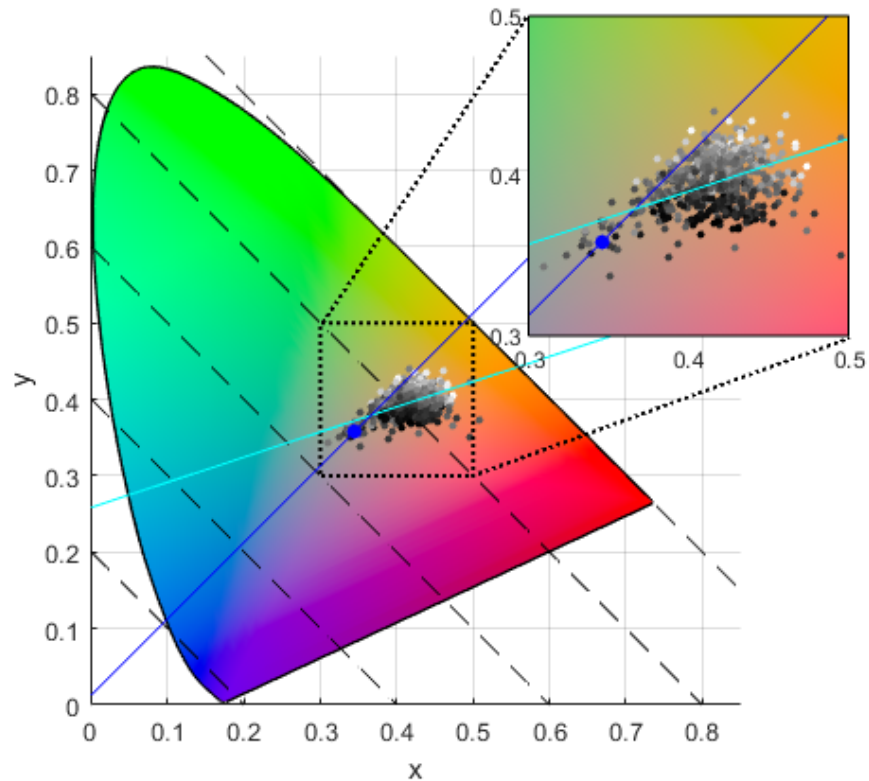
## 4.4 Modelling the Relationship between Chromaticity and TSB

Once device- and ambient-independent colour descriptors have been measured, the next step is to relate them to jaundice severity. The goal is to estimate the TSB as accurately as possible via the tissue colour measurement. In this work, the ambient-subtracted sclera chromaticity is the focus. The TSB estimate derived from sclera colour will be referred to as the scleral-conjunctival bilirubin or SCB. This concept is introduced in analogy with the TSB estimate derived from a transcutaneous bilirubinometer, which is the transcutaneous bilirubin or TcB. As noted in Section 3.3.4, bilirubin accumulates more in the conjunctiva than the sclera itself [95]. Scleral-conjunctival bilirubin is used here as a catch-all term.

The aim of this section is to create a prediction model for TSB based on chromaticity values. Simple linear models are considered first, followed by multiple linear models. Various candidate models are compared using cross validation to quantify the likely out-of sample prediction error. The DJAN study dataset is used for training and validating the model, while the UCLH Smartphone dataset is used as an independent dataset to test the resulting prediction model.

### 4.4.1 Simple Linear Regression

By considering simple linear models it is possible to compare the explanatory power of different features separately, which can help guide the development of more complicated models. One discussed in Chapter 3 is the Jaundice Eye Colour Index (JECI), which is proportional to z chromaticity. We also consider regression against x and y chromaticity. The final simple linear model is based on a total least squares (TLS) fit to the data in xy space. This is equivalent to the axis in xy space that ex-



**Figure 4.5:** Sclera xy chromaticity for 595 neonates measured by the app with ambient subtraction. The marker colour corresponds to TSB, with higher TSB corresponding to lighter markers. The total least squares and JECI axes are included.

plains the most variation in the chromaticity, or the first principle component. The predictor variable in this model corresponds to the position of the datapoint when projected onto this axis.

Figure 4.5 shows the xy chromaticity of 595 neonates from the DJAN study using the ambient subtraction method. The marker colour corresponds to the bilirubin level, with black representing the minimum value ( $10\mu\text{mol/L}$ ), and white representing the maximum value ( $472\mu\text{mol/L}$ ). The cyan line shows the axis of maximal variation found from a total least squares fit. The blue line shows the JECI axis, with the zero point for JECI, which is the D50 white point, shown as a blue circle. The dashed black lines indicate lines of equal z chromaticity, and thus equal JECI.

For each variable, a simple linear model regression model is trained against TSB. Table 4.4 reports the correlation with TSB and the mean absolute error using 10-fold cross validation. The best performing model is based on y chromaticity, while the worst is based on x chromaticity.

**Table 4.4:** Performance of simple linear models for TSB prediction.

| Model Predictor Variable | Correlation with TSB | Cross-Validated Mean Absolute Error ( $\mu\text{mol/L}$ ) |
|--------------------------|----------------------|---|
| x                        | 0.22 (p<0.01)        | 71  |
| y                        | <b>0.67</b> (p<0.01) | <b>55</b>   |
| z or JECI                | 0.47 (p<0.01)        | 65  |
| TLS Score, f(x,y)        | 0.33 (p<0.01)        | 69  |

It is noteworthy that the principle axis of chromaticity variation does not correlate well with TSB, as it is expected that bilirubin is the main source of variation in newborns. It is possible that performing a TLS analysis in a more perceptually uniform chromaticity space such as u'v' may change this result.

#### 4.4.2 Effect of Phototherapy

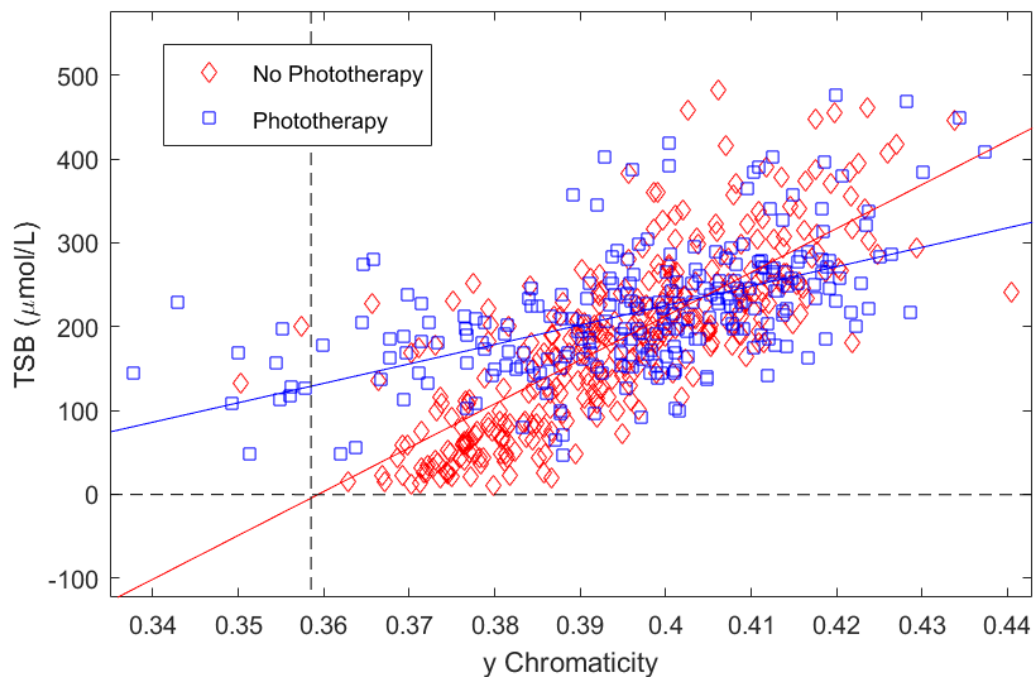
The relationship between sclera chromaticity and TSB may be confounded by the phototherapy status of the neonate. In the DJAN study infants undergoing phototherapy were imaged alongside infants not undergoing phototherapy. Of the 595 neonates imaged with valid TSB readings and suitable image pairs, 248 had under-

gone phototherapy at some point, and 347 had not.

TcB readings are not recommended for babies exposed to phototherapy [36]. Several studies have noted that the exposure of the skin to phototherapy causes a subsequent underestimation of TSB by TcB examination due to a bleaching effect of the light on the skin [113] [114]. Hulzebos et al. found that the TcB estimate of bilirubin decreases faster than the TSB in babies undergoing phototherapy and that the discrepancy persists for a while after the end of treatment [25]. They also note a large variation in the magnitude of this underestimation that is independent of age and ethnicity.

For this reason, several imaging-based studies have excluded babies undergoing phototherapy [77] [78]. In this study, the sclera is the tissue of interest rather than the skin. It is plausible that the bleaching effect does not effect the sclera in the same way as the skin, not least because the eyes are protected during phototherapy.

Figure 4.6 shows the y chromaticity regressed against TSB for both photother-



**Figure 4.6:** y chromaticity against TSB for phototherapy (n=248) and non-phototherapy (n=347) treated neonates. The vertical dashed line is the y chromaticity of the D50 white point.

apy (blue squares) and non-phototherapy (red diamonds) neonates. The overall correlation is 0.67, as quoted in Table 4.4. The phototherapy-exposed babies tend to have a wider variation in y chromaticity for a given TSB value. The sclera colour saturation - a measure of chromaticity distance from the white point - of phototherapy babies is significantly lower than non-phototherapy babies in this dataset (one-sided, unpaired t test,  $p < 0.01$ ), even though there are relatively few low-TSB phototherapy-exposed babies. This is consistent with the observation of Hulzebos et al. that phototherapy causes TSB underestimation and an increase in the estimation variance [25]. Although this is not conclusive, it would appear a bleaching-like effect, whereby the superficial tissues decrease in bilirubin quicker than systemic bilirubin, is seen in the sclera as well as the skin.

To determine if the relationship between chromaticity and TSB is significantly different for babies undergoing phototherapy, the difference between the regression slopes of the two groups is tested. For the y chromaticity, z chromaticity, and TLS score, the interaction with a dummy variable representing the phototherapy status was significant via an F test at  $p < 0.01$ . This implies a statistically different relationship with these colour metrics and TSB concentration. Phototherapy status was not found to influence the regression line between x chromaticity and TSB at the 0.01 significance level. Table 4.5 summarises these findings and shows the correlation between the individual variables and TSB for phototherapy (n=248) and non-phototherapy (n=347) groups. Based on these analyses the remainder of this chapter will focus on non-phototherapy babies.

**Table 4.5:** Comparing single explanatory variable correlation with TSB for phototherapy and non-phototherapy groups.

| Model Predictor Variable | Correlation vs TSB: Phototherapy Babies | Correlation vs TSB: Non-Phototherapy Babies | Significant Effect of Phototherapy on Fit? |
|--------------------------|---|---|--|
| x                        | 0.43 ( $p < 0.01$ )                     | 0.20 ( $p < 0.01$ )                         | No   |
| y                        | <b>0.54</b> ( $p < 0.01$ )              | <b>0.78</b> ( $p < 0.01$ )                  | Yes  |
| z or JECI                | 0.52 ( $p < 0.01$ )                     | 0.58 ( $p < 0.01$ )                         | Yes  |
| TLS Score, $f(x,y)$      | 0.47 ( $p < 0.01$ )                     | 0.37 ( $p < 0.01$ )                         | Yes  |

### 4.4.3 Multiple Linear Regression

Chromaticity is a two-dimensional feature, so using only one predictor variable necessarily involves a loss of information. Including more terms in the prediction model may improve prediction accuracy. Table 4.6 shows the correlation with TSB and cross-validated mean absolute error (MAE) for various multiple linear regression models for SCB with increasing numbers of terms. The correlation rises as more terms are included, but more complex models are penalised by the cross validation procedure.  $SCB(x,y,x^2)$  has the equal-best MAE with fewest terms. However, the  $SCB(y)$  model is almost as predictive and only uses one variable. Over the 10-fold cross validation the  $MAE \pm SE$  is  $51 \pm 2$  for  $SCB(y)$  and  $47 \pm 2$  for  $SCB(x,y,x^2)$ , and the  $RMSE \pm SE$  is  $64 \pm 3$  for  $SCB(y)$  and  $59 \pm 3$  for  $SCB(x,y,x^2)$ . Given the small difference in these figures of merit, the more parsimonious model  $SCB(y)$  is preferred.

**Table 4.6:** Performance of multiple linear models for TSB prediction.

| Number of Features | SCB Model             | Correlation with TSB       | Cross-Validated Mean Absolute Error ( $\mu\text{mol/L}$ ) |
|--------------------|-----------------------|----------------------------|---|
| 1                  | $SCB(y)$              | 0.78 ( $p < 0.01$ )        | 51  |
| 2                  | $SCB(x,y)$            | 0.78 ( $p < 0.01$ )        | 49  |
| 3                  | $SCB(x,y,xy)$         | 0.79 ( $p < 0.01$ )        | 48  |
| 3                  | $SCB(x,y,x^2)$        | <b>0.81</b> ( $p < 0.01$ ) | <b>47</b>   |
| 3                  | $SCB(x,y,y^2)$        | 0.79 ( $p < 0.01$ )        | 50  |
| 4                  | $SCB(x,y,xy,x^2)$     | <b>0.81</b> ( $p < 0.01$ ) | <b>47</b>   |
| 4                  | $SCB(x,y,xy,y^2)$     | 0.79 ( $p < 0.01$ )        | 49  |
| 4                  | $SCB(x,y,x^2,y^2)$    | <b>0.81</b> ( $p < 0.01$ ) | <b>47</b>   |
| 5                  | $SCB(x,y,xy,x^2,y^2)$ | <b>0.82</b> ( $p < 0.01$ ) | <b>47</b>   |

Henceforth, unless otherwise mentioned, SCB will refer to the model derived from a regression between ambient-subtracted y chromaticity and TSB for the 347 non-phototherapy babies. This prediction model is given in Equation 4.8.

$$SCB = -1882 + 5236y \quad (4.8)$$

In the remainder of this chapter this model will be evaluated.



## 4.5 SCB Model Evaluation and Validation

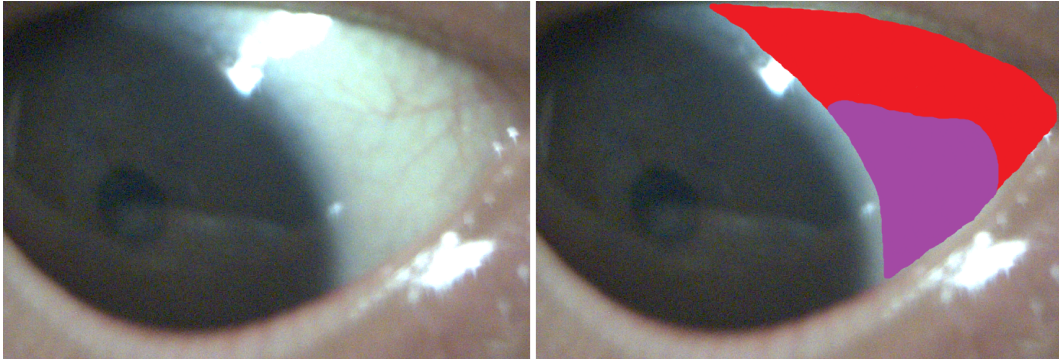
### 4.5.1 Effect of Segmentation

The method for sclera segmentation could greatly effect the chromaticity estimation and thus the TSB estimation. In this work, the sclera was segmented manually. It is interesting to investigate how much variation might result if the approach to segmentation was changed. For example, for a given sclera image, one user may exclude and include different parts of the image from another, and an automatic segmentation algorithm would be different again.

To investigate variation due to segmentation strategy, we compare two manual approaches. In the first approach, the entire sclera is included, regardless of unrepresentative regions such as eyelashes, blood vessels, or specular regions. In the second approach, the user is more discerning, and attempts to draw a contiguous segmentation that does not include any of these unrepresentative regions. These shall be referred to the full and judicious segmentation methods, respectively. In the analyses thus far, the judicious segmentation method was used. In the following, the chromaticity and TSB estimation results for the judicious and full methods are compared. The aim is to quantify the variability introduced by segmentation and determine if one method is superior in its TSB estimation accuracy.

Figure 4.7 shows an example sclera and a judicious and full segmentation of that sclera. The purple region is the judicious segmentation, which avoids the specular region of the upper left side and the blood vessels on the right side. The full segmentation includes all the area of the judicious segmentation as well as the red area.

Two points related to the manual segmentation should be emphasised. First, the judicious segmentation does not fully mitigate the influence of specular reflection and blood vessels, but rather minimises their effect. Second, both the full and judicious segmentations are subjective. The judicious segmentation requires more judgement from the user, and would therefore likely vary more between users asked to segment the same image. A full segmentation would likely lead to greater intra-user repeatability. However, even the location of the edge between the sclera and



**Figure 4.7:** Illustration of sclera ROI segmentation methods. Left: the original image. Right: The segmented image. In purple is the judicious segmentation. The full segmentation comprises the red and purple regions together.

iris is not completely objective.

In the following comparison we consider image pairs for 141 non-phototherapy subjects for which both a full and judicious segmentation was performed and that resulted in a sufficient SSNR. A single image pair per subject was considered, with no averaging between repeats.

Judicious segmentations comprised on average  $25,000 \pm 16,000$  pixels, while full segmentations comprised on average  $44,000 \pm 28,000$  pixels. Judicious segmentations were on average 61% the size of full segmentations.

y chromaticity for full segmentations was significantly lower than for judicious segmentations (two-sided, paired t test,  $p < 0.01$ ). The 95% confidence interval for the y bias was  $[-0.0017, -0.0032]$ , which leads to a  $[-9 \mu\text{mol/L}, -17 \mu\text{mol/L}]$  offset SCB value for full segmentations as compared to judicious segmentations via Equation 4.8.

For the 141 subjects under consideration, using SCB Equation 4.8, the bias in TSB estimation worsens from  $-24 \mu\text{mol/L}$  to  $-37 \mu\text{mol/L}$  when using full rather than judicious segmentation, and the MAE increases from  $63 \mu\text{mol/L}$  to  $70 \mu\text{mol/L}$ .

These results could indicate that specular areas are causing an underestimation in the saturation of the sclera colour when full segmentation is used. This would lead to a subsequent underestimation of TSB. However, the saturation is not significantly different between the two segmentation approaches at the 0.01 level ( $p = 0.41$ ). Furthermore, comparison of the SCB bias and MAE is not strictly fair

as the SCB equation was derived based on judicious segmentation, not the full segmentation. Instead, it is possible to compare the correlation of the y chromaticity with TSB directly. Full segmentation y chromaticity correlated with TSB with a coefficient of 0.46 compared to judicious segmentation at 0.50. The difference in the proportion of variance explained is not significant. The low correlation coefficient in both cases compared to previously quoted values can be explained by the fact that we are considering a restricted subsample of the dataset ( $n=141$ ), and the TSB range is limited at the low end, with the lowest value recorded being  $101\mu\text{mol/L}$ .

Even though the chromaticity estimates are different, it is not possible to conclude from these data alone that judicious segmentation is significantly better than full segmentation for predicting TSB. Using a less robust summary statistic than the median may increase the difference between the two approaches and indicate one as superior. We expect the median to have mitigated the influence of unrepresentative regions of the sclera (vessels, specular reflections, eyelashes, and blemishes).

The other quantity that it is useful to estimate is the variability introduced by different segmentation strategies at the per-subject level. In the absence of inter- and intra-user repeat segmentation data, comparing full and judicious segmentations may provide something approaching a practical upper bound for the variability introduced by manual segmentation.

The mean absolute difference between SCB as defined in Equation 4.8 using full and judicious segmentations is  $18\mu\text{mol/L}$ . This is approximately one third of the 10-fold cross validated MAE for the SCB model predictions,  $51\mu\text{mol/L}$ , implying that variance due to segmentation could constitute a significant fraction of the prediction error.

Another measure of the spread is the standard deviation of the differences, which is  $23\mu\text{mol/L}$ . Together with a bias of  $-13\mu\text{mol/L}$ , we find that the 95% limits of agreement between the two segmentation methods are  $[-59\mu\text{mol/L}, 33\mu\text{mol/L}]$ .

In this section, the aim has been to determine the effect of segmentation on TSB estimation accuracy and repeatability. A judicious segmentation did not provide significantly higher predictive accuracy for TSB than a full segmentation of

the sclera, although the proportion of variance explained was found to be slightly higher. Full sclera segmentation may vary less between users and is easier to describe to untrained individuals. Further investigations are recommended into automatic segmentation algorithms and pixel filtering after segmentation.

#### 4.5.2 Effect of Repeats

In the previous section, the SCB variability due to segmentation was analysed by comparing two different approaches. In this section, the variation between repeat captures of the same subject within the same capture session is investigated. The effect on the SCB accuracy of averaging repeats over several captures is quantified.

For this investigation, the 185 capture sessions for which there were three usable image pairs of the subject were included (non-phototherapy subjects, with sufficient SSNR). Judicious segmentations were applied.

To estimate the accuracy in the case of a single image pair, one pair is chosen at random from the three pairs for each of the 185 subjects using MATLAB's *datasample* function. The MAE and correlation are calculated using these pairs and the SCB equation. This random sample was performed 15 times. The mean MAE was found to be  $52\mu\text{mol/L}$ , and the mean correlation coefficient was 0.77. When instead the chromaticity estimates from all three image pairs were averaged, the resulting SCB estimate gave a MAE of  $46\mu\text{mol/L}$  and a correlation coefficient of 0.82. This suggests that averaging repeated measures can improve prediction accuracy.

The within-subject standard deviation of repeat measurements is  $36\mu\text{mol/L}$ . The variance of the difference between two consecutive measurements is double the within-subject variance. From this we can estimate the maximum (worst-case) likely difference to be  $99\mu\text{mol/L}$ , and 95% of repeats will have a significantly smaller difference: The mean absolute deviation from the average value is  $19\mu\text{mol/L}$ .

These data indicate that there is considerable variation in the SCB estimate even for repeated measures of the same subject on the same occasion. Averaging repeats is therefore highly valuable as a means of reducing this variance, and has

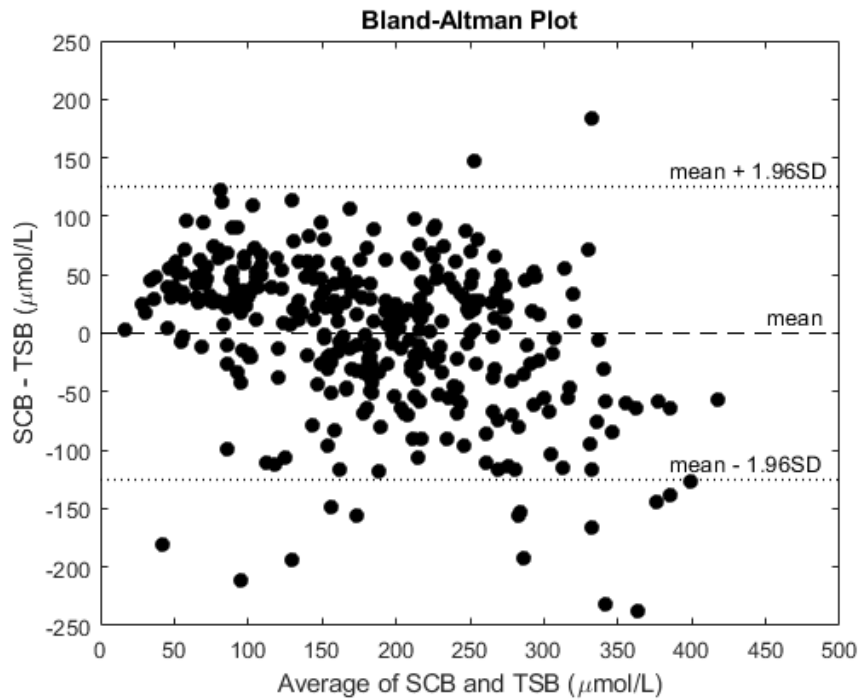
been shown to improve the estimate correlation with TSB. The only disadvantage of repeated measurements is the longer capture time. Capture time data are investigated in Chapter 5, where the practical usability of the app is considered.

### 4.5.3 Performance Consistency

Having analysed the repeat variability, effect of segmentation, and effect of phototherapy, the consistency of the technique with respect to various clinical properties must also be evaluated. In the following analysis ( $n=347$ ), non-phototherapy babies with sufficient SSNR are included. Judicious segmentations are used and where repeats are available, an average in chromaticity space is performed. The consistency of the performance with respect to TSB level, postnatal age, gender, birth weight, and gestational age is tested.

It is desirable to have a consistent SCB accuracy for babies of both high and low TSB level. A Bland-Altman plot is shown in Figure 4.8 for 347 subjects. The limits of agreement are  $[-125\mu\text{mol/L}, 125\mu\text{mol/L}]$ . Clearly, this is a clinically significant potential for disagreement, and so the technique can not replace serum bilirubin measurement as a diagnostic tool. There is a tendency for underestimation at high TSB, above approximately  $300\mu\text{mol/L}$ . This may be a result of a non-linear relationship between TSB and bilirubin pigment in the superficial tissues. After a certain point, there may be a saturation effect or at least a significant diminishing of the rate of change of pigmentation. Some evidence for this exists in the literature. TcBs have been found to progressively underestimate TSB at higher levels [115] [116] [117]. The Draeger JM-105 does not quote a TcB value above  $340\mu\text{mol/L}$  [118], presumably because estimation accuracy decreases. Sufian et al. fit a separate linear model between chromaticity and TSB above approximately  $220\mu\text{mol/L}$  [85]. A similar trend is visible in the Bland Altman analysis of Bili-Cam [77] [76]. Dong et al. showed that the BiliScan app is unreliable above a TSB  $18\text{mg/dL}$  ( $308\mu\text{mol/L}$ ), and that the result was never quoted above that value [80]. Although these studies relate to skin-based estimation of bilirubin, a similar effect is plausible in the sclera.

There is no significant correlation between the SCB error, that is SCB-TSB,



**Figure 4.8:** Bland-Altman plot for 347 DJAN subjects.

and postnatal age for the 344 subjects with a postnatal age recorded ( $p=0.07$ ).

There is no significant difference in the mean SCB error between male ( $n=192$ ) and female ( $n=155$ ) groups (unpaired two-sample  $t$  test,  $p=0.31$ ).

There is no significant correlation between the SCB error and birth weight for the 347 subjects with a birth weight recorded ( $p=0.11$ ).

For the 345 subjects with a gestational age recorded, there is a significant correlation with the SCB error ( $p<0.01$ ). The correlation is 0.2, with a 95% confidence interval of [0.1,0.3]. This implies that SCB is providing an underestimate for babies with shorter gestational ages. The SCB error for premature babies in our sample (less than 37 weeks gestational age,  $n=46$ ) is significantly biased compared to the term babies ( $n=299$ ), with a mean relative offset of  $-44\mu\text{mol/L}$  (95% confidence interval  $[-64\mu\text{mol/L}, -25\mu\text{mol/L}]$ ). One possible explanation for this relationship relates to the thickness of the sclera. It is often observed that premature babies have a blue tint to their sclera. This is because the immature sclera is relatively thin, allowing the choroidal pigment underneath to affect the colour. In the estimation of jaundice, a more blue sclera will counteract the yellowing due to bilirubin and lead

to an underestimate. A factor to correct for gestational age, where gestational age is known, may improve estimation accuracy. Alternatively, guidance to avoid using the app on premature babies could be provided.

In the remainder of this thesis, premature babies of less than 35 weeks gestational age are excluded from the analysis. Near-term babies between 35 and 37 weeks remain in the analysis along with term babies. This is partly due to the observed underestimation in premature babies described above, and partly to facilitate fair comparison with results from TcBs: NICE guidelines recommend against the use of TcBs in premature babies with a gestational age of less than 35 weeks [36]. This means the set of subjects with sufficient SSNR and no history of phototherapy reduces in size from 347 to 336. The subset of those subjects with three good repeats reduces in size from 185 to 179.

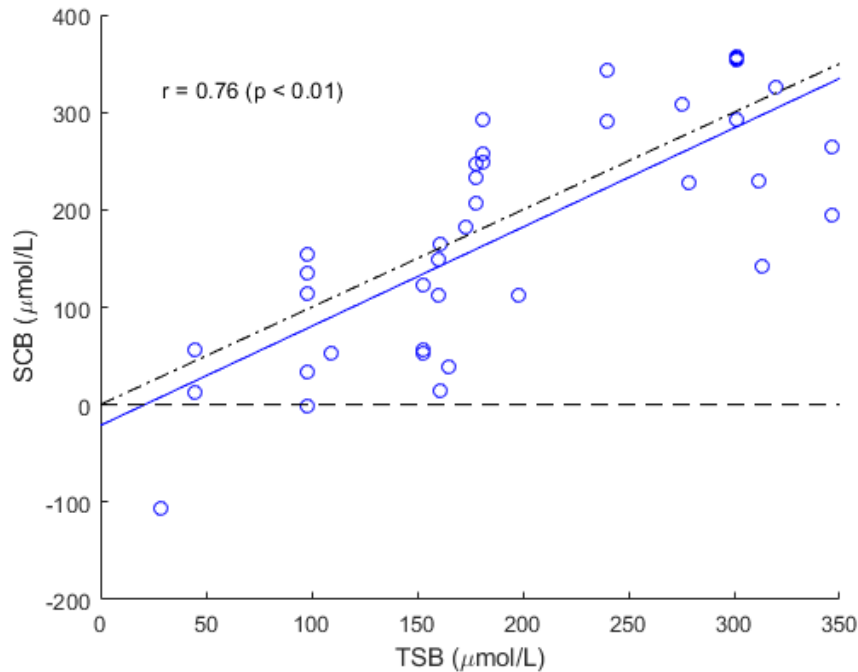
#### 4.5.4 Validation on UCLH Data

The development of the SCB model has relied on data from the DJAN study in Ghana. Specifically, it was trained on the 347 babies without phototherapy treated for whom there was sufficient SSNR in at least one image pair. 10-fold cross validation was used to simulate the out-of-sample performance. The cross-validated MAE was  $51\mu\text{mol/L}$ , the limits of agreement were  $[-125\mu\text{mol/L}, 125\mu\text{mol/L}]$ , and the correlation with TSB was 0.78 (95% confidence interval:  $[0.73, 0.81]$ ). The model will now be evaluated on data collected at UCLH to assess how well it generalises to data that has been collected in a different context and that was not involved in the model training process.

From 41 eligible subjects, 38 image sets had sufficient SSNR for inclusion. The image pairs from the UCLH study are subject to the same analysis as the DJAN data, and the SCB model is not altered. The resulting correlation plot is shown in Figure 4.9. A correlation of 0.76 is observed (95% confidence interval:  $[0.58, 0.89]$ ). This is in agreement with the training correlation performance. The MAE is  $65\mu\text{mol/L}$ .

A previously published analysis of the UCLH dataset by this author achieved a correlation of 0.75 ( $n=37$ ) [2]. That analysis was also based on ambient-subtracted

sclera chromaticity but did not employ the SSNR threshold method and used a different model to relate chromaticity to TSB.

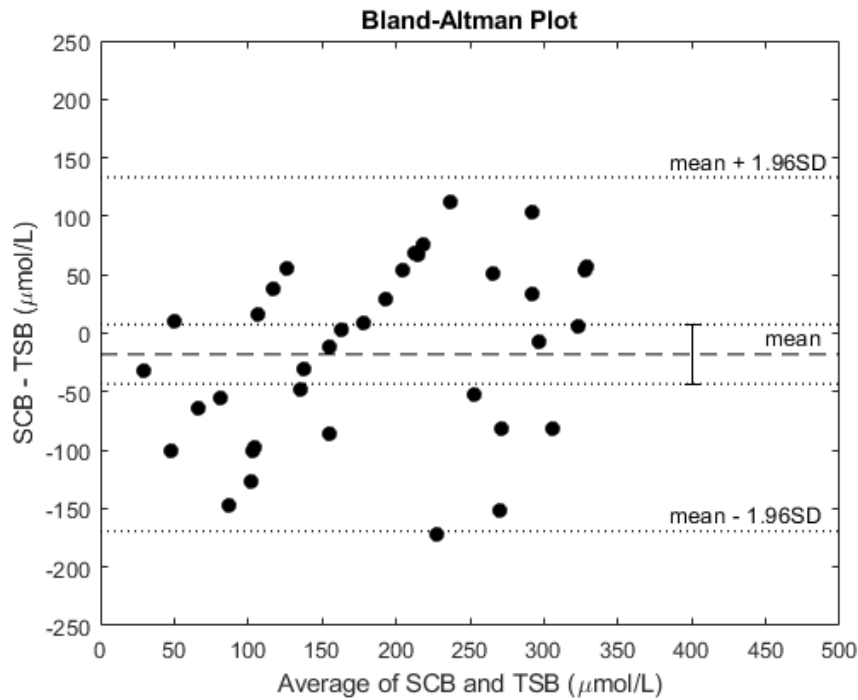


**Figure 4.9:** SCB correlation with TSB for UCLH data (n=38). The identity line (dot-dashed) and the line of best fit (blue) are included for reference.

Figure 4.10 shows the Bland-Altman plot for the UCLH SCB data. The limits of agreement ( $[-169\mu\text{mol/L}, 133\mu\text{mol/L}]$ ) and the bias with 95% confidence interval are indicated. The results are consistent with zero bias in the model for this test dataset.

The UCLH dataset was captured in a completely different context to the DJAN dataset. Most notably, a different modality of capture was employed. For the UCLH dataset the ambient subtraction was achieved with a screen-as-illumination modality with front-facing camera. For the DJAN dataset the back-facing camera and LED flash was used with a diffuser. Furthermore, different smartphones were used (LG Nexus 5X and Samsung S8). The two datasets were collected in different countries and the majority ethnicity of the two subject groups differed. The SCB model based on ambient-subtracted  $y$  chromaticity after SSNR threshold performed equally well on the UCLH dataset. This validates the performance consistency of the app.





**Figure 4.10:** Bland-Altman agreement plot for SCB and TSB for UCLH data (n=38). The 95% confidence interval for the bias contains zero.

### 4.5.5 Comparison with Literature

Table 4.7 shows a comparison between the neoSCB app performance and the performance of three recent smartphone app studies. All apps apart from the neoSCB app involve the use of a colour card and measured the skin colour, however, they are the most well-documented and validated smartphone methods in the literature and so are a suitable class with which to compare.

The study by Aune et al. included only term neonates. Those by Taylor et al. and Ren et al. included both term and near-term neonates. None of the studies involved neonates who had undergone phototherapy. These criteria align with the decisions made in the analysis of the DJAN and UCLH datasets. One point of divergence between the datasets is that the majority ethnicities were different in each case. The study by Taylor et al. involved a multi-ethnic sample, the UCLH study and the study by Aune et al. were majority Caucasian, the study by Ren et al. was majority Chinese, and the DJAN study was majority Black African. It is essential that a screening app is validated on a range of ethnicities to ensure

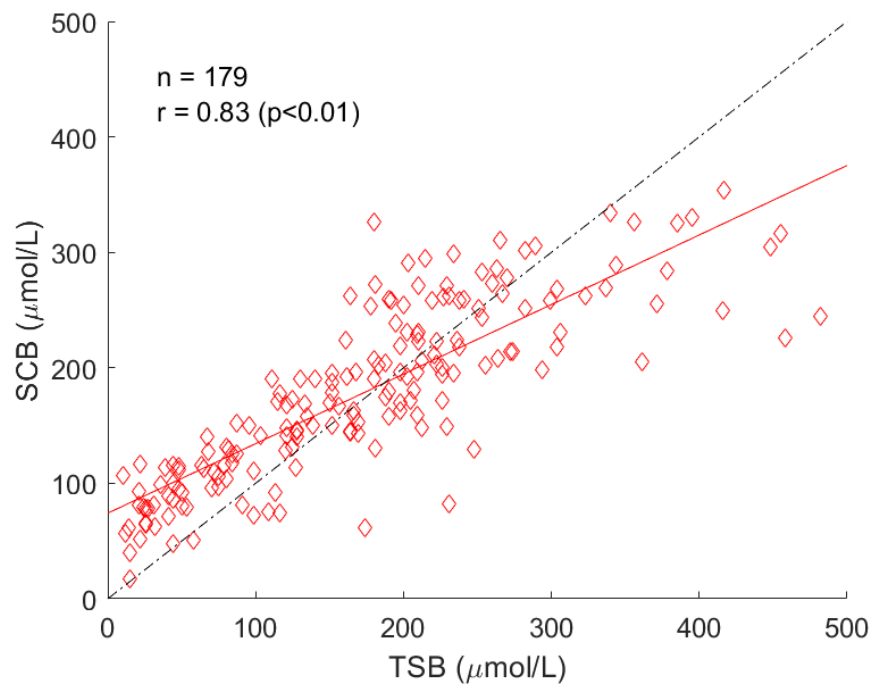
consistent performance. BiliCam showed consistent performance across all ethnic groups [77]. On the other hand, Picterus was found to have significantly higher correlation with TSB among Caucasian subjects than non-Caucasian subjects in the study by Aune et al. [78], and the learning-based BiliScan was found to have a lower correlation in a non-Chinese population by Swarna et al. [83]. Although neoSCB is not a skin-based app and uses a parsimonious model, it is important to demonstrate consistent performance between populations. The model trained on the majority Black African (DJAN) dataset performs comparably well on the majority Caucasian (UCLH) dataset, but wider validation is still needed.

**Table 4.7:** Table comparing quoted TSB prediction performance for various jaundice detection apps alongside SCB app performance.

| Study: App                         | Pearson's r (95% Confidence Interval) | n                        | Limits of Agreement ( $\mu\text{mol/L}$ ) |
|------------------------------------|---------------------------------------|--------------------------|---|
| Taylor et al. (2017): BiliCam [77] | 0.91 (0.89 – 0.92)                    | 530                      | [-62, 62]                                 |
| Aune et al. (2020): Picterus [78]  | 0.84 (0.79 – 0.88)                    | 185                      | [-84, 84]                                 |
| Ren et al. (2020): BiliScan [82]   | 0.78 (0.92/ 0.73 day/ night)          | 247 (183/ 64 day/ night) | [-81, 98]                                 |
| DJAN study: neoSCB                 | 0.78 (0.74 – 0.82)                    | 336                      | [-122, 126]                               |
| DJAN study: 3-repeat neoSCB        | 0.83 (0.78 – 0.87)                    | 179                      | [-110, 121]                               |
| UCLH study: neoSCB                 | 0.76 (0.58 – 0.89)                    | 38                       | [-169, 133]                               |

The neoSCB app correlation is comparable to that of BiliScan as evaluated by Ren et al. [82], and marginally lower than the result quoted by Aune et al. [78]. The best performance is reported by Taylor et al. (BiliCam), both in terms of correlation with TSB and limits of agreement with TSB measurement [77].

The methods of Aune et al. and BiliCam involve several image captures. If we consider only DJAN data for which an average chromaticity from three image pairs is available, the SCB correlation with TSB is 0.83 (95% confidence interval: [0.78, 0.87]), with limits of agreement of [-110 $\mu\text{mol/L}$ , 121 $\mu\text{mol/L}$ ] (n=179). Figure 4.11 shows the scatter plot of the correlation between TSB and three-repeat SCB.



**Figure 4.11:** SCB correlation with TSB for three-repeat DJAN data.  $n=179$ . The identity line (dot-dashed) and the line of best fit (red) are included for reference.

## 4.6 Summary

In this chapter, a simple model relating TSB to  $y$  chromaticity was developed and validated on unseen data. Ambient subtraction following a one-time calibration was shown to recover chromaticity better than some common white balance algorithms. The between-repeat variability was shown to be high, suggesting that multiple repeat measurements would improve performance accuracy. When averaging across three repeats, a correlation between SCB and TSB of 0.83 was achieved. This is competitive with other studies that have estimated TSB from RGB smartphone images.

All jaundice detection app studies in Table 4.7 have concluded that smartphone imaging is not sufficiently accurate to provide a measurement that could replace a serum bilirubin measurement, and neoSCB is no different. As a screening method, smartphones do, however, show promise. In the next chapter, the neoSCB app is evaluated as a screening tool and compared to existing screening methods: the TcB and visual inspection.

## **Chapter 5**

# **Screening Performance of the neoSCB App**

### **5.1 Overview**

In Chapter 4, the neoSCB app estimation accuracy was quantified. The limits of agreement between SCB with TSB were too wide for them to be used interchangeably. In this chapter, the potential of the neoSCB app as a screening tool is analysed.

The chapter begins with a discussion of the requirements of a screening tool and quantifies the performance using two different approaches to classification: thresholding SCB and discriminant analysis in chromaticity space. The neoSCB screening performance on DJAN and UCLH datasets is compared to that of other recently published smartphone camera-based jaundice screening tools.

The neoSCB app is then compared to TcB and visual inspection using data collected as part of the DJAN study. The effect of phototherapy on TcB performance and on app performance is shown. Visual inspection responses are analysed and some of the comparative advantages of smartphone screening are expounded.

Finally, the real-world applicability of the neoSCB screening app is discussed. A model for the effective deployment of the app in existing healthcare pathways is suggested. The practicality of three-repeat image capture protocol is considered with an analysis of the total time taken. Data from real-time results are used to compare real-time ROI selection to offline sclera segmentation.

## 5.2 Screening Neonates at Risk

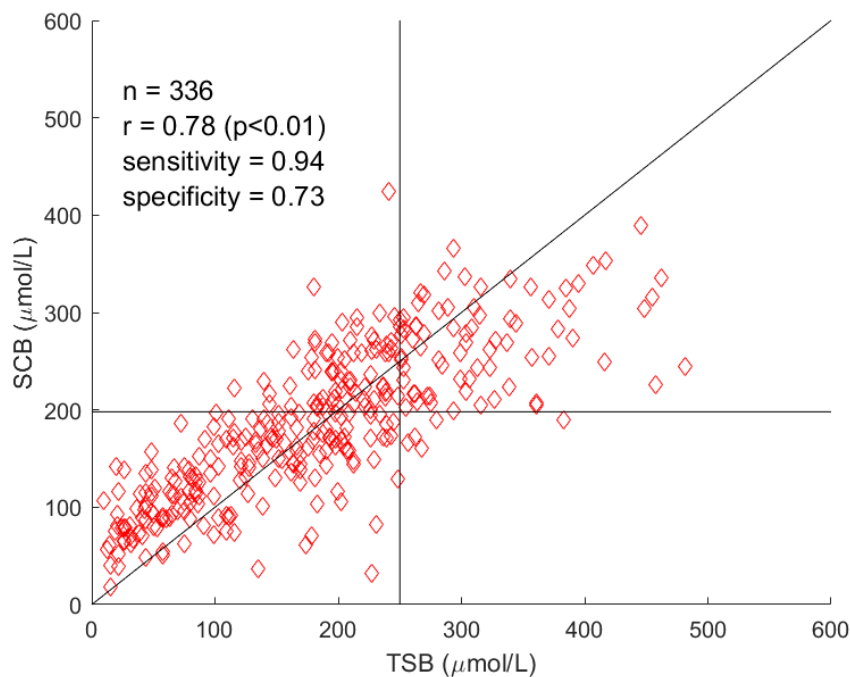
### 5.2.1 Screening and Decision Thresholds

The aim of screening for neonatal jaundice is to identify at-risk neonates and refer them in a timely manner for further diagnosis and treatment. The two most widely used approaches for screening are visual inspection and TcB. Visual inspection is not reliable as the sole means of screening for jaundice [36]. For a quantitative screening tool such as the TcB, an appropriate numerical decision threshold must be applied. The TSB level at which a neonate becomes at risk of kernicterus depends on several factors. Prematurity is one factor that increases the vulnerability of the newborn to the negative consequences of hyperbilirubinaemia. The postnatal age is also relevant in interpreting the risk associated with a TSB reading, as a high value in the first one or two days of life can be indicative of a pathological form of jaundice or a rapidly climbing TSB level. Given these and other factors, there is no one threshold of interest that can be used to screen newborns at risk. NICE guidelines recommend against using the TcB for neonates less than one day old or with a gestational age of less than 35 weeks [36]. For term and near-term babies greater than one day old, the decision threshold adopted is  $250\mu\text{mol/L}$ . This is equivalent to the phototherapy treatment threshold for a neonate of 35 weeks gestational age.

In the following analysis, the screening threshold used is  $250\mu\text{mol/L}$ . As well as being clinically relevant for the above reasons, it facilitates easy comparison between TcB screening studies and other jaundice detection studies, which often quote screening performance for this threshold. This *screening* threshold must be distinguished from the *decision* threshold. The screening threshold is the TSB level above which we would like to refer neonates. The decision threshold is the numerical output of the TcB, app, or other device above which we refer a neonate. Sometimes the decision threshold may be lower than the screening threshold to account for the measurement error of the screening device and reduce the false negative rate.

Figure 5.1 shows the correlation between SCB and TSB for the DJAN results ( $n=336$ ). The screening threshold is marked with a vertical line at  $250\mu\text{mol/L}$ . The decision threshold is shown as a horizontal line at  $198\mu\text{mol/L}$ . Data points in the

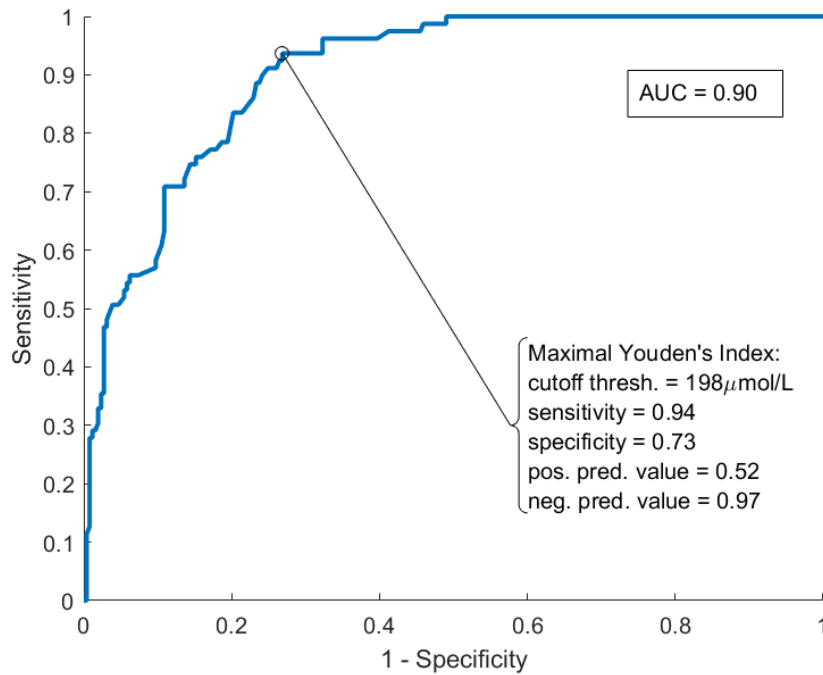
upper left and lower right quadrants formed by these lines are misclassified, the others are correctly classified. The sensitivity is 0.94, which means of the 79 babies with TSB greater than or equal to  $250\mu\text{mol/L}$ , 94% (74) are correctly identified as such. The specificity is 0.73: approximately one in four babies with TSB less than  $250\mu\text{mol/L}$  would be referred by the neoSCB unnecessarily. The decision threshold can be varied to preferentially increase either the sensitivity or specificity at the expense of the other.



**Figure 5.1:** Correlation plot for SCB and TSB for 336 DJAN subjects. Screening and decision thresholds are indicated with vertical and horizontal lines, respectively.

### 5.2.2 ROC Curves

A receiver operating characteristic (ROC) curve is used to graph the screening performance of a technique in terms of sensitivity and specificity for the entire range of possible decision thresholds. For a decision threshold at the highest end of the range, no subjects test positive. The specificity is one, as all babies not at risk are correctly identified as such, but the sensitivity is zero, as all at-risk babies are missed. At the other end of the range, the sensitivity is one but the specificity is zero.



**Figure 5.2:** ROC curve for neoSCB from DJAN data (n=336). Brace annotation from Ref. [119].

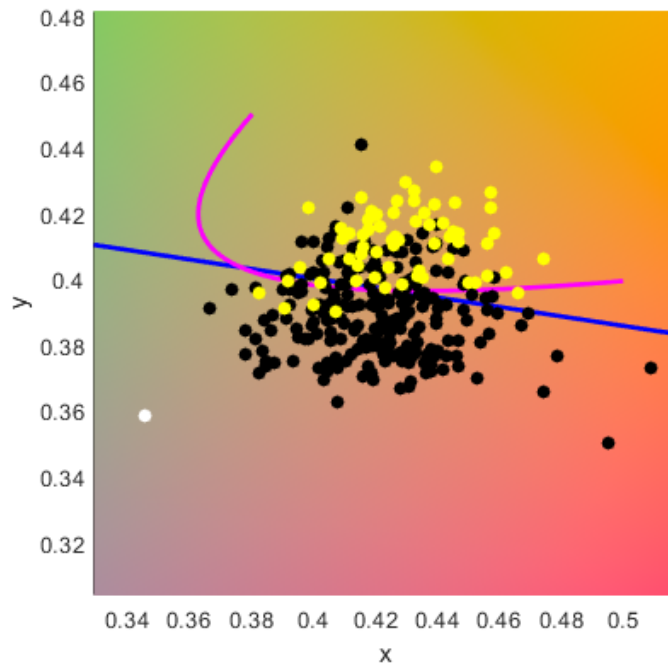
Figure 5.2 shows the ROC curve generated from the same data as in Figure 5.1. The point on the curve corresponding to a decision threshold of  $198\mu\text{mol/L}$  is indicated. This point was chosen to maximise Youden's Index, which is a measure of screening performance given by sensitivity + specificity - 1. Positive and negative predictive values express the probability that the test result is correct given the test result outcome. For example, given the result is negative, the probability that the newborn is not at risk is 0.97.

The area under curve (AUC) of the ROC curve is used to summarise screening performance in a decision threshold-agnostic way. A perfect screening technique would have an AUC of one. A device that randomly output a value would have an AUC of 0.5. The AUC measured in Figure 5.2 is 0.90.

### 5.2.3 Discriminant Analysis Approach

Thus far we have considered an approach analogous to the TcB. Like a TcB value, an SCB value is an estimate for TSB, and the screening decision is made by choosing an appropriate threshold for the SCB value. An alternative approach would

be to classify the newborns based on their sclera chromaticity: a one-step process rather than a two-step process. One such way to achieve this is discriminant analysis, which predicts class (less than or more than  $250\mu\text{mol/L}$ ) based on continuous variables (x and y chromaticity).



**Figure 5.3:** Discriminant analysis in  $xy$  chromaticity space. Neonates with TSB less than  $250\mu\text{mol/L}$  are represented in black, at-risk neonates in yellow. The linear and quadratic discriminant lines are in blue and magenta, respectively. The D50 white point is indicated in white.

Figure 5.3 shows the DJAN data labelled by their class, with yellow circles representing babies with a TSB value above  $250\mu\text{mol/L}$ . A linear discriminant and quadratic discriminant are fitted. The linear discriminant is a straight line, whereas the quadratic discriminant is a curve. The circles above these lines would be classified as being at risk and test positive. In fitting these models, the cost of misclassification of babies not at risk was down-weighted to 0.17 the cost of misclassifying a baby at risk to match the sensitivity of the SCB approach. The sensitivity for both the linear and quadratic discriminant models was 0.94. The specificities were 0.70 and 0.72 for linear and quadratic discriminant models, respectively. The AUC of the discriminant approach is 0.90. The performance of the discriminant model



approach is almost the same as that achieved using the SCB threshold approach. As there is no improvement in performance, the SCB method is preferred because it can be used in an analogous way to the TcB. It is also easier to compare the neoSCB performance to other screening apps, which typically estimate bilirubin concentration and quote the correlation with TSB as a performance metric.

### 5.2.4 Comparison with Literature

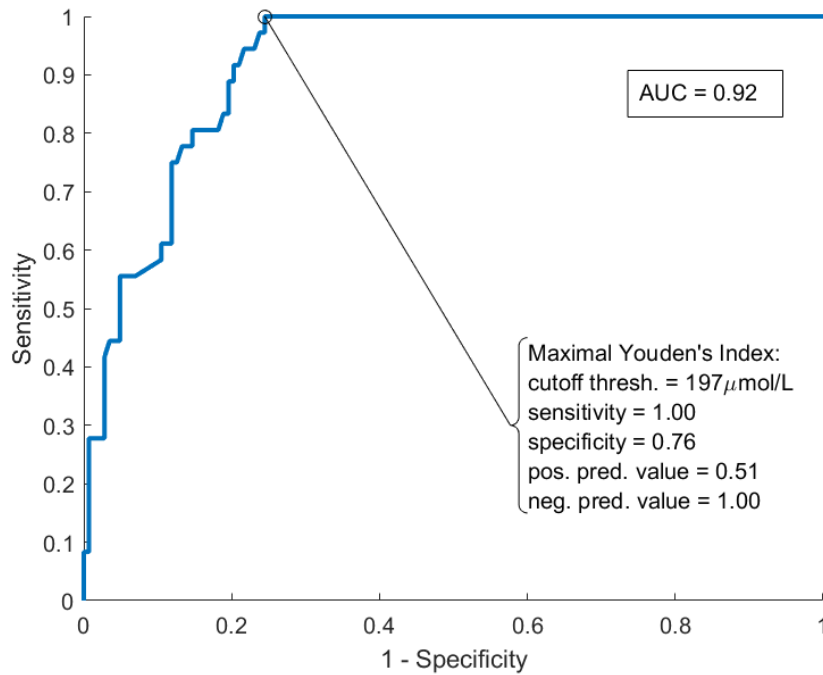
Table 5.1 shows the screening performance of neoSCB for the DJAN (training and validation) and UCLH (test) datasets alongside that of three other jaundice screening apps. The screening performance in the UCLH dataset is consistent with that of the DJAN dataset when the sensitivity and specificity confidence intervals are taken into account.

**Table 5.1:** Table comparing quoted screening performance for various jaundice detection apps alongside SCB app performance.

| Study: App                                    | Sensitivity (95% CI);<br>Specificity (95% CI)    | Screening<br>Threshold             | ROC<br>AUC |
|---|--|------------------------------------|------------|
| Taylor et al. (2017):<br>BiliCam (n=530) [77] | <b>1.00;</b><br>0.76                             | 291 $\mu\text{mol/L}$<br>(17mg/dL) | 0.99       |
| Aune et al. (2020):<br>Picterus (n=185) [78]  | <b>1.00;</b><br>0.69                             | 250 $\mu\text{mol/L}$              | 0.93       |
| Ren et al. (2020):<br>BiliScan (n=247) [82]   | <b>0.75;</b><br>0.87                             | 256 $\mu\text{mol/L}$<br>(15mg/dL) | 0.89       |
| DJAN study: neoSCB<br>(n=336)                 | <b>0.94 (0.91 – 0.97);</b><br>0.73 (0.68 – 0.78) | 250 $\mu\text{mol/L}$              | 0.90       |
| DJAN study: 3-repeat<br>neoSCB (n=179)        | <b>1.00 (0.95 – 1.00);</b><br>0.76 (0.70 – 0.82) | 250 $\mu\text{mol/L}$              | 0.92       |
| UCLH study: neoSCB<br>(n=38)                  | <b>0.89 (0.79 – 0.99);</b><br>0.79 (0.65 – 0.92) | 250 $\mu\text{mol/L}$              | 0.84       |

Figure 5.4 shows the ROC curve for the subset of the DJAN data with three suitable image pairs (n=179). SCB values are calculated via an average of the three chromaticity estimates. The sensitivity and specificity are quoted for a decision threshold that maximises Youden's Index (197  $\mu\text{mol/L}$ ).

The three-repeat neoSCB protocol has the same sensitivity/ specificity as Bili-Cam (1.00/ 0.76). However, it is difficult to directly compare screening performance based on the sensitivity and specificity results as different heuristics for choosing a



**Figure 5.4:** ROC curve for three-repeat neoSCB from DJAN data (n=179).

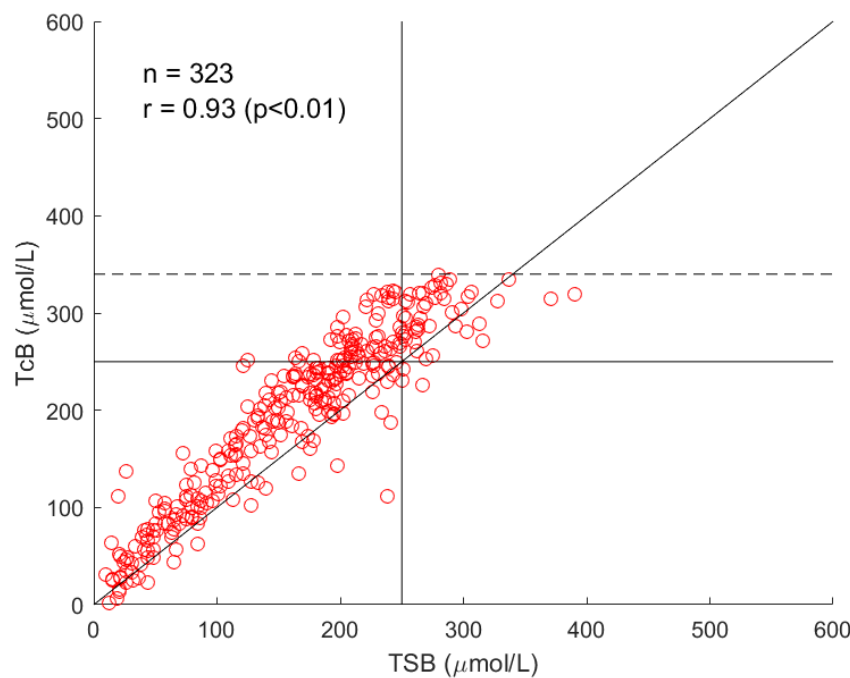
decision threshold are used in each study. Aune et al. report 100% sensitivity with 69% specificity, while Ren et al. report 75% sensitivity with 87% specificity. In the former case, sensitivity is increased at the expense of specificity, while in the latter case the specificity is higher than the sensitivity. The AUC is a more suitable metric for comparison in this case. Based on AUC, BiliCam has the best screening performance, BiliScan and neoSCB have similar performance, and Picterus has marginally better performance than these two. The AUC of three-repeat neoSCB (0.92) is comparable to that of Picterus (0.93). It is important to note that BiliCam uses a higher screening threshold of  $291\mu\text{mol/L}$  ( $17\text{mg/dL}$ ), which further complicates direct comparison. However, the comparison based on AUC screening metrics is consistent with the relative TSB estimation accuracy of these studies laid out in Table 4.7 of Chapter 4.

### 5.3 Comparison to TcB

The screening performance of the neoSCB app must be evaluated in the context of the existing screening methodologies to determine if it is potentially valuable. In

this section, the neoSCB app performance will be compared to TcB performance on the DJAN dataset.

JM-105 (Draeger UK) measurements were taken alongside neoSCB measurements for all DJAN subjects enrolled. Figure 5.5 shows the correlation between the TcB and TSB readings for eligible subjects. Eligible subjects were term or near-term neonates who had not undergone phototherapy and for which a TSB value was successfully recorded. The Pearson's  $r$  was 0.93 for these 323 subjects. The decision threshold is the same as the  $250\mu\text{mol/L}$  screening threshold for standard usage. These thresholds are shown as solid lines in Figure 5.5. The JM-105 displays a warning rather than display values greater than  $340\mu\text{mol/L}$ . This is evident in Figure 5.5, and a dotted line indicates this value.



**Figure 5.5:** Correlation between TSB and JM-105 TcB values. The solid lines represent the screening and decision thresholds. The dotted line at  $340\mu\text{mol/L}$  is the line above which the JM-105 does not display a numerical value.

To compare neoSCB and the JM-105, the subjects for which a TcB warning was displayed are considered as positive tests, as per the usage protocol. Term and near-term subjects with a suitable image pair with sufficient SSNR and no history of phototherapy were included. Table 5.2 shows the confusion matrices for the JM-105

and neoSCB app for screening 336 DJAN neonates for a TSB  $\geq 250\mu\text{mol/L}$ . For the neoSCB app, a decision threshold of  $198\mu\text{mol/L}$  maximised Youden's Index. The neoSCB sensitivity (95% C.I.) is 0.94 (0.91–0.97) and the JM-105 sensitivity is 0.96 (0.94–0.98). The neoSCB specificity (95% C.I.) is 0.73 (0.68–0.78) and the JM-105 specificity is 0.81 (0.76–0.85). This indicates that the neoSCB performance is slightly inferior to that of the JM-105.

**Table 5.2:** neoSCB and JM-105 confusion matrices for predicting neonates with TSB  $\geq 250\mu\text{mol/L}$ .

| neoSCB                    | test+ | test– | total | JM-105                    | test+ | test– | total |
|---------------------------|-------|-------|-------|---------------------------|-------|-------|-------|
| $\geq 250\mu\text{mol/L}$ | 74    | 5     | 79    | $\geq 250\mu\text{mol/L}$ | 76    | 3     | 79    |
| $< 250\mu\text{mol/L}$    | 69    | 188   | 257   | $< 250\mu\text{mol/L}$    | 50    | 207   | 257   |
| total                     | 143   | 193   | 336   | total                     | 126   | 210   | 336   |

If we restrict the dataset to include only those 179 subjects for which three repeat image pairs were available, the performance of the neoSCB app improves relative to the JM-105. The sensitivity of the neoSCB is 1.00 (0.95–1.00) and the JM-105 sensitivity is 0.97 (0.95–1.00). The neoSCB specificity (95% C.I.) is 0.76 (0.70–0.82) and the JM-105 specificity is 0.79 (0.73–0.85). By averaging three repeats, the neoSCB app screening performance is statistically indistinguishable from that of the JM-105 TcB.

Table 5.3 compares the screening performance of the neoSCB app with the screening performance of the JM-105 for the DJAN dataset and three-repeat DJAN dataset. It also includes the equivalent data from Taylor et al. (2017), as this study contained a similar direct comparison between TcB and app sensitivity and speci-

**Table 5.3:** Table comparing TcB and jaundice screening app sensitivity and specificity for this study and that of Taylor et al. [77].

| Study                     | n   | Screening Threshold  | Screening Method | Sensitivity; Specificity |
|---------------------------|-----|----------------------|------------------|--------------------------|
| Taylor et al. (2017) [77] | 312 | $291\mu\text{mol/L}$ | BiliCam          | <b>1.00</b> ; 0.76       |
|                           |     |                      | TcB              | <b>1.00</b> ; 0.75       |
| DJAN study                | 336 | $250\mu\text{mol/L}$ | neoSCB           | <b>0.94</b> ; 0.73       |
|                           |     |                      | TcB              | <b>0.96</b> ; 0.81       |
| DJAN study (3 repeats)    | 179 | $250\mu\text{mol/L}$ | neoSCB           | <b>1.00</b> ; 0.76       |
|                           |     |                      | TcB              | <b>0.97</b> ; 0.79       |

ficity, although using a different screening threshold. Taylor et al. used two models of TcB: Philips BiliCheck and Draeger JM-103 [77].

### 5.3.1 Effect of Phototherapy

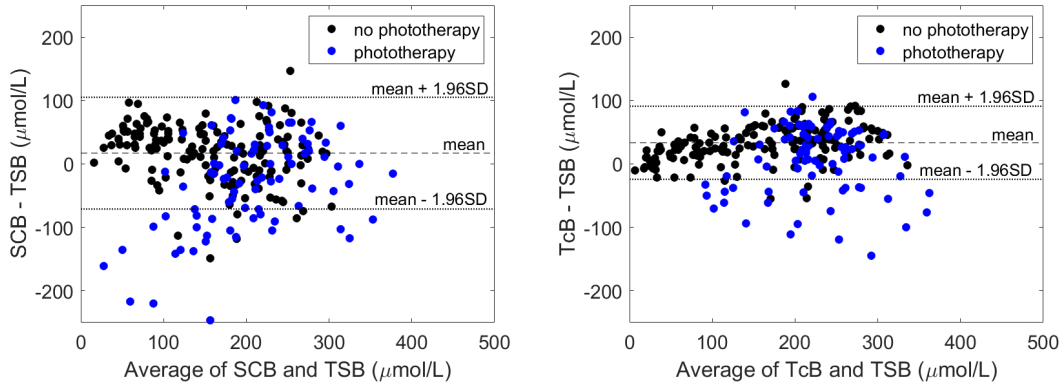
TcB measurement is not advised for babies undergoing phototherapy. However, several studies have investigated this space because of the potential advantages of being able to monitor the effect of phototherapy without repeated blood tests [114] [29]. One of the hypothesised advantages of imaging the sclera rather than the skin was that the bleaching effect would not be observed in the sclera, as the eyes are covered during phototherapy. In Chapter 4, it was shown the phototherapy-exposed babies had significantly different relationship between sclera chromaticity and TSB than non-phototherapy babies, leading to the conclusion that the sclera-based TSB estimation was also affected by phototherapy.

Figure 5.6 shows Bland-Altman plots for both neoSCB and JM-105 DJAN data. Both phototherapy and non-phototherapy babies are included. Subjects without a numerical TcB value or sufficient image pair SSNR were excluded, as were premature babies with a gestational age of less than 35 weeks. Only subjects with three suitable image pair repeats were used. The remaining 245 data points are colour coded according to their phototherapy status.

The bias and limits of agreement in Figure 5.6 are calculated from only non-phototherapy subjects. The majority of all data points falling outside these limits of agreement, which ought to encompass 95% of all differences, are from subjects who have undergone phototherapy. Almost all of these outliers represent underestimations on the part of the screening technique.

The bias among non-phototherapy babies is higher in the JM-105 data ( $33\mu\text{mol/L}$ , 95% C.I.: [ $31\mu\text{mol/L}$ ,  $36\mu\text{mol/L}$ ]) than the neoSCB data ( $17\mu\text{mol/L}$ , 95% C.I.: [ $14\mu\text{mol/L}$ ,  $21\mu\text{mol/L}$ ]). This is consistent with the observations of Taylor et al. and Olunsaya et al. that TcB devices (BiliChek and JM-103 in their cases) overestimate TSB in Black African neonates [117] [34]. However, without similar JM-105 data from neonates with less pigmented skin, it is impossible to say whether this is an overestimation due to ethnicity. The neoSCB data has a wider limits of

agreement than the JM-105 data. The neoSCB limits of agreement are  $-71\mu\text{mol/L}$  to  $105\mu\text{mol/L}$ , while the JM-105 limits of agreement are  $-24\mu\text{mol/L}$  to  $91\mu\text{mol/L}$ .



**Figure 5.6:** Bland-Altman plots for neoSCB and JM-105 data ( $n=245$ ). Neonates exposed to phototherapy are indicated in blue. The horizontal lines representing the mean bias and limits of agreement are calculated for the non-phototherapy neonates only.

In this section, the neoSCB app was found to have comparable screening performance in terms of sensitivity and specificity to the JM-105, provided three repeats were averaged. Phototherapy was shown to have a detrimental effect on both neoSCB app and TcB accuracy, causing potentially dangerous underestimations of bilirubin levels.

## 5.4 Comparison to Visual Inspection

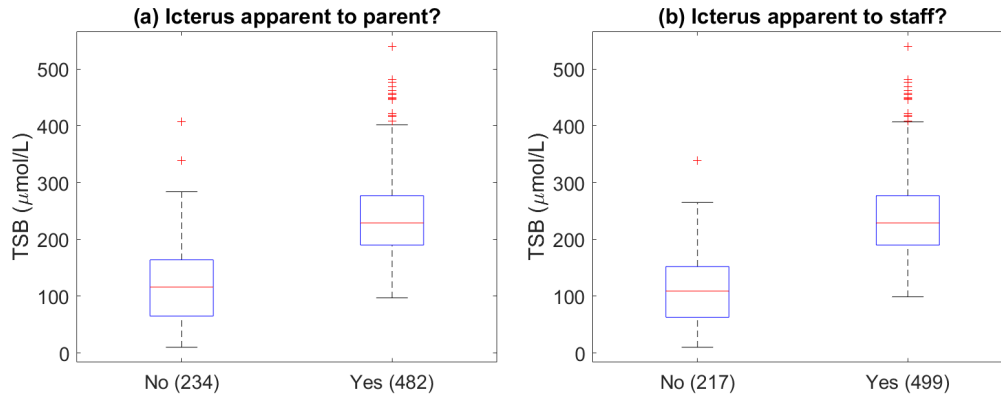
In cases where a TcB is available and jaundice is suspected, visual inspection should not be used as the sole method to screen for jaundice. However, in many cases, a TcB is not available. This may be due to economic constraints on healthcare provision, or simply because the newborn is in the home setting following discharge from hospital. It is in these circumstances that a smartphone-based solution stands to have the greatest impact.

NICE guidelines state that visual inspection should be performed by carers, healthcare professionals and parents, and should involve examining the naked baby in bright light. The examiner may lightly blanch the skin by pressing, and should also check the sclerae and gums [36].

In the DJAN study, the staff member using the app and a parent of the baby were separately asked, “Do the baby’s eyes seem yellow?”. The response was recorded as a binary yes/no. While this question does not capture whether or not the parent would have sought medical advice based on the sclera colour (or, indeed, whether they already had done), it may provide a simple upper bound on the sensitivity of the untrained eye to discolouration due to jaundice. Likewise, asking the healthcare professional conducting the study whether the eyes appear yellow does not capture whether they would have referred the neonate based on their observations. Indeed, healthcare professionals incorporate feeding patterns, behaviour, patient risk factors and history into such decisions in the absence of a TcB. However, by asking the same question to both groups it permits an analysis of any discrepancy between lay and professional examiners.

Figure 5.7 shows two box plots of the range of TSB values for the Yes and No responses to the question, “Do the baby’s eyes seem yellow?” with Figure 5.7 (a) for the parent’s response and Figure 5.7 (b) for the staff member’s response. From the 724 subject capture sessions, 716 had a response recorded from both a parent and a staff member. For the parent’s No responses, the median TSB was  $116\mu\text{mol/L}$ , with a minimum value of  $10\mu\text{mol/L}$  and a maximum of  $407\mu\text{mol/L}$ , while Yes responses had a TSB range of  $97\mu\text{mol/L}$  to  $540\mu\text{mol/L}$  with a median of  $229\mu\text{mol/L}$ . For the staff opinion, the ranges were similar. The No staff response subjects had TSB values between  $10\mu\text{mol/L}$  and  $339\mu\text{mol/L}$  (median  $109\mu\text{mol/L}$ ). The Yes staff response subjects had TSB values between  $99\mu\text{mol/L}$  and  $540\mu\text{mol/L}$  (median  $229\mu\text{mol/L}$ ). A similar proportion of subjects were identified as having yellow eyes by both staff and parents (70% and 67%, respectively). In 45 cases (6.3%) there was a discrepancy between the staff member and parent’s opinion. Staff were slightly more likely to notice icterus.

The range of No values for both inexperienced and experienced examiners covers the whole range of clinically relevant TSB values, although the majority are below  $200\mu\text{mol/L}$ . No examiner determined the sclera to be yellow for values below  $97\mu\text{mol/L}$ . There is little agreement on the threshold beyond which conjunc-



**Figure 5.7:** Box plots of TSB for responses to visual assessment question (n=716). (a) Parental opinion. (b) Staff member opinion.

tival icterus becomes visible. In adults, Ruiz et al. estimated the threshold to be 3.1mg/dL, or 53 $\mu\text{mol/L}$  [120]. Although their study involved 62 examiners, there were only six subjects. In neonates, Azzuqa and Watchko found that conjunctival icterus became visible at approximately 14mg/dL, or 239 $\mu\text{mol/L}$  [121]. However, only one of the four examiners was blinded to the TSB value, and no neonates with TSB less than 171 $\mu\text{mol/L}$  were enrolled.

If one naively takes the presence of conjunctival icterus as a positive screening test result for  $\text{TSB} \geq 250 \mu\text{mol/L}$ , the sensitivity would be 0.98 for staff and 0.96 for parents. However, the specificity is low at 0.41 and 0.43 for staff and parents, respectively. If this were the only heuristic for blood test referral the number of false alarms would be impractical. In the home setting, variable lighting conditions could make visual inspection less reliable still.

Comparing visual inspection with the three-repeat neoSCB protocol where responses were available (n=177), the sensitivity is similar (1.00 for neoSCB, 1.00 for staff, 0.94 for parents) and the specificity is higher for neoSCB (0.75 for neoSCB, 0.58 for staff, 0.62 for parents). Given the availability of smartphone devices in the home setting, the neoSCB app could bring peace of mind to parents and carers, and reduce the number of false alarms, which lead to costly and sometimes dangerous travel in remote parts of the world. Furthermore, the objective nature of the app recommendation may lead to better compliance than an instruction to seek medical advice “at the first sign of a yellow sclera”.



## 5.5 Applicability and Usability

In this chapter, we have evaluated the screening performance of the neoSCB app and shown it to be superior to naive visual inspection of the sclera and comparable to the JM-105 TcB. To show that it has potential in real-world scenarios, it is important to identify the precise clinical pathways in which the app might feature, as well as the environment in which it might be used and the users who might take advantage of it.

An existing model for the use of a screening device is that laid out by NICE for the use of the TcB. Term or near-term neonates who have not undergone phototherapy and are at least one day old can be screened at the  $250\mu\text{mol/L}$  level. This may take place in the postnatal ward, an outpatient clinic, or the home setting. In all cases, a trained healthcare professional is the user. A similar usage model could be implemented for the neoSCB app. A key advantage would be the relative accessibility and low expense of the app relative to the TcB. This would mean all visiting midwives and poorly resourced hospitals would have access to an objective means of screening for jaundice. Restricting users to healthcare professionals would carry benefits: one could ensure that the smartphone was well calibrated at all times, that it was used according to a protocol, and that the recommendation was acted on with the proper urgency.

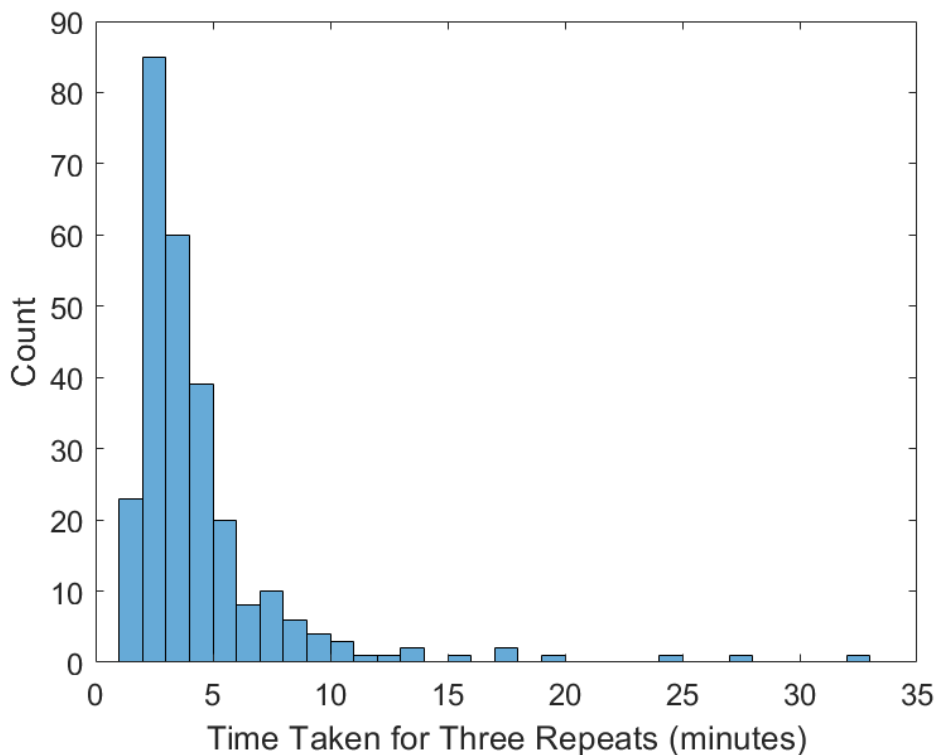
Given the ubiquity of smartphones, it would be possible for anyone to be a user of the app in principle. After verifying the suitability of their personal smartphone, new parents could download the neoSCB app and have it calibrated before the mother and neonate are discharged from hospital. This would allow for monitoring at will during the first two weeks of life, allaying anxiety for parents and flagging dangerous hyperbilirubinaemia sooner. This would require an intuitive and user-friendly app experience, including instructions for appropriate usage at each stage, including repeat captures. In the event of a positive test, the app could automatically contact relevant healthcare professionals to follow up. A history of tests could be stored and discussed with a visiting midwife, who would be better placed to interpret trends in bilirubin over time. Further studies are required to determine if

the app is usable with minimal instruction. Further app development could integrate the neoSCB output with a patient ID and history, as TcB devices now offer.

In the DJAN study all subjects were imaged by two dedicated research nurses. Their feedback helped improve the app usability to a great extent. For example, a shutter sound that accompanied the capture was removed because it made the subject flinch, blurring the second image in the image pair, and the gap between the flash and no-flash images was reduced to make the pair quicker to capture.

In a clinical environment, time is often constrained. To be practical, the app should not take too long to capture images. While the image capture itself is brief, it is also necessary to count the time taken to line up a suitable capture, which is complicated in the case of a moving or sleeping neonate. Furthermore, we have determined that competitive performance requires three repeat captures. Collecting all three must not be impractically time consuming.

272 subjects in the DJAN study had three repeats captured where a real-time



**Figure 5.8:** Histogram of time taken for three repeats and real-time result generation (n=269).

ROI selection lead to three quoted SCB values. 269 of these had the time recorded at the beginning and the end of the capture session. The histogram of the time taken to complete the process (capture, ROI selection, and recording the result successfully three times) is shown in Figure 5.8. The median time taken was three minutes. 95% of measurement sets took place within nine minutes. The longest time taken was 32 minutes. It appears that in the majority of cases the time taken to achieve a screening result is feasible.

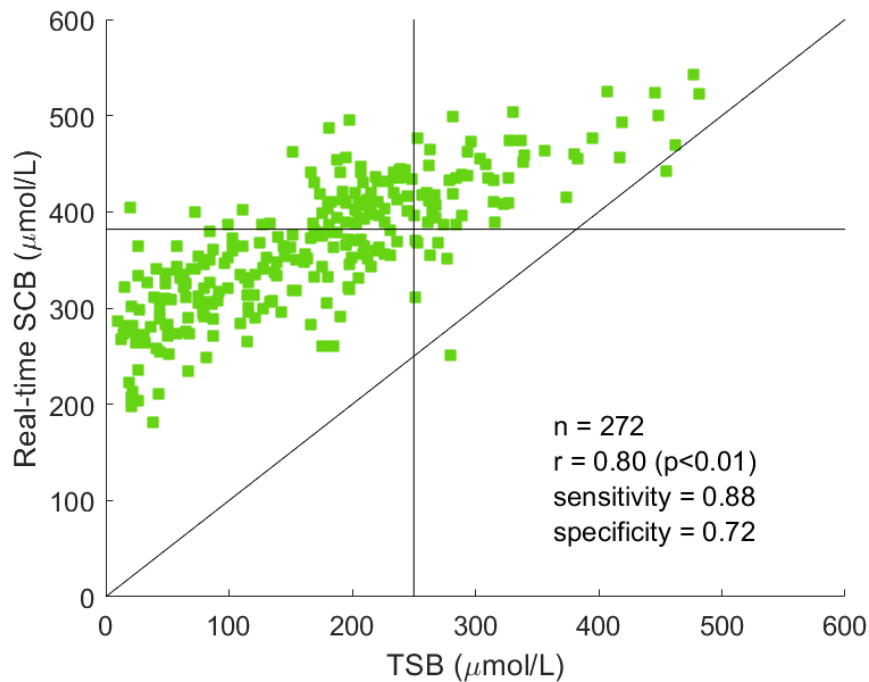
### **5.5.1 Real-time ROI Selection and Result**

Thus far, the sclera ROI has been drawn in a post-hoc analysis of the image data captured. Performance variability and trade-offs between a judicious and full sclera segmentation approach were discussed in Chapter 4. In the following, the accuracy of results generated through a real-time ROI selection are analysed.

The latest version of the neoSCB app supports real time ROI selection by panning and zooming of the captured image to position a square ROI box in the sclera. Raw pixel data are extracted from the flash and no-flash images according to this selection, and a SCB result is quoted in real-time. If the SSNR is insufficient, the result is not displayed and the user is prompted to take the image pair again. Instructions were given to the DJAN study research nurses capturing the data to position the ROI box such that it occupied the largest region of sclera possible, whilst ensuring that the box was fully inside the sclera in both the flash and no-flash images. See Appendix B for the full neoSCB usage guide, which includes examples.

The position of the ROI box on the flash image is also used on the no-flash image, which is displayed at the same time. By forcing the ROI box to be in the same position in each image, the user is made aware if excessive movement occurs between the two captures, as it will be impossible to position the box within the sclera in both images.

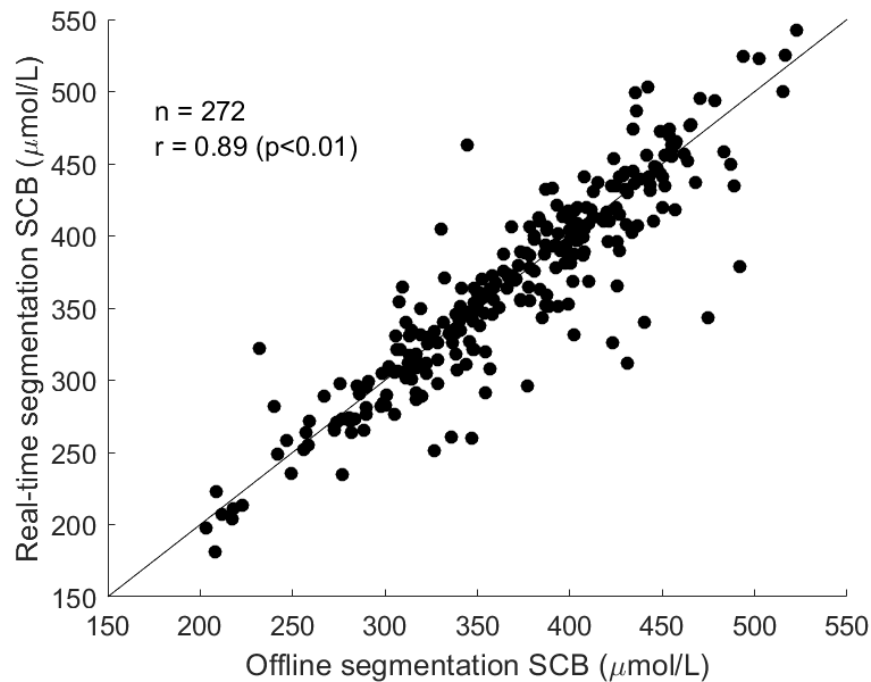
Figure 5.9 shows the correlation between TSB and the real-time estimate given by the neoSCB app and recorded on the datasheet. Three repeats were taken for each subject and have been subsequently averaged. Clearly, SCB overestimates TSB consistently. This is because an earlier SCB model trained on insufficient data



**Figure 5.9:** Real-time SCB averaged across three repeats vs TSB ( $n = 272$ ).

was being trialled in the neoSCB app. Nevertheless, the correlation is strong at 0.80 ( $p < 0.01$ ), and given an appropriately adjusted decision threshold, sensitivity and specificity are comparable to the latest SCB model with judicious segmentation.

In order to determine if real-time ROI selection is viable, it is necessary to isolate the effect of segmentation from the effect of the under-trained model. To this end, Figure 5.10 shows the correlation between the real-time SCB values and the SCB values using the same model (and same images) but with offline, judicious segmentations. Points are distributed along the line of equality. The mean absolute difference between SCB values using offline and in-app ROI selection is  $20\mu\text{mol/L}$ . This is only slightly higher than the mean absolute difference between judicious and full segmentation SCB values observed in Section 4.5.1 ( $18\mu\text{mol/L}$ ). This shows that the real-time ROI selection using the app interface is not significantly decreasing the screening performance despite the relative lack of fine control compared to offline segmentation.



**Figure 5.10:** SCB values calculated from users' real-time in-app segmentation correlation with SCB values calculated from offline (judicious) segmentation (n=272).

## 5.6 Summary

In this chapter the screening performance of the neoSCB app is shown to be as good as the JM-105 if three image pair repeats are averaged. The sensitivity was 1.00 (0.95-1.00) and the specificity was 0.76 (0.70-0.82).

The effect of phototherapy on SCB was to cause significant underestimation, negating the hypothesis that the sclera would not be affected by phototherapy while covered. As expected, phototherapy also caused significant underestimation in JM-105 measurements.

Visual inspection of the sclera for conjunctival icterus by both parents and staff was found to be effective in identifying neonates above the  $250\mu\text{mol/L}$ . However, specificity was lower than that quoted by TcB and image-based screening methods.

The neoSCB app was shown to be practical and effective in screening. The time taken to generate a result over three image pair repeats was less than 9 minutes in 95% of cases. Real-time ROI selection was shown to be convenient and deliver performance only slightly inferior to an offline judicious segmentation.

## Chapter 6

# General Conclusions

### Summary

In this thesis, the conception, development, and demonstration of the neoSCB jaundice screening app has been described. The app has screening performance comparable to a latest-generation TcB device while costing a fraction of the price. It has been trained on the largest database of images of jaundiced sclerae ever collected, which is also the largest study on image-based jaundice screening in a majority Black African population.

In developing the neoSCB app, the central focus has been to truly leverage the key advantage of the smartphone in a global context: accessibility. This has informed several decisions: the use of the Android operating system; the use of either front or back-facing camera; the lack of dependence on internet connectivity; the lack of need for accessories such as light-blocking phone housing or colour cards. Each of these decisions helps maximise the potential reach of the app, which is essential to reducing jaundice-related death and disability in parts of the world without reliable distribution networks, internet connectivity, or high-end phones.

The central challenge has been achieving device- and ambient-independent colour measurements without light-blocking phone housing or colour cards. To this end, a novel ambient subtraction approach was implemented. Colour card-based approaches such as BiliCam, BiliScan, and Picterus must contend with the cost, maintenance, and distribution of purpose-made colour cards, or risk inaccurate colour using commercial printers. Ambient subtraction has been shown to provide

accurate chromaticity measurements following a one-time calibration.

### **Contributions and findings**

The primary contributions of this work are the development of the neoSCB app and the demonstration of its screening efficacy. The correlation of TSB with three-repeat SCB values was 0.83 (0.78-0.87). The sensitivity in identifying newborns with  $TSB \geq 250 \mu\text{mol/L}$  was 1.00, with a lower bound of 0.95. The specificity was 0.76 (0.70-0.82). For the same 179 subjects, the sensitivity (specificity) of the Draeger JM-105 was found to be 0.97 (0.79). This result shows a comparable performance of the neoSCB app to the state-of-the-art TcB device.

It was shown that the app would result in significantly fewer false alarms than if parents were instructed to check for a yellow sclera discolouration. In our sample, parents were able to notice icterus in 96% of babies with  $TSB \geq 250$ . However, they also observed icterus in a large fraction of babies safely below the threshold. If parents were to seek medical attention based on their visual inspection of the eye, there would be a significant number of false alarms (false positive rate 0.57,  $n=716$ ). These findings are consistent with other studies on the visual identification of jaundice [39] [44]. There is a health and economic cost associated with unnecessary hospital visits, especially in remote areas and in times of pandemic. Conversely, when a newborn is in need of jaundice treatment, the objectivity of an app decision may translate to better care-seeking behaviour from parents compared to a subjective visual assessment. In particular, the COVID-19 pandemic may lead to poorer neonatal health outcomes because parents are reticent to seek medical attention for fear of viral exposure. An app such as neoSCB may mitigate this risk [122].

The neoSCB app is the first standalone app to screen for neonatal jaundice via the sclera. This obviates the need to account for skin pigmentation, which has been a source of error in similar apps using either machine learning or physics-based models to correct for it [82] [78]. Furthermore, the DJAN study results show that sclera chromaticity is a better predictor of TSB than skin chromaticity in Black African neonates. This work also demonstrates that it is feasible to capture newborn sclera images in a reasonable time frame in a clinical setting.

By imaging the sclera, we have shown a simple linear chromaticity-based TSB prediction model delivers competitive results. The SCB model derived from DJAN study data was chosen among a set of candidate models using 10-fold cross validation. It was shown to perform equally well on unseen data collected using a different model of smartphone in a different population (UCLH Smartphone study,  $n=38$ ). It is essential that a model for TSB estimation generalises to unseen data. Other smartphone app-based approaches have noticed a loss of performance among different ethnic groups or times of the day [82], or have employed cross validation in training their model but not yet tested on an independent test dataset [77].

Another contribution of this work is the image processing pipeline used to derive ambient- and device-independent chromaticity data from flash/ no-flash image pairs. Explicit subtraction of the ambient light contribution with a mapping to xy chromaticity provides an image feature invariant to ambient illumination and shading. Our novel contribution is the application of this approach to a medical imaging problem. Further, it is shown that a consumer smartphone is a practical device for this image processing system, and that flash-as-illumination and screen-as-illumination modalities are both feasible. This enables future research into applications of low-cost chromaticity measurement in situ.

This work has provided further evidence that phototherapy leads to a subsequent underestimation from TcBs. The DJAN study is the first to investigate the effect of phototherapy on the relationship between sclera colour and TSB. Like TcB measurements, SCB measurements following phototherapy were more variable and on average suffered from an underestimation bias.

The key findings of this work can be summarised as follows:

- It is feasible to capture a flash/ no-flash image pair of the newborn sclera using a smartphone in a clinical setting (using both flash-as-illumination and screen-as-illumination modalities)
- Ambient subtraction can be used to obtain accurate chromaticity measurements following a one-time calibration, without the need for a reference card in frame



- Sclera chromaticity is better correlated with TSB than skin chromaticity in Black African neonates
- As with TcB measurements, SCB measurements based on the sclera chromaticity have a greater variability and tend to underestimate TSB following phototherapy
- For babies with gestational age  $\geq 35$  weeks, the neoSCB app has screening performance comparable to the JM-105 in identifying babies with TSB  $\geq 250 \mu\text{mol/L}$

### **Future work**

The greatest limitation of the neoSCB app is the estimate variability. While this is mitigated by taking three repeats, more study is required into the relative contribution to this variability from various factors (segmentation approach, SSNR level, etc.), so that it can be reduced to a minimum.

Real-time segmentation by the app user was shown to be adequate: correlation with TSB was not significantly worse than when segmentation was performed manually afterwards. However, minimising the number of subjective judgements may decrease estimate variability and lead to fewer errors. More research is needed into automatic sclera identification algorithms, and what level of user involvement in the segmentation process is optimal for repeatability and accuracy. Possible simple user input could include drawing a box around the eye, seeding the segmentation at any point in the sclera, or confirming the output segmentation does not include non-sclera pixels. Machine vision could be used to trigger image capture automatically by dynamically monitoring the camera preview for the presence of the sclera and the quality of the illumination. This could make image capture using the front-facing neoSCB modality much easier. In addition, pre-capture tracking of the sclera would improve image quality if the scene metering for auto exposure and auto focusing was based on the region of interest.

Filtering the pixels in the segmented ROI to exclude specular regions and blood vessels could improve the SCB estimate accuracy. This could be done using a

simple threshold on the histogram of recorded colour values, as investigated by Mariakakis et al. [88]. At present, a bulk median estimate for the sclera colour is used. While this limits the influence of small regions that are unrepresentative of the sclera colour, it would fail in cases where bleeds or specular reflections constitute a larger fraction of the total ROI. The comparison of judicious and full segmentations did not show a large difference in performance, implying that the bulk median approach is robust to the proportions of unrepresentative pixels typically observed. The colour metric derived from the ROI could be further improved by making use of the spatial information recorded in the image. This idea is based on the observation that unrepresentative pixels are not randomly distributed through the ROI: a pixel is more likely to be affected by specular reflection if it is adjacent to another specular pixel. In future, superior TSB estimates could be made by combining both spatial and colour information.

The definition of SSNR proposed in this thesis can be used to improve ambient-subtracted chromaticity estimates [3], and the use of a real-time feedback on SSNR improved the quality of images captured with the neoSCB app. Future work could investigate ways of refining the definition of SSNR. Currently, SSNR is calculated based on the sum of RGB channels. However, measurements from different colour sensors have different SNR. Treating these separately could improve the SSNR calculation. Secondly, the current definition of SSNR does not take into account the fact that values are averaged over a region of interest. Estimates derived from more pixels should have higher SNR.

One way in which the neoSCB app could be improved to increase SSNR would be to use different exposure settings for the flash and no-flash images. At present, the ISO and exposure time are determined for the flash image and then fixed for the no-flash image. While this guarantees that the linear values can be directly subtracted, it would improve the SNR in the no-flash image if it had its own exposure settings. For subtraction, the no-flash image values could be scaled down by a factor based on the ratio of exposure times (and ISO values) in the two images.

More and more smartphones provide access to raw capture capabilities. How-

ever, the requirement for raw capture imposes some limit of the potential for scalability. Being able to use both front and back camera mitigates this problem to some extent (some smartphones only provide raw access capabilities on the back facing camera). The neoSCB app should notify the user if their smartphone does not support raw capture.

The smartphone active illumination and camera spectral sensitivity functions determine the space of possible recorded chromaticities. This work showed that the neoSCB app worked on two different smartphones: the LG Nexus 5X using the front-facing camera and screen-as-illumination, and the Samsung S8 using the back-facing camera and flash-as-illumination. Although this indicates that the method can be used on a variety of devices, more work is needed to identify the effect that varying CSS and illumination SPD can have on estimation accuracy so that the app only runs on suitable smartphones. The one-time calibration step could be modified so that it also verifies the suitability of the smartphone hardware.

More studies are required to determine the usability of the app among parents and healthcare professionals unfamiliar with it. These studies should be used to inform the app interface in future, including any instructions and disclaimers. Future versions could automatically alert a healthcare professional following a positive test, alleviating the need for the user to interpret and act on the test result. Future app versions should also include a guided calibration process with adequate instruction for the untrained user to carry out the calibration.

In this work, a framework for jaundice estimation based on a two-step process was presented. In the first step, an image-based colour measurement is made. In the second step, this colour measurement is related to jaundice severity. These two steps are independent. It would be fruitful to measure the spectral reflectance of the jaundiced newborn sclera at a range of bilirubin levels to quantify and potentially improve the accuracy of the first step. A set of spectral reflectance measurements would also help improve the mapping between native RGB and XYZ, which could be optimised for objects in that region of the colour space.

Future work should investigate how the relationship between sclera colour and

TSB may be different in premature babies and babies with high TSB. In this work, we have noticed an SCB underestimation of TSB in premature babies consistent with the observation that premature babies often have thinner, bluish sclerae. The results also suggest a biological saturation effect at high TSB whereby increasing blood serum bilirubin concentrations no longer translate into increasingly saturated colour measurements in the sclera. These physiological hypotheses must be further investigated and, if confirmed, incorporated into future SCB models.

**Table 6.1:** Summary of future work.

| Category  | Description   |
|---|---|
| Technical Improvements  | <ul style="list-style-type: none"> <li>• Allow ISO and exposure time to change between flash and no-flash captures then scale before subtraction</li> <li>• Optimise CCM for spectra of jaundiced sclerae</li> <li>• Improve definition and calculation of SSNR</li> </ul>  |
| Understanding and modelling bilirubin-chromaticity relationship | <ul style="list-style-type: none"> <li>• Understand and model the effect of prematurity on sclera colour to avoid underestimation in premature babies</li> <li>• Investigate correcting SCB underestimation at high TSB</li> </ul>  |
| App improvements  | <ul style="list-style-type: none"> <li>• App to guide user through one-time calibration</li> <li>• App to store history to infer TSB trend rate and direction</li> <li>• App to alert healthcare professional automatically in case of positive screening result</li> <li>• Develop real-time automatic region of interest selection</li> </ul> |
| Further validation  | <ul style="list-style-type: none"> <li>• Test on a wide variety of smartphone models</li> <li>• Study to determine usability of neoSCB app by un-trained individuals</li> <li>• Study to determine if access to neoSCB app translates to better neonatal outcomes</li> </ul>  |

### **Outlook and uptake of technology**

Results from this work indicate that the neoSCB app is an effective screening tool at the  $250\mu\text{mol/L}$  threshold for term and near-term babies who have not undergone phototherapy. These are the same parameters within which NICE guidelines recommend the usage of TcB devices. A usage protocol very similar to that of the TcB could be trialled in hospitals, which could significantly increase the numbers of healthcare professionals with access to an objective means of screening *in situ*. There would be some advantages in restricting use to healthcare professionals: It would ensure that the result of the screening decision could be acted on in a prompt fashion, and that the smartphones remained properly calibrated. Ultimately, empowering parents to screen for jaundice using their own devices could have an even greater impact. Pre-discharge smartphone vetting and guidance in the use of the neoSCB app could allow parents to make objective measurements in the home setting during the first two weeks of life. This would be invaluable in identifying problematic cases of jaundice early and avoiding false alarms.

Smartphones cameras represent a untapped opportunity to measure jaundice objectively in situations where the only available approach is visual inspection. It is hoped that this thesis can contribute to decreasing jaundice-related death and disability worldwide by putting the power to screen newborns in the hands of parents and caregivers.

## **Appendix A**

# **Data Collection Sheets**

Figure A.1 shows the data collection sheet for the UCLH Smartphone Study. Figure A.2 shows the data collection sheet for the DJAN Study. This is the final version of the DJAN data collection sheet (created August 2019). Real time neoSCB app SCB and SSNR results can be recorded for three repeats per capture session. One individual can be imaged on four separate occasions, with consecutive capture sessions represented by columns labelled A, B, C, and D.

|  | <i>Data</i>  | <i>Comments</i>                        |
|--|--|--|
| <i>Date</i>  | <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>                            |  |
| <i>Patient Study ID</i>  | <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>                            |  |
| <i>Beginning Time</i>  | <input type="text"/> <input type="text"/> : <input type="text"/> <input type="text"/>  |  |
| <i>End Time</i>  | <input type="text"/> <input type="text"/> : <input type="text"/> <input type="text"/>  |  |
| <i>Number of photos taken</i>  | <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>                            |  |
| <i>Ethnicity</i>   |  |  |
| <i>Gender</i>  | <b>M F</b>   |  |
| <i>Postnatal Age (days)</i>  | <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>                            |  |
| <i>Gestational Age (weeks)</i>   | <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>                            |  |
| <i>Birth Weight (kg)</i>   | <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>                            |  |
| <i>Point of Care<br/>Bilirubinometer (TSB)<br/>(<math>\mu\text{mol/L}</math>)</i>  | <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>                            |  |
| <i>Transcutaneous<br/>Bilirubinometer (TcB)<br/>(<math>\mu\text{mol/L}</math>)</i> | <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>                            |  |
| <i>Phototherapy</i>  | <b>Y N</b>   | <i>Do eyes look yellow?</i> <b>Y N</b> |
|  | <i>Biliblanket</i> <b>Y N</b>  | _____                                  |
|  | <i>Lights</i> <b>Y N</b><br>1<br>2<br>3  | _____                                  |
| <i>Location of patient</i>   | <i>Room:</i>   | <i>Baby's position:</i>                |
|  | <i>Surrounding Environment:</i><br>Day / Night<br>Natural Light <input type="checkbox"/><br>Ceiling Fluorescent Light <input type="checkbox"/><br>Other: | _____                                  |

**Figure A.1:** Data collection sheet for UCLH Smartphone Study.

## DJAN Project: Data Collection Sheet

|   |  |   |   |   |   |   |   |   |   |   |   |
|---|--|---|---|---|---|---|---|---|---|---|---|
| Patient Study ID                          | <table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="width: 20px; text-align: center;">D</td> <td style="width: 20px; text-align: center;">J</td> <td style="width: 20px; text-align: center;">_</td> <td style="width: 20px; text-align: center;">#</td> <td style="width: 20px; text-align: center;">#</td> <td style="width: 20px; text-align: center;">#</td> <td style="width: 20px; text-align: center;">#</td> </tr> </table> |   |   |   | D | J | _ | # | # | # | # |
| D   | J  | _   | #   | #   | # | # |   |   |   |   |   |
| Gender                                    | M / F  |   |   |   |   |   |   |   |   |   |   |
| Birth Weight (kg)                         | _ . _ _  |   |   |   |   |   |   |   |   |   |   |
| Gestational Age (weeks)                   | _ _  |   |   |   |   |   |   |   |   |   |   |
| Capture Session                           | A  | B   | C   | D   |   |   |   |   |   |   |   |
| Date Today                                | DD / MM / YY   | DD / MM / YY  | DD / MM / YY  | DD / MM / YY  |   |   |   |   |   |   |   |
| Postnatal Age (days + hours)              | _ _ d _ _ h  | _ _ d _ _ h   | _ _ d _ _ h   | _ _ d _ _ h   |   |   |   |   |   |   |   |
| Currently Undergoing Phototherapy?        | Y / N  | Y / N   | Y / N   | Y / N   |   |   |   |   |   |   |   |
| Phototherapy within last 24h?             | Y / N  | Y / N   | Y / N   | Y / N   |   |   |   |   |   |   |   |
| Location                                  |  |   |   |   |   |   |   |   |   |   |   |
| Phone ID                                  |  |   |   |   |   |   |   |   |   |   |   |
| Surrounding Light Environment             | <input type="checkbox"/> Natural<br><input type="checkbox"/> Ceiling<br><input type="checkbox"/> Other:  | <input type="checkbox"/> Natural<br><input type="checkbox"/> Ceiling<br><input type="checkbox"/> Other: | <input type="checkbox"/> Natural<br><input type="checkbox"/> Ceiling<br><input type="checkbox"/> Other: | <input type="checkbox"/> Natural<br><input type="checkbox"/> Ceiling<br><input type="checkbox"/> Other: |   |   |   |   |   |   |   |
| Appearance: Do eyes appear yellow?        | Parent:<br>Y / N<br>Staff:<br>Y / N  | Parent:<br>Y / N<br>Staff:<br>Y / N   | Parent:<br>Y / N<br>Staff:<br>Y / N   | Parent:<br>Y / N<br>Staff:<br>Y / N   |   |   |   |   |   |   |   |
| Photo Capture Beginning Time              | _ _ : _ _  | _ _ : _ _   | _ _ : _ _   | _ _ : _ _   |   |   |   |   |   |   |   |
| Photo Capture Ending Time                 | _ _ : _ _  | _ _ : _ _   | _ _ : _ _   | _ _ : _ _   |   |   |   |   |   |   |   |
| TcB (Jaundice Meter) (µmol/L)             | _ _ _ _  | _ _ _ _   | _ _ _ _   | _ _ _ _   |   |   |   |   |   |   |   |
| Point of Care Bilimeter (SBR) (µmol/L)    | _ _ _ _  | _ _ _ _   | _ _ _ _   | _ _ _ _   |   |   |   |   |   |   |   |
| Lab Result (total / conj.) (µmol/L)       | _ _ / _  | _ _ / _   | _ _ / _   | _ _ / _   |   |   |   |   |   |   |   |
| neoSCB App Results: SCB (µmol/L) and SSNR | Repeat<br>1  | (SSNR: )  | (SSNR: )  | (SSNR: )  |   |   |   |   |   |   |   |
|   | Repeat<br>2  | (SSNR: )  | (SSNR: )  | (SSNR: )  |   |   |   |   |   |   |   |
|   | Repeat<br>3  | (SSNR: )  | (SSNR: )  | (SSNR: )  |   |   |   |   |   |   |   |

**Figure A.2:** Final version of data collection sheet for DJAN Study.



## **Appendix B**

# **neoSCB User Guide**

The following pages reproduce the neoSCB app user guide created in August 2019. The guide was created by Fiona Young (UCL Natural Sciences) with contributions from this author and Miranda Nixon-Hill (UCL Centre for Doctoral Training in Medical Imaging). It includes a guide for using and navigating the app, examples of what constitutes a poor image pair compared to a good image pair, capture tips, how to improve the SSNR if it is insufficient, and a listing of known bugs at the time of writing.

# NEOSCB USER GUIDE

AUGUST 2019

## 1 GENERAL INFORMATION

This guide applies to version 1.0 of the neoSCB app (neonatal Scleral-Conjunctival Bilirubinometer).

The app allows the user to choose a region of interest (ROI) in the sclera and receive a total serum bilirubin concentration estimate in real time. We call this estimate the SCB (Scleral-Conjunctival Bilirubin).

**It is important to note that the SCB estimate must not be used to inform clinical decisions in any way. The app is intended for research purposes only.**

This document describes how to use the app (Section 2) and how to take the best possible images (Section 3).

The app is still under development and the developers are aware of some intermittent bugs. Section 4 describes these bugs and how to deal with them.

## 2 USING THE APP

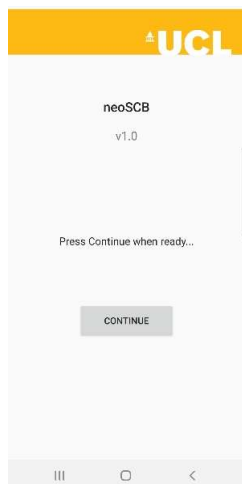


Figure 1 Start screen



Figure 2 Camera screen

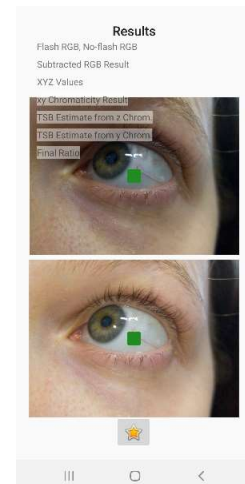




Figure 3 Selection and analysis screen

### START SCREEN

The starting screen (Figure 1) is what you see when you open the app. Click  when ready to take a pair of photos.

### CAMERA SCREEN

Here you can position the camera and take a pair of photos (Figure 2). Click the camera icon  or press either of the volume keys when ready to capture. Ensure that the subject does not move until the display changes to the selection display. Two photos are taken, one with and one without flash. There is a short delay between images. It is very important that the camera and subject do not move during this time.

---

## SELECTION AND ANALYSIS SCREEN

This is where region of interest (ROI) selection happens and the results of the analysis are displayed. Both the non-flash (top) and flash (bottom) images are shown.

---

### ROI selection

Zoom and pan around the **LOWER** image until the green square is positioned over an appropriate region of sclera. The green square represents the region of interest (ROI) that will be used for the pixel colour analysis. To zoom, press two fingers together on lower image and move them away from each other as if stretching them apart. To pan, simply drag your finger across the lower image.

ROI selection can only be performed on the bottom (flash) image. The top image is displayed to show where the selected ROI appears on the non-flash image (see Figure 4). This is important if some movement occurs (see section 3). The ROI should be chosen such that **ONLY** sclera is present in the ROI for **BOTH** images.

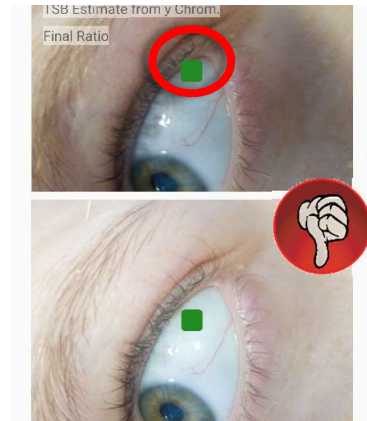



Figure 4 Example of bad ROI selection, where ROI includes non-sclera regions in one of the images

---

### Analysis

Once happy with the ROI position, tap the button at the bottom with the star  icon to perform pixel analysis

If analysis was successful, the results will be displayed above the images.

Analysis may be unsuccessful if the subtracted signal-to-noise ratio (SSNR) is too low. If SSNR is too low a warning message is displayed (Figure 5) and the results are not calculated.

If this is the case, first try moving the ROI, if possible, to a different region, as SSNR is calculated for the ROI only. If the warning message persists, retake the images with either

- Reduced ambient lighting (e.g. moving to a darker room or more shaded spot) or
- The camera closer to the subject

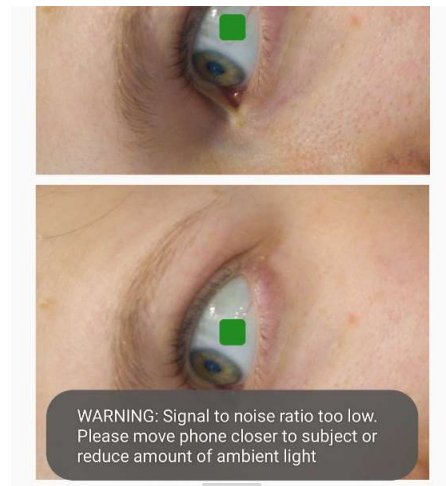


Figure 5 Low SSNR warning

---

## GOING BACK

To take a new image pair, return to the camera screen from the selection screen by tapping the back button once. Alternatively, the entire process can be restarted from the start screen by tapping back twice. Only have the camera screen open when ready to take photos, as the flash will drain battery power if the camera screen is left open for prolonged periods.

### 3 TAKING GOOD IMAGES

#### MOVEMENT

Ensure that the camera and subject remain still while the two images are being taken. If there is too much movement between the two images then they are unusable and new images should be taken (Figure 6). If it is difficult to select an ROI that is fully sclera for both images then there is too much movement and new photos must be taken. Also ensure that the subject does not blink during image acquisition.

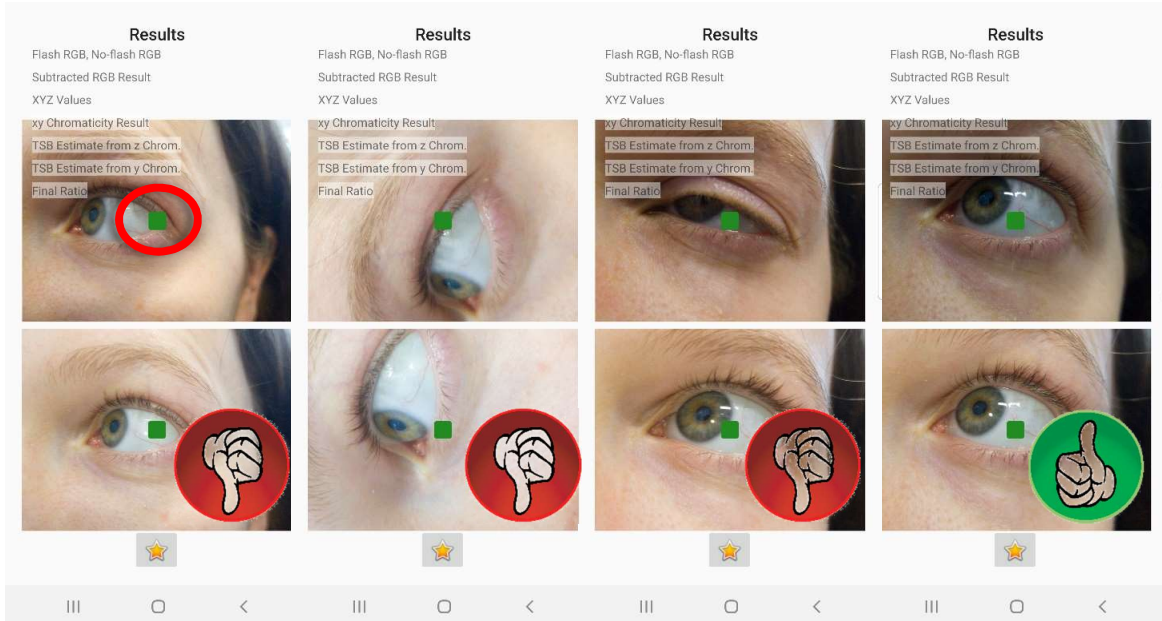


Figure 6 Examples of good and bad image pairs

#### LIGHTING

Ambient lighting should be kept to a minimum if possible, to maximise the difference between flash and non-flash. If there is too much ambient light, the SSNR warning will show (Figure 7). In this case, retake the photos in a darker environment or with the camera closer to the subject.

#### CAMERA POSITION

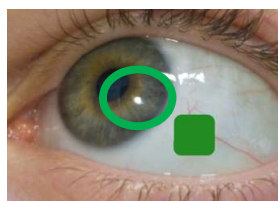


Figure 8 Specular reflection positioned over the iris

Hold the camera horizontally if possible and ensure that the specular reflection cause by the flash is positioned over the iris (as shown in Figure 8) and not over the sclera.

Holding the camera as close to the subject as possible will improve the signal to noise ratio.

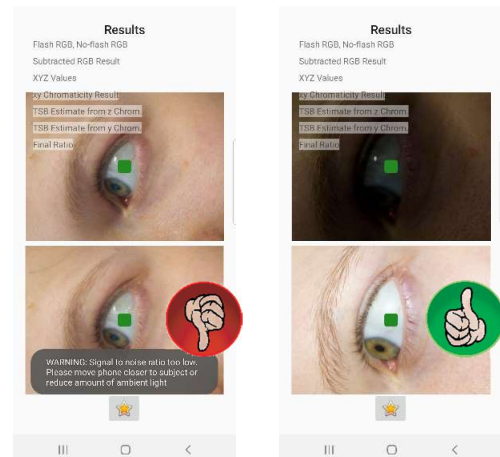


Figure 7 Bad and good lighting conditions. The greater the difference between flash and non-flash the better.

## 4 BUGS

### 1. Non-Flash image brighter than flash image

**Cause** Different exposure settings for flash and non-flash images

**Symptoms**

- Non-flash image (top image on ROI selection display) appears brighter than flash image
- SSNR too low warning displayed

**Solution** Go back and retake image pair.

### 2. Files misnamed (Blank non-flash image)

**Cause** ?

**Symptoms** Non-flash image will be blank / won't show up on the Selection screen (see Figure 9).

**Solution** Go back and retake image pair.

### 3. Empty byte file

**Solution** Go back and retake image pair.

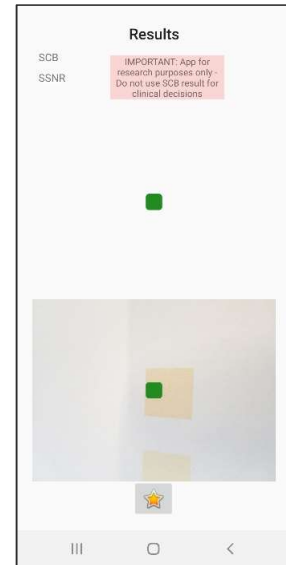
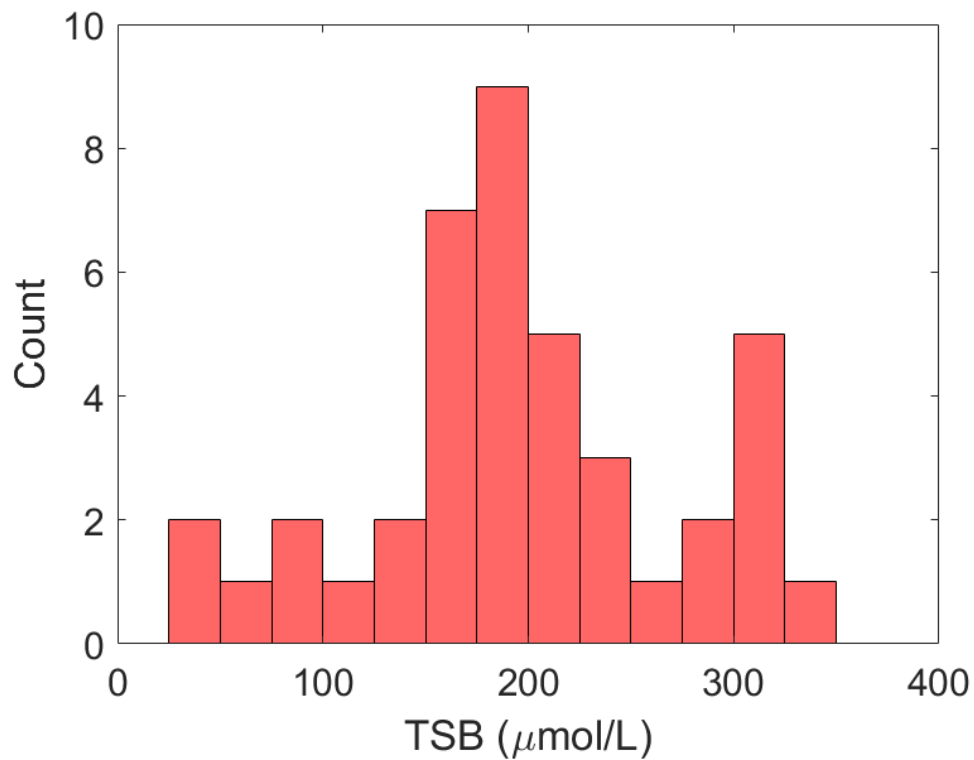


Figure 9 Blank non-flash image

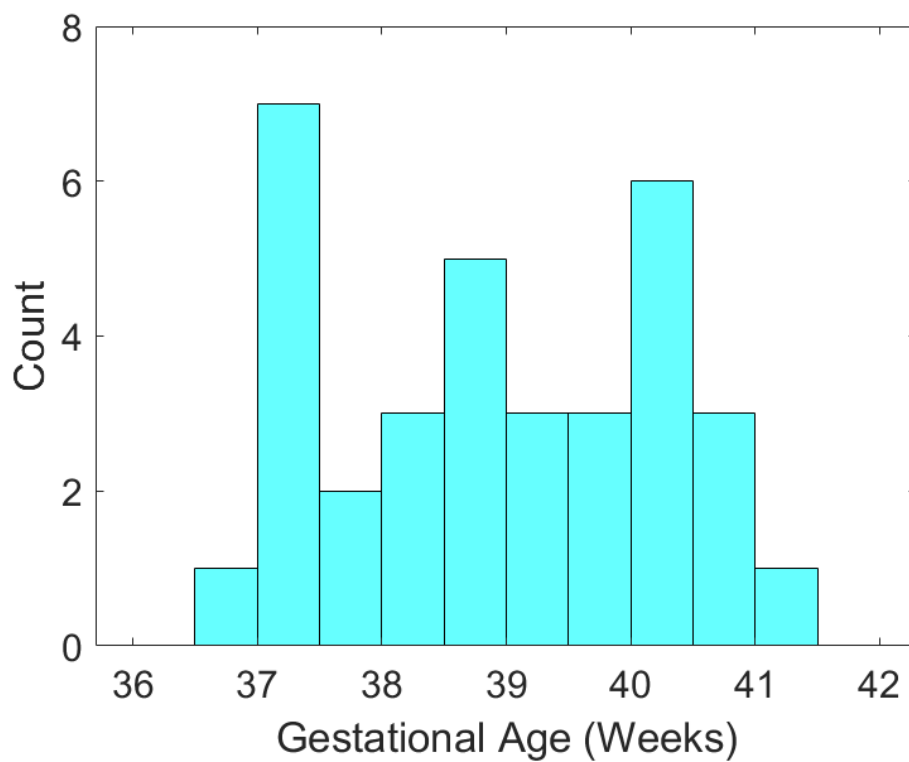
## Appendix C

# UCLH Smartphone Study Statistics

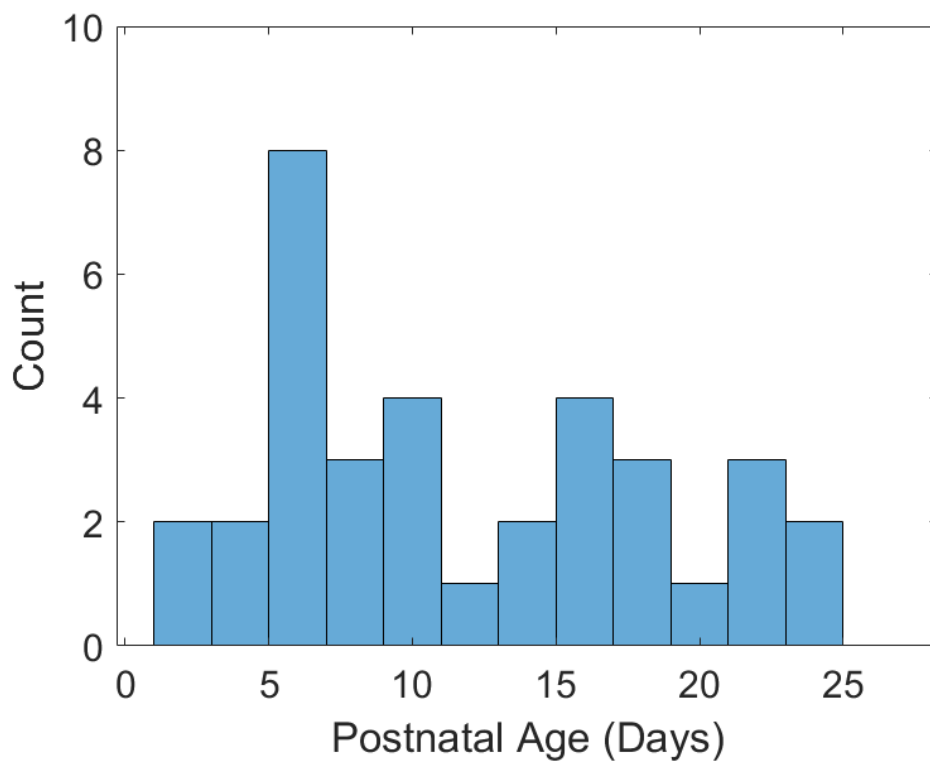
Histograms for the data collected in the UCLH Smartphone Study follow. Figure C.1 shows the histogram of TSB values, Figure C.2 shows the histogram of gestational ages, and Figure C.3 shows the histogram of postnatal ages.



**Figure C.1:** Histogram of TSB values for UCLH Smartphone Study (n=41).



**Figure C.2:** Histogram of gestational age for UCLH Smartphone Study (n=41).



**Figure C.3:** Histogram of postnatal age for UCLH Smartphone Study (n=41).

# Bibliography

- [1] Terence S Leung, Felix Outlaw, Lindsay W MacDonald, and Judith Meek. Jaundice Eye Color Index (JECI): quantifying the yellowness of the sclera in jaundiced neonates with digital photography. *Biomedical optics express*, 10(3):1250–1256, 2019.
- [2] Felix Outlaw, Miranda Nixon, Oluwatobiloba Odeyemi, Lindsay W MacDonald, Judith Meek, and Terence S Leung. Smartphone screening for neonatal jaundice via ambient-subtracted sclera chromaticity. *Plos one*, 15(3):e0216970, 2020.
- [3] Miranda Nixon, Felix Outlaw, and Terence S Leung. Accurate device-independent colorimetric measurements using smartphones. *Plos one*, 15(3):e0230561, 2020.
- [4] Felix Outlaw, Judith Meek, Lindsay W MacDonald, and Terence S Leung. Screening for neonatal jaundice with a smartphone. In *Proceedings of the 2017 International Conference on Digital Health*, pages 241–242, 2017.
- [5] Felix Outlaw, Miranda Nixon, Nana Okai Brako, Lindsay W MacDonald, Judith Meek, Christabel Enweronu-Laryea, and Terence S Leung. Smartphone colorimetry using ambient subtraction: application to neonatal jaundice screening in Ghana. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 172–175, 2019.



- [6] Miranda Nixon, Felix Outlaw, Lindsay W MacDonald, and Terence S Leung. The importance of a device specific calibration for smartphone colorimetry. In *27th Color and Imaging Conference Final Program and Proceedings*, pages 49–54, 2019.
- [7] Chulhyun Ahn et al. Advances in wound photography and assessment methods. *Advances in skin & wound care*, 21(2):85–93, 2008.
- [8] Lan B Van, KM Sicotte, RR Lassiter, KA Jablonski, DA Crean, JC Jeng, and MH Jordan. Digital photography: enhancing communication between burn therapists and nurses. *The Journal of burn care & rehabilitation*, 25(1):54–60, 2004.
- [9] National Collaborating Centre for Cancer. Melanoma: assessment and management. *National Institute for Health and Care Excellence (UK)*, 2015.
- [10] Shaun Collings, Oliver Thompson, Evan Hirst, Louise Goossens, Anup George, and Robert Weinkove. Non-invasive detection of anaemia using digital photographs of the conjunctiva. *PloS one*, 11(4):e0153286, 2016.
- [11] Graham E Quinn and Anand Vinekar. The role of retinal photography and telemedicine in ROP screening. In *Seminars in Perinatology*, volume 43, pages 367–374. Elsevier, 2019.
- [12] Josef Huemer, Siegfried K Wagner, and Dawn A Sim. The evolution of diabetic retinopathy screening programmes: A chronology of retinal photography from 35 mm slides to artificial intelligence. *Clinical Ophthalmology (Auckland, NZ)*, 14:2021, 2020.
- [13] Judith Shore, Michelle Green, Andrew Hardy, and Deborah Livesey. The compliance and cost-effectiveness of smartphone urinalysis albumin screening for people with diabetes in England. *Expert review of pharmacoeconomics & outcomes research*, 20(4):387–395, 2020.

- [14] Janet Rennie, Shona Burman-Roy, M Stephen Murphy, Guideline Development Group, et al. Neonatal jaundice: summary of NICE guidance. *BMJ*, 340(7757):c2409, 2010.
- [15] Pew Research Center. Smartphone ownership is growing rapidly around the world, but not always equally. *Pew Research Center's Global Attitudes Project*, 2019.
- [16] Vinod K Bhutani, Alvin Zipursky, Hannah Blencowe, Rajesh Khanna, Michael Sgro, Finn Ebbesen, Jennifer Bell, Rintaro Mori, Tina M Slusher, Nahed Fahmy, et al. Neonatal hyperbilirubinemia and rhesus disease of the newborn: incidence and impairment estimates for 2010 at regional and global levels. *Pediatric research*, 74(Suppl 1):86, 2013.
- [17] Alex Mariakakis, Edward Wang, Shwetak Patel, and Mayank Goel. Challenges in realizing smartphone-based health sensing. *IEEE Pervasive Computing*, 18(2):76–84, 2019.
- [18] R.A. Joshi and B. Jain. *Multiple Choice Questions in Physiology*. B. Jain BHMS Solved Papers. B Jain Publishers Pvt Limited, 2002.
- [19] Steven M Shapiro. Bilirubin toxicity in the developing nervous system. *Pediatric neurology*, 29(5):410–421, 2003.
- [20] David A Lightner and Antony F McDonagh. Molecular mechanisms of phototherapy for neonatal jaundice. *Accounts of Chemical Research*, 17(12):417–424, 1984.
- [21] M Jeffrey Maisels and Antony F McDonagh. Phototherapy for neonatal jaundice. *New England Journal of Medicine*, 358(9):920–928, 2008.
- [22] Thomas Hegyi, I Mark Hiatt, Ian M Gertner, Robert Zanni, and Theodorore Tolentino. Transcutaneous bilirubinometry II. Dermal bilirubin kinetics during phototherapy. *Pediatric research*, 17(11):888–891, 1983.

- [23] L Murli, A Thukral, MJ Sankar, S Vishnubhatla, AK Deorari, VK Paul, A Sakariah, R Agarwal, et al. Reliability of transcutaneous bilirubinometry from shielded skin in neonates receiving phototherapy: a prospective cohort study. *Journal of Perinatology*, 37(2):182–187, 2017.
- [24] KL Tan and F Dong. Transcutaneous bilirubinometry during and after phototherapy. *Acta paediatrica*, 92(3):327–331, 2003.
- [25] Christian V Hulzebos, Deirdre E Vader-van Imhoff, Arend F Bos, and Peter H Dijk. Should transcutaneous bilirubin be measured in preterm infants receiving phototherapy? The relationship between transcutaneous and total serum bilirubin in preterm infants with and without phototherapy. *PloS one*, 14(6):e0218131, 2019.
- [26] R Fonseca, R Kyralessa, M Malloy, J Richardson, and SK Jain. Covered skin transcutaneous bilirubin estimation is comparable with serum bilirubin during and after phototherapy. *Journal of Perinatology*, 32(2):129–131, 2012.
- [27] Shakuntala Nanjundaswamy, Anna Petrova, Rajeev Mehta, and Thomas Hegyi. Transcutaneous bilirubinometry in preterm infants receiving phototherapy. *American journal of perinatology*, 22(03):127–131, 2005.
- [28] Enrico Zecca, Giovanni Barone, Daniele De Luca, Rosa Marra, Eloisa Tiberi, and Costantino Romagnoli. Skin bilirubin measurement during phototherapy in preterm and term newborn infants. *Early human development*, 85(8):537–540, 2009.
- [29] U Costa-Posada, A Concheiro-Guisán, MF Táboas-Ledo, E González-Colmenero, ML González-Durán, M Suarez-Albo, C Duran Fernández-Feijoo, M Pumarada-Prieto, Cristina Martínez-Reglero, and JR Fernández-Lorenzo. Accuracy of transcutaneous bilirubin on covered skin in preterm and term newborns receiving phototherapy using a JM-105 bilirubinometer. *Journal of Perinatology*, pages 1–6, 2019.

- [30] Anne L Gordon, Michael English, J Tumaini Dzombo, Mary Karisa, and Charles RJC Newton. Neurological and developmental outcome of neonatal jaundice and sepsis in rural Kenya. *Tropical Medicine & International Health*, 10(11):1114–1120, 2005.
- [31] Tina M Slusher, Alvin Zipursky, and Vinod K Bhutani. A global need for affordable neonatal jaundice technologies. In *Seminars in perinatology*, volume 35, pages 185–191. Elsevier, 2011.
- [32] Thomas B Newman and Mark A Klebanoff. Neonatal hyperbilirubinemia and long-term outcome: another look at the collaborative perinatal project. *Pediatrics*, 92(5):651–657, 1993.
- [33] Thomas B Newman, Petra Liljestrand, and Gabriel J Escobar. Infants with bilirubin levels of 30 mg/dl or more in a large managed care organization. *Pediatrics*, 111(6):1303–1311, 2003.
- [34] Bolajoko O Olusanya, Donald O Imosemi, and Abieyuwa A Emokpae. Differences between transcutaneous and serum bilirubin measurements in Black African neonates. *Pediatrics*, page e20160907, 2016.
- [35] Eunice Adei-Atiemo, Onike Rodrigues, and Ebenezer Badoe. Classification and risk factors for cerebral palsy in the Korle Bu Teaching Hospital, Accra: A case–control study. *Pediatrics*, 135(Supplement 1):S7–S7, 2015.
- [36] Rachel C Amos, Hannah Jacob, and Wynne Leith. Jaundice in newborn babies under 28 days: NICE guideline 2016 (CG98). *Archives of Disease in Childhood*, 102(4):207–209, 2017.
- [37] Samar N El-Beshbishi, Karen E Shattuck, Amin A Mohammad, and John R Petersen. Hyperbilirubinemia and transcutaneous bilirubinometry. *Clinical Chemistry*, 55(7):1280–1287, 2009.

- [38] Emily House. A multi-spectral digital imaging technique for the non-contact screening of neonatal jaundice - a pilot study. Master's thesis, UCL, London, 2016.
- [39] Virginia A Moyer, Chul Ahn, and Stephanie Sneed. Accuracy of clinical judgement in neonatal jaundice. *Archives of pediatrics & adolescent medicine*, 154(4):391–394, 2000.
- [40] Daniele De Luca, Enrico Zecca, Antonio Alberto Zuppa, and Costantino Romagnoli. The joint use of human and electronic eye: visual assessment of jaundice and transcutaneous bilirubinometry. *The Turkish journal of pediatrics*, 50(5):456, 2008.
- [41] American Academy of Pediatrics Subcommittee on Hyperbilirubinemia. Management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation. *Pediatrics*, 114(1):297, 2004.
- [42] Tina M Slusher, Ishaya A Angyo, Fidela Bode-Thomas, Francis Akor, Sunday D Pam, Adedotun A Adetunji, Donald W McLaren, Ronald J Wong, Hendrik J Vreman, and David K Stevenson. Transcutaneous bilirubin measurements and serum total bilirubin levels in indigenous African infants. *Pediatrics*, 113(6):1636–1641, 2004.
- [43] Mickael Afanetti, Sergio Eleni dit Trolli, Nadya Yousef, Ikram Jrad, and Mostafa Mokhtari. Transcutaneous bilirubinometry is not influenced by term or skin color in neonates. *Early human development*, 90(8):417–420, 2014.
- [44] Arie Riskin, Ada Tamir, Amir Kugelman, Miri Hemo, and David Bader. Is visual assessment of jaundice reliable as a screening tool to detect significant neonatal hyperbilirubinemia? *The Journal of pediatrics*, 152(6):782–787, 2008.
- [45] N Kevin Ives. Management of neonatal jaundice. *Paediatrics and Child Health*, 21(6):270–276, 2011.

- [46] Maria MP Petrou and Costas Petrou. *Image processing: the fundamentals*. John Wiley & Sons, 2010.
- [47] Tim Hayes. Next-generation cell phone cameras. *Optics and Photonics News*, 23(2):16–21, 2012.
- [48] Lindsay W MacDonald. Determining camera spectral responsivity with multispectral transmission filters. In *Color and Imaging Conference*, volume 2015, pages 12–17. Society for Imaging Science and Technology, 2015.
- [49] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.
- [50] Tuan V.-D. Joel M. Optical properties of tissue. In *Biomedical Photonics Handbook*, chapter 2, pages 1–76. CRC Press, 2003.
- [51] Optical properties spectra. Oregon Medical Laser Center (OMLC), Portland, Oregon. <http://www.omlc.org/spectra/> (accessed: 2015.10.03).
- [52] Michael F. Sasha M., Jelena M. Reflectance spectroscopy. In Nimmi Ramanujam David A. Boas, Constantinos Pitris, editor, *Handbook of Biomedical Optics*, chapter 6, pages 103–130. CRC Press, 2011.
- [53] Robert William Gainer Hunt and Michael R Pointer. *Measuring colour*. John Wiley & Sons, 2011.
- [54] Rob Sumner. Processing raw images in MATLAB, 2014. [http://www.rcsumner.net/raw\\_guide/RAWguide.pdf/](http://www.rcsumner.net/raw_guide/RAWguide.pdf/) (accessed: 2020.11.03).
- [55] Steven A Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985.
- [56] H-C Lee, Edwin J. Breneman, and Carl P. Schulte. Modeling light reflection for computer color vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(4):402–409, 1990.

- [57] Dohyoung Lee and Konstantinos N Plataniotis. A taxonomy of color constancy and invariance algorithm. In *Advances in Low-Level Color Image Processing*, pages 55–94. Springer, 2014.
- [58] Samuel W Hasinoff. Photon, poisson noise. In *Computer Vision*, pages 608–610. Springer, 2014.
- [59] Ce Liu, Richard Szeliski, Sing Bing Kang, C Lawrence Zitnick, and William T Freeman. Automatic estimation and removal of noise from a single image. *IEEE transactions on pattern analysis and machine intelligence*, 30(2):299–314, 2008.
- [60] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, 2008.
- [61] EW Bogaart, W Hoekstra, IM Peters, ACM Kleimann, and JT Bosiers. Very low dark current CCD image sensor. *IEEE Transactions on Electron Devices*, 56(11):2462–2467, 2009.
- [62] Hugh S Fairman, Michael H Brill, and Henry Hemmendinger. How the CIE 1931 color-matching functions were derived from Wright-Guild data. *Color Research & Application*, 22(1):11–23, 1997.
- [63] Stephen Westland, Caterina Ripamonti, and Vien Cheung. *Computational colour science using MATLAB*. John Wiley & Sons, 2012.
- [64] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980.
- [65] Brian Funt and Lilong Shi. The rehabilitation of maxRGB. In *Color and Imaging Conference*, volume 2010, pages 256–259. Society for Imaging Science and Technology, 2010.
- [66] David A Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–35, 1990.

- [67] Graham D Finlayson, Steven D Hordley, and Ingeborg Tastl. Gamut constrained illuminant estimation. *International Journal of Computer Vision*, 67(1):93–109, 2006.
- [68] Graham D Finlayson and Steven D Hordley. Color constancy at a pixel. *JOSA A*, 18(2):253–264, 2001.
- [69] J-M Geusebroek, Rein Van den Boomgaard, Arnold W. M. Smeulders, and Hugo Geerts. Color invariance. *IEEE Transactions on Pattern analysis and machine intelligence*, 23(12):1338–1350, 2001.
- [70] Jeffrey M DiCarlo, Feng Xiao, and Brian A Wandell. Illuminating illumination. In *Color and Imaging Conference*, volume 2001, pages 27–34. Society for Imaging Science and Technology, 2001.
- [71] Laurence T Maloney. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *JOSA A*, 3(10):1673–1683, 1986.
- [72] Cheng Lu and Mark S Drew. Practical scene illuminant estimation via flash/no-flash pairs. In *Color and Imaging Conference*, volume 2006, pages 84–89. Society for Imaging Science and Technology, 2006.
- [73] Somsak Leartveravat. Transcutaneous bilirubin measurement in full term neonate by digital camera. *Medical Journal of Srisaket Surin Buriram Hospitals*, 24(1):105–118, 2009.
- [74] Terence S Leung, Karan Kapur, Ashley Guillian, Jade Okell, Bee Lim, Lindsay W MacDonald, and Judith Meek. Screening neonatal jaundice based on the sclera color of the eye using digital photography. *Biomedical optics express*, 6(11):4529–4538, 2015.
- [75] Moattar Raza Rizvi, Farah Mansoor Alaskar, Raid Saleem Albaradie, Noor Fatima Rizvi, and Khaled Al-Abdulwahab. A novel non-invasive tech-



- nique of measuring bilirubin levels using Bilicapture. *Oman medical journal*, 34(1):26, 2019.
- [76] Lilian De Greef, Mayank Goel, Min Joon Seo, Eric C Larson, James W Stout, James A Taylor, and Shwetak N Patel. Bilicam: using mobile phones to monitor newborn jaundice. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 331–342, 2014.
- [77] James A Taylor, James W Stout, Lilian de Greef, Mayank Goel, Shwetak Patel, Esther K Chung, Aruna Koduri, Shawn McMahon, Jane Dickerson, Elizabeth A Simpson, et al. Use of a smartphone app to assess neonatal jaundice. *Pediatrics*, 140(3):e20170312, 2017.
- [78] Anders Aune, Gunnar Vartdal, Håkon Bergseng, Lise Lyngsnes Randeberg, and Elisabeth Darj. Bilirubin estimates from smartphone images of newborn infants' skin correlated highly to serum bilirubin levels. *Acta Paediatrica*, 2020.
- [79] ZH Rong, F Luo, LY Ma, L Chen, L Wu, W Liu, LZ Du, and XP Luo. Evaluation of an automatic image-based screening technique for neonatal hyperbilirubinemia. *Zhonghua er ke za zhi= Chinese journal of pediatrics*, 54(8):597–600, 2016.
- [80] Xiaoyue Dong, Xiaofan Sun, Zhangbin Yu, and Shuping Han. Image-based neonatal hyperbilirubinemia screening after hospital discharge. *Iranian Journal of Public Health*, 49(6):1079–1086, 2020.
- [81] Bo Yang, Di Huang, Xiangyu Gao, SU Min, LI Min, Honglin Lei, Yi Ren, Chunming Shi, and Dandan Zhao. Neonatal and early infantile jaundice: assessment by the use of the smartphone. *Chinese Journal of Neonatology*, 33(4):277–282, 2018.

- [82] Yi Ren, Di Huang, Bo Yang, and Xiangyu Gao. The effects on accuracy of image-based estimating neonatal jaundice with a smartphone app in the different conditions. *Pediatric Medicine*, 3(0), 2020.
- [83] Srujana Swarna, Sekar Pasupathy, Balaji Chinnasami, DN Manasa, and B Ramraj. The smart phone study: assessing the reliability and accuracy of neonatal jaundice measurement using smart phone application. *International Journal of Contemporary Pediatrics*, 5(2):285–289, 2018.
- [84] Mustafa Aydın, Fırat Hardalaç, Berkan Ural, and Serhat Karap. Neonatal jaundice detection system. *Journal of medical systems*, 40(7):1–11, 2016.
- [85] AT Sufian, GR Jones, HM Shabeer, EY Elzagzoug, and JW Spencer. Chromatic techniques for in vivo monitoring jaundice in neonate tissues. *Physiological measurement*, 39(9):095004, 2018.
- [86] Sarah B Munkholm, Tobias Krøgholt, Finn Ebbesen, Pal B Szecsi, and Søren R Kristensen. The smartphone camera as a potential method for transcutaneous bilirubin measurement. *PloS one*, 13(6):e0197938, 2018.
- [87] Pouria Padidar, Mohammadamin Shaker, Hamid Amoozgar, Mohammadhossein Khorraminejad-Shirazi, Fariba Hemmati, Khadijeh Sadat Najib, and Shahnaz Pourarian. Detection of neonatal jaundice by using an Android OS-based smartphone application. *Iranian Journal of Pediatrics*, 29(2), 2019.
- [88] Alex Mariakakis, Megan A Banks, Lauren Phillipi, Lei Yu, James Taylor, and Shwetak N Patel. Biliscreen: smartphone-based scleral jaundice monitoring for liver and pancreatic disorders. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(2):1–26, 2017.
- [89] Md Messal Monem Miah, Rafat Jamal Tazim, Fatema Tuj Johora, Md Ibrahim Al Imran, Sanzida Sayedul Surma, Fariba Islam, Rashid Shabab, Celia Shahnaz, and Arik Subhana. Non-invasive bilirubin level quantification and jaundice detection by sclera image processing. In *2019 IEEE Global Humanitarian Technology Conference (GHTC)*, pages 1–7. IEEE, 2019.

- [90] Md Redwan Sammir, Kazi Md Towhidul Alam, Pradipta Saha, Md Mushfiqur Rahaman, and Quazi Delwar Hossain. Design and implementation of a non-invasive jaundice detection and total serum bilirubin measurement system. In *2018 10th International Conference on Electrical and Computer Engineering (ICECE)*, pages 137–140. IEEE, 2018.
- [91] Amit Laddi, Sanjeev Kumar, Shashi Sharma, and Amod Kumar. Non-invasive jaundice detection using machine vision. *IETE Journal of Research*, 59(5):591–596, 2013.
- [92] Grand Challenges Canada. Saving Lives at Birth: A Grand Challenge for Development. <https://www.grandchallenges.ca/programs/saving-lives-at-birth/> (accessed: 2020.10.09).
- [93] Timothy York and Raj Jain. Fundamentals of image sensor performance. [jain/cse567-11/ftp/imgsens/index.html](http://jain/cse567-11/ftp/imgsens/index.html), 2011.
- [94] Zhuo Hui, Aswin C Sankaranarayanan, Kalyan Sunkavalli, and Sunil Hadap. White balance under mixed illumination using flash photography. In *2016 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2016.
- [95] Ramesh C Tripathi and Linas A Sidrys. Conjunctival icterus, not scleral icterus. *JAMA*, 242(23):2558–2558, 1979.
- [96] John J Kuiper. Conjunctival icterus. *Annals of internal medicine*, 134(4):345–346, 2001.
- [97] Anthony S Fauci et al. *Harrison's principles of internal medicine*, volume 2. McGraw-Hill, Medical Publishing Division New York, 2008.
- [98] D Coffin. DCRAW application, 2016.
- [99] Stephen Westland. Computational colour science using MATLAB 2e, 2020. <https://www.mathworks.com/matlabcentral/fileexchange/40640->

computational-colour-science-using-matlab-2e/, MATLAB Central File Exchange (accessed: 2020.10.24).

- [100] International Commission on Non-Ionizing Radiation Protection et al. Light-Emitting Diodes (LEDS): Implications for safety. *Health Physics*, 118(5):549–561, 2020.
- [101] International Commission on Non-Ionizing Radiation Protection et al. IC-NIRP guidelines on limits of exposure to incoherent visible and infrared radiation. *Health Physics*, 105(1):74–96, 2013.
- [102] William T Ham, Harold A Mueller, and David H Sliney. Retinal sensitivity to damage from short wavelength light. *Nature*, 260(5547):153–155, 1976.
- [103] S Point. Blue light hazard: are exposure limit values protective enough for newborn infants. *Radioprotection*, 53(3):219–224, 2018.
- [104] Helen Mactier, Sanjay Maroo, Michael Bradnam, and Ruth Hamilton. Ocular biometry in preterm infants: implications for estimation of retinal illuminance. *Investigative ophthalmology & visual science*, 49(1):453–457, 2008.
- [105] Luis J Haddock, David Y Kim, and Shizuo Mukai. Simple, inexpensive technique for high-quality smartphone fundus photography in human and animal eyes. *Journal of ophthalmology*, 2013, 2013.
- [106] Shan-Jiun Lin, Chung-May Yang, Po-Ting Yeh, and Tzyy-Chang Ho. Smartphone fundoscopy for retinopathy of prematurity. *Taiwan Journal of Ophthalmology*, 4(2):82–85, 2014.
- [107] Andrew Bastawrous. Smartphone fundoscopy. *Ophthalmology*, 119(2):432–433, 2012.
- [108] Anubhav Goyal, Mahesh Gopalakrishnan, Giridhar Anantharaman, Dhileesh P Chandrashekhara, Thomas Thachil, and Ashish Sharma. Smartphone guided wide-field imaging for retinopathy of prematurity in

- neonatal intensive care unit—a Smart ROP (SROP) initiative. *Indian journal of ophthalmology*, 67(6):840, 2019.
- [109] David Y Kim, François Delori, and Shizuo Mukai. Smartphone photography safety. *Ophthalmology*, 119(10):2200–2201, 2012.
- [110] Sheng Chiong Hong, Giles Wynn-Williams, and Graham Wilson. Safety of iPhone retinal photography. *Journal of medical engineering & technology*, 41(3):165–169, 2017.
- [111] Ophthalmic instruments – Fundamental requirements and test methods – Part 2: Light hazard protection. ISO 15004-2:2007, International Organization for Standardization, Geneva, CH, February 2007.
- [112] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. *ACM transactions on graphics (TOG)*, 23(3):664–672, 2004.
- [113] Gaurav Nagar, Ben Vandermeer, Sandy Campbell, and Manoj Kumar. Effect of phototherapy on the reliability of transcutaneous bilirubin devices in term and near-term infants: a systematic review and meta-analysis. *Neonatology*, 109(3):203–212, 2016.
- [114] L Casnocha Lucanova, K Matasova, M Zibolen, and P Krcho. Accuracy of transcutaneous bilirubin measurement in newborns after phototherapy. *Journal of Perinatology*, 36(10):858–861, 2016.
- [115] William D Engle, Gregory L Jackson, Dorothy Sendelbach, Denise Manning, and William H Frawley. Assessment of a transcutaneous device in the evaluation of neonatal hyperbilirubinemia in a primarily Hispanic population. *Pediatrics*, 110(1):61–67, 2002.
- [116] William D Engle, Gregory L Jackson, Elizabeth K Stehel, Dorothy M Sendelbach, and M Denise Manning. Evaluation of a transcutaneous jaundice meter

- following hospital discharge in term and near-term neonates. *Journal of perinatology*, 25(7):486–490, 2005.
- [117] James A Taylor, Anthony E Burgos, Valerie Flaherman, Esther K Chung, Elizabeth A Simpson, Neera K Goyal, Isabelle Von Kohorn, Nui Dhepyasuwan, et al. Discrepancies between transcutaneous and serum bilirubin measurements. *Pediatrics*, 135(2):224–231, 2015.
- [118] Medicines and Healthcare Products Regulatory Agency. JM103 and JM105 jaundice meters – risk of misinterpretation of measurement in hyperbilirubinemia cases. *Medical Device Alerts*, 2018. <https://www.gov.uk/drug-device-alerts/jm103-and-jm105-jaundice-meters-risk-of-misinterpretation-of-measurement-in-hyperbilirubinemia-cases/> (accessed: 2020.11.03).
- [119] Pal Naevelid Saevik. Curly brace annotation, 2020. <https://www.mathworks.com/matlabcentral/fileexchange/38716-curly-brace-annotation/>, MATLAB Central File Exchange (accessed: 2020.10.24).
- [120] Mario A Ruiz, Sammy Saab, and Leland S Rickman. The clinical detection of scleral icterus: observations of multiple examiners. *Military medicine*, 162(8):560–563, 1997.
- [121] Abeer Azzuqa and Jon F Watchko. Bilirubin concentrations in jaundiced neonates with conjunctival icterus. *The Journal of pediatrics*, 167(4):840–844, 2015.
- [122] Xiao-Lu Ma, Zheng Chen, Jia-Jun Zhu, Xiao-Xia Shen, Ming-Yuan Wu, Li-Ping Shi, Li-Zhong Du, Jun-Fen Fu, and Qiang Shu. Management strategies of neonatal jaundice during the coronavirus disease 2019 outbreak. *World Journal of Pediatrics*, pages 1–4, 2020.