JAMA Network Open

Original Investigation | Pediatrics

Development and Validation of a Smartphone Application for Neonatal Jaundice Screening

Alvin Jia Hao Ngeow, MMed; Aminath Shiwaza Moosa, MMed; Mary Grace Tan, MD; Lin Zou, MSc; Millie Ming Rong Goh, MITB; Gek Hsiang Lim, MSc; Vina Tagamolila, MD; Imelda Ereno, MD; Jared Ryan Durnford, MMed; Samson Kei Him Cheung, MMed; Nicholas Wei Jie Hong, MMed; Ser Yee Soh, MMed; Yih Yann Tay, MSc; Zi Ying Chang, MMed; Ruiheng Ong, MMed; Li Ping Marianne Tsang, MMed; Benny K. L. Yip, BSocSc; Kuok Wei Chia, BSc; Kelvin Yap, BEng; Ming Hwee Lim, MBBS; Andy Wee An Ta, BSc; Han Leong Goh, PhD; Cheo Lian Yeo, MMed; Daisy Kwai Lin Chan, MMed; Ngiap Chuan Tan, MMed; for the BiliSG Study Group

Abstract

IMPORTANCE This diagnostic study describes the merger of domain knowledge (Kramer principle of dermal advancement of icterus) with current machine learning (ML) techniques to create a novel tool for screening of neonatal jaundice (NNJ), which affects 60% of term and 80% of preterm infants.

OBJECTIVE This study aimed to develop and validate a smartphone-based ML app to predict bilirubin (SpB) levels in multiethnic neonates using skin color analysis.

DESIGN, SETTING, AND PARTICIPANTS This diagnostic study was conducted between June 2022 and June 2024 at a tertiary hospital and 4 primary-care clinics in Singapore with a consecutive sample of neonates born at 35 or more weeks' gestation and within 21 days of birth.

EXPOSURE The smartphone-based ML app captured skin images via the central aperture of a standardized color calibration sticker card from multiple regions of interest arranged in a cephalocaudal fashion, following the Kramer principle of dermal advancement of icterus. The ML model underwent iterative development and *k*-folds cross-validation, with performance assessed based on root mean squared error, Pearson correlation, and agreement with total serum bilirubin (TSB). The final ML model underwent temporal validation.

MAIN OUTCOMES AND MEASURES Linear correlation and statistical agreement between paired SpB and TSB; sensitivity and specificity for detection of TSB equal to or greater than 17mg/dL with SpB equal to or greater than 13 mg/dL were assessed.

RESULTS The smartphone-based ML app was validated on 546 neonates (median [IQR] gestational age, 38.0 [35.0-41.0] weeks; 286 [52.4%] male; 315 [57.7%] Chinese, 35 [6.4%] Indian, 169 [31.0%] Malay, and 27 [4.9%] other ethnicities). Iterative development and cross-validation was performed on 352 neonates. The final ML model (ensembled gradient boosted trees) incorporated yellowness indicators from the forehead, sternum, and abdomen. Temporal validation on 194 neonates yielded a Pearson *r* of 0.84 (95% CI, 0.79-0.88; $P < .001$), 82% of data pairs within clinically acceptable limits of 3 mg/dL, sensitivity of 100%, specificity of 70%, positive predictive value of 10%, negative predictive value of 100%, positive likelihood ratio of 3.3, negative likelihood ratio of 0, and area under the receiver operating characteristic curve of 0.89 (95% CI, 0.82-0.96).

CONCLUSIONS AND RELEVANCE In this diagnostic study of a new smartphone-based ML app, there was good correlation and statistical agreement with TSB with sensitivity of 100%. The screening tool has the potential to be an NNJ screening tool, with treatment decisions based on TSB

(continued)

Key Points

Question Can the merging of domain knowledge (Kramer principle of dermal advancement of icterus) with current machine learning (ML) techniques create a unique smartphone-based screening tool for neonatal hyperbilirubinemia?

Findings In this diagnostic study of 546 neonates, the smartphone-based ML application, which used an ML model that incorporated yellowness indicators from the forehead, sternum, and abdomen, underwent internal-external validation against total serum bilirubin (TSB). Pearson *r* was 0.84, sensitivity was 100%, specificity was 70%, and area under the receiver operating characteristic curve was 0.89.

Meaning These findings suggest the screening tool has good correlation and statistical agreement with TSB, as well as excellent sensitivity.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Abstract (continued)

(reference standard). Further prospective studies are needed to establish the generalizability and cost-effectiveness of the screening tool in the clinical setting.

JAMA Network Open. 2024;7(12):e2450260. doi:10.1001/jamanetworkopen.2024.50260

Introduction

Neonatal jaundice (NNJ) affects 60% of term and 80% of preterm infants,[1-7] with early detection critical to prevent bilirubin encephalopathy. The criterion standard diagnostic method, total serum bilirubin (TSB) measurement,[8] is invasive and costly. Noninvasive methods like inspection[7,9,10] and icterometry[11-14] have shown variable accuracy.

Transcutaneous bilirubinometry (TcB) is a rapid, noninvasive tool for screening bilirubin levels in health care settings, using optical spectroscopy to measure bilirubin through the skin.[15,16] A Cochrane review[17] of 23 studies with 5058 participants indicated that TcB was generally effective, with varying accuracy due to different study conditions. In Singapore, TcB is the standard method for screening neonates with NNJ in hospitals and outpatient settings. However, its use is limited to health care facilities, requiring calibration, maintenance, and trained personnel. This can lead to increased health care costs, as well as travel expenses and inconvenience for families.

Recently, smartphone applications (apps) have emerged as alternatives for estimating bilirubin levels using digital images of the skin. However, existing apps like Biliscan[18-21] have had inconsistent performance across different regions. To the authors' knowledge, no existing smartphone-based NNJ apps[19-28] concurrently use digital images from multiple regions of interest arranged in a cephalocaudal fashion, specifically the forehead, sternum, and abdomen, as predictors for a single bilirubin estimate, which is grounded in the Kramer principle[29] of cephalocaudal advancement of dermal icterus that underpins bedside NNJ evaluation.

The study team aimed to develop and validate a new artificial intelligence (AI)–based smartphone app for estimating bilirubin levels in a multiethnic neonatal population using skin and/or scleral color images. The study also assessed the association between skin tone, as classified by the Fitzpatrick[30] skin phototype, and the accuracy of the app. The study team hypothesized that concurrently incorporating skin yellowness from multiple regions of interest, specifically the forehead, sternum, and abdomen, as predictors into the app's machine learning (ML) model and acquiring a training dataset from a multiethnic neonatal population would enhance its performance.

Methods

This diagnostic study adhered to the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guideline[31] and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD)+AI[32] updated guidance for reporting clinical prediction models. A dual-phase prospective study from June 2022 to June 2024 was conducted to develop and validate the BiliSG app. We recruited a consecutive sample that included term and late preterm Asian neonates born at 35 or more weeks' gestation aged within 21 days who were clinically stable regardless of ethnicity. Neonates were excluded if they had skin lesions that interfered with image acquisition or were undergoing phototherapy. Ethical approval for the study was granted by the SingHealth institutional review board.

The study took place at Singapore General Hospital (SGH) and 4 SingHealth Polyclinics located at Bedok, Bukit Merah, Punggol, and Sengkang in Singapore. Parents provided informed consent, and data collected included demographic (gestational age, birthweight, sex, electronic health record–reported ethnicity [Chinese, Indian, Malay, and other ethnicities]), and clinical (cephalohematoma, ABO and Rhesus blood group,[33,34] glucose-6-phosphate dehydrogenase

deficiency status, and phototherapy status) information from electronic health records, as well as skin tone assessed using the Fitzpatrick[30] scale by study members. Type of feeding (breastmilk, formula, or mixed) and qualitative end-user acceptability information were self-reported. ABO incompatibility in our study was defined as group O mothers with non–group O newborns, group A mothers with group B or AB newborns, and group B mothers with group A or AB newborns.

The study was divided into 2 phases: phase 1 (June 2022 to October 2023) focused on app development and initial ML model creation, while phase 2 (November 2023-June 2024) concentrated on validating the model. There was no public involvement in the study design.

Methods of NNJ Assessment

Eligible neonates underwent several bilirubin measurements or estimations within the same hour using smartphone-predicted bilirubin (SpB), TcB, and TSB methods. TcB was measured with the Dräger JM-105 bilirubinometer (Dräger Medical GmbH),[35] which was calibrated on a daily basis. Three measurements from the sternal area were taken, and the mean was recorded.

TSB was measured using capillary heel prick samples analyzed on the Unistat analyzer through direct spectrophotometry[36] in all study centers. Calibration of the Unistat analyzers occurred every 6 months, along with standardized maintenance at SingHealth Polyclinics and SGH. Quality control was conducted twice daily as per manufacturer standards, with trends tracked using Levy-Jennings plots and Westgard rules. All analyzers took part in proficiency testing to ensure measurement accuracy through peer group and accuracy-based surveys. For neonates who underwent multiple SpB-TSB tests, all results contributed to model development and cross-validation, while only the initial measurement was used for temporal validation.

SpB estimates were obtained using the smartphone app on an Apple iPhone 12 model (Apple Inc) by placing a color calibration sticker card with a central aperture on specific areas (forehead, sternum, and abdomen) or near the eye (eFigure 1 in Supplement 1). Forehead, sternal, and abdominal images were selected based on the Kramer principle,[29] which describes the cephalocaudal progression of dermal icterus with increasing hyperbilirubinemia. Initially, images were taken from zones 1 to 3, with plans to include limb images (zones 4 and 5) in future studies. Scleral imaging was optional as a backup for participants with darker skin tones,[37] which may affect accuracy. The app captured and analyzed images of these regions of interest (eFigure 1 in Supplement 1). High-quality images were captured in ambient light and screened for artifacts, with those covering over 35% of the regions of interest excluded. A small group of trained study members (A.J.H.N., M.G.T., V.T., I.E., J.R.D., S.K.H.C., N.W.J.H., and S.Y.S.) performed SpB measurements to minimize variability and were blinded to the results. SpB and TSB measurements were taken within an hour of each other. Clinical management was guided by TcB and TSB measurements as per local clinical practice guidelines.[38]

Development of the Color Calibration Sticker Card

The color calibration sticker card underwent several iterations to create a final version with a central aperture for skin color assessment and surrounding colored squares. It was designed to correct for light intensity and temperature variations and was printed on matte paper to minimize reflections. The sticker design (eFigure 2 in Supplement 1) also reduced shadows that could affect image accuracy. In the app, the card and skin patches were segmented for analysis, enabling precise color adjustments[39] based on the card's patches and improving measurement reliability.

Inclusion of Predictors in the ML Model

The initial selection of predictors for the ML model was informed by clinical knowledge, using yellowness-related features from the forehead, sternum, and abdomen based on the Kramer principle.[29] Key predictors included various color metrics (yellow channel in the cyan-magenta-yellow-key [CMYK] color model, blue chromaticity, jaundice eye color index,[40] and the B channel for yellow and blue components from the LAB color space) and the neonate's age, as TSB levels typically

peak around days 3 to 5.[41,42] Additional predictors included risk factors for jaundice, such as blood group incompatibility, glucose-6-phosphate dehydrogenase deficiency, preterm birth, cephalohematoma, and exclusive breastfeeding,[43] as well as factors affecting visual estimation accuracy like skin tone and ethnicity.[44] The final selection of predictors was automated during the training of the gradient boosted trees model, which sequentially built decision trees that learned from each other's errors, assessing feature importance based on their contribution to reducing the loss function.

Iterative Development and Internal-External Validation of the ML Model

During the iterative development and validation of the ML model for predicting bilirubin levels, various techniques—including support vector machine, random forest, extreme gradient boosting, light gradient-boosting machine, gradient boosted trees, and extra trees—were used for their ability to handle nonlinear relationships, their robustness against overfitting, and their ability to provide insights into feature importance. The model output was continuous. The model development employed a 5-fold *k*-folds cross-validation method,[45,46] dividing the dataset into 5 subgroups. Each fold served once as a test set while the others trained the model, ensuring no overlap between training and testing data for an unbiased evaluation. Missing data were addressed by using the median for continuous variables and the mode for categorical ones, with one-hot encoding for nominal and label encoding for ordinal variables. Feature engineering was performed to create variables like preterm and age above 2 weeks. Data were scaled using a standard scaler before model development. Model performance was assessed using metrics such as root mean squared error (RMSE),[47] Pearson correlation, and Bland-Altman plots. The final model underwent temporal validation on prospectively recruited neonates. The primary outcomes studied were the linear correlation and statistical agreement between paired SpB and TSB measurements.

Statistical Analysis

Descriptive analysis summarized the demographic and clinical data of neonates. Pearson correlation coefficient was used to assess the linear relationship between SpB values and TSB. A Bland-Altman plot evaluated the agreement between SpB and TSB within clinically acceptable limits of 50 µmol/L (approximately 3 mg/dL), as prior studies have demonstrated differences of between 2 to 3 mg/dL[24,48] between TcB, the widely accepted method of screening,[43] and TSB.

Subgroup analysis by skin tone and sensitivity analysis on poor-quality images was conducted to explore accuracy factors. Sensitivity, specificity, likelihood ratios, and predictive values were calculated using a predefined decision rule (SpB ≥13 mg/dL to predict TSB ≥17 mg/dL) for comparison with prior TcB[49,50] and smartphone app validation studies.[51]

The smartphone-based ML app's utility was compared with TcB using receiver operator characteristic curves and the area under the curve (AUC) based on the same decision rule of TcB 13 mg/dL or greater to predict TSB 17 mg/dL or greater. All analyses were performed using Python version 3.9.17 (Python Software Foundation), following the prespecified study protocol.

The sample size was estimated based on clinically acceptable limits of agreement for TSB (±3 mg/dL). Using data from Aune et al,[23] a conservative mean difference of 0.5 mg/dL and a SD of 1.3 mg/dL indicated that 463 paired measurements were needed to detect agreement based on a 95% CI for the limits of agreement at 80% power and maximum allowable difference of 3 mg/dL. Accounting for a 15% loss of information due to image quality or participant withdrawals, 545 babies would need to be recruited, contributing to 545 pairs of measurements for model development (70%) and prospective validation (30%). A 2-sided *P* value less than .001 was considered significant.

Results

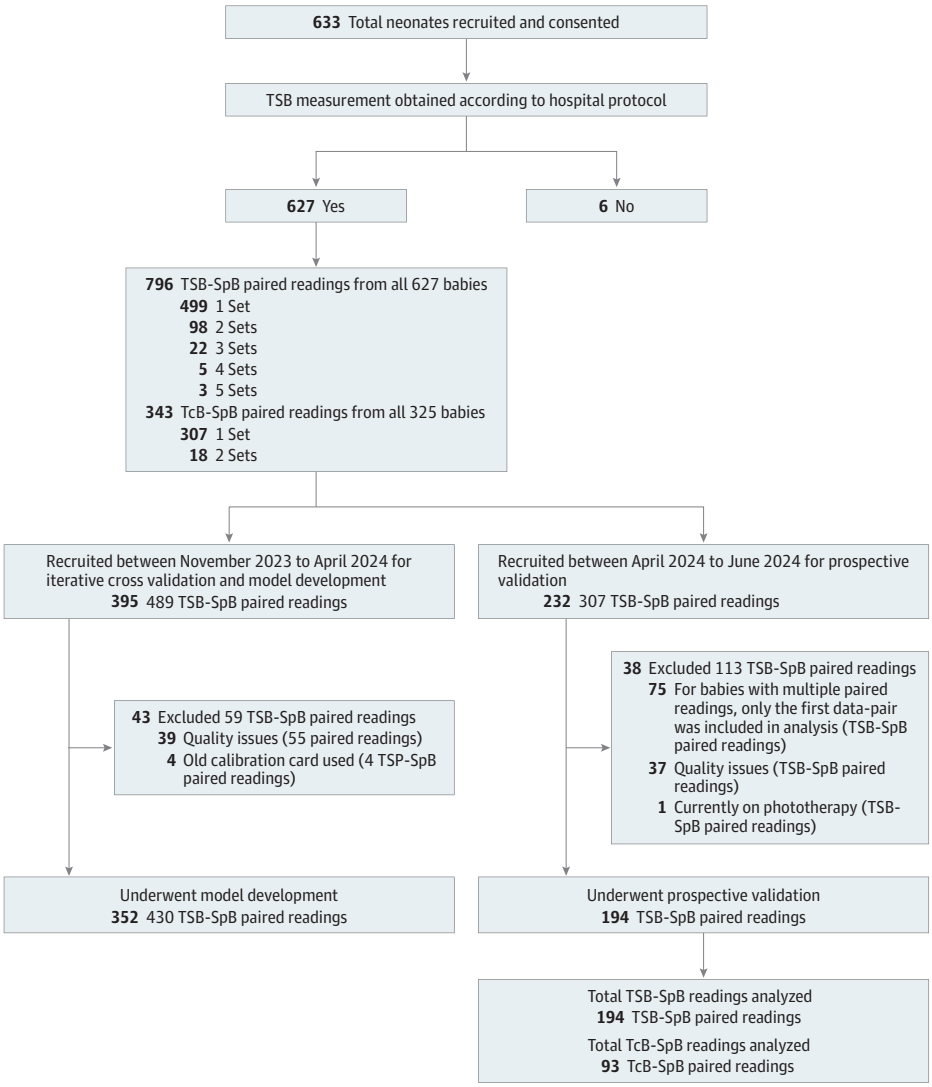
Between November 2023 and June 2024, 633 unique neonates were recruited at SGH and 4 primary care clinics (**Figure 1**). A total of 627 neonates had paired TSB and SpB readings for analysis, with 499

(79.6%) having 1 reading, 98 (15.6%) 2 readings, 22 (3.5%) 3 readings, 5 (0.8%) 4 readings, and 3 (0.5%) 5 readings.

During the iterative development and cross-validation phase, 395 neonates were recruited, but 43 were excluded for poor image quality, leaving 352 for cross-validation. After finalizing the ML model in mid-April 2024, 232 additional neonates were recruited, with 37 excluded for image quality and 1 for phototherapy, resulting in 194 for temporal validation.

Demographic data (**Table**) showed an ethnic distribution similar to the national population (315 [57.7%] Chinese, 35 [6.4%] Indian, 169 [31.0%] Malay, and 27 [4.9%] other ethnicities), with a median (IQR) gestational age of 38.0 (35.0–41.0) weeks and birth weight of 3045 (1975-4514) grams and 285 male neonates (52.4%). Glucose-6-phosphate deficiency was present in 15 neonates (2.7%), and ABO group incompatibility was documented in 149 (27.7%). TSB level distribution (**Figure 2**) showed comparable patterns between the prospective validation dataset and the training set, particularly for TSB values 10 mg/dL or less and 14 mg/dL or greater, despite a slight leftward skew in the validation dataset.

Figure 1. Consort Diagram of Patient Recruitment



SpB indicates screening tool bilirubin; TcB, transcutaneous bilirubinometry; TSB, total serum bilirubin.

Predictors

In the final ML model, yellowness-related predictors from the forehead, sternal, and abdominal regions were among the top predictors. The discernible yellowness gradient between the sternal and abdominal regions decreased at higher TSB levels, particularly higher than 12 mg/dL (eFigure 3 in

Table. Demographic and Clinical Characteristics

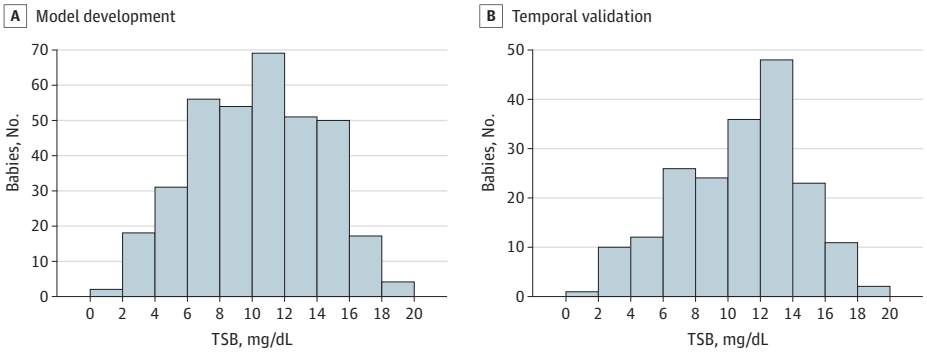
Characteristic	Patients, No. (%)		
	All (N = 546)	In model development (n = 352)	In prospective validation (n = 194)
Gestational age, median (range), wk	38.0 (35.0-41.0)	38.0 (35.0-41.0)	38.0 (35.0-41.0)
Full term	514 (94.1)	326 (92.6)	188 (96.9)
Preterm	32 (5.9)	26 (7.4)	6 (3.1)
Hour of life, median (range), h	120.0 (24.0-504.0)	120.0 (24.0-504.0)	118.5 (24.0-504.0)
Birthweight, median (range), g	3044.5 (1975.0-4514.0)	3046.5 (1975.0-4284.0)	3032.5 (2094.0-4514.0)
Sex			
Female	260 (47.6)	168 (47.7)	92 (47.4)
Male	286 (52.4)	184 (52.3)	102 (52.6)
Ethnicity			
Chinese	315 (57.7)	194 (55.1)	121 (62.4)
Indian	35 (6.4)	22 (6.3)	13 (6.7)
Malay	169 (31.0)	118 (33.5)	51 (26.3)
Others[a]	27 (4.9)	18 (5.1)	9 (4.6)
TSB, mean (SD), mg/dL	10.51 (3.83)	10.34 (3.84)	10.82 (3.80)
Range of TSB, mg/dL	0.88-20.46	1.64-19.94	0.88-20.46
Skin tone type			
I	131 (24.0)	41 (11.6)	90 (46.4)
II	340 (62.3)	241 (68.5)	99 (51.0)
III	68 (12.5)	63 (17.9)	5 (2.6)
IV	7 (1.3)	7 (2.0)	0
Mode of delivery			
Spontaneous	252 (46.2)	151 (42.9)	101 (52.1)
Cesarean	187 (34.2)	112 (31.8)	75 (38.7)
Operative vaginal	107 (19.6)	89 (25.3)	18 (9.3)
Presence of cephalohematoma	11 (2.0)	3 (0.9)	8 (4.1)
ABO incompatibility	149/538 (27.7)	101345 (29.3)	48/193 (24.9)
Glucose-6-phosphate dehydrogenase deficiency	15 (2.7)	11 (3.1)	4 (2.1)
RH incompatibility	4 (0.7)	4 (1.1)	0
Scleral images obtained	10 (1.8)	7 (2.0)	3 (1.5)

Abbreviation: TSB, total serum bilirubin.

SI conversion factor: To convert bilirubin to µmol/L, multiply values by 17.1.

[a] No subcategories were collected under Other.

Figure 2. Distribution of Total Serum Bilirubin (TSB) Levels



SI conversion factors: To convert bilirubin to µmol/L, multiply values by 17.1.

Supplement 1), while no consistent gradient was observed between the forehead and other regions. Scleral images were excluded from analysis due to successful capture in only 10 participants.

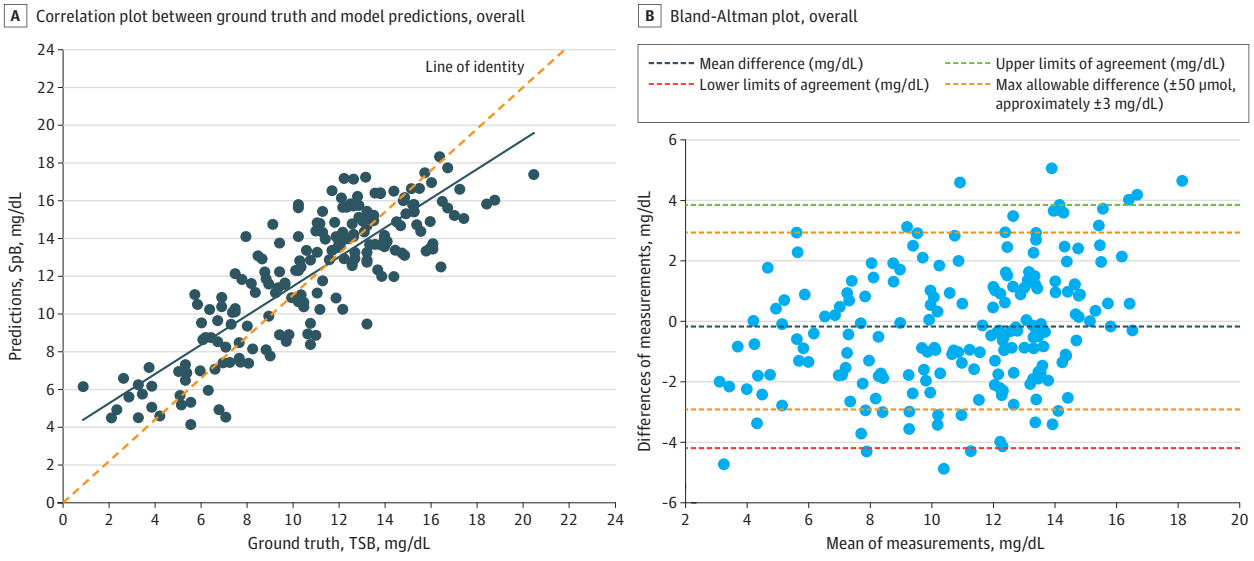
To improve interpretability of the model, Shapley additive explainability tools (refer to eMethods and eFigure 4 in Supplement 1) were employed, revealing that the neonate's hour of life was consistently 1 of the top predictors. This aligns with clinical knowledge that TSB levels typically rise in the first few days of life.[41,42] Other significant features included yellowness-related attributes from skin images, such as the CMYK yellow channel, the LAB B channel, and the jaundice eye color index value, indicating the model's reliance on the degree of yellowness in the images for predictions.

Cross-Validation and Temporal Validation of the Final ML Model

The gradient boosted trees model was selected as the final algorithm due to its consistent performance, achieving an RMSE of 2.41 mg/dL and a Pearson correlation of 0.77 ($P < .001$) between SpB and TSB, with 76% of data pairs within a clinically acceptable difference of 50 µmol/L (approximately 3 mg/dL). After finalizing the ML model, 194 individuals were prospectively recruited for temporal validation, which revealed a strong correlation between SpB and TSB (refer to eTable 1 in Supplement 1; **Figure 3**A), with a Pearson coefficient of 0.84 (95% CI, 0.79-0.88; $P < .001$). For ethnic groups, the coefficients were 0.86 (95% CI, 0.80-0.90; $P < .001$) for Chinese, 0.91 (95% CI, 0.73-0.97; $P < .001$) for Indian, and 0.81 (95% CI, 0.69-0.89; $P < .001$) for Malay neonates. For Fitzpatrick skin phototype II and III, coefficients were 0.85 (95% CI, 0.79-0.90; $P < .001$) and 0.92 (95% CI, 0.20-0.99; $P = .03$), respectively. A sensitivity analysis of 231 babies (194 without artifacts and 37 with significant artifacts affecting >35% of regions of interest) showed a Pearson coefficient of 0.81 (95% CI, 0.76-0.85; $P < .001$).

The Bland-Altman plot (Figure 3B) indicated that 82% of paired measurements were within the maximum acceptable difference, with SpB readings slightly lower than TSB, with a mean difference of −0.18 mg/dL (95% limits of agreement [LoA], −4.20 to 3.84 mg/dL). The mean differences among the largest ethnic groups were similar: −0.21 (95% LoA, −4.01 to 3.59 mg/dL) for Chinese, −0.30 (95% LoA, −3.55 to 2.95 mg/dL) for Indian, and −0.46 (95% LoA, −4.72 to 3.79 mg/dL) for Malay. The final model's RMSE was 2.06 mg/dL.

Figure 3. Pearson Correlation and Bland-Altman Plot Between Smartphone-Predicted Bilirubin (SpB) and Total Serum Bilirubin (TSB) (n = 194)



A, Each dot represents a paired observation between the reference standard TSB and the SpB for individual measurements. The scatter of the dots shows the relationship between the actual and predicted values. The solid line represents the line of best fit through the data points, showing the trend in the relationship between reference standard TSB and SpB. B, Each dot represents the difference between the SpB and reference standard TSB levels for each observation, plotted against the mean of the SpB and reference standard TSB values for that observation.

Calibration of ML Model

To assess the calibration of the ML model across the TSB range, the prospective validation set (TSB levels from 0.88 mg/dL to 20.46 mg/dL) was divided into 3 equidistant groups: (1) TSB less than 6.5 mg/dL, (2) TSB between 6.5 mg/dL and 13.0 mg/dL, and (3) TSB greater than 13.0 mg/dL. In each group, 83% (24 of 29), 85% (88 of 104), and 77% (47 of 61) of cases, respectively, fell within the clinically acceptable range of ±3 mg/dL, suggesting an alignment of predictions and underscoring the model's reliability.

Diagnostic Accuracy of SpB and TcB Measurements

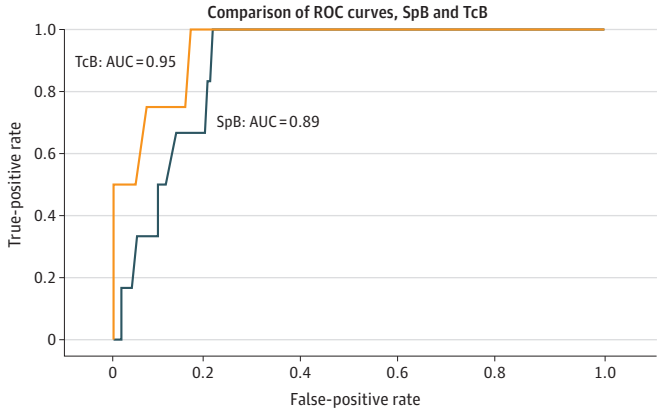
Using TSB as the criterion standard, the sensitivity and specificity of SpB were 100% (95% CI, 100%-100%) and 70% (95% CI, 63%-76%), respectively, while for TcB, they were 100% (95% CI, 100%-100%) and 51% (95% CI, 40%-61%) (eTable 2 and eTable 3 in Supplement 1). The positive likelihood ratios were 3.30 for SpB and 2.02 for TcB, with both having a negative likelihood ratio of 0.0. The positive predictive value for SpB was 10% (95% CI, 2%-17%), and for TcB, it was 8%(95% CI, 1%-16%), while both had a negative predictive value of 100% (95% CI, 100%-100%). The areas under the receiver operating characteristic curve were 0.89 (95% CI, 0.82-0.96) for SpB and 0.95 (95% CI, 0.87-1.00) for TcB, suggesting diagnostic accuracy for both measurements (**Figure 4**).

Discussion

The smartphone-based ML app, developed and validated in the Singapore population, demonstrated a strong correlation (Pearson coefficient of 0.84) between SpB and TSB. It achieved 100% sensitivity and 70% specificity based on the decision rule of SpB 13 mg/dL or greater to predict TSB 17mgl/dL or greater. Its diagnostic accuracy was comparable with TcB.[17] Screening tool–derived SpB demonstrated a stronger correlation with TSB compared with the pooled correlation coefficient of 0.77 reported by Hedge et al,[52] which analyzed data from 1733 neonates across 10 studies on smartphone jaundice screening apps that assessed skin yellowness.

The smartphone-based ML app is an innovative NNJ screening tool that has demonstrated strong sensitivity and moderate specificity across a multiethnic population, improving accuracy by using multiple skin regions and color calibration stickers. It can potentially allow remote screening of NNJ by means of home-based SpB measurements and teleconsultations with health care professionals. This approach could reduce the need for frequent clinic visits, especially during epidemics, and has been positively received by parents for its convenience.[58,59] Future plans involve validating the app's effectiveness across different smartphone models, different camera specifications, varying lighting conditions, and use by different end-users, as well as exploring various

Figure 4. Receiver Operating Characteristic (ROC) Curves of Screening Tool Bilirubin (SpB) and Transcutaneous Bilirubin (TcB)



AUC indicates area under the curve.

SpB thresholds for guiding the need for TSB assay. A pilot study will assess the app's safety and acceptability, while a parallel health economics study will evaluate its cost-effectiveness and environmental impact. Ultimately, the goal is to integrate the screening tool with teleconsultation services in primary care clinics in Singapore, establishing a decentralized care model for NNJ screening.

Strengths and Limitations

The screening tool integrates clinical knowledge, specifically the Kramer principle of cephalocaudal advancement of dermal icterus, with ML by analyzing images from multiple regions, including the forehead, sternum, and abdomen. This approach sets it apart from other smartphone apps that rely on a single region of interest for bilirubin estimation, which can lead to inconsistent results. For instance, using only the sternal area may overlook more severe jaundice that has progressed to the abdomen, affecting the accuracy of predicted bilirubin levels. The study found no consistent yellowness gradient between the forehead and sternal/abdominal regions. This aligns with previous research indicating that bilirubin measurements from the forehead, an exposed area,[53,54] tend to be underestimated. The authors suggest that natural phototherapy may contribute to this underestimation.

Second, the design of the color calibration sticker card enhanced image quality by eliminating shadows that can affect accuracy of bilirubin estimates. Sensitivity analyses suggested that these shadow artifacts had a significant negative association with correlation, underscoring the importance of this design feature for accuracy.

Third, the large validation cohort of 546 neonates minimized the risk of overfitting and bias, thereby enhancing accuracy.[55] The app was trained on a diverse dataset which included neonates of different skin tone who were recruited from various ethnic groups in Singapore. The Fitzpatrick phototype scale[30] provided a more objective classification of skin tone than reliance on ethnic grouping which is often used as a proxy to describe skin tone. This is crucial in a multiethnic country like Singapore.

Fourthly, the study conducted both internal (cross validation-TRIPOD Type 1b) and external temporal validation (TRIPOD Type 2b) of the ML model.[56] This approach ensured reproducibility and effectiveness of the model in real-world scenarios, providing an objective measure of performance compared with internal validation alone.

The study had several limitations that may affect the generalizability of the smartphone-based ML app. First, only a small number of neonates had Fitzpatrick phototype IV or higher, belonged to minority ethnic groups, or were undergoing phototherapy during recruitment, making it uncertain how well the app applies to these populations without further data from a larger sample.

Additionally, practical challenges in capturing scleral images meant that only a few participants were able to undergo this imaging. Many neonates had their eyelids closed due to drowsiness or sleep, which poses a significant hurdle for NNJ screening apps relying on scleral images, as newborns can sleep for up to 20 hours a day.[57] Future development may need to consider video-based imaging to address this issue.

The smartphone-based ML app was tested solely on the iPhone 12 due to local cybersecurity policies. Broader validation across various smartphone models, including those running on Android, is necessary to determine its accessibility on different devices.

The study focused on the first 3 zones in the Kramer principle as regions of interest, which are commonly used for TcB measurements.[53] Zones 4 and 5, which involve limb imaging, were not evaluated due to potential concerns about motion artifacts. Further research is needed to assess whether including these zones could enhance accuracy, particularly for bilirubin levels that approached 20 mg/dL.

While the smartphone-based ML app achieved a sensitivity of 100%, its specificity was only 70%, leading to potential false positives and unnecessary blood draws for TSB testing. This could result in additional costs and emotional stress for parents. Nevertheless, its specificity was still higher

than that of TcB, which had a specificity of 51%. As the app is intended to be used as a screening tool, type I errors—misclassifying negative cases as NNJ—are less critical than type II errors, which involve missing actual jaundice cases. Future studies should carefully select SpB thresholds to predict the respective TSB thresholds as per current age-based, risk-stratified local clinical practice guidelines,[38] emphasizing sensitivity over specificity, to minimize missed cases of severe NNJ.

SpB readings were slightly lower than TSB measurements, with a mean difference of −0.18 mg/dL. This underestimation could lead to missed cases of NNJ that require phototherapy. Further analysis is needed to determine optimal SpB thresholds for the respective TSB thresholds to reduce false negatives. The Bland-Altman plot indicated that while 82% of data pairs fell within a set limit of 3 mg/dL, 9% were over 3 mg/dL lower than the actual TSB. To minimize false negatives, a proposed SpB threshold should be more than 3 mg/dL below TSB levels. For example, using a SpB threshold of 13 mg/dL or greater to predict TSB 17 mg/dL or greater resulted in no false negatives and a sensitivity of 100%.

While 82% of patient data pairs fell within the clinically acceptable limits, 35 cases (18%) exceeded those limits, with half being overestimations and half underestimations. A majority (71%) of these patients had Fitzpatrick skin tone phototype I. The underrepresentation of this skin tone type in the training dataset may have contributed to these discrepancies during temporal validation. Future model refinement will focus on minority skin tones (Types I and IV) and racial groups (Indian and others) to enhance accuracy and inclusivity.

Additionally, although screening was performed by a small group of trained personnel to minimize interoperator variability, operator-specific data were not collected. Future studies will evaluate this variability to bolster confidence in the app's robustness and reliability in real-world scenarios, particularly for use by parents and caregivers who may be less experienced with such tools. Understanding how the app performs with various operators will help improve its usability and accuracy across diverse settings.

Conclusions

In this diagnostic study of a new smartphone-based ML app, there was good correlation and statistical agreement with TSB, with sensitivity of 100%. The app has the potential to be an NNJ screening tool, with treatment decisions based on TSB (the reference standard). Further prospective studies are needed to establish the generalizability, and cost-effectiveness of the screening tool in the real-world setting.

ARTICLE INFORMATION

Accepted for Publication: October 20, 2024.

Published: December 11, 2024. doi:10.1001/jamanetworkopen.2024.50260

Corresponding Author: Alvin Jia Hao Ngeow, MMed (Paediatric Medicine), Department of Neonatal and Developmental Medicine, Singapore General Hospital, 20 College Rd, c/o Singapore General Hospital, Academia, Singapore 169856 (alvin.ngeow.j.h@singhealth.com.sg).

Author Affiliations: Department of Neonatal and Developmental Medicine, Singapore General Hospital, Singapore (Ngeow, M. G. Tan, Tagamolila, Ereno, Durnford, Cheung, Hong, Soh, Yeo, Chan); SingHealth Polyclinics, Singapore (Moosa, Chang, Ong, Tsang, N. C. Tan); Nursing Division, Singapore General Hospital, Singapore (Tay); Synapxe (formerly Integrated Health Information Systems, IHiS), Singapore (Zou, M. M. R. Goh, Ta, H. L. Goh); Health Services Research Unit, Singapore General Hospital, Singapore (G. H. Lim); Yong Loo Lin School of Medicine, National University of Singapore, Singapore (Ngeow, Yeo, Chan); Paediatrics Academic Clinical Programme, Duke-NUS Medical School, Singapore (Ngeow, Durnford, Cheung, Hong, Soh, Yeo, Chan); Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore (Ngeow, Yeo, Chan); Family Medicine

Academic Clinical Programme, Duke-NUS Medical School, Singapore (Moosa, Chang, Ong, Tsang, N. C. Tan); Department of Future Health System, Singapore General Hospital, Singapore (Yip, Chia); Axrail Private Limited, Singapore (Yap); Department of Clinical Pathology, Singapore General Hospital, Singapore (M. H. Lim).

Author Contributions: Dr Ngeow had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Ngeow, Moosa, G. Lim, Tay, Chang, Ong, Yip, Chia, Yap, M. Lim, Ta, H. Goh, Yeo, N. Tan.

Acquisition, analysis, or interpretation of data: Ngeow, Moosa, M. Tan, Zou, M. Goh, G. Lim, Tagamolila, Ereno, Durnford, Cheung, Hong, Soh, Tay, Ong, Tsang, Yip, H. Goh, Yeo, Chan, N. Tan.

Drafting of the manuscript: Ngeow, Moosa, M. Tan, M. Goh, G. Lim, Tagamolila, Durnford, Ong, Yip, M. Lim, Yeo, Chan, N. Tan.

Critical review of the manuscript for important intellectual content: Ngeow, Moosa, Zou, M. Goh, G. Lim, Ereno, Cheung, Hong, Soh, Tay, Chang, Ong, Tsang, Chia, Yap, Ta, H. Goh, Chan, N. Tan.

Statistical analysis: Zou, M. Goh, Ta, H. Goh.

Obtained funding: Ngeow, Moosa.

Administrative, technical, or material support: Ngeow, M. Tan, Zou, Tagamolila, Ereno, Durnford, Cheung, Tay, Chang, Tsang, Yip, Chia, Yap, Ta, Yeo, N. Tan.

Supervision: Ngeow, Tay, Chang, Ta, H. Goh, Chan, N. Tan.

Conflict of Interest Disclosures: Drs Ngeow and N. Tan reported a patent pending filed with Intellectual Property Office of Singapore (IPOS) on June 13, 2024. No other disclosures were reported.

Funding/Support: The study is funded by Ministry of Health (Singapore) Health Innovation (MHI) Fund (grant No. MH 110:12/12-30).

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Group Information: The BiliSG Study Group members are listed in Supplement 2.

Data Sharing Statement: See Supplement 3.

Additional Contributions: The BiliSG Study Group thanks the senior management of SingHealth, including Kenneth Kwek, MMed (Chief Executive Officer, SGH), Ruban Poopalalingam, MMed (Chairman Medical Board, SGH) and Chian Min Loo, MRCP (Division of Medicine Chair, SGH) for their unwavering support throughout this project. We also thank the residents and senior residents, clinical research coordinators, physicians and nurses who assisted in patient recruitment. Above all, we thank parents and study participants for their participation.

REFERENCES

1. Sarici SÜ, Serdar MA, Korkmaz A, et al. Incidence, course, and prediction of hyperbilirubinemia in near-term and term newborns. *Pediatrics*. 2004;113(4):775-780. doi:10.1542/peds.113.4.775

2. Cohen SM. Jaundice in the full-term newborn. *Pediatr Nurs*. 2006;32(3):202-208.

3. Ng MCW, How CH. When babies turn yellow. *Singapore Med J*. 2015;56(11):599-602. doi:10.11622/smedj.2015167

4. Newman TB, Xiong B, Gonzales VM, Escobar GJ. Prediction and prevention of extreme neonatal hyperbilirubinemia in a mature health maintenance organization. *Arch Pediatr Adolesc Med*. 2000;154(11):1140-1147. doi:10.1001/archpedi.154.11.1140

5. Mukherjee D, Coffey M, Maisels MJ. Frequency and duration of phototherapy in preterm infants <35 weeks gestation. *J Perinatol*. 2018;38(9):1246-1251. doi:10.1038/s41372-018-0153-4

6. Bhutani VK, Zipursky A, Blencowe H, et al Neonatal hyperbilirubinemia and Rhesus disease of the newborn: incidence and impairment estimates for 2010 at regional and global levels. *Pediatr Res*. 2013;74(Suppl 1):86-100. doi:10.1038/pr.2013.208

7. Keren R, Tremont K, Luan X, Cnaan A. Visual assessment of jaundice in term and late preterm infants. *Arch Dis Child Fetal Neonatal Ed*. 2009;94(5):F317-F322. doi:10.1136/adc.2008.150714

8. Hulzebos CV, Camara JE, van Berkel M, et al; IFCC Working Group Neonatal Bilirubin. Bilirubin measurements in neonates: uniform neonatal treatment can only be achieved by improved standardization. *Clin Chem Lab Med*. 2024;62(10):1892-1903. doi:10.1515/cclm-2024-0620

9. Bredemeyer SL, Polverino JM. Assessment of jaundice in the term infant–a clinical challenge: part I. *Neonatal Paediatr Child Health Nurs*. 2006;9(3):15-20.

10. Moyer VA, Ahn C, Sneed S. Accuracy of clinical judgment in neonatal jaundice. *Arch Pediatr Adolesc Med*. 2000;154(4):391-394. doi:10.1001/archpedi.154.4.391

11. Gupta PC, Kumari S, Mullick DN, Lal UB. Icterometer: a useful screening tool for neonatal jaundice. *Indian Pediatr*. 1991;28(5):473-476.

12. Hulzebos CV, Vitek L, Coda Zabetta CD, et al. Screening methods for neonatal hyperbilirubinemia: benefits, limitations, requirements, and novel developments. *Pediatr Res*. 2021;90(2):272-276. doi:10.1038/s41390-021-01543-1

13. Hamel BC. Usefulness of icterometer in black newborns with jaundice. *Trop Doct*. 1982;12(4 Pt 2):213-214. doi:10.1177/004947558201200429

14. Luu MN, Le LT, Tran BH, et al. Home-use icterometry in neonatal hyperbilirubinaemia: cluster-randomised controlled trial in Vietnam. *J Paediatr Child Health*. 2014;50(9):674-679. doi:10.1111/jpc.12611

15. Cheng NY, Lin YL, Fang MC, Lu WH, Yang CC, Tseng SH. Noninvasive transcutaneous bilirubin assessment of neonates with hyperbilirubinemia using a photon diffusion theory-based method. *Biomed Opt Express*. 2019;10(6):2969-2984. doi:10.1364/BOE.10.002969

16. Paul HA, Adams BJ, Venner AA. Improving quality of transcutaneous bilirubin measurements: value of in-house developed quality control. *Pract Lab Med*. 2021;24:e00206. doi:10.1016/j.plabm.2021.e00206

17. Okwundu CI, Olowoyeye A, Uthman OA, et al Transcutaneous bilirubinometry versus total serum bilirubin measurement for newborns. *Cochrane Database Syst Rev*. 2023;5(5):CD012660. doi:10.1002/14651858.CD012660.pub2

18. Ngeow AJH, Tan MG, Dong X, et al. Validation of a smartphone-based screening tool (Biliscan) for neonatal jaundice in a multi-ethnic neonatal population. *J Paediatr Child Health*. 2023;59(2):288-297. doi:10.1111/jpc.16287

19. Huang D, Yang B, Gao X, et al. Influences on accuracy of automated image-based estimation of neonatal serum bilirubin level using smartphone application under different circumstances. *Chin J Perinat Med*. 2019;12:269-277.

20. Rong ZH, Luo F, Ma LY, et al. Evaluation of an automatic image-based screening technique for neonatal hyperbilirubinemia. Article in Chinese. *Zhonghua Er Ke Za Zhi*. 2016;54(8):597-600.

21. Lingaldinna S, Konda KC, Bapanpally N, Alimelu M, Singh H, Ramaraju M. Validity of bilirubin measured by biliscan (smartphone application) in neonatal jaundice–an observational study. *J Nepal Paediatr Soc*. 2021;41(1):93-98. doi:10.3126/jnps.v40i3.29412

22. De Greef L, Goel M, Seo MJ, et al. Bilicam: using mobile phones to monitor newborn jaundice. Paper presented at: UBICOMP; September 13-17, 2014; Seattle, WA. Accessed October 30, 2024. https://ubicomplab.cs.washington.edu/pdfs/bilicam.pdf

23. Aune A, Vartdal G, Bergseng H, Randeberg LL, Darj E. Bilirubin estimates from smartphone images of newborn infants' skin correlated highly to serum bilirubin levels. *Acta Paediatr*. 2020;109(12):2532-2538. doi:10.1111/apa.15287

24. Taylor JA, Burgos AE, Flaherman V, et al; Better Outcomes through Research for Newborns Network. Discrepancies between transcutaneous and serum bilirubin measurements. *Pediatrics*. 2015;135(2):224-231. doi:10.1542/peds.2014-1919

25. Aydın M, Hardalaç F, Ural B, Karap S. Neonatal jaundice detection system. *J Med Syst*. 2016;40(7):166-166. doi:10.1007/s10916-016-0523-4

26. Mansor MN, Hariharan M, Basah SN, Yaacob S. New newborn jaundice monitoring scheme based on combination of pre-processing and color detection method. *Neurocomputing Amst*. 2013;120:258-261. doi:10.1016/j.neucom.2012.10.034

27. Swarna S, Pasupathy S, Chinnasami B, Manasa D, Ramraj B. The smart phone study: assessing the reliability and accuracy of neonatal jaundice measurement using smart phone application. *Int J Contemp Pediatrics*. 2018;5(2):285-289. doi:10.18203/2349-3291.ijcp20175928

28. Munkholm SB, Krøgholt T, Ebbesen F, Szecsi PB, Kristensen SR. The smartphone camera as a potential method for transcutaneous bilirubin measurement. *PloS One*. 2018;13(6):e0197938. doi:10.1371/journal.pone.0197938

29. Kramer LI. Advancement of dermal icterus in the jaundiced newborn. *AJDC 1960*. 1969;118(3):454-458. doi:10.1001/archpedi.1969.02100040456007

30. Oakley A. Fitzpatrick skin phototype. 2012. Accessed May 10, 2024. https://dermnetnz.org/topics/skin-phototype

31. Bossuyt PM, Reitsma JB, Bruns DE, et al; STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology*. 2015;277(3):826-832. doi:10.1148/radiol.2015151516

32. Collins GS, Moons KG, Dhiman P, et al. TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;e078378. doi:10.1136/bmj-2023-078378

33. The Royal Children's Hospital Melbourne. Blood groups and compatibilities. Accessed October 30, 2024. https://www.rch.org.au/bloodtrans/about_blood_products/blood_groups_and_compatibilities/

34. Murray NA, Roberts IAG. Haemolytic disease of the newborn. *Arch Dis Child Fetal Neonatal Ed*. 2007;92(2):F83-F88. doi:10.1136/adc.2005.076794

35. Drager. Instructions for use Drager JM-105. Published online 2020. Accessed October 30, 2024. https://www.draeger.com/Content/Documents/Products/jm-105-sw-120-ifu-9510905-en.pdf

36. Barko HA, Jackson GL, Engle WD. Evaluation of a point-of-care direct spectrophotometric method for measurement of total serum bilirubin in term and near-term neonates. *J Perinatol*. 2006;26(2):100-105. doi:10.1038/sj.jp.7211436

37. Szabo P, Wolf M, Bucher HU, Haensse D, Fauchere JC, Arlettaz R. Assessment of jaundice in preterm neonates: comparison between clinical assessment, two transcutaneous bilirubinometers and serum bilirubin values. *Acta Pædiatrica Oslo*. 2004;93(11):1491-1495. doi:10.1111/j.1651-2227.2004.tb02635.x

38. Academy of Medicine Singapore. Guidelines on Evaluation and Management of Neonatal Jaundice. October 2018. Accessed October 20, 2021. https://www.ams.edu.sg/view-pdf.aspx?file=media%5C4572_fi_961.pdf&ofile=CPCHS+Guidelines+on+Evaluation+and+Management+of+Neonatal+Jaundice+FINAL.pdf

39. McCamy CS, Marcus H, Davidson JG. A color-rendition chart. *J Appl Photogr Eng*. 1976;2(3):95-99.

40. Leung TS, Outlaw F, MacDonald LW, Meek J. Jaundice Eye Color Index (JECI): quantifying the yellowness of the sclera in jaundiced neonates with digital photography. *Biomed Opt Express*. 2019;10(3):1250-1256. doi:10.1364/BOE.10.001250

41. Bhutani VK, Johnson L, Sivieri EM. Predictive ability of a predischarge hour-specific serum bilirubin for subsequent significant hyperbilirubinemia in healthy term and near-term newborns. *Pediatrics*. 1999;103(1):6-14. doi:10.1542/peds.103.1.6

42. Ding G, Zhang S, Yao D, et al. An epidemiological survey on neonatal jaundice in China. *Chin Med J (Engl)*. 2001;114(4):344-347.

43. Kemper AR, Newman TB, Slaughter JL, et al. Clinical practice guideline revision: management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation. *Pediatrics*. 2022;150(3):1. doi:10.1542/peds.2022-058859

44. Dionis I, Chillo O, Bwire GM, Ulomi C, Kilonzi M, Balandya E. Reliability of visual assessment of neonatal jaundice among neonates of black descent: a cross-sectional study from Tanzania. *BMC Pediatr*. 2021;21(1):383-383. doi:10.1186/s12887-021-02859-x

45. Jung Y. Multiple predicting K-fold cross-validation for model selection. *J Nonparametr Stat*. 2018;30(1):197-215. doi:10.1080/10485252.2017.1404598

46. Wong TT, Yeh PY. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans Knowl Data Eng*. 2020;32(8):1586-1594. doi:10.1109/TKDE.2019.2912815

47. Plevris V, Solorzano G, Bakas NP, Ben Seghier MEA. Investigation of performance metrics in regression analysis and machine learning-based prediction models. Paper presented at: European Community on Computational Methods in Applied Sciences; June 5-9, 2022; Oslo, Norway. Accessed November 20, 2024. https://www.scipedia.com/public/Plevris_et_al_2022a

48. Maisels MJ, Ostrea EM Jr, Touch S, et al. Evaluation of a new transcutaneous bilirubinometer. *Pediatrics*. 2004;113(6):1628-1635. doi:10.1542/peds.113.6.1628

49. Ercan S, Ozgun G. The accuracy of transcutaneous bilirubinometer measurements to identify the hyperbilirubinemia in outpatient newborn population. *Clin Biochem*. 2018;55:69-74. doi:10.1016/j.clinbiochem.2018.03.018

50. Engle WD, Jackson GL, Stehel EK, Sendelbach DM, Manning MD. Evaluation of a transcutaneous jaundice meter following hospital discharge in term and near-term neonates. *J Perinatol*. 2005;25(7):486-490. doi:10.1038/sj.jp.7211333

51. Taylor JA, Stout JW, de Greef L, et al. Use of a smartphone app to assess neonatal jaundice. *Pediatrics*. 2017;140(3):e20170312. doi:10.1542/peds.2017-0312

52. Hegde D, Rath C, Amarasekara S, Saraswati C, Patole S, Rao S. Performance of smartphone application to accurately quantify hyperbilirubinemia in neonates: a systematic review with meta-analysis. *Eur J Pediatr*. 2023;182(9):3957-3971. doi:10.1007/s00431-023-05073-2

53. Poland RL, Hartenberger C, McHenry H, Hsi A. Comparison of skin sites for estimating serum total bilirubin in in-patients and out-patients: chest is superior to brow. *J Perinatol*. 2004;24(9):541-543. doi:10.1038/sj.jp.7211141

54. Yamauchi Y, Yamanouchi I. Factors affecting transcutaneous bilirubin measurement: effect of daylight. *Acta Paediatr Jpn*. 1991;33(5):658-662. doi:10.1111/j.1442-200X.1991.tb01882.x

55. Rajput D, Wang WJ, Chen CC. Evaluation of a decided sample size in machine learning applications. *BMC Bioinformatics*. 2023;24(1):48. doi:10.1186/s12859-023-05156-9

56. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BJOG*. 2015;122(3):434-443. doi:10.1111/1471-0528.13244

57. Galland BC, Taylor BJ, Elder DE, Herbison P. Normal sleep patterns in infants and children: a systematic review of observational studies. *Sleep Med Rev*. 2012;16(3):213-222. doi:10.1016/j.smrv.2011.06.001

58. Moosa AS, Ngeow AJH, Yang Y, et al. A novel smartphone app for self-monitoring of neonatal jaundice among postpartum mothers: qualitative research study. *JMIR MHealth UHealth*. 2023;11:e53291. doi:10.2196/53291

59. Yan Q, Gong Y, Luo Q, et al. Effects of a smartphone-based out-of-hospital screening app for neonatal hyperbilirubinemia on neonatal readmission rates and maternal anxiety: randomized controlled trial. *J Med Internet Res*. 2022;24(11):e37843. doi:10.2196/37843

SUPPLEMENT 1.

eFigure 1. Use of BiliSG Application to Acquire Images of Sclera, Forehead, Sternum, and Abdomen
eFigure 2. Color Sticker Eliminates Shadow Within Central Aperture
eFigure 3. Sternal-Abdominal Yellowness Gradient Across TSB Range
eMethods. Selection of Predictors
eFigure 4. SHAP Analysis of Final Machine Learning Model
eTable 1. Correlation and Agreement With Total Serum Bilirubin (TSB) for Smartphone-Predicted Bilirubin (SpB) and Transcutaneous Bilirubin (TcB)
eTable 2. Diagnostic Accuracy of SpB and TcB
eTable 3. Cross-Tabulation of SpB and TSB Results

SUPPLEMENT 2.

Nonauthor Collaborators

SUPPLEMENT 3.

Data Sharing Statement