Non-Invasive Skin Image Analysis for Enhanced Detection of Neonatal Jaundice

March 29, 2026

Abstract

Neonatal jaundice affects 60–80% of newborns globally and, if untreated, may lead to severe complications such as kernicterus. Current diagnostic methods present significant challenges: total serum bilirubin (TSB) measurement is invasive and resource-intensive, visual assessment is subjective, and transcutaneous bilirubinometers (TeB) are costly and sensitive to skin tone variations. This study proposes a deep learning-based framework for non-invasive bilirubin estimation using smartphone-acquired skin images. Using a dataset of 2,235 images from 745 neonates, we implemented preprocessing techniques including color calibration and automated skin segmentation to reduce variability due to lighting and background noise. Three convolutional neural networks—ResNet50, VGG16, and EfficientNet—were adapted for regression using transfer learning. EfficientNet achieved the best performance with a mean absolute error (MAE) of 2.23 mg/dL and R^2 of 0.67, demonstrating strong agreement with clinical TSB measurements. The proposed approach offers a low-cost, scalable alternative for neonatal jaundice screening, particularly in resource-limited settings. Future work will focus on improving model performance in severe cases and enhancing dataset diversity.

keywords: Neonatal jaundice, Deep learning, Computer vision, Healthcare AI, Low-resource settings

1 Introduction

Neonatal jaundice is a common clinical condition characterized by elevated bilirubin levels in newborns. While often benign, severe cases can result in irreversible neurological damage if not detected early. Conventional diagnostic approaches, including total serum bilirubin (TSB) testing and transcutaneous bilirubinometry (TeB), face limitations in accessibility, cost, and reliability, particularly in low-resource settings. Visual inspection, still widely used in such settings, lacks objectivity and performs poorly across diverse skin tones.

Recent advances in computer vision and deep learning provide an opportunity to develop scalable and cost-effective alternatives using smartphone imaging.

However, challenges such as lighting variability, skin tone diversity, and lack of robust preprocessing pipelines limit practical deployment. This study aims to develop a deep learning framework for estimating bilirubin levels from skin images, addressing these limitations through improved preprocessing and model design.

2 Method(s)

2.1 Dataset and Preprocessing

The study utilized a retrospective dataset comprising 2,235 images from 745 neonates, with three anatomical regions captured per subject (head, face, and chest). To ensure consistency, a color calibration card was included in each image.

Preprocessing involved two main steps. First, color correction was performed using HSV-based segmentation of calibration patches to normalize illumination differences. Second, automated skin segmentation was achieved by combining HSV and YCrCb color space thresholds to isolate relevant skin regions while minimizing background noise.

2.2 Model Development

Three pre-trained convolutional neural networks—ResNet50, VGG16, and EfficientNet—were adapted for regression tasks using transfer learning. The final classification layers were replaced with fully connected regression heads to predict bilirubin concentration.

The dataset was split into training (70%), validation (15%), and test (15%) sets. Model performance was evaluated using mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2). The regression formulation is given as:

$$\hat{y} = f(x) \tag{1}$$

where x represents the processed skin image and $\hat{y}$ the predicted bilirubin level.

3 Results

EfficientNet achieved the best overall performance, with an MSE of 8.49, MAE of 2.23 mg/dL, and R^2 of 0.67, outperforming ResNet50 (MSE = 8.81, MAE = 2.35) and VGG16 (MSE = 11.00, MAE = 2.55). The model demonstrated stable convergence with minimal overfitting, indicated by close alignment between training and validation loss.

Error analysis showed that residuals were approximately normally distributed (mean $\approx$ 0.03 mg/dL, standard deviation = 3.02). However, prediction errors

increased for severe jaundice cases (bilirubin > 15 mg/dL), with MAE rising to 2.98 mg/dL compared to 1.94 mg/dL for moderate cases, reflecting dataset imbalance.

Table 1: Model Performance Comparison

Model	**MSE**	**MAE (mg/dL)**	$\mathbf{R^2}$
ResNet50	8.81	2.35	0.64
VGG16	11.00	2.55	0.58
EfficientNet	8.49	2.23	0.67

4 Discussion and Conclusion

This study demonstrates the feasibility of using deep learning and smartphone imaging for non-invasive neonatal jaundice detection. Compared to traditional methods, the proposed approach reduces reliance on invasive procedures and expensive equipment, making it particularly suitable for low-resource settings.

The integration of color calibration and automated skin segmentation improves robustness against environmental variability, while deep learning models enable automated feature extraction beyond handcrafted approaches. Despite promising results, limitations include reduced accuracy in severe cases and limited dataset diversity, which may affect generalizability.

Future work will focus on expanding the dataset across multiple populations, incorporating synthetic data augmentation, and integrating explainable AI techniques to enhance clinical interpretability. Ultimately, this approach has the potential to be deployed as a mobile health tool, enabling accessible, real-time screening and improving neonatal outcomes globally.

References

[1] World Health Organization. Neonatal jaundice: Clinical guidelines and management, 2020.

[2] LeCun, Y., Bengio, Y., Hinton, G. Deep learning. *Nature*, 521(7553):436–444, 2015.

[3] Esteva, A., et al. A guide to deep learning in healthcare. *Nature Medicine*, 25:24–29, 2019.