

Non-Invasive Detection of Neonatal Jaundice Using Deep Learning and Image Processing: A DenseNet121-Based Approach

Aida Abdi

Amirkabir University of Technology

Mohammad Rabiee

`mrabiee@aut.ac.ir`

Amirkabir University of Technology

Mohammad Hossein Alizadeh Roknabadi

Amirkabir University of Technology

Article

Keywords: Neonatal jaundice, Bilirubin, Non-invasive screening, Deep learning, Image processing, DenseNet

Posted Date: May 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9494347/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Neonatal jaundice is a common condition that requires timely assessment to prevent severe complications. Conventional bilirubin measurement relies on invasive blood sampling, which can cause discomfort and limits frequent monitoring. A non-invasive, image-based approach was developed for estimating total serum bilirubin (TSB) levels using deep learning regression. The study population consisted of neonates aged between 1 day and 3 months. A dataset of 3550 neonatal and infant skin images was collected at the Children's Medical Center, with reference TSB values obtained using a Hitachi 912 analyzer. All images were adjusted to a standardized dimension of 224×224 pixels and preprocessed by converting RGB to CIE (L^*a^*b), involving contrast-limited adaptive histogram equalization applied to the L^* channel, along with noise suppression and normalization. In this research, four different models were evaluated, including a machine learning model (Decision Tree) and three deep learning architectures (MobileNetV2, ResNet50, and DenseNet121). A DenseNet121 model pretrained on ImageNet was adapted for regression by replacing the classification head with a regression head and trained with optimization performed via AdamW and training guided by a mean squared error loss function, along with early stopping and learning rate scheduling. DenseNet121 demonstrated the best predictive performance, achieving a mean absolute error (MAE) of 3.27 mg/dL, mean squared error (MSE) of 17.6 mg/dL², root mean squared error (RMSE) of 4.2 mg/dL, and R^2 of 0.81. The present work presents a robust, non-invasive, and low-cost solution for neonatal jaundice screening and monitoring.

1. Introduction

Contrary to common belief, neonatal jaundice [1] is not considered a disease, but rather a medical sign. Neonatal jaundice, clinically termed neonatal hyperbilirubinemia, is a condition in neonates characterized by elevated total serum bilirubin (TSB) levels [2, 3, 4]. Clinically, the disorder is identified by yellow discoloration involving the cutaneous surface, sclerae, and mucosal surfaces. This condition typically arises within the initial 14 days of life and is a common reason for neonatal return to the hospital for additional care [5]. In the first week following birth, roughly 60% of full-term neonates and close to 80% of preterm infants exhibit jaundice. [6, 7]. Most cases of neonatal jaundice are mild, transient, and resolve on their own without treatment, known as "physiological jaundice" [8]. However, it is important to differentiate this from more severe "pathological jaundice," as untreated cases may ultimately lead to bilirubin-induced encephalopathy and related neurological complications [9]. Severe neonatal jaundice can be classified into unconjugated and conjugated hyperbilirubinemia. Unconjugated hyperbilirubinemia (UHB), which represents the predominant type, results from excessive unconjugated bilirubin and may appear physiologically on the second or third day, resolving within a week, or pathologically within the first 24 hours, often indicating infections, genetic disorders, or internal bleeding [10, 11]. Conjugated hyperbilirubinemia (CHB), though less common, is always pathological, arising from elevated conjugated bilirubin due to conditions such as biliary atresia, neonatal hepatitis, or metabolic disorders, and typically requires medical or surgical intervention [9]. Neonatal jaundice typically begins on the face and upper trunk and then gradually progresses downward toward the lower limbs, with preterm infants exhibiting a higher susceptibility to its onset [12]. As jaundice occurs in more than half of newborns and their immune systems are still insufficiently developed, swift and precise recognition is critical to avert severe and potentially life-threatening complications [13, 14]. Post-discharge follow-up assessments are often difficult to implement, which can result in delayed recognition and management of the condition [15]. Delays of this nature may prolong hospitalization and necessitate clinical interventions, including phototherapy or exchange transfusion, while increasing the probability of significant adverse outcomes such as kernicterus [16, 17], cognitive impairment, hearing loss, and reduced IQ [18–20]. Visual evaluation, the technique most commonly applied in outpatient contexts such as homes or resource-limited clinics, does not consistently provide reliable diagnostic accuracy [21]. Studies also indicate that even highly experienced physicians may have difficulty precisely assessing jaundice severity through visual observation alone [22]. The reference standard for bilirubin assessment is measurement from blood samples; however, repeated sampling is invasive and may be burdensome for newborns and caregivers. As a result, non-invasive screening and monitoring methods have received increasing attention. Optical and imaging-based [23–25] approaches are particularly attractive because they enable rapid and painless screening and can potentially support more frequent monitoring [26]. Such approaches have been explored in tools including BiliCapture, BiliCam, and Biliscan, which estimate bilirubin by analyzing color changes in skin. Nevertheless, performance may be affected by confounding factors such as ambient illumination, device variability, skin pigmentation, and calibration requirements [27–29]. Although several non-invasive methods for neonatal jaundice screening exist, they often suffer from limited accuracy, sensitivity to lighting conditions, and poor generalization across diverse populations.

This study addresses these limitations by applying a DenseNet121-based deep learning regression model to neonatal skin images, providing a more robust, accurate, and practical approach for early jaundice detection. This research seeks to establish a non-invasive

framework for estimating neonatal bilirubin levels through the integration of image processing and deep learning techniques. In this approach, neonatal photographs are transformed into multiple color spaces and processed using a DenseNet121 architecture adapted for regression analysis. DenseNet121 was selected because its dense connectivity promotes effective feature propagation, mitigates gradient-vanishing issues, and enables extraction of subtle color and texture cues from neonatal skin. This framework seeks to provide accurate bilirubin estimation, support earlier recognition and the clinical management of neonatal jaundice, minimize dependence on invasive diagnostic procedures, and ultimately improve clinical outcomes in newborns. The subsequent sections of this manuscript are arranged as follows. Section 2 describes the Materials and Methods, including the data collection process, image preprocessing steps, and the DenseNet121-based regression model. Section 3 presents the experimental results. Sections 4 and 5 discuss the findings and provide the conclusions of the study, respectively.

2. Materials and Methods

In this paper, a DenseNet121 deep learning architecture is applied to generate bilirubin level predictions as part of the neonatal jaundice detection process. During the preprocessing stage, neonatal skin images are initially acquired in the RGB color representation and subsequently transformed into the CIE $L^*a^*b^*$ color space. The proposed network performs a regression-based estimation of bilirubin concentration.

2.1. Study Design and Population

Data were collected in a hospital setting using a specialized optical sensor system. Images of the entire face and selected regions, including the eyes, nose, lips, and tongue, were captured, and variables related to optical deviations from normal conditions, potentially indicative of neonatal jaundice, were analyzed. The imaging system was mounted on an adjustable stand at a standardized distance from each neonate. To ensure measurement accuracy, all ambient light sources were removed, and imaging was performed under controlled lighting conditions. Target regions were extracted according to predefined protocols, and reference points and calibration cards were recorded. All study procedures were performed in compliance with approved ethical principles and established clinical standards. The investigation included neonates admitted during a defined study period to the Children's Medical Center affiliated with Tehran University of Medical Sciences. Infants with suspected jaundice were identified by a pediatric specialist and subsequently underwent bilirubin measurement. The study cohort comprised both healthy and jaundiced neonates of both sexes, who were enrolled following an initial clinical evaluation, confirmation of eligibility based on the predefined criteria, and receipt of written informed consent obtained from their parents or legal guardians. The control group was composed of neonates without clinical evidence of jaundice. A detailed overview of the eligibility criteria applied for participant inclusion is presented in Table 1.

Table 1. Overview of the criteria defining eligibility for the study population.	
Inclusion Criteria	Description
Neonatal age	1 day to 3 months
Gestational maturity	Only full-term neonates (35–42 weeks) with birth weight > 2 kg (maximum observed 4.35 kg)
Clinical status confirmation	Health status clearly documented and clinically confirmed as healthy or jaundiced
Physiological stability	Only neonates in stable general condition; neonates requiring intensive interventions such as intravenous therapy, mechanical ventilation, continuous monitoring, or other life-support equipment were excluded
Parental consent	Participation was preceded through written informed consent secured from the parents or legal guardians

2.2. Image Acquisition Protocol

Images of neonates were acquired using the iPhone 6S smartphone equipped with a 12-megapixel camera, selected for its stability, ease of use, and compatibility with mobile-based non-invasive jaundice detection systems. The setup included an adjustable camera stand, a 24-color calibration card, storage devices, and a personal computer. Neonates were placed in a dedicated enclosure with dark, anti-reflective walls to minimize ambient light and maintain consistent illumination. All images were captured in the supine position

from a frontal view to ensure data comparability and precise extraction of visual features. Images were also taken from the face, abdomen, and arms to examine the relationship of each region with non-invasive assessment of neonatal jaundice. Nonetheless, the main emphasis of the study was placed on the chest region. All images were captured without flash and saved directly in RAW form. For each neonate, multiple consecutive images were captured, and the highest-quality image was selected for analysis. Figure 3 shows (a) example images of the chest and face regions of neonates captured during image acquisition, and (b) the color calibration card containing 24 reference colors, used to standardize lighting and ensure accurate color analysis.

2.3. Reference Standard Measurement

To determine the actual bilirubin levels and diagnose neonatal jaundice, total serum bilirubin (TSB) measurement was employed as the reference standard, with all blood collection and laboratory analyses conducted in the Children’s Medical Center laboratory according to established clinical protocols. Blood samples were drawn by experienced laboratory personnel from the neonates’ antecubital vein, with approximately 2 mL ($\approx$ 2 cc) collected using a scalp device and transferred into anticoagulant-free tubes. Following centrifugation, the separated serum was used for bilirubin measurement, which was performed on a Hitachi 912 analyzer using spectrophotometric methods and validated Kitman diagnostic kits, and the resulting values were considered as the reference measurements for all subsequent analyses. To minimize the effect of time-dependent bilirubin variations, skin imaging was conducted prior to blood sampling, ensuring a maximum interval of five minutes between imaging and sample collection, while laboratory results were available and recorded approximately twenty minutes after sampling.

2.4. Dataset Characteristics

The dataset employed in this investigation comprised 3,550 images from 500 neonates, aged between one day and three months. A total serum bilirubin (TSB) value greater than 15 mg/dL was considered the criterion for defining jaundice. Following preprocessing and quality assessment, 3422 images were included in the final analysis, with 2409 used for training and 1013 reserved for testing and model evaluation. The demographic and clinical characteristics of the study population, including sex and health status, are presented in Table 2.

Table 2. Distribution of neonates by sex and health status.	
Characteristic	Number of neonates
Boys	287
Girls	213
Healthy	307
Jaundiced	193
Total	500

2.5. Image Preprocessing

To improve model accuracy and ensure the data were prepared for subsequent analysis, all preprocessing procedures were carried out in Python using the Pandas library. In this stage, the consistency between the image data and their corresponding labels was verified to ensure data integrity. First, the path to the label file (Labels), stored as a CSV file, was defined and loaded using the Pandas library. This file contained the image filenames along with their corresponding numeric values (Tcb-Value), which served as target labels. Next, the path to the folder containing the images was specified, and a helper function was defined to check the physical existence of each image file at the given path. This function generates the full path for a given image name and verifies whether the file exists. The function was then applied to the image name column to determine for each row whether the corresponding image exists in the folder. The result of this check was added as a new column called Image-Exists. Subsequently, any missing images were identified and, if present, reported as warnings to the user. This step is crucial to detect inconsistencies between labeled data and image files and prevent errors during model training. If all images were successfully found, a confirmation message was displayed. After this validation, the full path of each image was saved in a new column named Image-Path. Finally, only the required information, including the image paths and the numeric label (Tcb-Value), was selected and stored in a final DataFrame. This structured dataset serves as the input to the deep learning model, ensuring accurate and error-free loading throughout the convolutional neural network training phase.

For more reliable characterization of skin color and texture, the preprocessed images were resized to a uniform resolution of 224×224 pixels, and their pixel intensities were normalized to the $[0, 1]$ range. The images were subsequently transformed from the RGB color space to CIE Lab to decouple the luminance component (L^*) from the chromatic channels (a^* and b^*), thereby minimizing the impact of ambient illumination. CLAHE (Contrast-Limited Adaptive Histogram Equalization) was applied to the L^* channel to correct non-uniform lighting and enhance fine-grained texture information. After isolating the skin regions, the region of interest (ROI) was extracted to ensure that the model focused on diagnostically relevant areas while excluding unrelated background content. To further expand dataset diversity and mitigate overfitting, a series of augmentation techniques including controlled rotation, minor translation, brightness adjustment, scaling, and horizontal flipping were applied to generate additional samples without altering the intrinsic characteristics of the original images.

Following ROI extraction, images were converted from RGB to the CIE ($L^*a^*b^*$) color model, separating luminance (L^*) from chromatic components (a^* and b^*) to improve stability under varying illumination. The L^* channel was processed using Contrast Limited Adaptive Histogram Equalization (CLAHE) to increase local contrast while limiting noise enhancement. Noise reduction was performed to suppress imaging artifacts, and all images were standardized and reformatted to 224×224 pixels in order to meet the input specifications of the DenseNet121 architecture.

2.6. Model Architectures

DenseNet121 is an advanced convolutional neural network (CNN) architecture that employs dense connections across its layers. In this model, the output from every layer is provided as input to all subsequent layers, allowing the network to extract only new features instead of repeatedly learning similar ones. This mechanism reduces the number of parameters, improves information flow, and prevents gradient vanishing. Despite its depth, DenseNet121 is lightweight and efficient, making it highly reliable for applications requiring precise feature extraction, such as medical image-based disease detection [30]. The overall structure of DenseNet121, shown in Fig. 2, consists of convolutional blocks with dense connections, allowing each layer to access feature maps from all preceding layers. Each convolutional block in DenseNet121 consists of 1×1 and 3×3 convolutional layers, used sequentially for feature dimensionality reduction and spatial pattern extraction. The 1×1 convolution compresses the feature maps to reduce computational cost, followed by the 3×3 convolution to learn new features from the combination of outputs from preceding layers. Batch normalization and ReLU activation are applied at all stages to maintain computational stability and reduce training fluctuations. Additionally, the output layer includes one neuron with a linear activation function, enabling the model to output a numerical estimate of neonatal bilirubin concentration. These layers form the fundamental building blocks of DenseNet121, enabling the network to extract deep image features with relatively low computational overhead. DenseNet121 employs an advanced residual learning mechanism, where the outputs of all preceding layers are made available to each subsequent layer, and only new features are generated. This approach prevents gradient vanishing, enhances information flow across the network, and stabilizes the training process. The network's 121-layer depth, despite being lightweight, allows the extraction of complex patterns and subtle color and texture details of neonatal skin. Using pretrained weights on ImageNet provides the model with a solid initial understanding of general image features, improving learning accuracy on medical datasets.

The DenseNet121 architecture exhibits high generalization capability, indicating diminished responsiveness to fluctuations in lighting conditions, imaging angle, and skin tone differences. Moreover, the reduced number of trainable parameters helps manage overfitting effectively and lowers the need for large training datasets. These characteristics make DenseNet121 particularly suitable for medical applications, where subtle changes in color and texture are critical (Fig. 6) .

2.7. Model Training and Validation

For training the DenseNet121 model, the images and labeled data were first loaded from a CSV file, and only existing images in the specified path were retained. The dataset was partitioned into three subsets, allocating 70% for training, 15% for testing, and 15% for validation. If the validation set did not include both classes, minority class samples were duplicated to maintain balance. To increase data diversity and reduce overfitting, data augmentation techniques including limited rotations, zooming, brightness adjustments, and horizontal and vertical shifts were applied in a controlled manner. The DenseNet121 model was loaded, and its fully connected layer was modified for regression output with a linear activation function. Training was performed on a GPU to accelerate processing, and the model underwent training for a total of 20 epochs. In each epoch, the model's predictions were compared with the true values to calculate errors, followed by evaluation on the validation data. The Mean Squared Error (MSE) loss function was used for training; squaring the errors increases the sensitivity of the model to large deviations, ensuring more accurate predictions in critical regions.

The AdamW optimizer with a learning rate of 10^{-4} and a weight decay of 10^{-6} was employed to prevent overfitting, encouraging the model to learn only generalizable and relevant patterns from the data. For model evaluation, key metrics including MSE, MAE, RMSE, and the coefficient of determination (R^2) were calculated. These metrics provide a quantitative assessment of prediction accuracy, deviation magnitude, and model generalization capability. The combination of careful preprocessing, controlled data augmentation, optimized model settings, and comprehensive metric evaluation enabled DenseNet121 to achieve high accuracy, stability, and reliable generalization in predicting the target values.

2.8. Evaluation Metrics

To assess the performance of the DenseNet121 model, four main regression metrics were used: MAE, MSE, RMSE, and R^2 . Each metric provides a different perspective on the model's prediction accuracy, and their combination allows for a comprehensive evaluation of predictive quality. Regression performance was assessed using the following standard metrics:

Mean Absolute Error (MAE)

measures the average magnitude of prediction errors, regardless of their direction as shown in Eq. 1

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad \text{Eq. 1}$$

- **Mean Squared Error (MSE)**: calculates the average squared difference between predicted and actual values, giving higher weight to larger errors as shown in Eq. 2:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad \text{Eq. 2}$$

- **Root Mean Squared Error (RMSE)**: the square root of the MSE, reported in the same units as the target variable, as shown in Eq. 3:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad \text{Eq. 3}$$

- **Coefficient of Determination (R^2)**: the proportion of variance in the dependent variable that can be explained by the model is measured using the coefficient of determination, as presented in Eq. 4:

$$R^2 = 1 - \frac{(\sum (y_i - \hat{y}_i)^2)}{(\sum (y_i - \bar{y})^2)} \quad \text{Eq. 4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

These metrics were applied to both test and validation datasets, demonstrating the accuracy, stability, and generalization capability of the model. The selection of these metrics is justified given the clinical relevance of prediction errors and the importance of penalizing large deviations in medical decision-making.

2.9. Statistical Analysis

A comprehensive statistical analysis was performed to rigorously assess the predictive performance, robustness, and generalization ability of the proposed DenseNet121-based framework. Since the primary focus of this study is quantitative prediction accuracy rather than hypothesis testing, evaluation was based on well-established regression performance metrics. Model performance on the test dataset was assessed using (MAE), (MSE), and (R^2). MAE captures the mean absolute deviation between the estimated and reference bilirubin measurements, serving as a clear and interpretable measure of the model's overall accuracy. MSE and RMSE capture the magnitude of errors and assign higher weight to larger deviations, while R^2 denotes the proportion of variability in the target variable

that the model successfully explains. To complement numerical evaluation, visual analyses including histograms of prediction errors and scatter plots of predicted versus actual bilirubin values were employed. These analyses enable a clear understanding of error distribution and model performance trends. The detailed results and figures are presented in Section 3.

3. Results

3.1. Model Training and Validation Dynamics

The training and validation performance of the DenseNet121 model were assessed across 20 epochs. The training loss exhibited a gradual decline throughout the training process, indicating effective learning, although minor fluctuations in certain epochs were likely attributable to challenging input samples. A comparable downward pattern was observed in the validation loss, suggesting stable model behavior and the absence of notable overfitting. The accuracy obtained in both the training and validation stages showed consistent improvement, with validation accuracy exceeding 0.80 by the third epoch, indicating that the model progressively captured meaningful patterns within the dataset. Figures 7 and 8 present the accuracy and loss curves for the training and validation stages, respectively, demonstrating model convergence and robust generalization to previously unseen data. The simultaneous reduction in loss and increase in accuracy indicate that DenseNet121 successfully extracted informative features from neonatal images for reliable bilirubin level prediction.

3.2. Model Performance Comparison

The performance of the models employed in this study—including Decision Tree, MobileNetV2, ResNet50, and DenseNet121 was evaluated and compared using multiple regression performance metrics. Table 3 provides a comparative summary of the principal evaluation parameters for each model. Model performance was assessed using MAE, MSE, RMSE, and R^2 , which quantify prediction accuracy and the model’s capacity to represent the underlying data.

Table 3. Comparison of evaluation metrics for different models.				
	MAE	MSE	RMSE	R^2
Decision Tree	3.9	26	5.1	0.68
MobileNetV2	3.6	22	4.7	0.76
ResNet50	3.6	21	4.6	0.78
DenseNet121	3.27	17.6	4.2	0.81

The results indicate that the Decision Tree exhibited the weakest performance among the models, with an MAE of 3.9 mg/dL, RMSE of 5.1 mg/dL, and R^2 of 0.68. Compared to deep learning networks, this classical machine learning model struggles to capture the nonlinear and complex patterns in image data, resulting in lower predictive accuracy. ResNet50 also showed acceptable performance, recording an MAE of 3.6 mg/dL and R^2 of 0.78. MobileNetV2 showed acceptable performance, recording an MAE of 3.6 mg/dL and R^2 of 0.76. However, its simpler architecture, while computationally efficient, limits its capacity to model the complex relationships present in image data. DenseNet121 demonstrated outstanding performance on all evaluation metrics. With the lowest MAE of 3.27 mg/dL, the lowest MSE of 17.6 mg/dL², and the lowest RMSE of 4.2 mg/dL, DenseNet121 shows higher accuracy in predicting actual neonatal bilirubin levels and a significant reduction in prediction errors. Furthermore, an R^2 value of 0.81 indicates that DenseNet121 effectively explains a substantial portion of the variance in the target variable. Overall, the findings show that models based on deep learning, particularly DenseNet121, significantly outperform classical models like Decision Tree in predicting neonatal bilirubin levels. This superiority can be attributed to the deep architecture and dense connectivity of DenseNet121, which enable more effective high-level feature extraction and maintain information flow across layers.

3.3. Scatter Plot Analysis

The following section presents a visual evaluation of the performance of the trained DenseNet121 model on the test dataset. The stored model is first loaded, and predictions for neonatal bilirubin levels are obtained for all test samples, allowing direct comparison

with the actual measurements. To complement numerical evaluation, predicted bilirubin levels are plotted against the actual values along a reference line with a slope of one, representing ideal agreement, while the distribution of prediction errors is displayed in a histogram. The closeness of the points to the reference line and the density of errors around zero indicate that the model performs accurately, stably, and without systematic bias across different bilirubin levels (Figs. 9 and 10).

Overall, these analyses enable a detailed appraisal of the model's predictive accuracy, stability, and generalization capability, serving as a strong foundation for comparison with other approaches.

3.4. Error Distribution Analysis

This section evaluates the performance exhibited by the model by examining the discrepancies between the predicted and measured bilirubin concentrations. The prediction error is defined as the numerical difference between the model's output and the corresponding ground-truth value, calculated as: $\text{Prediction Error} = \text{Predicted} - \text{Actual}$. Positive error values indicate that the model overestimates the true value, while negative values indicate underestimation. To better understand the behavior of these errors, the distribution of prediction errors referred to as the error distribution is analyzed. Error distribution illustrates how frequently different error values occur across all samples and is typically visualized using a histogram. This representation helps reveal the spread, central tendency, and possible skewness of the errors. The distribution of errors, calculated as the difference between predictions and ground truth values, allows for the assessment of error patterns, dispersion, and potential systematic bias. A concentration of errors around zero indicates balanced and unbiased predictions. This analysis, together with the numerical evaluation of the model, provides a comprehensive view of its accuracy, stability, and generalization capability, forming a solid basis for comparison with other proposed methods.

3.5. Performance by Bilirubin Level

The DenseNet121 model was further assessed across varying ranges of neonatal bilirubin levels to evaluate its stability and reliability. Analysis indicates that the model consistently provides accurate predictions throughout the spectrum of bilirubin concentrations, maintaining a low deviation from actual values even at higher levels. This demonstrates that the DenseNet121 architecture effectively captures subtle color and texture variations in neonatal skin, enabling robust estimation of bilirubin levels regardless of their magnitude. Overall, the model shows reliable performance across all clinically relevant ranges, reinforcing its potential utility for early, non-invasive assessment and the surveillance of neonatal jaundice.

4. Discussion

4.1. Primary Findings

The main aim of this study was to develop a non-invasive framework designed for prediction neonatal bilirubin levels based on image processing and deep learning. The DenseNet121 model outperformed conventional machine learning and other convolutional neural network architectures. Across all evaluation metrics, including MAE, MSE, RMSE, and R^2 , DenseNet121 consistently achieved the highest predictive accuracy, highlighting its ability to extract and utilize subtle color and texture information from neonatal skin. The model maintained robust performance across varying bilirubin concentrations, suggesting reliability for clinical application. These findings underscore the feasibility of using high-resolution images in combination with advanced deep learning architectures for early detection of neonatal jaundice. The controlled image acquisition protocol and rigorous preprocessing, including ROI extraction and color space transformation, contributed to the model's stability and generalization, even when tested on previously unseen samples. Overall, the results validate the DenseNet121-based framework as a promising tool for non-invasive neonatal bilirubin monitoring.

Table 4. Non-invasive neonatal jaundice detection methods using image processing : a comparative summary.

Ref	Body part	Processing method	Region	Dataset	Results	Distinction from Our Work
[31]	face	Deep learning technique (SSGNN)	India	Not mentioned	Accuracy: 93%	Our work develops and evaluates a DenseNet121 framework using a neonatal dataset comprising a larger number of cases, with no feature-fusion component included
[32]	sternum	Deep learning technique (ImageNet)	Chiayi, Taiwan	228	Accuracy: 90.56%, Sensitivity: 67.8%, Specificity: 96.9%, Precision: 85.08%, AUC: 92.6	Improved performance was achieved using the DenseNet121 architecture, trained using a dataset including both jaundiced and non-jaundiced neonates collected under a standardized and controlled acquisition protocol
[33]	face	Deep learning technique (ResNet50)	Mosul, Iraq	145	Accuracy: 84.09%	DenseNet121 was employed, more comprehensive evaluation metrics were utilized, and improved results were achieved
[34]	face	Deep learning technique (SPAGNN, SPEGNN)	Chennai, India	Not mentioned	Accuracy: 96.5%, sensitivity: 96.6%, precision: 95%, AUC: 97.2	We applied DenseNet121; used a larger dataset and performed a more comprehensive evaluation
[35]	sclera	Transfer learning and Hybrid model techniques	Bangalore, India	201	Accuracy: 88.57%	Our framework operates purely on image inputs and is inherently suitable for implementation on smartphone platforms
[36]	Not mentioned	Vision Transformer + ML classifier	India	760	Accuracy: 96.97%, sensitivity: 96.97%, precision: 97.54%, specificity: 96.97%	We applied DenseNet121; used a larger dataset and performed a more comprehensive evaluation
[37]	skin	Deep learning technique	Internet database	2235	Accuracy: 96.87% % precision: 94.75% recall = 92.69% F1 = 93.72%.	We employ DenseNet121; use more comprehensive metrics

4.2. Clinical Implications

The DenseNet121-based framework proposed in this study has notable implications for the field of neonatal care. As detailed in Table 5, the method demonstrates several advantages over conventional techniques, including reduced financial burden, accelerated result generation, and the absence of invasive procedure. By providing rapid, accurate, and non-invasive evaluation of bilirubin levels, the system can support timely diagnosis and intervention, potentially reducing the need for invasive collection of blood samples. Early identification of high-risk neonates may facilitate more efficient use of phototherapy and hospital resources, improving patient outcomes and reducing caregiver burden. Moreover, the portability and simplicity of image-based bilirubin assessment make it suitable for deployment in settings with limited resources, where access to laboratory testing is constrained. Integration of this technology into routine clinical workflows could enhance monitoring, enable remote screening, and support continuous assessment of neonatal jaundice without disrupting standard care procedures.

Table 5. Comparison of the proposed image-based deep learning method with conventional bilirubin assessment techniques.

Feature / Criteria	Conventional Methods (TcB, Blood Sampling)	Proposed Image-Based Deep Learning Method
Invasiveness	Invasive (blood draw) or semi-invasive	Completely non-invasive
Required Equipment	Laboratory devices, bilirubinometers	Camera + software only
Cost	Moderate to high	Low operational cost
Time to Result	Several minutes to hours (lab delay)	Instant or near-real-time
User Skill Needed	Requires trained clinical staff	Easy to use; minimal training needed
Accuracy & Consistency	Variable across populations; needs calibration	High accuracy; robust feature extraction (DenseNet121)
Suitability in Low-Resource Settings	Limited due to equipment and cost	Highly suitable; portable and low-cost
Frequency of Monitoring	Limited (due to discomfort or cost)	Enables frequent or continuous monitoring
Remote Screening	Not feasible	Fully feasible (telemedicine capable)
Caregiver Burden	Higher (blood draws, hospital visits)	Reduced burden; faster and simpler

4.3. Methodological Considerations

This study emphasizes several methodological strengths, including standardized image acquisition, rigorous preprocessing, and comprehensive evaluation metrics. The consistent imaging protocol minimized variability due to lighting and positioning, ensuring that the model learned relevant features rather than spurious correlations. Data augmentation further enhanced the robustness of the model by simulating realistic variations in neonatal images. Additionally, the use of multiple regression metrics and visual evaluation provided a multi-faceted assessment of performance. By combining MAE, MSE, RMSE, and R^2 with scatter plot and histogram analyses, the study ensured that both numerical accuracy and error distribution were carefully examined. These methodological choices contribute to the reliability, reproducibility, and generalizability of the results.

The training process was monitored over epochs to assess convergence and learning behavior. Figure 11 illustrates the epoch-wise progression of training and validation losses (MSE). The continuous reduction in both curves demonstrates robust convergence and effective model learning, without any indication of overfitting. The close alignment of the two curves demonstrates that the model generalized well to unseen data during training. Overall, these results confirm that the DenseNet121 architecture, combined with the proposed preprocessing and regression head, provides a reliable and robust framework for estimating neonatal serum bilirubin levels using non-invasive skin images.

4.4. Limitations & Future Direction

Despite the encouraging results, several limitations need to be acknowledged. The dataset, although diverse, was collected from a single clinical center, which may limit generalizability across different populations and imaging devices. In addition, certain environmental factors, such as minor variations in ambient lighting or patient movement, could still influence predictions, despite rigorous preprocessing. Furthermore, while DenseNet121 demonstrated robust performance, it necessitates significant computational resources during training and inference, potentially restricting its implementation in low-resource settings. Future work should consider cross-center validation, multi-device evaluation, and lightweight model adaptations to address these constraints.

Future studies may benefit from enlarging the dataset by incorporating data from multiple clinical centers and more diverse populations to improve generalizability. Incorporating different imaging devices, including smartphones and low-cost cameras, can further validate the robustness of the proposed method. Additionally, exploring transfer learning and lightweight model architectures may allow real-time deployment in low-resource settings without compromising accuracy. Integration of multimodal data, such as combining image-based predictions with demographic and clinical parameters, could further improve predictive performance. Longitudinal studies tracking bilirubin levels over time may provide insights into early intervention strategies and contribute to more

personalized neonatal care. Overall, these proposed directions are intended to mature the underlying framework into deployable, scalable systems suitable for routine clinical practice.

5. Conclusion

This study demonstrated that deep neural networks, particularly DenseNet121, outperform classical methods including the Decision Tree algorithm and other deep models including ResNet50 and MobileNetV2 in non-invasive prediction of neonatal jaundice from skin images. DenseNet121 outperformed the other models by exhibiting minimal prediction errors and the highest coefficient of determination, accurately modeling complex relationships between image features and bilirubin levels. ResNet50 and MobileNetV2 also delivered satisfactory performance but were slightly less accurate than DenseNet121. The Decision Tree exhibited the weakest performance, highlighting the limitations of classical approaches when handling complex image data. A notable strength of this study lies in the use of authentic data collected locally. Images were directly collected from neonates visiting the clinic, reflecting actual clinical conditions. This allowed the models to perform reliably when applied to real-world data. Furthermore, the dataset includes a varied range of skin pigmentation Iranian neonates, enhancing the model's generalizability within the target population. The substantial dataset size also contributed to more stable model training and improved generalization capability. Overall, the findings indicate that DenseNet121, as the proposed optimal model, provides reliable predictions of neonatal jaundice from skin images. Its accurate, stable, and generalizable performance demonstrates competitiveness with previous studies and highlights that employing advanced deep neural networks can significantly enhance non-invasive neonatal jaundice detection systems. DenseNet121 demonstrated the best predictive performance, achieving a mean absolute error (MAE) of 3.27 mg/dL, mean squared error (MSE) of 17.6 mg/dL², root mean squared error (RMSE) of 4.2 mg/dL, and R² of 0.81. This study presents a robust, non-invasive, and cost-effective approach for the screening and monitoring of neonatal jaundice.

Declarations

Acknowledgements

The authors extend their sincere appreciation to Dr. Razieh Sangesari for her valuable contributions during the initial phases of this project. The team also gratefully acknowledges the support and cooperation of the Children's Medical Center and Tehran University of Medical Sciences throughout the study. We further recognize the neonatal care staff and laboratory teams for their assistance in coordinating and facilitating the collection of clinical images and data. Our deepest gratitude is reserved for the parents and authorized guardians of the participating infants whose consent and trust made this research possible.

Statement of Author Contributions

A.A. and M.R. were responsible for the conceptual design of the study and contributed to manuscript preparation. The initial draft was written by A.A., who also developed the DenseNet121-based deep learning model, conducted the experimental analyses, and oversaw project management. M.R. provided clinical guidance and participated in data collection, validation, and neonatal skin imaging. M.H.A. contributed to data annotation and the review of relevant literature. All authors critically reviewed and approved the final manuscript.

Funding Statement

This study did not receive any external financial support. All work was carried out independently as an academic initiative, with institutional backing provided by Amirkabir University of Technology and TUMS (Tehran University of Medical Sciences).

Conflict of Interest

No known financial interests or personal relationships exist that could be perceived as having influenced the research presented in this study.

Ethical Approval and Informed Consent

Ethical clearance for this investigation was granted by the Institutional Review Board of the Children's Medical Center at Tehran University of Medical Sciences (IR.TUMS.CHMC.REC.1399.001). The research was conducted in strict adherence to the ethical standards outlined in the Declaration of Helsinki. Written informed consent was obtained from the legal guardians of all participating neonates prior to enrollment; no participant was included in the study without first providing informed and voluntary consent. To

ensure confidentiality, all data were anonymized before analysis. Furthermore, permission for the publication of de-identified images in an open-access scientific journal was obtained.

Access to Data and Study Materials

Disclosure of the datasets associated with this investigation is restricted, given the stringent patient-confidentiality safeguards and institutional ethical directives governing their use. Nevertheless, de-identified data and supplementary materials may be accessed through the corresponding author upon the submission of a justified request and contingent on obtaining the necessary ethical clearance.

References

1. Slusher, T. M. et al. Burden of severe neonatal jaundice: a systematic review and meta-analysis. *BMJ paediatrics open*. **1** (1), e000105 (2017).
2. Hardalac, F., et al., Classification of neonatal jaundice in mobile application with noninvasive imageprocessing methods. *Turkish Journal of Electrical Engineering and Comput. Sci.*, **29**(4): pp. 2116–2126. (2021).
3. Aune, A. et al. Bilirubin estimates from smartphone images of newborn infants' skin correlated highly to serum bilirubin levels. *Acta Paediatr.* **109** (12), 2532–2538 (2020).
4. Munkholm, S. B. et al. The smartphone camera as a potential method for transcutaneous bilirubin measurement. *PloS one*. **13** (6), e0197938 (2018).
5. Smitherman, H., A.R. Stark, and V.K. Bhutan. Early recognition of neonatal hyperbilirubinemia and its emergent management. in *Seminars in fetal and neonatal medicine*. 2006. Elsevier.
6. Olusanya, B. O., Kaplan, M. & Hansen, T. W. Neonatal hyperbilirubinaemia: A global perspective. *Lancet Child. Adolesc. Health*. **2** (8), 610–620 (2018).
7. Maheshwari, V., Díaz-González de Ferris, M. E., Filler, G. & Kotanko, P. Novel extracorporeal treatment for severe neonatal jaundice: A mathematical modelling study of allo-hemodialysis. *Sci. Rep.* **14** (1), 21910 (2024).
8. Awe, O., et al., Prevalence of jaundice among neonates admitted in a tertiary hospital in Southwestern Nigeria. *Adv Pediatr Neonatol Care*, 2021. 121(10.29011).
9. Ansong-Assoku, B., Shah, S. D. & Adnan, M. AnkolaPA. Neonatal jaundice. (2018).
10. Diala, U. M. et al. Global Prevalence of Severe Neonatal Jaundice among Hospital Admissions: A Systematic Review and Meta-Analysis. *Journal of Clinical Medicine*, 12(11), p.3738. (2023).
11. Kinshella, M. L. W. et al. Challenges and recommendations to improve implementation of phototherapy among neonates in Malawian hospitals. *BMC Pediatr.* **22** (1), 367 (2022).
12. Rennie, J., Burman-Roy, S. & Murphy, M. S. Neonatal jaundice: Summary of NICE guidance. *Bmj* **340** (2010).
13. van der Geest, B. A. et al. Assessment, management, and incidence of neonatal jaundice in healthy neonates cared for in primary care: a prospective cohort study. *Sci. Rep.* **12** (1), 14385 (2022).
14. Singla, R. & Singh, S. A framework for detection of jaundice in new born babies using homomorphic filtering-based image processing. In *International Conference on Inventive Computation Technologies (ICICT)* Vol. 3, 1–5 (IEEE, 2016). (2016).
15. Ahmadpour-Kacho, M. et al. Cord blood alkaline phosphatase as an indicator of neonatal jaundice. *Iran. J. Pediatr.* **25**(5) (2015).
16. Althnian, A., Almanea, N. & Aloboud, N. Neonatal Jaundice Diagnosis Using a Smartphone Camera Based on Eye, Skin, and Fused Features with Transfer Learning. *Sensors*, 21, p.7038. (2021).
17. Han, K. et al. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **45** (1), 87–110 (2022).
18. Seidman, D. S. et al. Neonatal hyperbilirubinemia and physical and cognitive performance at 17 years of age. *Pediatrics* **88** (4), 828–833 (1991).
19. Boskabadi, H., Zakerihamidi, M., Moradi, A. & Bakhshaei, M. Risk factors for sensorineural hearing loss in neonatal hyperbilirubinemia. *Iran. J. Otorhinolaryngol.* **30** (99), 195 (2018).
20. Tsao, P. C. et al. Long-term neurodevelopmental outcomes of significant neonatal jaundice in Taiwan from 2000–2003: A nationwide, population-based cohort study. *Sci. Rep.* **10** (1), 11374 (2020).
21. Porter, M. L. & Dennis, B. L. Hyperbilirubinemia in the term newborn. *Am. Fam Phys.* **65** (4), 599–607 (2002).

22. Riskin, A., Tamir, A., Kugelman, A., Hemo, M. & Bader, D. Is visual assessment of jaundice reliable as a screening tool to detect significant neonatal hyperbilirubinemia? *J. Pediatr.* **152** (6), 782–787 (2008).
23. De Luca, D. et al. Skin bilirubin nomogram for the first 96 h of life in a European normal healthy newborn population, obtained with multiwavelength transcutaneous bilirubinometry. *Acta Paediatr.* **97** (2), 146–150 (2008).
24. 15 et al. Large scale validation of a new non-invasive and non-contact bilirubinometer in neonates with risk factors. *Sci. Rep.* **10** (1), 11149 (2020).
25. 16 et al. Non-invasive estimation of hemoglobin, bilirubin and oxygen saturation of neonates simultaneously using whole optical spectrum analysis at point of care. *Sci. Rep.* **13** (1), 2370 (2023).
26. Kihara, T. et al. Identification and Quantification of Jaundice by Trans-Conjunctiva Optical Imaging Using a Human Brain-like Algorithm: A Cross-Sectional Study. *Diagnostics* **13** (10), 1767 (2023).
27. Dionis, I. et al. Reliability of visual assessment of neonatal jaundice among neonates of black descent: a cross-sectional study from Tanzania. *BMC Pediatr.* **21**, 1–6 (2021).
28. 10 et al. Is visual assessment of jaundice reliable as a screening tool to detect significant neonatal hyperbilirubinemia? *J. Pediatr.* **152** (6), 782–787 (2008). e2.
29. Miah, M. M. M. et al. Non-invasive bilirubin level quantification and jaundice detection by sclera image processing. in 2019 IEEE Global Humanitarian Technology Conference (GHTC). IEEE. (2019).
30. Albelwi, S. A. Deep architecture based on DenseNet-121 model for weather image recognition. *International J. Adv. Comput. Sci. Applications*, **13**, 10, (2022).
31. Saini, N., A. Kumar, and P. Khera, Non-invasive bilirubin detection technique for jaundice prediction using smartphones. *International Journal of Computer Science and Information Security*, 2016. 14(8): p. 1060.
32. Hsu, W.-Y. and H.-C. Cheng. A fast and effective system for detection of neonatal jaundice with a dynamic threshold white balance algorithm. in *Healthcare*. 2021. MDPI.
33. Abdulrhman, A.I., M.S. Jarjees, and N.A. Kasim. A review study of newborn bilirubin monitoring systems based on image processing techniques. in *AIP Conference Proceedings*. 2023. AIP Publishing.
34. Madhusundar, N. and R. Surendran. Neonatal jaundice identification over the face and sclera using graph neural networks. in 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE. (2023).
35. Sreedha, B., Nair, P. R. & Maity, R. Non-invasive early diagnosis of jaundice with computer vision. *Procedia Comput. Sci.* **218**, 1321–1334 (2023).
36. Eliazer, M. et al. Integrating vision transformer-based deep learning model with kernel extreme learning machine for non-invasive diagnosis of neonatal jaundice using biomedical images. *Sci. Rep.* **15** (1), 25493 (2025).
37. Makhlooghi, F., Artificial intelligence-based non-invasive bilirubin prediction for neonatal jaundice using 1D convolutional neural network. *Scientific Reports*, 2025. 15(1): p. 11571.

Figures

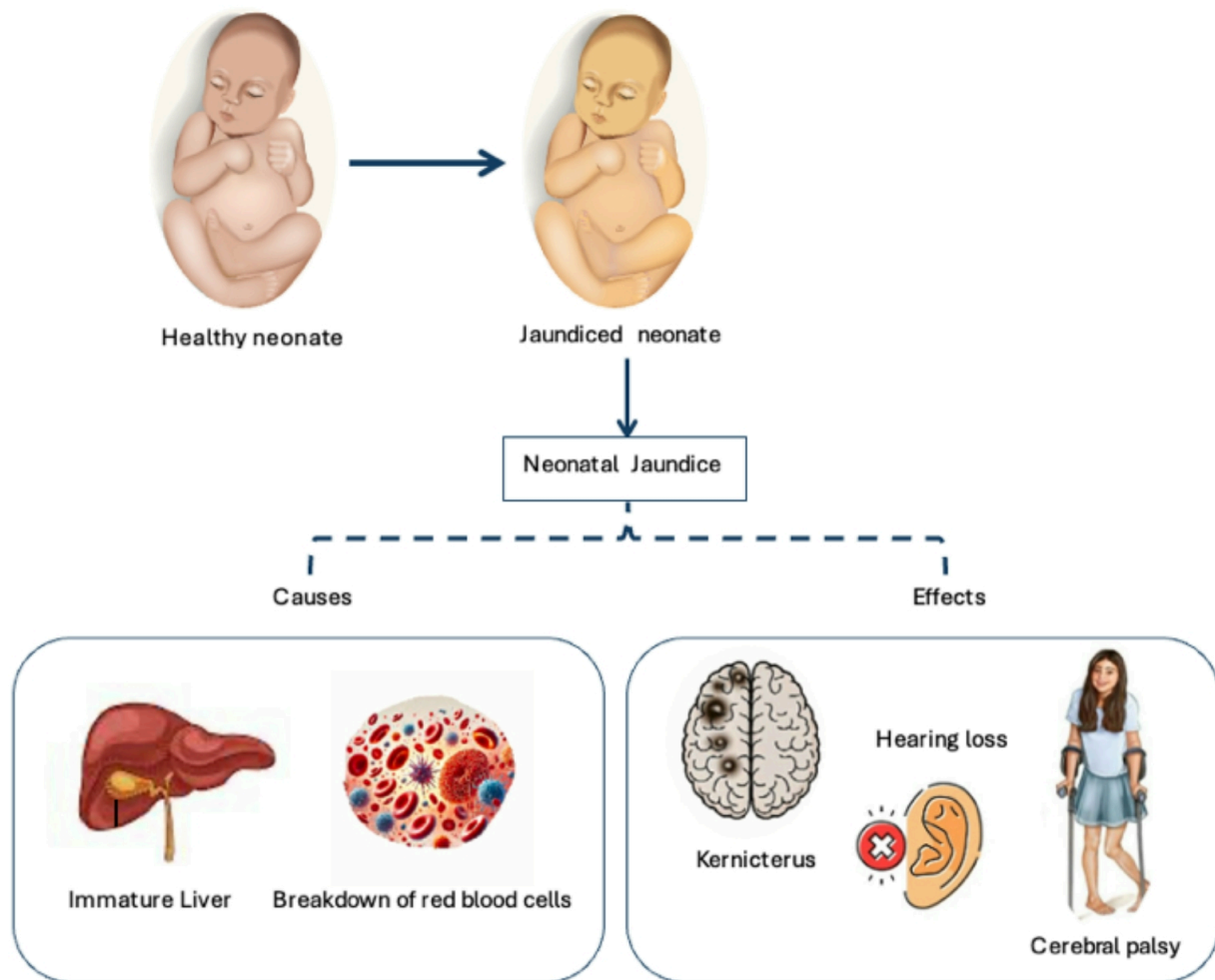


Figure 1

Overview of the primary etiological factors and resulting effects of neonatal jaundice.

Neonatal Jaundice detection Using a DenseNet-based Deep Learning Model

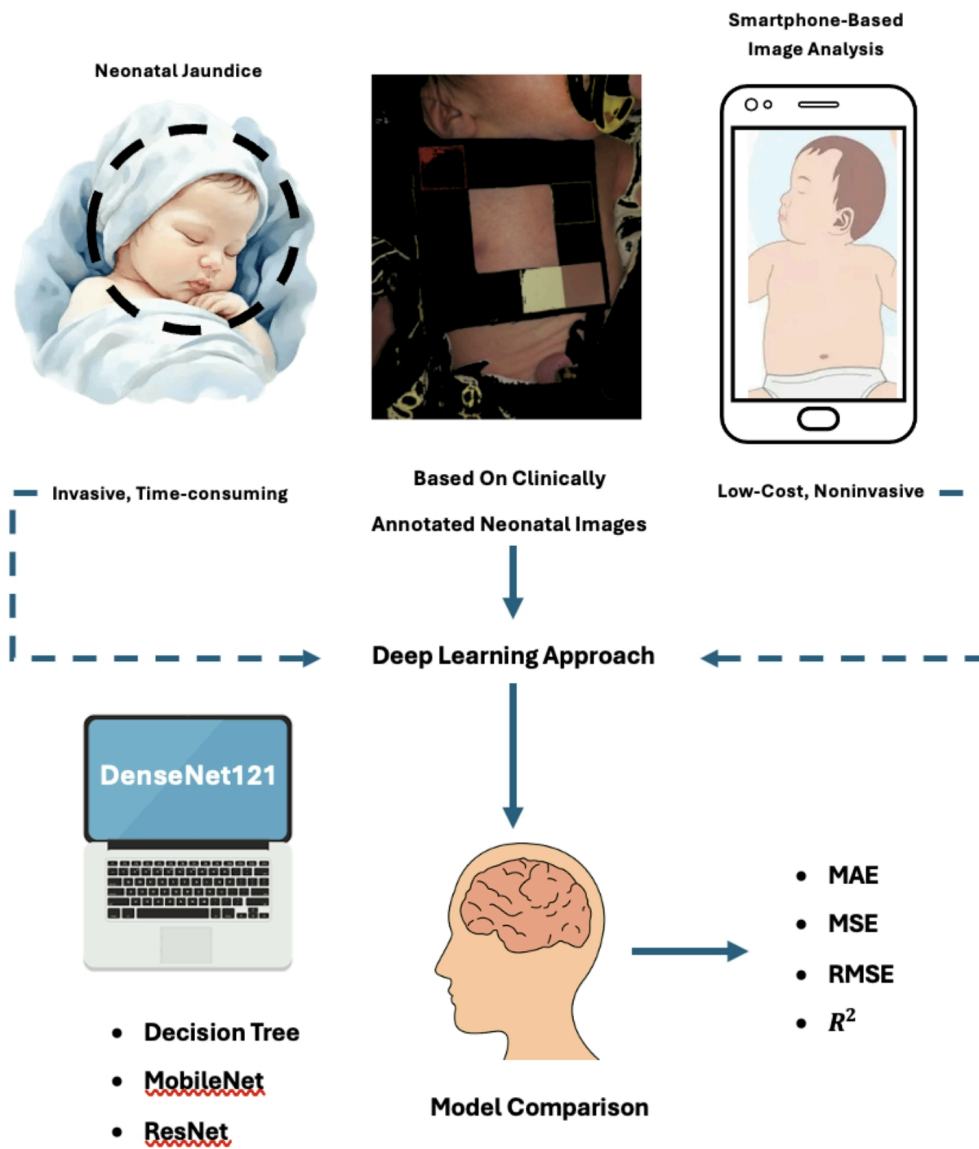


Figure 2

A schematic depiction of the proposed approach for detecting neonatal jaundice is illustrated. Clinically labeled neonatal skin images obtained with a smartphone camera are preprocessed and later examined using a deep learning architecture derived from DenseNet121.

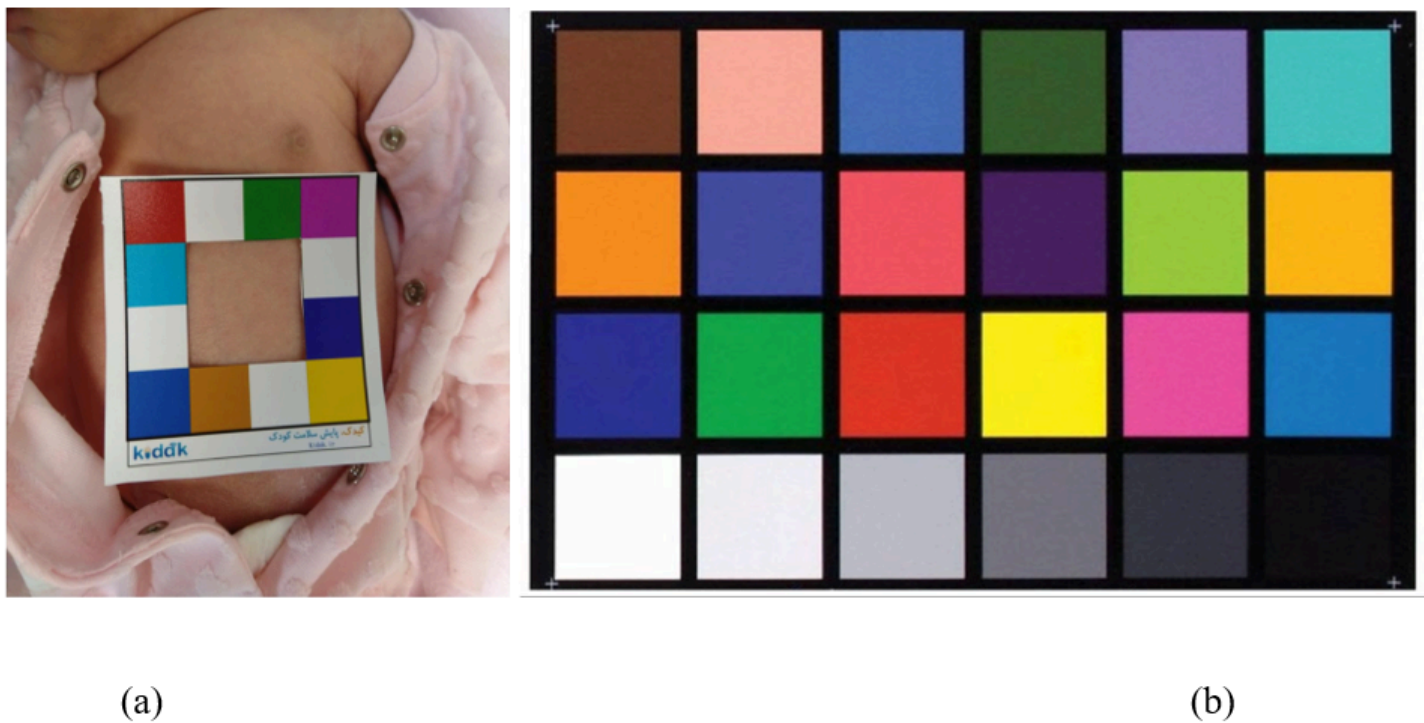


Figure 3

Representative image of the neonatal upper-torso region (a), and the 24-color calibration card (b).

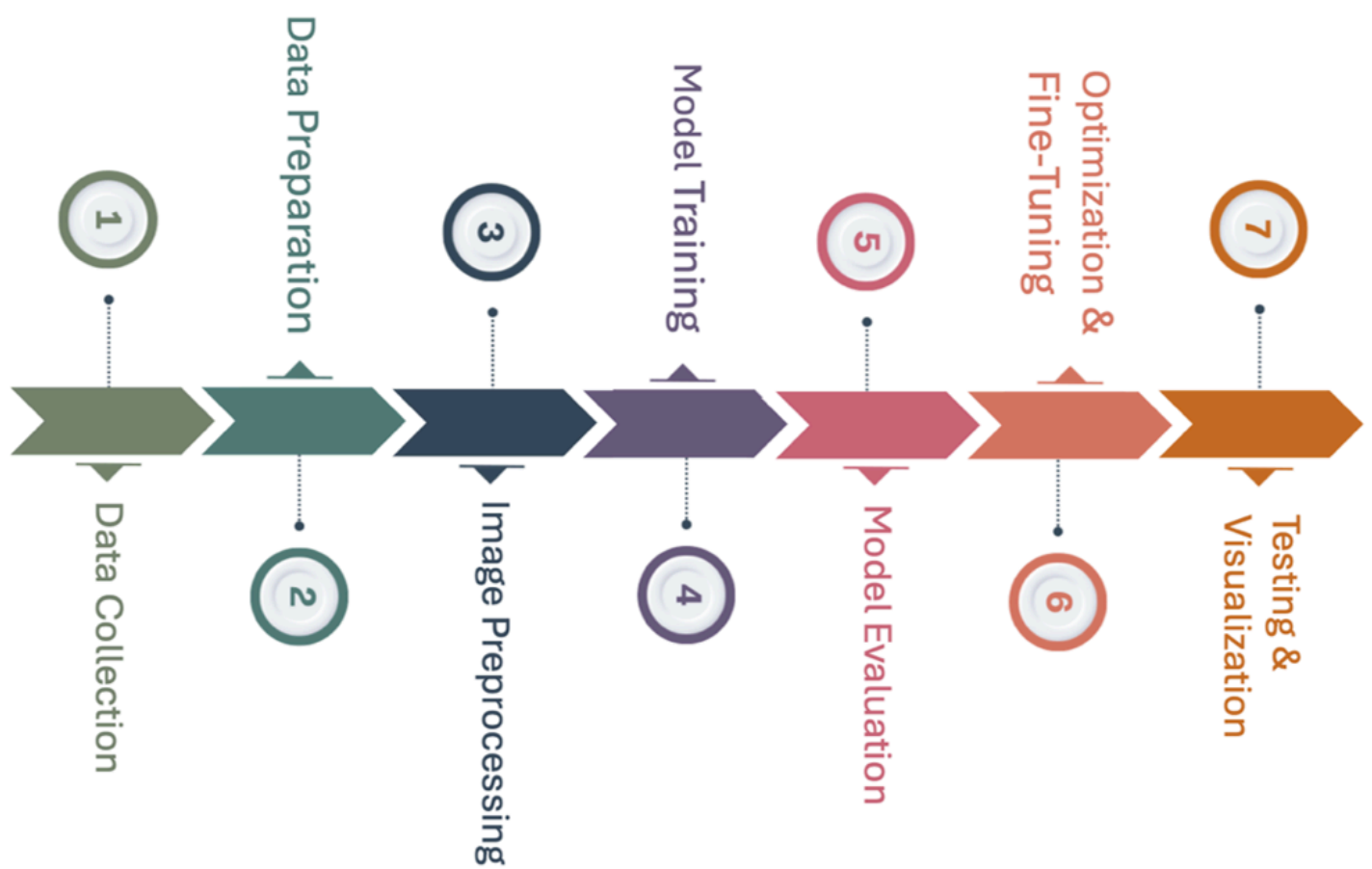


Figure 4

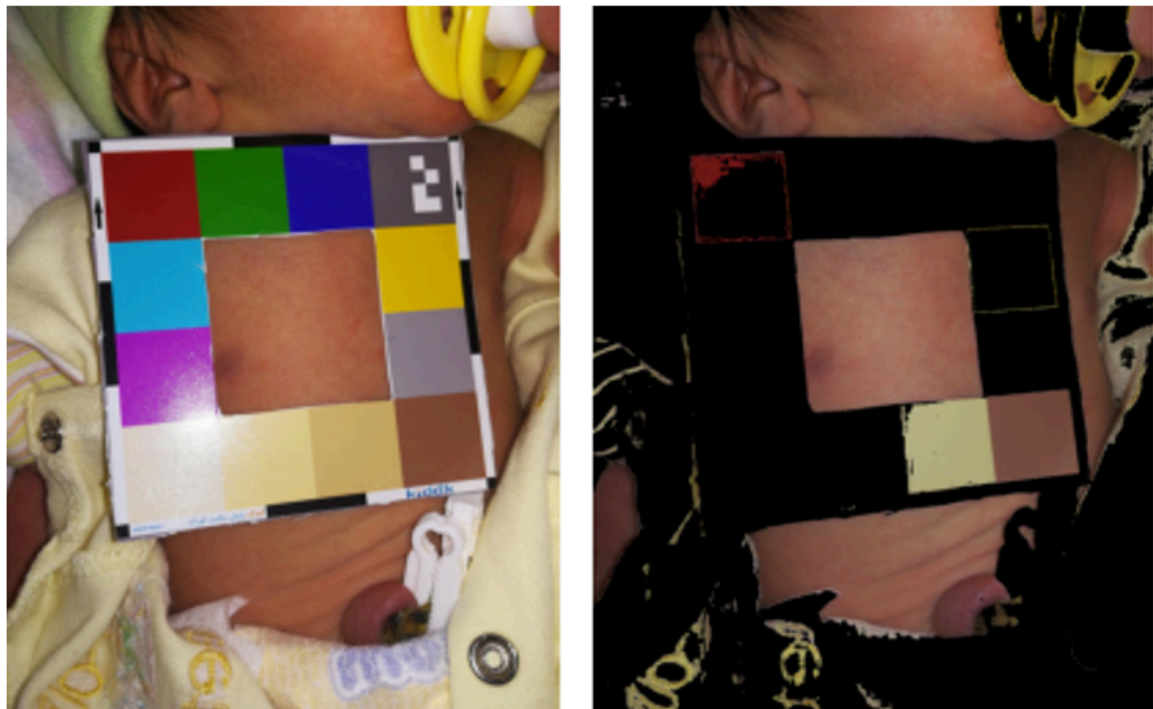


Figure 5

The skin detection and ROI extraction stage isolates the infant’s skin areas by eliminating surrounding non-skin regions. This process enhances the precision of the analytical steps that follow. The figure contrasts the unprocessed image with its ROI-filtered counterpart, offering a clear before-and-after view.

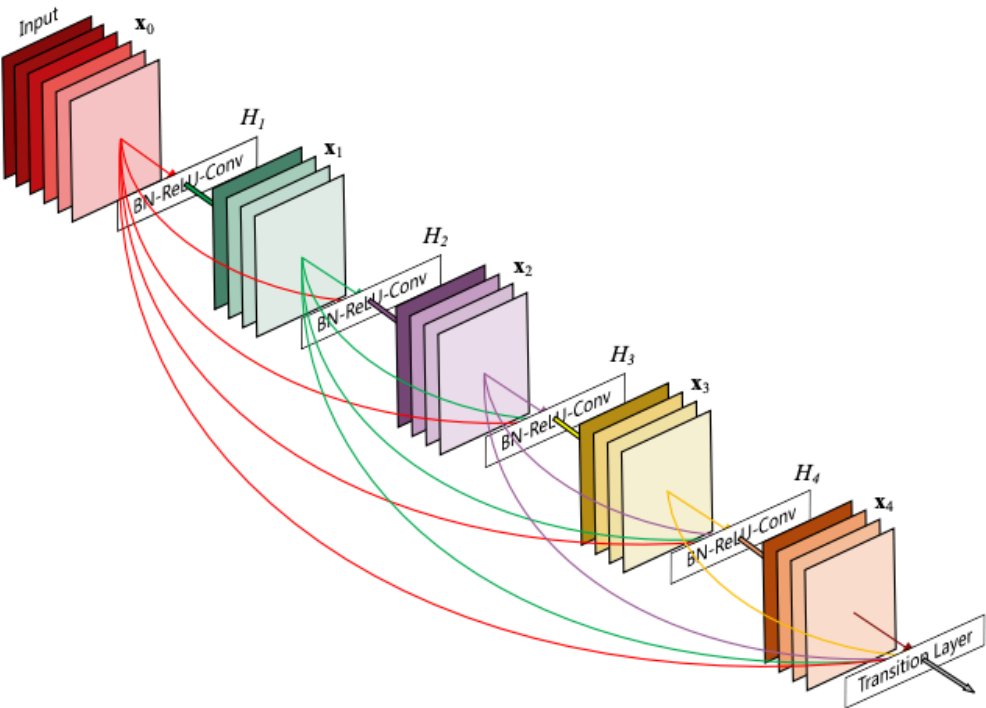
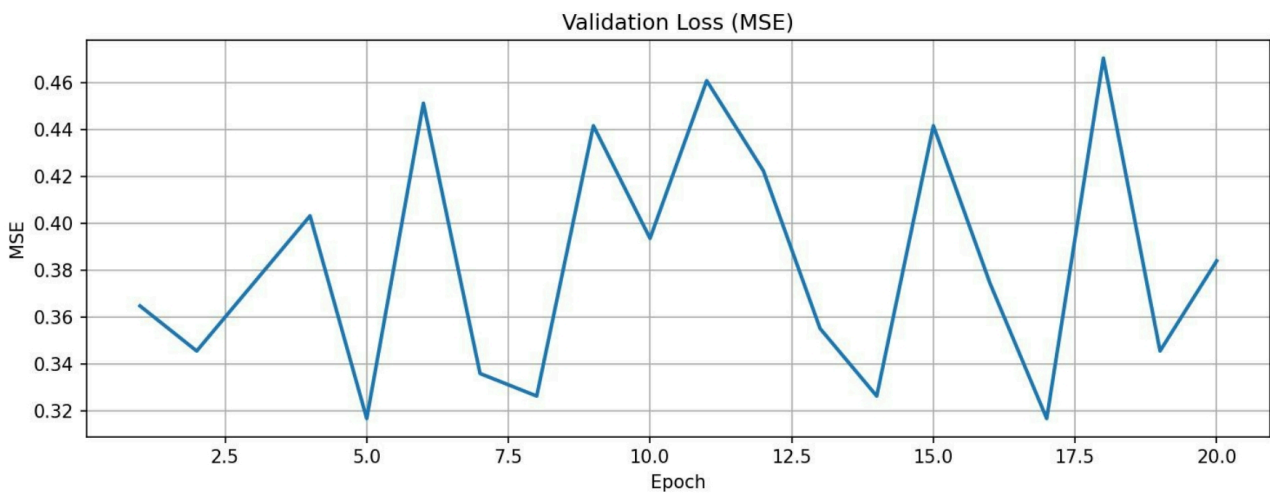


Figure 6

Architecture of the DenseNet121 network [30]. The network consists of convolutional blocks with dense connections, allowing each layer to access feature maps from all preceding layers.



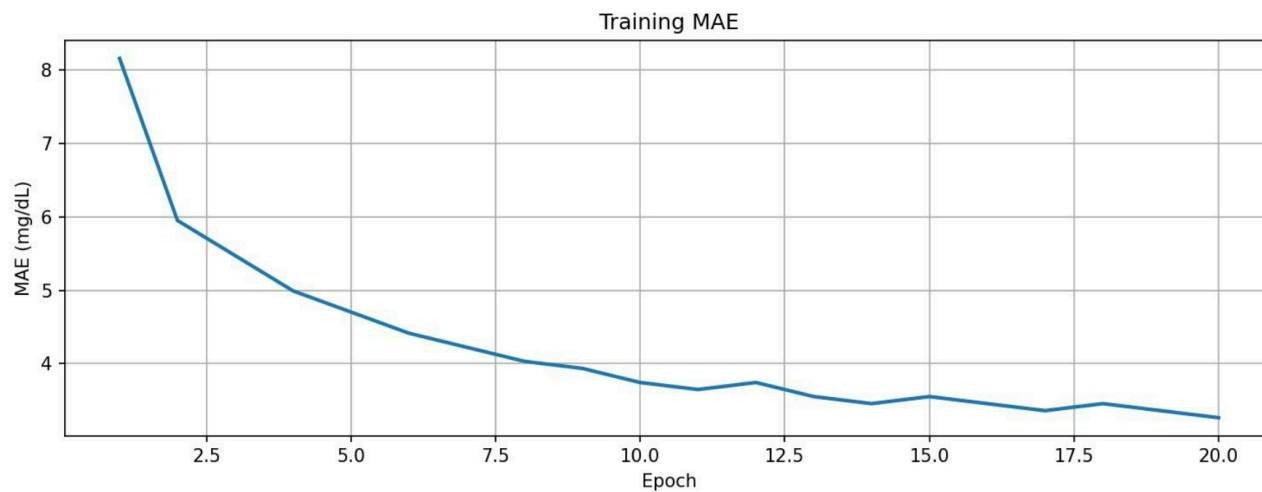
(a)



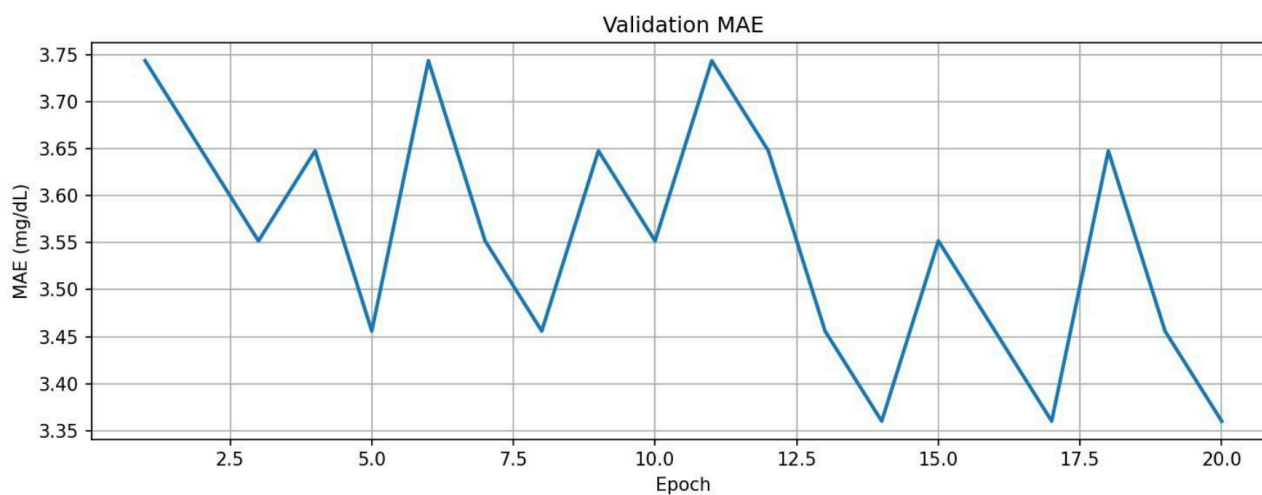
(b)

Figure 7

Presents the loss curves for the training (a) and validation (b) phases over 20 epochs.



(a)



(b)

Figure 8

Illustrates the training (a) and validation (b) accuracy curves across 20 epochs.

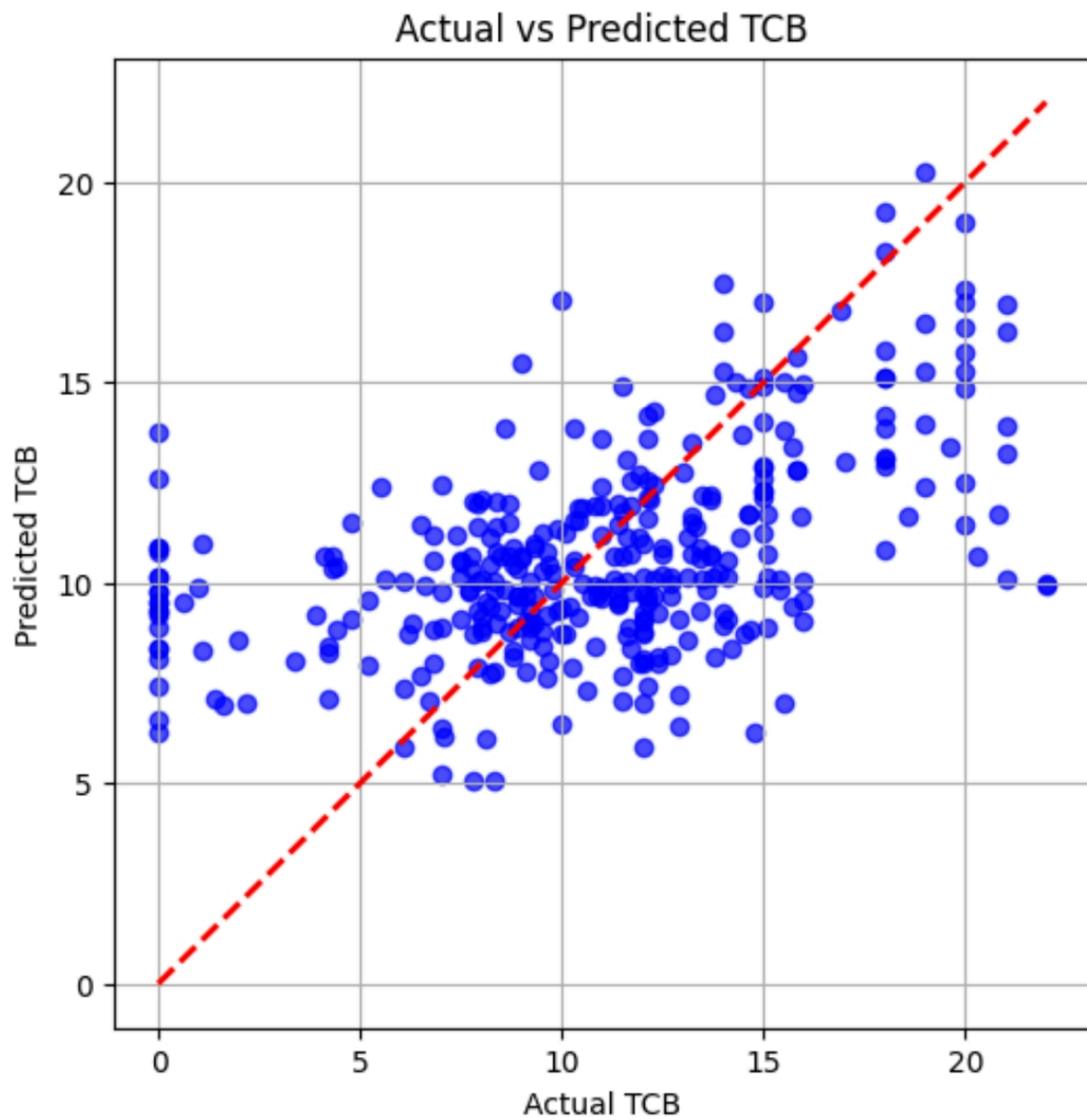


Figure 9

Scatter plot of predicted vs. actual bilirubin values.

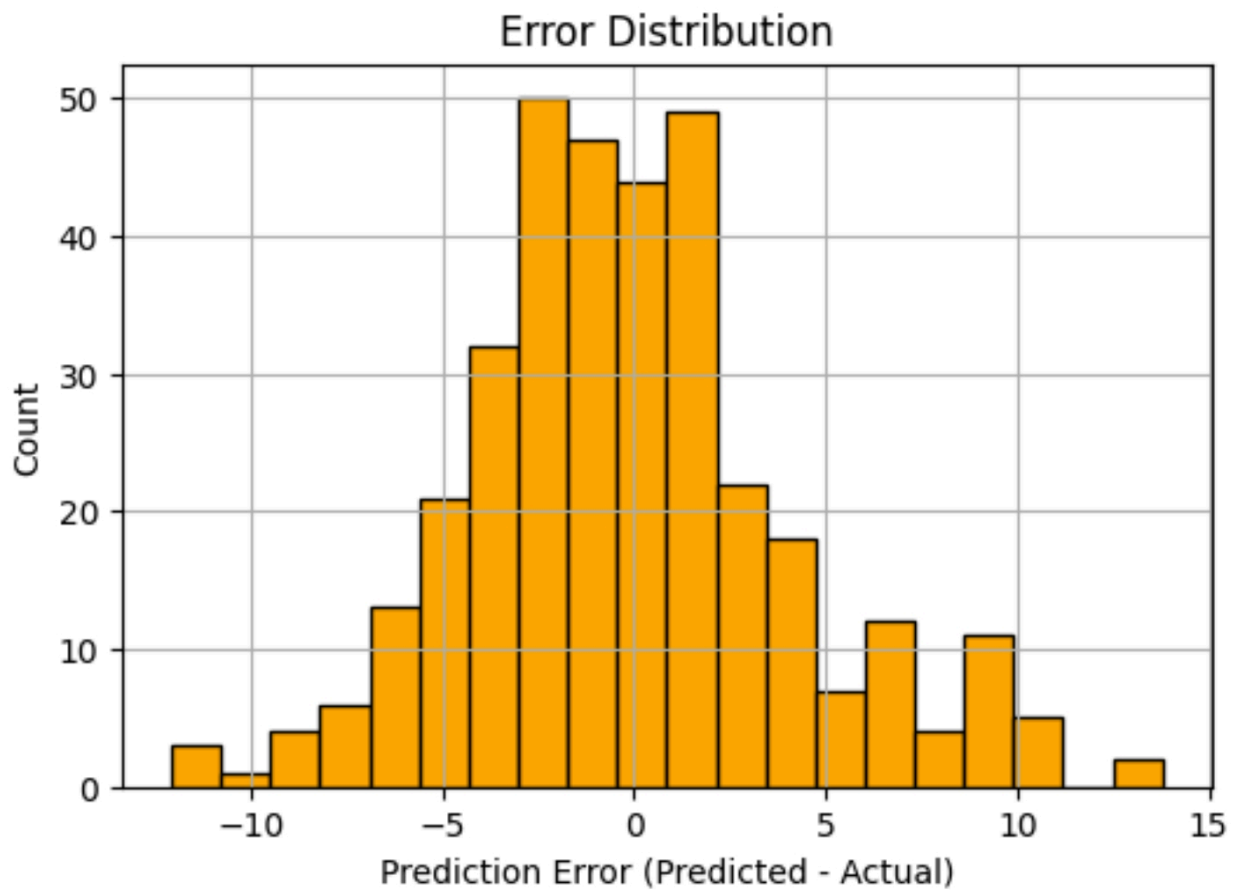


Figure 10

Histogram of prediction errors.



Figure 11

Training and validation mean squared error (MSE) curves for the DenseNet121 regression model over 20 epochs.