ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct1211

Automated Calibration-Free Neonatal Jaundice Detection Using Machine Learning

Mohamed E. Mongy[1,2*]; Ahmed Taha[2]; Mai K. Galab[2]

[1]Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt
[2]Osim Higher Institute of Administrative Science, Giza, Egypt

Corresponding Author: Mohamed E. Mongy[*]

Publication Date: 2026/10/30

Abstract: Neonatal jaundice is a common clinical condition during the first week of life and requires early detection to reduce the risk of severe complications associated with elevated bilirubin levels. Conventional diagnostic approaches, including invasive blood sampling and device-based screening, may be costly, time-consuming, or unsuitable for continuous monitoring in resource-constrained settings. This paper presents an automated, non-invasive, and calibration-free framework for neonatal jaundice detection from smartphone images. The proposed framework first detects neonatal facial regions using YOLOv8 and then applies the Segment Anything Model (SAM) to obtain accurate facial skin segmentation. The segmented regions are enhanced using Contrast Limited Adaptive Histogram Equalization (CLAHE), followed by image augmentation to improve model robustness. Color-based features are extracted from RGB, YCbCr, HSV, and LAB color spaces and evaluated using six machine learning classifiers: Multi-Layer Perceptron (MLP), Decision Tree (DT), K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). The experiments were conducted on 447 neonatal images. The proposed segmentation pipeline achieved 100% accuracy in neonatal facial region segmentation on the evaluated dataset. Moreover, a comparative evaluation was conducted across three facial regions: full face, eye, and forehead. The results showed that eye-based segmentation achieved the best classification outcomes, reaching 94.51% accuracy when combined with CLAHE, image augmentation, and the KNN classifier. Overall, the findings suggest that the proposed framework can support practical, low-cost, and calibration-free screening of neonatal jaundice using smartphone images.

Keywords: *Neonatal Jaundice, Bilirubin, Hyperbilirubinemia, Total Serum Bilirubin, Transcutaneous Bilirubin.*

How to Cite: Mohamed E. Mongy; Ahmed Taha; Mai K. Galab (2026) Automated Calibration-Free Neonatal Jaundice Detection Using Machine Learning. *International Journal of Innovative Science and Research Technology*, 11(10), 2196-2206. https://doi.org/10.38124/ijisrt/26oct1211

I. INTRODUCTION

Neonatal jaundice is a frequent disease affecting babies in their first few weeks of life [1]. According to worldwide data, it affects over 80% of preterm babies, 60% of full-term newborns, and 10% of breastfeeding infants [2]. Every year, 24 million newborns suffer from jaundice [3].

Jaundice, also known as hyperbilirubinemia, is the yellowing of the skin and sclera caused by an excess of bilirubin in the bloodstream [4]. Jaundice usually appears when blood bilirubin levels surpass 2.0 mg/dL [5]. The symptoms of newborn jaundice are caused by a buildup of bilirubin, a water-soluble yellow tetrapyrrolic pigment produced during the breakdown of red blood cells. Red blood cells in the human body have a lifetime of roughly 120 days in adults, although this is much shorter in newborns. Consequently, newborns have a greater red blood cell turnover, which causes these cells to disintegrate more quickly and raises bilirubin levels accordingly. Normally, the liver processes bilirubin secreted by damaged red blood cells, aiding its removal. Bilirubin is then excreted through urine and bile [6]. A brief excess of bilirubin is typically harmless and resolves spontaneously. However, elevated bilirubin levels in newborns can lead to neurotoxicity and irreversible brain damage, referred to as kernicterus. This condition can result in developmental delays, speech impairments, hearing loss, cerebral palsy, and, in severe cases, death [6, 7].

Early detection and treatment of jaundice or hyperbilirubinemia are critical in preventing kernicterus. Phototherapy is a commonly used treatment for reducing excessive bilirubin levels by exposing the affected infant to specific wavelengths of blue light, which convert bilirubin into a safer, excretable form [8]. However, in cases of very high bilirubin levels, exchange transfusions (ET) may be required. ET is an invasive procedure that involves replacing a portion of the infant's blood with donor blood to lower bilirubin levels. Nevertheless, phototherapy remains the most

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct1211

widely used treatment for jaundice or hyperbilirubinemia [8, 9].

There are many invasive and non-invasive procedures available for diagnosing jaundice. Total Serum Bilirubin (TSB) measurement is a typical invasive procedure that involves drawing a tiny blood sample with a syringe and needle to determine the bilirubin levels in the blood. This method relies on spectrophotometry [10]. However, due to its invasive nature, the procedure may cause pain to the newborn, risk of infection, anemia, and increased anxiety for parents [11]. Figure 1 illustrates the process for testing TSB levels.



Fig 1 Illustrates the Process for Testing TSB Levels.

Visual examination is the most basic non-invasive method for detecting hyperbilirubinemia in newborns. However, this approach is unreliable for identifying jaundice since it is dependent on the operator's perception and may be affected by the infant's skin tone [12, 13]. Transcutaneous bilirubinometer (TCB) is a non-invasive alternative to TSB measurement that has the potential to reduce the number of blood draws by around 30%. Nonetheless, TCB treatments come with large ongoing expenditures, making them quite pricey. TCB testing is only provided to healthcare professionals in hospital settings, not to the general population. Furthermore, TCB applies only to newborns who are 35 weeks or older during gestation and at least 24 hours old. As a result, it is not appropriate for newborns born before 35 weeks gestation or during the first 24 hours of life [14]. Figure 2 illustrates the process for testing TCB levels.



Fig 2 Measurement of Jaundice using TCB Measurement.

Advanced technologies such as digital cameras and smartphones have enabled the development of a more portable, cost-effective, and efficient method for non-invasive jaundice monitoring, thereby reducing stress for both parents and newborns. These technologies are readily accessible, and inexpensive, and are capable of operating as with existing monitoring systems. These are most of the methods that are based on color analysis and image processing to detect jaundice [15].

Although there has been good progress in neonatal jaundice screening using mobile devices, many current methods are based on color calibration cards or on the analysis of scleral color, which might not be suitable for actual neonatal environments. Also, strong and completely automatic segmentation is still difficult due to occlusion (e.g., caregiver's hand, pacifiers) and lighting conditions. Thus, there is a need for an accurate, inexpensive and fully automated system which does not require the use of calibration cards, yet retains the reliability of region segmentation and the consistency of classification accuracy.

➢ *This Study has Made the Following Key Contributions:*

- A fully automated non-invasive framework for neonatal jaundice detection using smartphone images, eliminating the need for color calibration cards.
- A robust facial skin segmentation pipeline based on YOLOv8 for detection and SAM for segmentation, demonstrating reliable operation under partial occlusions.
- A systematic evaluation of preprocessing strategies (CLAHE enhancement and image augmentation) and their impact on classification performance across multiple machine learning models.
- A comparative study of segmentation regions (face, eye, and forehead), showing that eye-based segmentation yields the highest accuracy (94.51%) when combined with CLAHE and augmentation using KNN.
- Extensive performance reporting using standard metrics (accuracy, precision, recall, F1-score, and specificity) to support reproducibility and fair model comparison.

The paper is organized as follows: related work is reviewed in Section 2, the proposed methodology is described in Section 3, experimental results are presented in Section 4, and conclusions are drawn in Section 5.

II. RELATED WORKS

The literature presents various methods for detecting jaundice that utilize image processing and computer vision techniques. Different approaches to color analysis have employed a range of items, including skin, sclera, urine strips, and feces. However, collecting urine and feces samples can be challenging, making these procedures impractical. Additionally, diet may influence the colors of urine and feces, rendering these systems unreliable and misleading. [17].

Somsak Leartveravat (2009) [18] performed the first research on jaundice, evaluating 61 newborns who had the condition. The researcher used Photoshop to manually compute the CMYK components and measured bilirubin levels by subtracting the M component from the Y component. Pearson's product-moment correlation and linear regression analysis revealed a high association between the Y-M value and bilirubin levels as assessed by TSB. Despite its limits in accuracy, this procedure opened the path for the development of a variety of non-invasive jaundice diagnostic techniques.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct1211

Muhammad Mansour et al. (2012) [19] used photos from a random newborn monitoring database that they found on Google. They chose photos of healthy and jaundiced newborns recorded using the Matlab image acquisition toolkit at different angles and lighting conditions. The YCrCb color space approach was used, which divides the chrominance (CrCb) and brightness (Y) components into several channels. The researchers compared the skin tones of normal and neonate neonates displaying jaundice through standard deviation, mean and kurtosis calculation and found that the results of kurtosis was more than the other two. In the next study in 2013 [20] the same researchers applied the KNN algorithm to classify newborns photos as normal or jaundiced ones. This was recommended along with a more advanced and effective method of diagnosing and treating jaundice. The model of segmentation however, are dependent on the detection techniques of YCrCb and HSV, which reduces the potential of achieving 100% efficiency at segmenting all pictures. This dependency can lead to insufficient segmentation and hence affect the accuracy of the proposed method.

In 2015, Terence Leung et al. [21] Developed a technique for detecting infant jaundice using scleral pictures. This method was used to measure bilirubin levels based on the color of the sclera. The experimental study included 110 newborns and proved the method's potential as a screening tool for jaundice.

A study by Mustafa Aydın et al. (2016) [22] evaluated 80 images of newborns, half of the images were normal and half jaundiced, taken using a smartphone camera. The scientists employed an AI algorithm called a 'classifier' to achieve color balance and a segmentation strategy, which involved calibration of a specific area of the baby's skin using an 8 colored card. They changed the color of the baby's skin and got information from the calibration card in RGB, YCrCb and LAB color spaces. Then, the kNN and SVR methods were used to calculate the bilirubin levels. These AI approaches exhibited quicker processing speeds and better results.

Alex Mariakakis et al. (2017) [23] created an eye photo-Bili metering app for a smartphone that was used to capture photos of the eye and calculate the bilirubin concentration, even when the level is too high to be detected by the naked eye. In a clinical investigation of 70 subjects, the mean error of BiliScreen with attachment was -0.09 ± 2.76 mg/dl with a Pearson correlation value of 0.89 when predicting bilirubin levels. The app has a sensitivity of 89.7% and a specificity of 96.8% and is a useful screening tool to detect potentially troublesome instances.

In this regard, a computer-based sclera image-screening method for NBJ was proposed by Felix Outlaw et al. (2017) [24] which estimates TSB by using the RGB pixel values of raw images of the sclera. A study at the UCH Neonatal Unit revealed a very strong correlation ($r = 0.71$, $p < 0.01$) between the values obtained from the method and the actual values for TSB.

The method was tested on a relatively small, but demographically narrow sample of 64 Caucasian newborns and 11 of whom had TSB levels exceeding 205 μmol/L. Furthermore, the assessment was restricted to an iPhone 6 camera configuration, which might impact generalizability to other cameras, populations, and imaging settings.

Endang Juliastuti et al. 2019 [26] created the approach to determine the risk of jaundice by taking baby pictures using a digital camera. The researchers have captured 120 images and extracted RGB, HSV and YCbCr color space values as input for linear regression modelling and validation. The method was accurate and achieved an accuracy of 67%. It was however found to be limited by using K-Means clustering for segmentation, where non-skin regions are also included in the analysis. Furthermore, a color correction card was used with this method.

In 2019, Pouria Padidar et al. [3] introduced a new mobile application for detecting neonatal jaundice, which was developed using the Android operating system and detected the disease through the images captured by the smartphone. 113 images of newborns were used in the study as the goal was to develop a portable and easily accessible screening tool. The method, however, assumed the use of a forehead calibration card, and could be influenced by the number of images used in the training set and the variability in lighting conditions when the images were acquired.

Fırat Hardalaç et al. (2021) [27] proposed a simple algorithm using correlation analysis and multiple-input single-output (MISO) regression model for bilirubin prediction that is low-cost. A total of 196 subjects were included in the study and two-class classification accuracy of the proposed method was found to be 92.5% on test group. The method did, however, still need to be calibrated with a calibration card on the infant's skin, potentially reducing its utility for screening at home or routinely.

Alhanoof Althnian et al. (2021, [16]) used composite skin and eye images to detect neonatal jaundice by using deep transfer learning and a few classical machine learning models such as MLP, SVM, DT and RF. Additionally, the study has performed statistical analysis of the performance of the study (accuracy, precision, recall, F-score and AUC). The results indicated that classical models achieved better performance for eye and merged features; transfer learning model demonstrated better performance on skin images. But the method required a calibration card, and the face recognition stage was not successful on some of the occluded images, such as images with a mother's hand or a pacifier. Moreover, sclera-based acquisition can be challenging to be consistently applied on the newborns.

Warqaa Hashim et al., 2021 [28] suggested a computer vision-based GUI system for neonatal jaundice detection. The system consisted of skin color analysis, ROI selection, color range selection and practical acquisition setup with the help of digital camera and micro controller. Conditional rules for classification were used in the implementation in MATLAB. The results obtained were reasonably accurate, but the results

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct1211

were limited by the manual ROI selection and the controlled illumination conditions.

Supaporn Dissaneevate et al. (2022) [29] presented a cross-platform mobile computer-aided diagnosis system to diagnose neonatal hyperbilirubinemia based on data mining and image processing methods. The study included 178 pictures of infants' skin taken by two smartphones and one DSLR camera. The RGB, HSV, LCH, and LAB colour space models were extracted and applied to the classifiers DT, kNN and CNN. CNN outperformed other models with respect to the performance of classification. The method was still limited by the relatively small number of neonatal skin images, and the computational heavy burden of CNN-based classification.

Dhananjay Kalbande et al. (2022) [30] presented a deep learning-based approach to classifying jaundice from eye images. The study compared ResNet50 with Mask-RCNN and obtained better results for segmenting and identifying jaundiced eye images with Mask-RCNN. But the method was limited to eye images and the deep learning models used were computationally expensive and could have an impact on real-time implementation.

Bhagavathi Sreedha et al. (2023) [31] proposed a computer vision-based method for jaundice detection using machine learning models, color detection, mask segmentation, and image augmentation. The study proposed the use of weighted average yellowness (WAY) measure for quantifying the yellow color from scleral images in the CMYK color space. Nevertheless, the method was tested on 35 eye images from the Internet, and it is not possible to determine the generalizability of the results.

Jyoti Prajapati et al. (2023) [32] developed the jScan application for the estimation of bilirubin level using images of the sclera using the smartphone. The system comprised of sclera segmentation, extraction of RGB index values and extraction of yellow index values, prediction of bilirubin using a regression model. The results were promising, but the application was meant for adults, and needed more data to be more accurate in predicting the results, which made it not directly applicable to neonatal screening.

Although numerous significant advances have been achieved in neonatal jaundice detection based on image processing, computer vision and machine learning techniques, several recurring limitations can be pinpointed from the literature. First, several methods were based mainly on the analysis of the sclera or eyes for jaundice detection [21, 23, 24, 30–32]. While the sclera can be used to give clues to bilirubin-related yellow discoloration, eye opening may be challenging in neonatal screening when neonates are not able to keep their eyes open during image acquisition. This reduces the uniformity and applicability of the purely eye-based techniques in actual neonatal setting.

Second, several studies relied on color calibration card or external reference objects to standardize the process of image acquisition and minimize color variation [22]–[24], [26], [3]–[27], [16]. Calibration based methods can help to normalize the colors, but they decrease the practicality of the system, since they need to have other physical materials at the moment of shooting the image. Close proximity of such objects to the infant may also cause hygiene concerns, and restrict the feasibility of the approach in terms of its scalability for routine or home-based screening in neonatal settings.

Third, in several previous studies [25, 3, 29, 31, 32] the size of the datasets and generalizability are still significant problems. Several methods tested in this study were based on limited data sets, limited demographic populations, images taken from the Internet, or a specific camera model. All these factors might affect the reliability of the reported results when the system is applied to other skin tones, imaging devices, lighting conditions, and clinical environment.

Furthermore, there are several existing methods which are still difficult to achieve strong and automatic region segmentation. Manual or semi-automatic region-of-interest selection, traditional color-based segmentation, face/eye detection procedures that fail under practical imaging conditions [19],[20],[26],[28],[16],[32] were utilized for some studies. Segments can be occluded by hands, pacifiers, shadows or uncontrolled lighting by the caregivers, which can lead to a degradation of the segmentation quality and hence of the classification performance. In addition, the majority of previous studies examined a limited area of the face (e.g., sclera or skin), but have not performed a systematic comparison between various facial regions using the same preprocessing and classification methods.

To overcome this, in the present study, a new framework which is fully automated and calibration free for the detection of neonatal jaundice from smartphone images is proposed. This work combines the YOLOv8 model for neonatal face detection, the SAM for precise facial skin segmentation, CLAHE-based Contrast Enhancement, image augmentation, extraction of features in multiple color spaces, and comparative classification using machine learning. This work is unique in that it compares three facial regions (full face, eye, and forehead), while keeping all of the experimental parameters the same. This enables a fair and reproducible comparison of the discriminative ability of each region and demonstrates the best classification performance of the proposed framework based on eye-based segmentation.

III. MATERIAL AND METHODS

➢ *The Following are Six Steps of the Proposed Model:*

- Dataset Collection: Collect a wide variety and a representative set of facial images. Several facial expressions, orientations and lighting conditions should be taken into account, to make the model robust and generalizable.
- Face Detection and Segmentation: Face detection and segmentation algorithms are applied to images to detect and segment faces. This procedure ensures that only those facial features are analyzed that are relevant to the task.
- Contrast Enhancement – This is where the CLAHE technique is used to improve image quality. This

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct1211

enhancement emphasizes facial details and features, creating a more pronounced and detailed appearance.

- Image Augmentation: This step involves rotating, scaling, flipping and color changes to the image. Artificial expansion of diversity and size of the data set, which helps to prevent overfitting in the model.
- Feature Extraction: Various color spaces like RGB, YCbCr, Lab and HSV are used to extract the features from the enhanced images that are relevant. The multi-faceted method enables to preserve a variety of features that are essential for the correct classification.
- Finally, features extracted are classified using several machine learning algorithms like MLP, KNN, RF, SVM, DT, and XGBoost. This step is designed to build a strong model that is able to detect or classify human faces in accordance with the features extracted in the previous steps.

A block diagram of the proposed method is shown in figure 3. The approach starts with face detection and segmentation of the input images. These segmented face images are then passed through CLAHE image enhancement to generate the enhanced images and then passed through the image augmentation to increase the number of images. Features extracted from different color space, such as RGB, YcbCr, Lab, and HSV are done for the augmented images. The features extracted are subsequently classified using machine learning methods like MLP, KNN, RF, SVM, DT and XGBoost. Further details of each stage will be elaborated in the following subsections.

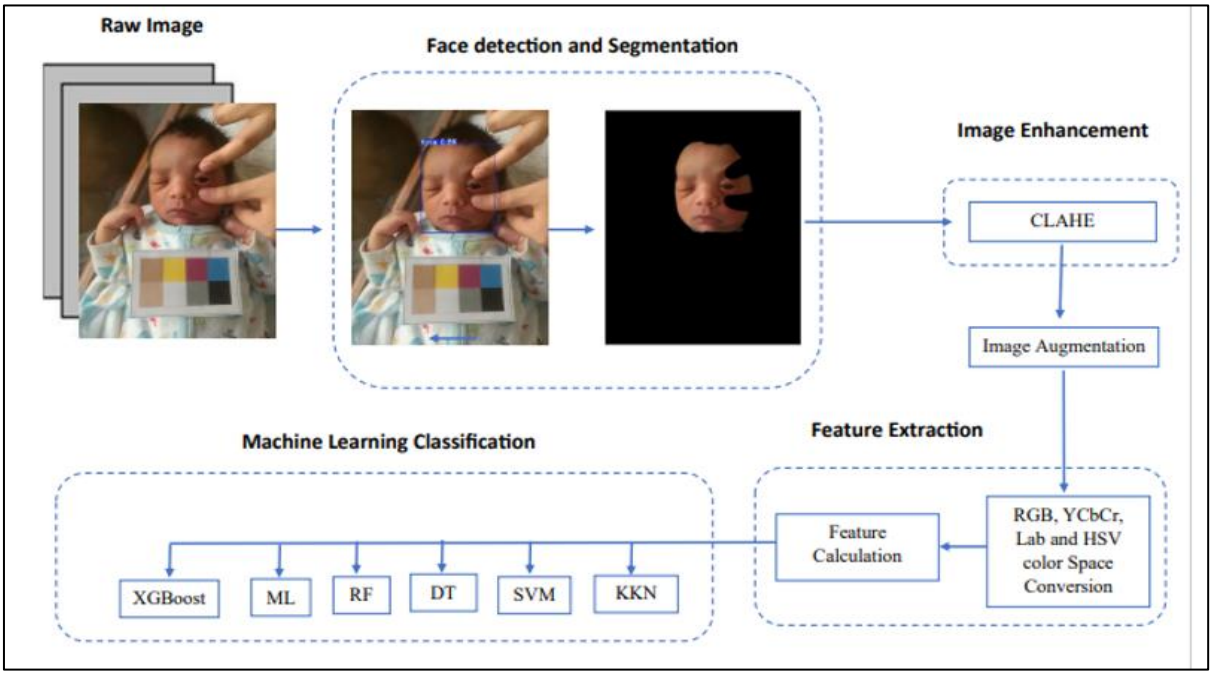


Fig 3 Block Diagram of Proposed Method

➢ *Dataset Collection*

The proposed method is based on a dataset of images acquired from Method [16], specifically designed for jaundice prediction. These images were captured at KKUH in Riyadh, Saudi Arabia, using the built-in camera of a Samsung S7 smartphone. The dataset comprises 62 male and 38 female neonates, of whom 67% were healthy and 33% were jaundiced. The average gestational age of the newborns was 38 weeks, and their average age at the time of capture was 1 day. During feature extraction, the average TSB level was 231 Mmol/m. The dataset consists of 447 normal and jaundiced infant images used for training machine learning models. The training tests were conducted on 282 normal and 165 jaundiced infants.

➢ *Face Detection and Segmentation*

The first step in extracting features from neonates was to detect and segment their faces. To achieve this, the pretrained face detector YOLOv8 [33-35] was utilized for face detection. Additionally, the SAM Model [36] was used to segment the faces of neonates.

- *The YOLOv8 Model*

A basic inference and training interface was developed using YOLOv8 for neonatal face detection. The model outputs bounding boxes, confidence scores, and class labels for each detected face. As illustrated in Figure 4, YOLOv8 successfully detects the neonatal face even in the presence of partial occlusion caused by the caregiver's hand, achieving a confidence score of 0.86.



Fig 4 Face Detection by Yolo v8.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct1211

- *SAM Model*

The face was detected by using YOLOv8 and the facial area was then segmented by SAM. First a segmentation technique based on threshold was tested but it did not separate the skin region of the mother and the face of the newborn because there was mixed information in the intensity value image. LangSAM was subsequently tested on the test sets, and found to be somewhat better, but to lack the ability to give a good delineation of fine boundaries and shaded areas.

SAM was eventually chosen because of its high segmentation accuracy, ability to deal with shadows and accurate boundary extraction. In Figure 5, the model is able to detect the neonatal face and segment the head region even if the face is partially covered by the caregiver's hand. The final facial skin segmentation result using SAM is shown in Figure 6, where only the neonatal face is kept, and the background areas are eliminated.

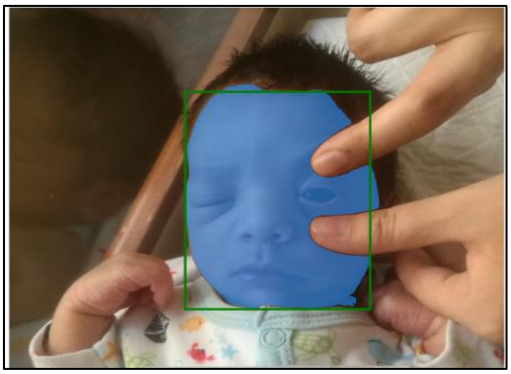


Fig 5 From Yolo v8 Using the SAM Model to Detect Neonates' Face



Fig 6 Segmentation Face by SAM Model.

➢ *Contrast Enhancement*

During this step, the photo is strengthened to create skin features and faces more noticeable. A number of contrast enhancement techniques were tested and evaluated to find the best technique for neonatal facial images taken under different illumination conditions.

Histogram equalization, color constancy, dynamic-threshold white balancing and Contrast Limited Adaptive Histogram Equalization (CLAHE) [38] were tested techniques. Histogram equalization can be useful to increase the contrast overall, but can also over-contrast areas of low contrast. Color constancy was effective at reducing color imbalance, but not so effective under extreme lighting changes, and at times produced color casts in more complex lighting situations, dynamic threshold white balance. The decision was made based on qualitative and visual comparison, which resulted in selection of CLAHE as the one providing a good balance between contrast enhancement, color preservation and detail retention without introducing artifacts of over amplification. The difference between the segmented images and CLAHE enhanced images for normal and jaundiced cases are shown in Figure 7, which shows that the contrast and facial details were enhanced after applying the CLAHE. The histograms before and after applying CLAHE are shown in figure 8, where the effects of the redistribution of pixel intensities and the improvement in image contrast are clearly visible.

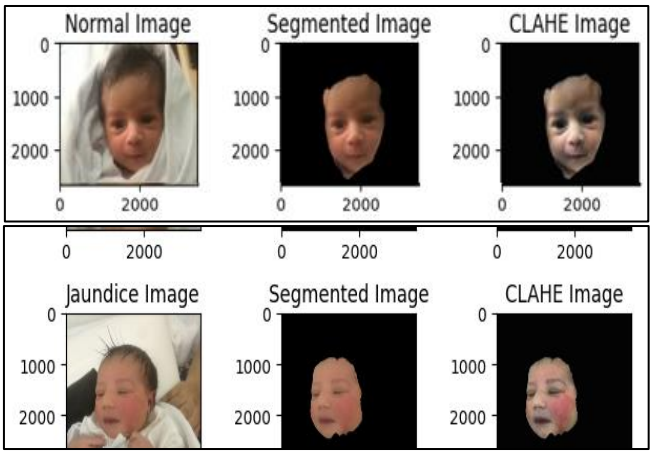


Fig 7¨Difference Between Images Segmented and CLAHE Images (Normal and Jaundice Images)

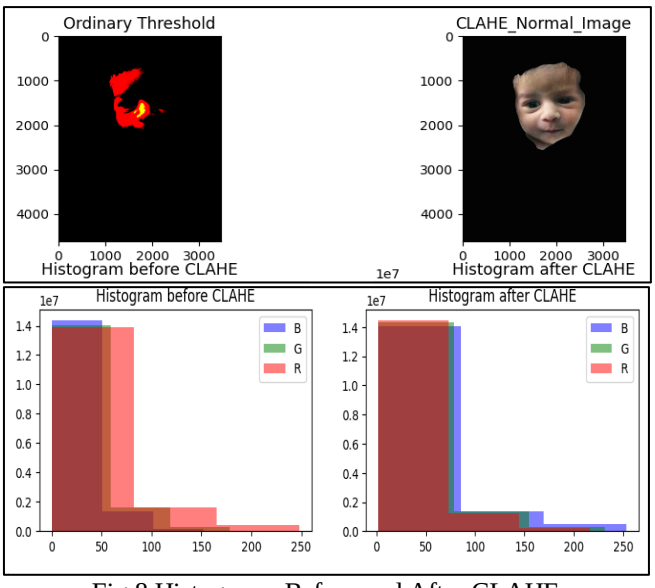


Fig 8 Histograms Before and After CLAHE

➢ *Image Augmentation*

In this step, the objective is to increase the diversity of the training data and improve model robustness while reducing overfitting. Image augmentation is applied to generate additional samples from the existing dataset, eliminating the need for manual data collection.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct1211

The augmentation process includes cropping up to 20% of the image area, rotating images within a range of −45° to 45°, vertical flipping with a probability of 0.5, and applying Gaussian blur with a sigma value ranging from 0 to 1. Figure 9 presents examples of the augmented images, illustrating the visual variations introduced through the augmentation process.

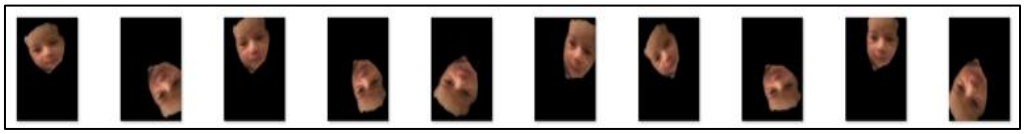


Fig 9 Samples of Image Augmentation.

➢ *Feature Extraction*

In this step, relevant features are extracted from the augmented images using multiple color spaces. Initially, color features are extracted from the RGB color space and subsequently converted to the YCbCr, Lab, and HSV color spaces. This multi–color space representation enables the capture of diverse chromatic characteristics relevant to neonatal jaundice detection.

Figure 10 illustrates examples of the extracted features, where part (a) represents feature samples from normal images and part (b) represents feature samples from jaundiced images.

	Filename	Y	Cb	Cr	H	S	V	R	G	B	L	a	b	Class
0	augmented_1-100_0.png	22.163727	131.930998	124.642366	2.269560	19.909515	27.668804	16.216102	20.512581	27.664334	23.543676	130.243308	131.477973	normal
1	augmented_1-100_1128.png	22.665616	132.020205	124.566919	2.340469	20.314484	28.294916	16.583502	20.977291	28.290795	24.076109	130.294355	131.556871	normal
2	augmented_1-100_1410.png	30.544686	133.415950	123.373568	3.132341	27.387134	38.128893	22.349620	28.269241	38.122843	32.445111	131.090740	132.791667	normal
3	augmented_1-100_1692.png	20.188199	131.687202	124.914374	2.016255	18.429709	25.353317	14.720982	18.616367	25.349567	21.441182	130.144961	131.201687	normal
4	augmented_1-100_1974.png	19.296379	131.423118	125.076960	1.987633	17.298614	24.089650	14.117723	17.858830	24.086089	20.497476	129.953586	131.028585	normal

(A)

	Filename	Y	Cb	Cr	H	S	V	R	G	B	L	a	b	Class
0	augmented_1-11_0.png	18.506671	132.265406	125.072695	1.529789	18.171996	24.486664	13.331475	16.471574	24.482489	19.666551	130.797126	131.046406	jaundice
1	augmented_1-11_1155.png	21.702491	133.131856	124.527056	1.731193	21.674405	28.897616	15.565145	19.236409	28.892774	23.076664	131.397453	131.621458	jaundice
2	augmented_1-11_1320.png	22.116912	133.096526	124.501286	1.817421	21.755799	29.262847	15.933210	19.685142	29.257372	23.503491	131.341721	131.639662	jaundice
3	augmented_1-11_1485.png	23.597427	133.438188	124.267872	1.955855	23.163862	31.221516	16.999467	21.002781	31.216316	25.076102	131.566281	131.883801	jaundice
4	augmented_1-11_165.png	24.060102	133.544405	124.193952	1.977710	23.643620	31.833862	17.332882	21.414617	31.828023	25.568148	131.635587	131.959623	jaundice

(B)

Fig 10 Feature Sample: (a) Features from Normal Images vs. (b) Features from Jaundice Images

➢ *Machine Learning Classification*

In this step, the extracted color features are classified using multiple classical machine learning models. Six classifiers are evaluated, namely MLP, KNN, SVM, DT, RF, and XGBoost.

These models were selected to provide a comprehensive comparison between different learning paradigms, including distance-based, tree-based, ensemble-based, and neural network approaches. All classifiers were trained and evaluated using the same feature sets and experimental settings to ensure a fair comparison of their performance.

➢ *Evaluation Metrics*

The performance of the proposed models was evaluated using five standard metrics: accuracy, precision, recall (sensitivity), F1-score, and specificity. These metrics were selected to provide a comprehensive assessment of classification performance, particularly for medical diagnosis tasks.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct1211

$$\text{Accuracy} = (TP + TN) / (TP + FP + TN + FN) \quad (1)$$

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

$$F1 = (2 \times (\text{precision} \times \text{recall}))/(\text{precision} + \text{recall}) \quad (4)$$

$$\text{Specificity} = TN / (TN + FP) \quad (5)$$

Where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section evaluates the performance of six machine learning models—MLP, SVM, KNN, RF, DT, and XGBoost—for classifying neonatal images into normal and jaundice classes. Feature vectors extracted from RGB, YCbCr, Lab, and HSV color spaces were used for model training and evaluation. Model performance was assessed using accuracy, precision, recall (sensitivity), F1-score, and specificity, as summarized in Table 1.

Table 1 Shows the Amount of Data Divided into Training and Testing, and Performance of the Machine Learning Models.

Split data	Metrics Technique	Accuracy	Precision	F1 Score	Recall = Sensitivity	Specificity
80% for training 20% for testing	MLP	80.92%	81.02%	85.57%	90.66%	64.77%
	KNN	93.09%	93.79%	94.5%	95.22%	89.55%
	RF	89.53%	91.36%	91.63%	91.91%	85.59%
	SVM	72.52%	71.55%	80.83%	92.88%	38.74%
	DT	85.67%	87.87%	88.61%	89.35%	79.55%
	XGBOOST	88.92%	89.35%	91.31%	93.37%	81.53%
85% for training 15% for testing	MLP	83.69%	83.93%	87.59%	91.59%	70.35%
	KNN	93.22%	93.61%	94.67%	95.76%	88.94%
	RF	90.06%	91.2%	92.18%	93.17%	84.81%
	SVM	71.82%	71.53%	80.33%	91.59%	38.4%
	DT	86.99%	88.81%	89.76%	90.73%	80.68%
	XGBOOST	88.8%	88.85%	91.33%	80.07%	80.07%
90% for training 10% for testing	MLP	78.93 %	77.65%	84.55%	92.8%	56.17%
	KNN	93.77%	94.12%	95.03%	95.97%	90.16%
	RF	90.04%	91.31%	92.05%	92.8%	85.51%
	SVM	71.34%	70.62%	80.0%	92.26%	37.03%
	DT	87.13%	88.55%	89.78%	91.06%	80.68%
	XGBOOST	88.89%	89.02%	91.29%	93.68%	81.04%
93% for training 7% for testing	MLP	84.22%	84.51%	88.04%	91.88%	71.05%
	KNN	94.0%	93.78%	95.33%	96.94%	88.95%
	RF	89.55%	91.99%	91.71%	91.42%	86.32%
	SVM	72.12%	71.91%	80.62%	91.73%	38.42%
	DT	87.51%	90.18%	90.11%	90.05%	83.16%
	XGBOOST	88.58%	89.4%	91.14%	92.96%	81.05%
94% for training 6% for testing	MLP	84.31%	87.37%	87.75%	88.14%	77.57%
	KNN	94.7%	94.81%	95.89%	96.99%	90.65%
	RF	89.62%	91.71%	91.87%	92.04%	85.36%
	SVM	73.14%	72.93%	81.38%	92.04%	39.88%
	DT	84.99%	89.13%	88.09%	87.08%	81.31%
	XGBOOST	88.83%	89.49%	91.43%	93.45%	80.69%
95% for training 5% for testing	MLP	83.33%	82.09%	87.74%	94.22%	64.58%
	KNN	94.85	94.59%	95.99%	97.43%	90.41%
	RF	89.84	91.7%	92.0%	92.29%	85.61%
	SVM	73.17	72.53%	81.39%	92.72%	39.48%
	DT	87.53%	89.64%	90.21%	90.79%	81.92%
	XGBOOST	88.35%	89.6%	90.93%	92.29%	81.55%
96% for training 4% for testing	MLP	68.87%	73.86%	75.98%	78.23%	52.97%
	KNN	94.08%	93.77%	95.38%	97.04%	89.04%
	RF	89.85%	91.27%	92.0%	92.74%	84.93%
	SVM	72.59%	72.34%	80.76%	91.4%	40.64%
	DT	87.31%	88.98%	90.04%	91.13%	80.82%
	XGBOOST	88.32%	89.15%	90.91%	92.74%	80.82%

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct1211

Table 1 gives the performance of the six machine learning classifiers for various training and testing split ratios. The results indicate that KNN was the best in terms of the consistency and balance of performance in all the experimental conditions. This superiority remained even if the ratio of the training set to the testing set was changed from 80%/20% to 96%/4%, which showed that KNN was robust to different data-splitting ratios. The highest classification accuracy was achieved by KNN for every split ratio than other models like MLP, RF, SVM, DT, and XGBoost. The most favorable performance was obtained when the training set is composed of 95% of the data and the testing 5%: KNN achieved an accuracy of 94.85%, a precision of 94.59%, an f1 score of 95.99%, a recall of 97.43% and a specificity of 90.41%.

But KNN is not only superior in terms of accuracy, it is also superior in terms of the other evaluation measures. KNN also achieved a very high recall and specificity simultaneously, which shows its ability to identify the cases of jaundice and also correctly identify the normal cases. RF and XGBoost showed promising outcome and their performance across the evaluated splits was competitive but was generally not as good as KNN. SVM was able to give a good recall for some configurations however its low specificity shows a poor performance in correctly identifying normal cases. Likewise, MLP was not as stable, particularly for the case where the test portion was very small. Thus, comparative results have shown that KNN was found to be an appropriate classifier for the proposed neonatal jaundice detection system based on color features.

From clinical point of view, the high recall achieved with KNN becomes significant in the context of neonatal jaundice screening as it is important to minimize false negative cases. So, the proposed framework can be viewed as screening support tool which can be used to identify a newborn who should be assessed further by standard bilirubin measurement methods, e.g., TSB or TCB. But the system proposed here is not for replacing the clinical diagnosis but for providing simple and calibration-free preliminary screening.

The confusion matrices of the six classifiers are shown in Fig. 11, which relate the distribution of true positives, true negatives, false positives and false negatives among the classifiers.

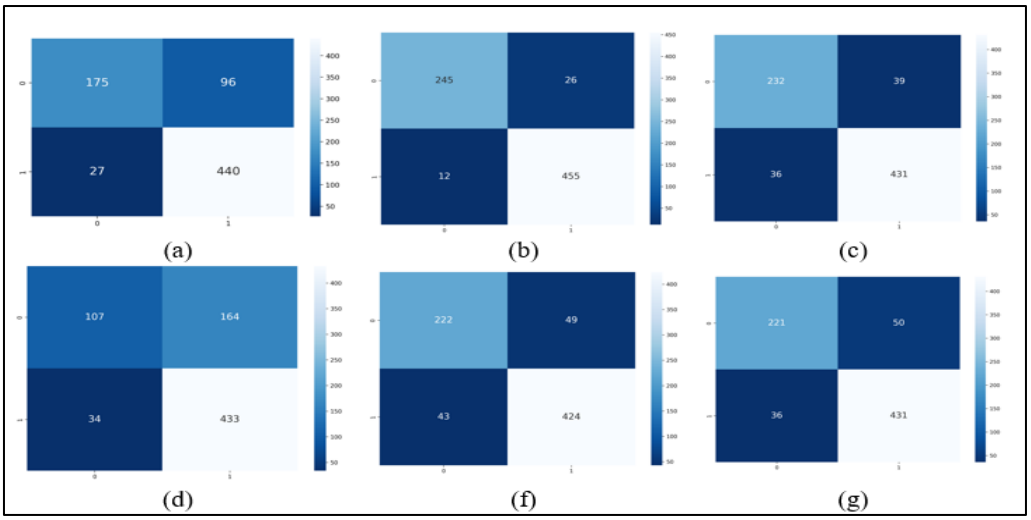


Fig 11 The Confusion Matrix Using (a) MLP, (b) KNN, (c) RF, (d) SVM, (f) DT, (g) XGBOOST Technique.

➢ *Comparison Across Facial Regions (Face, Eye, and Forehead)*

Apart from comparing various classifiers and preprocessing pipelines, a regional analysis was performed to check the effect of the chosen facial region on the performance of the neonatal jaundice detection. Three areas were studied: full facial skin, eye area, and forehead. The experimental conditions were the same in each region, such as contrast enhancement using CLAHE, image augmentation and machine learning classification.

The results show that eye-based segmentation was the most successful with a maximum accuracy of 94.51% obtained by the KNN classifier. Full-face segmentation also yielded good results, with 93.09% accuracy for KNN when tested in the same experimental setup. However, the results of forehead segmentation were not as promising as those obtained through the segmentation of the eyes, where the highest accuracy was 84.23% achieved by KNN. This indicates that the eye area had better discriminative color information for neonatal jaundice detection and also the forehead area was more sensitive to the illumination variation and segmentation noises.

In all the three evaluated facial regions, the accuracy obtained from KNN was better than the other classifiers. RF and XGBoost gave relatively good results, particularly in the eye region, but they were not as good as KNN. The performance of SVM was the lowest among all the regions, which means the color space features extracted in this experiment are not closer to the SVM to use. The classification accuracy achieved for the three facial regions with the six classical machine learning classifiers, MLP, KNN, RF, SVM, DT and XGBoost is summarized in Table 2.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct1211

Table 2 Lassification Accuracy (%) Obtained After Applying the Proposed Methodology to Different Facial Regions (face, eye, and Forehead) using Various Machine Learning Classifiers.

Face segmentation result:	Eye segmentation result:	Forehead segmentation result:
Accuracy: MLP: 81.80 %	Accuracy: MLP: 79.13 %	Accuracy: MLP: 77.54 %
KNN: 93.09 %	**KNN: 94.51 %**	**KNN: 84.23 %**
RF: 89.52 %	RF: 91.89 %	RF: 84.08 %
SVM: 71.68 %	SVM: 69.27 %	SVM: 67.86 %
DT: 86.77 %	DT: 88.69 %	DT: 76.12 %
XGBOOST: 87.53 %	XGBOOST: 90.15 %	XGBOOST: 81.47 %

V. CONCLUSION

This method introduced a color-calibration card-free, complete automatic system for detecting neonatal jaundice using the smartphone camera images. The proposed method combines powerful facial detection with YOLOv8 and fine segmentation with SAM, which allows for accurate facial skin extraction even in the presence of partial occlusion and changeable lighting. The segmentation pipeline was able to segment the facial regions in neonatal images with 100% accuracy, demonstrating the robustness and reliability of the segmentation pipeline.

Various preprocessing methods such as CLAHE based contrast enhancement and image augmentation were applied to the images, with the use of multiple classical machine learning classifiers to get a comprehensive evaluation. Furthermore, the comparison analysis between various parts of the face (face, eye and forehead) was carried out under the same experimental conditions. It was observed that eye-based segmentation showed the highest performance with a classification accuracy of 94.51% with the use of the CLAHE enhancement and the image augmentation technique using KNN classifier.

The KNN model outperformed the others machine learning models, and was the most accurate model (maximum accuracy of 94.85%) compared with the others models used in the related work. The findings indicate that effective segmentation, adequate pre-processing and traditional machine learning methods can effectively detect neonatal jaundice.

The proposed technique is generally an easy and cost-effective way of detecting jaundice early, which can be used both in the clinic and at home. Suitable testing of the method with larger and more diverse datasets and the feasibility of its real-time deployment on mobile devices is required.

ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Egypt, for academic support and valuable guidance. This research did not receive any external funding.

REFERENCES

[1.] L. M. Gartner, “Neonatal jaundice,” Pediatrics in Review, vol. 15, no. 11, pp. 422–432, 1994.

[2.] J. Rennie, S. Burman-Roy, and M. S. Murphy, “Neonatal jaundice: Summary of NICE guidance,” BMJ, vol. 340, 2010.

[3.] P. Padidar et al., “Detection of neonatal jaundice by using an Android OS-based smartphone application,” Iranian Journal of Pediatrics, vol. 29, no. 2, 2019.

[4.] F. A. Dzulkifli, M. Y. Mashor, and K. Khalid, “Methods for determining bilirubin level in neonatal jaundice screening and monitoring: A literature review,” J. Eng. Res. Educ., vol. 10, pp. 1–10, 2018.

[5.] P. Puppalwar, K. Goswami, and A. Dhok, “Review on evolution of methods of bilirubin estimation,” IOSR Journal of Dental and Medical Sciences, vol. 1, no. 3, pp. 17–28, 2012.

[6.] W. Chee et al., “Jaundice in infants and children: Causes, diagnosis and management,” Hong Kong Medical Journal, 2018.

[7.] S. Ullah, K. Rahman, and M. Hedayati, “Hyperbilirubinemia in neonates: Types, causes, clinical examinations, preventive measures and treatments,” Iranian Journal of Public Health, vol. 45, no. 5, pp. 558–568, 2016.

[8.] K. Mreihil et al., “Phototherapy is commonly used for neonatal jaundice but greater control is needed to avoid toxicity in the most vulnerable infants,” Acta Paediatrica, vol. 107, no. 4, pp. 611–619, 2018.

[9.] B. Katti, “Analytical review on technologies designed for detection and treatment of neonatal jaundice,” Int. J. Sci. Res., vol. 8, p. 2518, 2021.

[10.] A. R. Deepthi, C. Singh, and J. B. Jeeva, “A device to detect hyperbilirubin through non-invasive technique,” in Proc. i-PACT, 2019.

[11.] R. Singla and S. Singh, “A framework for detection of jaundice in newborn babies using homomorphic filtering,” in Proc. ICICT, 2016.

[12.] R. Keren et al., “Visual assessment of jaundice in term and late preterm infants,” Arch. Dis. Child. Fetal Neonatal Ed., vol. 94, no. 5, pp. F317–F322, 2009.

[13.] A. Riskin et al., “Is visual assessment of jaundice reliable as a screening tool?” J. Pediatrics, vol. 152, no. 6, pp. 782–787, 2008.

[14.] C. Aranda, D. Torrubia, and L. Álvarez, “Correlation of serum, capillary and transcutaneous bilirubin,” Biomed. J. Sci. Tech. Res., vol. 1, no. 3, pp. 722–726, 2017.

[15.] K. Srividya et al., “Review on jaundice detection in neonates using image processing,” in Proc. IC3IoT, 2022.

[16.] A. Althnian, N. Almanea, and N. Aloboud, “Neonatal jaundice diagnosis using a smartphone camera,” Sensors, vol. 21, no. 21, 2021.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct1211

[17.] W. Hashim et al., “Neonatal jaundice detection using a computer vision system,” Designs, vol. 5, no. 4, p. 63, 2021.
[18.] S. Leartveravat, “Transcutaneous bilirubin measurement by digital camera,” Medical Journal of Srisaket Surin Buriram Hospitals, vol. 24, no. 1, pp. 105–118, 2009.
[19.] M. N. Mansor et al., “Jaundice monitoring using color detection,” Procedia Engineering, vol. 29, pp. 1631–1635, 2012.
[20.] M. Mansor et al., “New newborn jaundice monitoring scheme,” Neurocomputing, vol. 120, pp. 258–261, 2013.
[21.] T. S. Leung et al., “Screening neonatal jaundice based on sclera color,” Biomedical Optics Express, vol. 6, no. 11, pp. 4529–4538, 2015.
[22.] M. Aydın et al., “Neonatal jaundice detection system,” Journal of Medical Systems, vol. 40, pp. 1–11, 2016.
[23.] A. Mariakakis et al., “BiliScreen: Smartphone-based scleral jaundice monitoring,” Proc. ACM IMWUT, vol. 1, no. 2, Art. no. 20, 2017.
[24.] F. Outlaw et al., “Screening for neonatal jaundice with a smartphone,” in Proc. Digital Health Conf., 2017.
[25.] S. B. Munkholm et al., “Smartphone camera for transcutaneous bilirubin,” PLOS One, vol. 13, no. 6, p. e0197938, 2018.
[26.] E. Juliastuti et al., “Risk zone estimation of newborn jaundice,” in Proc. ICA, IEEE, 2019.
[27.] F. Hardalac et al., “Classification of neonatal jaundice in mobile applications,” Turkish J. Electr. Eng. Comput. Sci., vol. 29, no. 4, pp. 2116–2126, 2021.
[28.] W. Hashim et al., “Computer vision for jaundice detection,” in IOP Conf. Series: Mater. Sci. Eng., 2021.
[29.] S. Dissaneevate et al., “Mobile computer-aided diagnosis of neonatal hyperbilirubinemia,” Int. J. Innov. Res. Sci. Stud., vol. 5, no. 1, pp. 10–17, 2022.
[30.] D. Kalbande et al., “Deep learning approach for early diagnosis of jaundice,” in Proc. ICICC, vol. 3, Springer, 2022.
[31.] B. Sreedha, P. R. Nair, and R. Maity, “Non-invasive early diagnosis of jaundice,” Procedia Computer Science, vol. 218, pp. 1321–1334, 2023.
[32.] J. Prajapati et al., “jScan: Smartphone-assisted bilirubin quantification,” IEEE Sensors Journal, 2023.
[33.] J.-H. Kim, N. Kim, and C. S. Won, “High-speed drone detection based on YOLOv8,” in Proc. ICASSP, IEEE, 2023.
[34.] D. Qi et al., “YOLO5Face: Why reinventing a face detector,” in ECCV, Springer, 2022.
[35.] W. Yang and Z. Jiachun, “Real-time face detection based on YOLO,” in Proc. ICKII, IEEE, 2018.
[36.] A. Kirillov et al., “Segment anything,” in Proc. IEEE/CVF ICCV, 2023.
[37.] J. Redmon et al., “You only look once: Unified, real-time object detection,” in Proc. CVPR, 2016.
[38.] A. M. Reza, “Realization of CLAHE for real-time image enhancement,” J. VLSI Signal Process., vol. 38, pp. 35–44, 2004.