Avances en Interacción Humano-Computadora, 2024, Volumen 9, No. 1, pp. 179-184
DOI: 10.47756/aihc.y9i1.164

Sclera Segmentation in Images for Bilirubin Level Measurement Using the U-Net Network

Diana A. Mendoza-Mora, Adriana H. Vilchis-González, Rigoberto Martínez-Méndez, Vianney Muñoz-Jiménez, Iván Francisco-Valencia

Published: 30 November 2024

Abstract
The sclera is a white membrane rich in collagen and elastin fibres, which gives it an affinity for bilirubin. This yellow substance is produced by the breakdown of the heme group, and when its levels are elevated, it causes jaundice a condition that leads to yellowing of the skin, mucous membranes, and sclera. The intensity of the yellowing in the sclera is directly related to bilirubin levels in the body. This relationship enables the extraction of features to infer these levels using machine learning techniques based on RGB images of the sclera. Sclera segmentation is a transcendent factor in achieving this goal. For this reason, this article presents the results of sclera segmentation using the U-Net network. This is a convolutional network composed of encoding and decoding layers and is used for medical image segmentation. The model was trained and validated with a set of 181 eye images and their corresponding binary masks. The results obtained during the training phase are Loss (0.006), Precision (0.976), Recall (0.973) and F1-score (0.974), and in the validation phase: Loss (0.145), Precision (0.897), Recall (0.863) and F1-Score (0.880). These results demonstrate the U-Net model's effectiveness in segmenting the sclera, particularly in the training phase where the metrics are highly favorable. However, the slight decrease in performance during the validation phase suggests the need for further refinement. Future work will focus on increasing the dataset size and introducing data augmentation techniques to improve generalization and robustness. Ultimately, accurate sclera segmentation is a critical step toward developing reliable Machine Learning models for non-invasive bilirubin level estimation.

Keywords:
Segmentation, Sclera, U-Net, RGB imaging, Bilirubin index.

Mendoza-Mora D. A., Vilchis-González A. H., Martínez-Méndez R., Muñoz-Jiménez V., Francisco-Valencia I.
Facultad de Ingeniería, Universidad Autónoma del Estado de México. Toluca, Estado de México.
Email: dmendozam279@alumno.uaemex.mx, {avilchisg, rmartinezme, vmunozj, ifranciscov}@uaemex.mx

1 Introduction
In Mexico, liver diseases, such as cirrhosis and liver cancer, cause the death of thousands of people each year [1]. Jaundice, characterized by a yellow coloration of the skin, mucous membranes, and the sclera of the eye, is a common symptom of these diseases, caused by an increase in bilirubin in the body [2]. Bilirubin is produced when red blood cells break down and is eliminated through the liver and bile [2]. An increase in bilirubin levels indicates a liver problem, becoming it a marker of liver function [2].

The standard method for measuring bilirubin levels is the diazo reaction [3] which requires a blood sample obtained through a puncture in the patient's arm vein [4]. However, this method is invasive and can cause complications such as bruising and infections [4]. To address this issue, bilirubinometers have been developed; these are non-invasive devices that measure bilirubin through transcutaneous measurements [4]. Although these devices are useful for newborns due to their thinner skin, they may present inconsistencies in individuals with thicker or darker skin [4].

The literature includes several articles that have developed non-invasive methods to infer bilirubin levels using sclera images and machine learning models. Their methodology involves acquiring an image of the eye using photographic cameras or smartphone cameras [5], [6], [7], [8], [9], [10]. These methods utilize the sclera, which turns yellow when bilirubin levels increase, allowing for the extraction of various features from the tones present in the image. The authors label the characteristics of the participating subjects with the bilirubin level obtained from the invasive test applied. The labelled data was used for the training and validation of classical machine learning models, such as linear regression or convolutional networks, resulting in the inference of bilirubin levels.

Systems based on sclera images for bilirubin measurement have shown promising results but still face challenges.

A significant challenge in detecting bilirubin levels from eye images is the segmentation of the sclera [8]. Various methods have been used to achieve this, such as thresholding [7] and the GrabCut algorithm [8]. However, these approaches can produce inconsistent results, especially when there are intense glares or varying lighting conditions in the images [8]. For instance, thresholding can be affected by intense glares [7], while GrabCut requires manual initialization and can be inconsistent under changing lighting conditions [8].

Sclera segmentation is considered a biometric data point because the veins present in it have a unique pattern in each person. The use of the sclera as bio-metric data has encouraged the development of models to segment it, as is the case of the Sclera Segmentation and Recognition Benchmarking Competition (SSRBC) [11], [12]. In the various editions of this competition, methods utilizing models such as k-means, DBSCAN, U-Net, ResNet50, ResNet34, and combinations of U-Net and VGG have been presented [11], [12]. The winning methods have achieved an accuracy above 0.90 [11], [12]. An encoder-decoder network called Sclera-Net was developed to maintain the finest structures in the image and achieved an error rate of 0.0093 and an F1-score of 96.2421 [13]. A sclera segmentation method based on the U-Net model called ScleraSegNet achieved an F1-score of 91.43% [14].

The sclera region contains data that allows inferring the level of bilirubin; therefore, an effective segmentation of all pixels belonging to it would increase the correlation between results from image-based inference systems and those from invasive tests. To achieve this goal, this article uses a U-Net model for sclera segmentation in RGB eye images acquired through a prototype device using two ESP32 cam modules [15].

This document is organized by the present introduction, a theoretical framework with the context of convolutional networks and the U-Net model for reader understanding, the methodology detailing the development, training, and validation of the model, the results obtained, a discussion, and a conclusion summarizing the work.

2 Methodology

In this section, the steps used to develop the proposed method for performing semantic segmentation of the sclera region in an RGB eye image using a U-Net network are detailed. The methodological steps are illustrated in Figure 1, which provides a systematic framework for the implementation and evaluation of the U-Net network.

The algorithms were implemented using the Python 3.10. programming language. For the modeling, training, and validation of the U-Net model, the TensorFlow y Keras library version 2.13 was used.

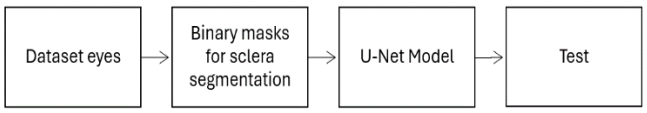

Figure 1. Steps for semantic segmentation of the sclera using a U-Net network.

2.1 Dataset eyes

A dataset composed of 118 RGB ocular images, obtained from 20 patients at the ISSSTE Hospital Clinic "Daniel Salazar Fernández" in the municipality of Huejutla de Reyes, Pachuca, Hidalgo, Mexico, was used. The participants were volunteers, and a series of photos of each eye were taken using an electronic device. A blood bilirubin test was also performed, and blood pressure, heart rate, and respiratory rate were recorded. The protocol didn't request information on liver conditions or other pathologies. Among these patients, 15 were women and 5 men, aged between 49 and 71 years. The images were captured using an electronic device [15] illustrated in Figure 2, with a googles type design, which consists of two ESP32-CAM modules equipped with OV2640 cameras of 2MP resolution and a series of LEDs for illumination. The images were acquired in a hospital room with constant lighting.

Figure 2. Electronic Device for Measuring Sclera Colouring [15]**.**

During the image acquisition, the subjects were standing and were asked to place the device at eye level, as shown in Figure 3. During the process, each patient adopted four eye different positions: looking up, to the left, to the right, and straight ahead. One image of each eye was captured from each position.

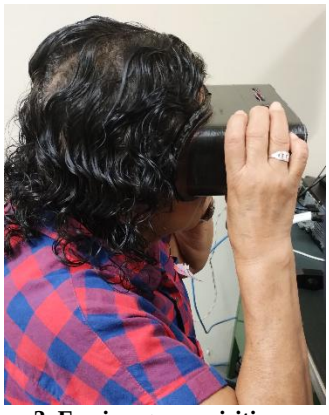

Figure 3. Eye image acquisition process.

Eight RGB images were obtained from each patient, each with dimensions of 1024x768 pixels and a resolution of 72 pixels per inch.

A total of 160 images were obtained, of which 118 were selected after discarding those with issues such as blurry images, images without the complete ocular region due to incorrect device placement, and those appearing black due to acquisition failures. The selected images were resized to various sizes using the nearest neighbour method with the OpenCV library.

After acquiring the images, a table was created to relate the patient data (age, gender, bilirubin levels) to their corresponding images. The images were labeled to indicate which eye they corresponded to (left or right) and their orientation. The images were considered independent, without regard to the relationship with the subject, for training the segmentation model.

Initial feedback about the device was obtained through oral questions to the participants. This approach was carried out without the use of a usability technique, aiming to gather a preliminary understanding of user opinions. The questions posed were: Did the device feel heavy? Was it uncomfortable when placed on the face? Did the internal light of the device bother you? Could you easily locate the button to take the photo?

The participants' responses regarding the weight and comfort of the device were unanimous: they considered the weight appropriate, and that the device fit the face without causing

discomfort. Regarding the light of the device, the subjects did not experience any discomfort during or after the capture.

In terms of usability, users encountered difficulties positioning the device at eye level and exhibited inconsistencies when pressing the button. On several occasions, participants pressed the button twice due to the lack of an indicator confirming that the capture had been successfully completed.

Nevertheless, the SUS (System Usability Scale) [16] tool is considered for evaluating the usability of the device. This tool assesses the usability of a device or system through a 10-question questionnaire that users respond to after interacting with the device or system. This questionnaire is designed to measure aspects such as usability, ease of use, efficiency, and overall satisfaction.

2.1.1 Binary masks for sclera segmentation

The set of eye images was labelled with their corresponding sclera region. The sclera regions were labeled by a single person, a PhD student in Biomedical Technology and Control at UAEM, using the points and polygons provided by the LabelMe tool. The sclera in the eye images was manually selected one by one. After labeling, the region was peer-reviewed to ensure it corresponded to the region of interest. To label the dataset, an image (sclera and background) was created in PNG format using the LabelMe application. This tool allowed manually selecting the region of interest through a series of points and polygons on the RGB image, as illustrated in Figure 4.A. After selecting the scleras in the entire dataset, the script labelme2voc.py was executed to mass-generate PNG images with the sclera region, an example of these images is illustrated in Figure 4.B.

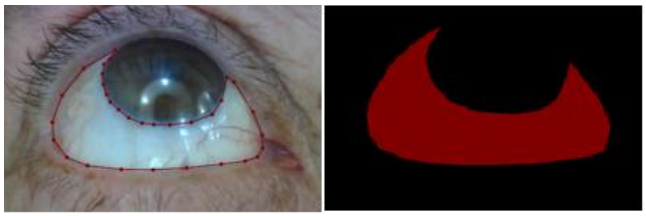

Figure 4. A) Manual segmentation of the sclera. B) Image with sclera region.

The set of images of the eyes and their sclerae were linked into a .CSV file one by one and divided into subsets of 80% for model training and 20% for validation. The images were resized using the nearest neighbour method to reduce the images by 25% of their size [13].

2.1.2 U-Net Model

The U-Net model is a convolutional network designed for semantic image segmentation [17]. It has a symmetrical architecture consisting of a series of convolutional layers for feature extraction followed by transposed convolutional layers that reconstruct the segmented image.

The encoding process, consists of a sequence of two 3x3 convolutional operations, followed by a 2x2 pooling operation [17]. This pattern is repeated four times, increasing the number of filters at each step. Additionally, a progression of two 3x3 convolutional operations connects the encoder with the decoder [17].

The decoder, first performs an up-sampling of the feature map using a 2x2 transposed convolution operation, halving the feature channels [17]. Then, a sequence of two 3x3 convolution operations is performed. This process is repeated four times, halving the number of filters at each step. Finally, a 1x1 convolution operation is applied to generate the final segmentation map [16].

In this architecture, all convolutional layers except the last one use the ReLU (Rectified Linear Unit) activation function, while the final convolutional layer employs a sigmoid activation function [17]. The architecture of the U-Net network used for the semantic segmentation of the sclera is illustrated in Figure 5.

The ReLU function is an activation function that is widely used in neural networks, particularly in deep learning models [13]. Its purpose is to introduce non-linearity in the output of a neuron, a crucial task for the neural network to learn and model complex relationships in the data. Its mathematical representation is as follows [13]:

$$ReLu = max(0, x) \quad (1)$$

This indicates that for any input value x:

if x is greater than zero, the output is equal to x.

- if x is less than or equal to zero, the output is zero.

The Sigmoid function is an activation function used in the final layer of neural networks for binary classification problems. Its goal is to map any real value into a range between 0 and 1, which allows the output to be interpreted as a probability. The mathematical definition of the sigmoid function is as follows:

$$Sigmoid(x) = \frac{1}{1+e^{x}} \quad (2)$$

Where:

x is the input to the function, which can be any real number.
e is the base of the natural logarithm (Euler number).

In the U-Net model, the Sigmoid functions was tested as activation function in the final layer. This function was chosen because can assign a probability of belonging to the class of interest to each pixel, such as sclera segmentation.

2.1.3 Evaluation metrics

The quantitative evaluation of the model was performed using the metrics of Loss, precision, recall and F1-Score [12], [13]. The Loss metric calculates how different the predictions, in this case, the sclera images obtained from the model, are from the manually segmented sclera images (labels). This loss function quantifies the model's error in terms of the difference between what it predicts and what it should predict. Its goal is to minimize the loss, which implies reducing the difference between the predictions and the actual labels as the model is trained with more data.

Precision measures the correctness of the model's positive predictions. It indicates how many of the positive predictions made by the model were correct. Recall is the measure that indicates the model's ability to find all the actual positive cases that exist. For both metrics, the ideal value expected as a result of a model is close to 1.

The F1-Score [13] represents the harmonic mean of precision and recall, balancing both metrics. A higher F1-score indicates that the model has achieved a balance between precision and recall.

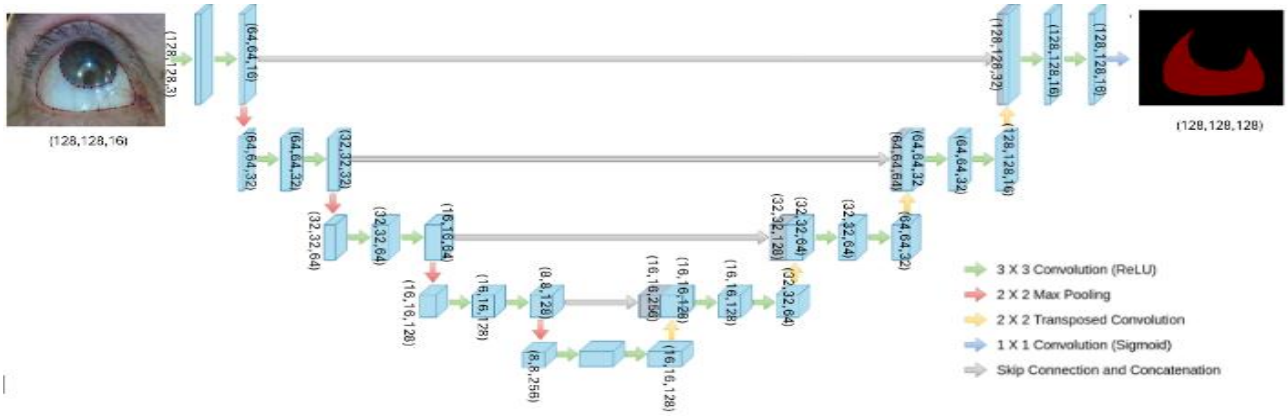

Figure 5. Structure of the U-Net model.

3 Results

During the model training phase, various combinations of hyperparameters for the U-Net model were tested to evaluate the sclera segmentation results. The set of 118 images was resized to a resolution of 256x192 pixels. The activation function used was ReLU, and the output function was sigmoid. The filter size in each layer of the model was 3x3. Other hyperparameters used during training and kept constant were Optimizer=Adam, Learning rate=0.001, Batch size=10 and loss function=binary_crossentropy. Batch Normalization layers were added to improve stability and help the model generalize the data. In training the model, 40 and 100 epochs were tested.

Table 1 presents the results of the loss, F1-Score, precision, and recall metrics for the different trainings and validations of the U-Net model.

Figures 6-9 show the values of the metrics Loss, Precision, Recall, and F1-Score obtained during training over 100 epochs. In Figures 6 to 9 show the plots of the metrics obtained during the training and validation stages of the model. A consistent decrease in the loss function is observed in both cases. However, in the model trained with 100 epochs, the loss plot is similar in both the training and validation phases, indicating a good model fit. In contrast, the model trained with 40 epochs shows a loss above 2.5 in the validation phase, suggesting an insufficient fit and possible overfitting to the training set.

Regarding the accuracy, recall, and F1-Score metrics, it is noteworthy that during the training phase, these metrics exceed 0.90 in both models. However, in the validation phase, the model trained with 40 epochs shows a significantly lower performance in recall and F1-Score compared to the 100-epoch model. This highlights the importance of a higher number of epochs to achieve better overall model performance.

When analyzing the graphical results of the model, it is evident that the segmentation obtained does not achieve adequate accuracy compared to the model trained with 100 epochs. As shown in Figure 10, the sclera region appears significantly smaller when using the model trained with fewer epochs. This indicates that the model has failed to effectively capture the features required for accurate segmentation in the validation phase. The discrepancy between the generated masks suggests that a higher number of epochs is crucial to improve the model's ability to generalize correctly and provide more accurate and consistent segmentation.

The results obtained with the U-Net model using the selected hyperparameters and the set of images generated by the device demonstrate superior performance in sclera segmentation compared to the U-Net model presented by Wang. These improved metrics suggest that the adjustments made have been effective in optimizing segmentation in the context of the images used. However, further testing by increasing the data volume and exploring different variations in the images is essential to ensure that the U-Net model maintains its performance under different conditions.

4 Conclusions

In this study, we demonstrated the effectiveness of the U-Net model for sclera segmentation using RGB images. The high-performance metrics achieved during the training phase, including a low loss of 0.006 and high values for precision, recall, and F1-score, underscore the model's capability in accurately segmenting the sclera. However, the slightly lower performance metrics observed during the validation phase, with a loss of 0.145 and reduced precision, recall, and F1-score, highlight the need for further improvements.

To enhance the model's robustness and generalization, future work will involve expanding the dataset and applying data augmentation techniques. This will help in refining the model's accuracy and ensuring consistent performance across diverse conditions. The successful segmentation of the sclera is a crucial step toward utilizing Machine Learning for reliable non-invasive bilirubin level estimation, which could have significant implications in medical diagnostics and patient care.

In the state of the art, there are models for scleral segmentation; however, at this time, it is not possible to compare the model developed in this work with those models, as images obtained with an electronic device are being used instead of the data employed by the models mentioned in the literature.

5 Future Work

As future work, it's proposed to generate a dataset of eye images acquired with the new version of the device. This dataset will include the results of the bilirubin levels obtained from the blood test applied to each participating subject.

The sclera region, segmented using the U-Net model, will be used to extract a feature vector through another convolutional model. This vector will be labelled with the bilirubin index values corresponding to each blood test. With this data, a dense neural network will be trained to infer bilirubin levels.

Table 1. Metrics resulting from the model training and validation phase

	Training phase				Validation phase			
Model	**loss**	**recall**	**presicion**	**F1-score**	**loss**	**recall**	**presicion**	**F1-score**
U-Net test 1	0.0349	0.9028	0.9312	0.9168	0.1151	0.6051	0.9777	0.7475
U-Net test 2	0.0103	0.9576	0.9759	0.9667	0.1158	0.8361	0.8876	0.8611

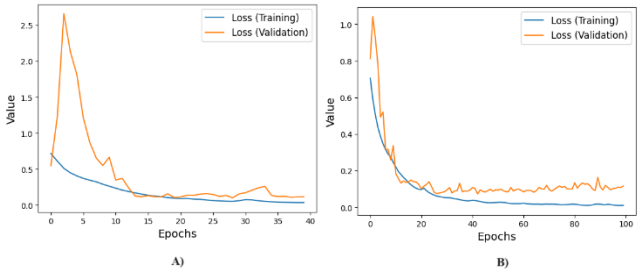

Figure 6. Loss A) 40 epochs, B) 100 epochs.

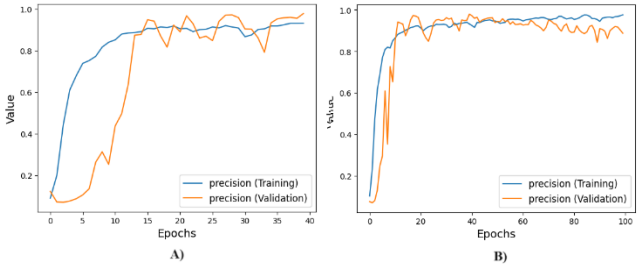

Figure 7. Precision A) 40 epochs, B) 100 epochs.

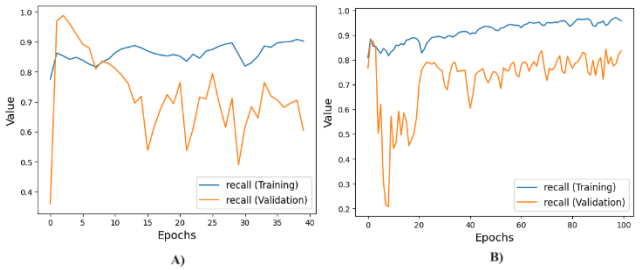

Figure 8. Recall A) 40 epochs, B) 100 epochs.

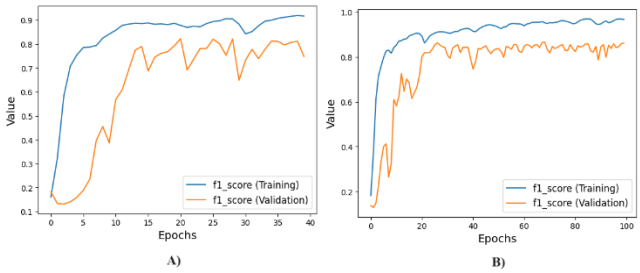

Figure 9. F1-score A) 40 epochs, B) 100 epochs.

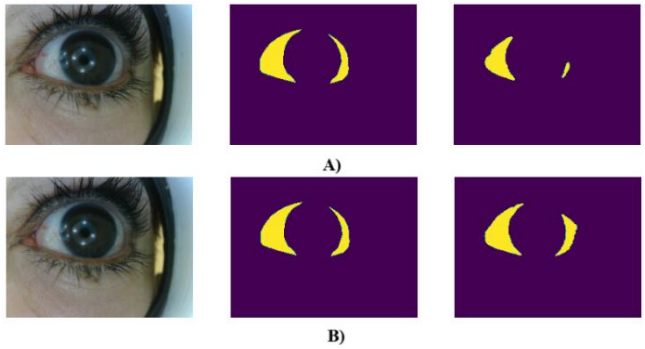

Figure 10. A) Eye image, manually segmented mask and mask predicted by the 40-epoch mode. B) Eye image, manually segmented mask and mask predicted by the 100-epoch mode.

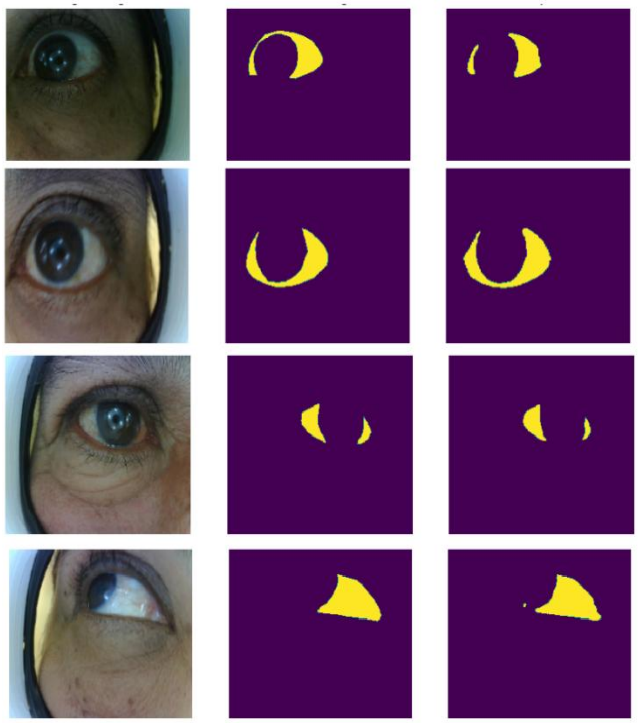

Figure 11. RBG image of the eye, manual binary mask and mask predicted by the 100-epoch mode.

6 References

[1] N. C. Flores-García, M. Dirac, H. Han, y D. Kershenobich-Stalnikowitz, «La carga de la enfermedad por cirrosis

hepática en México», *GMM*, vol. 159, n.º 6, p. 12605, dic. 2023, doi: 10.24875/GMM.23000370.

[2] A. R. Guerra-Ruiz *et al.*, «Bilirrubina: Medición y utilidad clínica en la enfermedad hepática», *Advances in Laboratory Medicine / Avances en Medicina de Laboratorio*, vol. 2, n.º 3, pp. 362-372, ago. 2021, doi: 10.1515/almed-2021-0016.

[3] L. Ngashangva, V. Bachu, y P. Goswami, «Development of new methods for determination of bilirubin», *Journal of Pharmaceutical and Biomedical Analysis*, vol. 162, pp. 272-285, ene. 2019, doi: 10.1016/j.jpba.2018.09.034.

[4] G. Alfieri, R. Mir Villamayor, L. E. Genes De Lovera, E. M. Otazo Arévalos, S. G. Miño Moreno, y J. P. G. Bordón Dure, «Aplicación del bilirrubinómetro no invasivo en recién nacidos», *Pediatr (Asunción).*, vol. 46, n.º 3, pp. 158-164, nov. 2019, doi: 10.31698/ped.46032019002.

[5] C. Enweronu-Laryea *et al.*, «Validating a Sclera-Based Smartphone Application for Screening Jaundiced Newborns in Ghana», *Pediatrics*, vol. 150, n.º 1, p. e2021053600, jun. 2022, doi: 10.1542/peds.2021-053600.

[6] T. Kihara *et al.*, «Identification and Quantification of Jaundice by Trans-Conjunctiva Optical Imaging Using a Human Brain-like Algorithm: A Cross-Sectional Study», *Diagnostics*, vol. 13, n.º 10, p. 1767, may 2023, doi: 10.3390/diagnostics13101767.

[7] Md. M. M. Miah *et al.*, «Non-Invasive Bilirubin Level Quantification and Jaundice Detection by Sclera Image Processing», en *2019 IEEE Global Humanitarian Technology Conference (GHTC)*, oct. 2019, pp. 1-7. doi: 10.1109/GHTC46095.2019.9033059.

[8] A. Mariakakis, M. A. Banks, L. Phillipi, L. Yu, J. Taylor, y S. N. Patel, «BiliScreen: Smartphone-Based Scleral Jaundice Monitoring for Liver and Pancreatic Disorders», *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, n.º 2, pp. 1-26, jun. 2017, doi: 10.1145/3090085.

[9] F. Outlaw, M. Nixon, O. Odeyemi, L. W. MacDonald, J. Meek, y T. S. Leung, «Smartphone screening for neonatal jaundice via ambient-subtracted sclera chromaticity», *PLoS ONE*, vol. 15, n.º 3, p. e0216970, mar. 2020, doi: 10.1371/journal.pone.0216970.

[10] Md. R. Sammir, K. Md. Towhidul Alam, P. Saha, Md. M. Rahaman, y Q. Delwar Hossain, «Design and Implementation of a Non-invasive Jaundice Detection and Total Serum Bilirubin Measurement System», en *2018 10th International Conference on Electrical and Computer Engineering (ICECE)*, Dhaka, Bangladesh: IEEE, dic. 2018, pp. 137-140. doi: 10.1109/ICECE.2018.8636801.

[11] A. Das *et al.*, «Sclera Segmentation Benchmarking Competition in Cross-resolution Environment», en *2019 International Conference on Biometrics (ICB)*, Crete, Greece: IEEE, jun. 2019, pp. 1-7. doi: 10.1109/ICB45273.2019.8987414.

[12] A. Das *et al.*, «Sclera Segmentation and Joint Recognition Benchmarking Competition: SSRBC 2023», en *2023 IEEE International Joint Conference on Biometrics (IJCB)*, Ljubljana, Slovenia: IEEE, sep. 2023, pp. 1-10. doi: 10.1109/IJCB57857.2023.10448601.

[13] R. A. Naqvi y W.-K. Loh, «Sclera-Net: Accurate Sclera Segmentation in Various Sensor Images Based on Residual Encoder and Decoder Network», *IEEE Access*, vol. 7, pp. 98208-98227, 2019, doi: 10.1109/ACCESS.2019.2930593.

[14] C. Wang, Y. He, Y. Liu, Z. He, R. He, y Z. Sun, «ScleraSegNet: an Improved U-Net Model with Attention for Accurate Sclera Segmentation», en *2019 International Conference on Biometrics (ICB)*, Crete, Greece: IEEE, jun. 2019, pp. 1-8. doi: 10.1109/ICB45273.2019.8987270.

[15] Arzate Cruz Karla Yaneli, «Desarrollo de un dispositivo electrónico para medir la coloración de la esclera y su relación con el nivel de bilirrubina en sangre», Universidad Autónoma del Estado de México, Toluca, Estado de México, 2024.

[16] M. Ríos-Hernández, J. M. Jacinto-Villegas, O. Portillo-Rodríguez, y A. H. Vilchis-González, «User-Centered Design and Evaluation of an Upper Limb Rehabilitation System with a Virtual Environment», *Applied Sciences*, vol. 11, n.º 20, p. 9500, oct. 2021, doi: 10.3390/app11209500.

[17] N. Ibtehaz y M. S. Rahman, «MultiResUNet : Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation», *Neural Networks*, vol. 121, pp. 74-87, ene. 2020, doi: 10.1016/j.neunet.2019.08.025.