



Pontificia Universidad  
**JAVERIANA**  
Cali

**Facultad de Ingeniería  
y Ciencias**

Ingeniería Biomédica

INFORME FINAL DE TRABAJO DE GRADO

Sistema de soporte al diagnóstico no invasivo de  
ictericia neonatal en Colombia: un enfoque desde la  
visión por computador

Camilo Salazar Barney  
Angelica Maria Ruiz Ortiz

*Director*

Dr. Cristian Alejandro Torres Valencia

29 de enero de 2026



Santiago de Cali, 29 de enero de 2026

Señores  
Pontificia Universidad Javeriana – Cali  
Dra. María Constanza Pabón Burbano  
Decano  
Facultad de Ingeniería y Ciencias  
Ciudad

Cordial Saludo.

Por medio de la presente nos permitimos presentarle el Trabajo de Grado titulado “Sistema de soporte al diagnóstico no invasivo de ictericia neonatal en Colombia: un enfoque desde la visión por computador”.

Esperamos que este trabajo reúna todos los requisitos académicos, cumpla el propósito para el cual fue creado y sirva de apoyo para futuros proyectos relacionados con la profesión.

Atentamente,



---

Camilo Salazar Barney



---

Angelica Maria Ruiz Ortiz



# Agradecimientos

Queremos expresar nuestro agradecimiento a todas las personas que hicieron posible el desarrollo de este proyecto. De manera especial, extendemos nuestro reconocimiento a nuestro director, Cristian Alejandro Torres, por su acompañamiento constante, su orientación académica y su capacidad para guiarnos con claridad en cada fase del proceso. Su compromiso y disponibilidad fueron fundamentales para llevar este trabajo a buen término.

Como equipo de trabajo, agradecemos el espacio de colaboración que permitió desarrollar este proyecto de manera conjunta. El esfuerzo compartido, la revisión mutua de cada sección y la construcción articulada de la metodología y los resultados fueron centrales para consolidar el documento y alcanzar los objetivos propuestos.

A nuestras familias, gracias por su apoyo incondicional, por la paciencia en los momentos más exigentes y por acompañarnos incluso en los periodos de mayor carga académica. Su respaldo emocional fue decisivo para culminar este trabajo.

Agradecemos de manera especial a los profesionales de neonatología Ana María Gómez y Andrés Lozano, quienes aportaron su experiencia clínica y sus observaciones durante el desarrollo del proyecto. Sus perspectivas desde la práctica asistencial nos permitieron orientar el sistema hacia necesidades reales del entorno hospitalario, enriqueciendo la dimensión clínica del estudio.

A todos quienes nos acompañaron en este proceso, gracias por su apoyo y por contribuir de forma significativa al desarrollo de este proyecto.



# Glosario

## Símbolos

$TP$	True Positive / Verdadero positivo
$TN$	True Negative / Verdader Negativo
$FP$	False Positive / Falso Positivo
$FN$	False Negative / Falso Negativo
$R^2$	Coefficiente de determinación del modelo de regresión
$K$	En validación cruzada, el número de iteraciones y el número de subconjuntos.
$n$	Número total de muestras en las fórmulas de métricas de desempeño
$y_i$	Valor real de bilirrubina sérica)
$\hat{y}_i$	Valor predicho de bilirrubina sérica
$\bar{y}$	Promedio de los valores reales de bilirrubina sérica
$TPR$	True Positive Rate (Tasa de verdaderos positivos)
$FPR$	False Positive Rate (Tasa de falsos positivos)

## Acrónimos y Abreviaturas

$TSB$	Bilirrubina Sérica Total
$BTC$	Bilirrubinometría Transcutánea
$ROI$	Region Of Interest / Región de interés
$CNN$	Convolutional Neural Network / Red Neuronal Convolutacional
$RGB$	Red-Green-Blue
$YCrCb$	Luma-Red difference- Blue difference / Luminancia-Crominancia roja-Crominancia Azul
$HSV$	Hue-Saturation-Value / Matiz-Saturación-Valor
$AUC$	Area Under the Curve / Área Bajo la Cuva
$ROC$	Receiver Operating Characteristic / Característica Operativa del Receptor
$MAE$	Mean Absolute Error / Error Absoluto Promedio
$RMSE$	Root Mean Square Error / Error Cuadrático Medio

## Términos

<i>Bilirrubina Sérica</i>	Es la concentración de bilirrubina en la sangre, expresada en mg/dL. Es el parámetro clínico utilizado para determinar la severidad de la ictericia neonatal de forma invasiva.
<i>Clasificación</i>	Tarea de aprendizaje automático que busca asignar una etiqueta a una entrada, en este caso, entre neonatos con o sin ictericia a partir de entradas de imágenes neonatales.
<i>Regresión</i>	Modelo que predicción de un valor numérico continuo. En este proyecto se utiliza para estimar el valor de bilirrubina sérica a partir de imágenes neonatales.
<i>Exactitud Top-1</i>	Métrica que indica si la clase verdadera de un problema de clasificación multiclase concuerda con la clase predicha con mayor probabilidad por el modelo. Esta métrica fue implementada en la clasificación con la base de datos ImageNet como métrica de desempeño.
<i>Exactitud Top-5</i>	Métrica que indica si la clase verdadera en un problema de clasificación multiclase se encuentra dentro de las cinco clases con mayor probabilidad predichas por el modelo. Esta métrica fue implementada en la clasificación con la base de datos ImageNet como métrica de desempeño.
<i>Hiperparámetro</i>	Parámetro que se configura previo al entrenamiento de un modelo de forma manual, como la tasa de aprendizaje, el batch size o el número de épocas.
<i>Umbral de decisión</i>	Valor que se aplica a las probabilidades predichas por un modelo de clasificación para asignar una etiqueta. Ajustar este umbral permite mejorar métricas de desempeño del modelo.
<i>Curva Precision-Recall</i>	Curva que muestra la relación entre las métricas de desempeño Precisión y Recall para diferentes umbrales de decisión.
<i>Robustez</i>	Propiedad de un modelo que describe su capacidad de mantener un desempeño estable aun enfrentando variaciones en los datos no previstas durante el entrenamiento.
<i>Pipeline</i>	Conjunto organizado y secuencial de pasos necesarios para procesar datos o generar modelos.
<i>Streamlit</i>	Framework de Python para construir interfaces web interactivas.
<i>Interfaz Gráfica</i>	Medio visual que utiliza elementos diversos como botones para generar una interacción entre un usuario y un sistema computacional.

# Resumen

La ictericia neonatal es una de las condiciones clínicas más comunes durante los primeros días de vida y su identificación temprana es fundamental para evitar complicaciones asociadas a niveles elevados de bilirrubina. En muchos entornos de atención en salud, en especial aquellos con recursos limitados, el diagnóstico continúa dependiendo de métodos invasivos o evaluaciones visuales sujetas a variabilidad, lo que puede prolongar los tiempos de respuesta y limitar la oportunidad del tratamiento. Este proyecto se desarrolló con el propósito de explorar una alternativa no invasiva basada en el procesamiento digital de imágenes y en la aplicación de modelos de aprendizaje profundo para apoyar la detección de ictericia neonatal y estimar la concentración aproximada de bilirrubina a partir de fotografías neonatales. El alcance del estudio se centró al análisis del conjunto de imágenes y datos presentes en la base de datos de acceso abierto NeoJaundice y no incluyó la validación clínica directa en instituciones de salud.

La metodología aplicada se dividió en cuatro etapas. En primer lugar, se realizó un análisis de las imágenes disponibles, determinando las características cromáticas y variaciones relevantes para el diagnóstico. Posteriormente, se implementó un pipeline de preprocesamiento que incluyó normalización de color, segmentación de la región de interés y la generación de una máscara adaptativa orientada a resaltar tonalidades amarillentas. En la tercera etapa, se entrenaron modelos basados en redes neuronales convolucionales para las tareas de clasificación y regresión, ajustando parámetros para obtener predicciones compatibles con las etiquetas del conjunto de datos. Finalmente, se desarrolló una interfaz interactiva que permitiera realizar análisis individuales y por lotes, integrando así el flujo completo de etapas desarrolladas y facilitando la visualización detallada de los análisis realizados.

Los resultados obtenidos muestran que el sistema propuesto logró procesar las imágenes del conjunto de datos de manera estable y generar predicciones coherentes con la información disponible, evidenciando que el enfoque planteado constituye una base sólida para avanzar hacia futuras evaluaciones en escenarios clínicos reales.

**Palabras clave:** procesamiento digital de imágenes, ictericia neonatal, aprendizaje profundo, bilirrubina, diagnóstico no invasivo.



# Abstract

Neonatal jaundice is a frequent condition during the first days of life, and its timely identification is essential to prevent complications associated with elevated bilirubin levels. In many clinical settings, especially those with limited resources, diagnosis still relies on invasive laboratory measurements or subjective visual evaluation, which affects response times and may delay medical decisions. This project proposes a non-invasive approach based on digital image processing and deep learning, aimed at supporting the classification of jaundice and estimating serum bilirubin concentration using neonatal photographs. The focus of the study was on analyzing the set of images and data in the NeoJaundice open-access database and did not include direct clinical validation in any clinical institution.

The work was structured in four stages: first, an analysis of the dataset images was performed to identify chromatic characteristics relevant to the diagnosis, so that a preprocessing pipeline could be implemented to ensure chromatic consistency. Subsequently, models based on convolutional neural networks were developed and then trained in classification and regression tasks, adjusting parameters to obtain predictions compatible with the dataset labels. Finally, a Streamlit interactive interface that enabled individual and batch analyzes was constructed to integrate the complete flow of stages developed and facilitate detailed visualization of the analyzes performed.

The results obtained indicate that the proposed approach operates consistently under the experimental conditions of this study and represents a feasible basis for future clinical validation.



# Índice general

<b>1. Introducción</b>	<b>1</b>
<b>2. Planteamiento del Problema</b>	<b>3</b>
<b>3. Justificación</b>	<b>7</b>
<b>4. Objetivos</b>	<b>9</b>
4.1. Objetivo General . . . . .	9
4.2. Objetivos Específicos . . . . .	9
<b>5. Marco de Referencia</b>	<b>11</b>
5.1. Áreas Temáticas . . . . .	11
5.2. Marco Teórico . . . . .	11
5.2.1. La ictericia neonatal . . . . .	11
5.2.2. Métodos existentes de diagnóstico de la ictericia neonatal . . . . .	14
5.2.3. Procesamiento digital de imágenes como herramienta para el apoyo al diagnóstico de ictericia neonatal . . . . .	16
5.3. Trabajos Relacionados . . . . .	29
<b>6. Materiales y Métodos</b>	<b>31</b>
6.1. Características relevantes para el procesamiento de imágenes de ictericia neonatal . .	31
6.1.1. Características en la literatura . . . . .	31
6.1.2. Base de datos seleccionada para el proyecto . . . . .	32
6.1.3. Preprocesamiento . . . . .	33
6.2. Implementación del procesamiento con deep learning . . . . .	37
6.2.1. Entorno de trabajo computacional . . . . .	37
6.2.2. Aumento de datos . . . . .	37
6.2.3. Modelo de clasificación . . . . .	38
6.2.4. Modelo de regresión . . . . .	41
6.2.5. Prueba adicional con imágenes externas . . . . .	42
6.2.6. Modelo combinado de clasificación y regresión . . . . .	43
6.3. Validación del sistema . . . . .	43
6.3.1. Métricas de validación para el modelo de clasificación . . . . .	43
6.3.2. Métricas de validación para el modelo de regresión . . . . .	44
6.3.3. Validación cruzada . . . . .	45
6.4. Desarrollo de la interfaz gráfica . . . . .	45
6.4.1. Características importantes para la interfaz . . . . .	45
6.4.2. Diseño de la interfaz . . . . .	46

6.4.3. Estructura del proyecto e implementación de la interfaz . . . . .	48
<b>7. Resultados y Discusión</b>	<b>51</b>
7.1. Modelo de clasificación . . . . .	51
7.1.1. Resultados de Arquitecturas de transfer learning . . . . .	51
7.1.2. Resultados de Arquitecturas 2DCNN . . . . .	52
7.1.3. Resultados de Cross-validation . . . . .	55
7.2. Modelo de Regresión . . . . .	55
7.2.1. Resultados con base de datos NJN . . . . .	56
7.3. Interfaz en streamlit . . . . .	58
7.3.1. Descripción general . . . . .	58
7.3.2. Módulo de diagnóstico individual . . . . .	58
7.3.3. Módulo de análisis por lotes . . . . .	59
7.3.4. Visualización y trazabilidad de resultados . . . . .	60
7.3.5. Análisis e interpretación final . . . . .	61
7.3.6. Acceso a la interfaz de Streamlit . . . . .	61
<b>8. Conclusiones</b>	<b>63</b>
<b>9. Trabajos futuros</b>	<b>65</b>
<b>10. Anexos</b>	<b>67</b>
<b>Anexos</b>	<b>67</b>
Anexo 1 – Rangos de umbralización de pixeles de piel utilizados en la literatura para los espacios de color YCrCb y HSV . . . . .	68
Anexo 2 – Estudios de clasificación con Transfer Learning . . . . .	69
Anexo 3 – Hiperparámetros para los modelos de clasificación . . . . .	70
Anexo 4 – Hiperparámetros para los modelos de regresión . . . . .	71
Anexo 5 – Estructura de la aplicación de streamlit . . . . .	72
Anexo 6 – Requisitos para desarrollo de la aplicación y esquema de exportación de resultados	74
Anexo 7 – Manual de usuario de la interfaz . . . . .	75
Anexo 8 – Tarjeta de Calibración . . . . .	85
<b>Bibliografía</b>	<b>87</b>

# Índice de figuras

5.1. Etapas del procesamiento de imágenes . . . . .	17
5.2. Representación conceptual de los espacios de color RGB, HSV y CIELab. . . . .	18
5.3. Ejemplo de aplicación de operación de convolución . . . . .	21
5.4. Ejemplo de aplicación de operación de Padding . . . . .	22
5.5. Estrategias de Pooling . . . . .	23
5.6. Capas densas o Fully connected . . . . .	24
5.7. Efecto de un dropout con $p = 0.5$ en una red neuronal . . . . .	24
5.8. Ejemplo clásico de CNN de clasificación . . . . .	25
5.9. Ejemplo de Validación cruzada . . . . .	28
6.1. Ejemplos de imagenes en la base de datos NeoJaundice . . . . .	32
6.2. Preprocesamiento de las imágenes de la base de datos NeoJaundice . . . . .	33
6.3. Ejemplo de los cambios de imagen en cada etapa del preprocesamiento . . . . .	36
6.4. Características del conjunto de datos de bilirrubina sérica de la base de datos NeoJaundice . . . . .	41
6.5. Ejemplos de las imágenes en la base de datos NJN . . . . .	43
6.6. Boceto preliminar de la interfaz . . . . .	47
6.7. Flujo funcional del sistema de soporte al diagnóstico implementado en la interfaz desarrollada. . . . .	49
7.1. Matriz de Confusión del modelo de clasificación con threshold tuning (Experimento 3) . . . . .	53
7.2. Curvas de aprendizaje del modelo de clasificación con threshold tuning (Experimento 3) . . . . .	53
7.3. Composición de la arquitectura 2DCNN seleccionada para la clasificación . . . . .	54
7.5. Diagrama de correlación del modelo de estimación de bilirrubina sérica . . . . .	56
7.4. Composición de la arquitectura final de estimación de bilirrubina sérica . . . . .	57
7.6. Vista del módulo de diagnóstico individual en la interfaz desarrollada. Se observa la predicción del modelo, la concentración estimada de bilirrubina y la interpretación clínica correspondiente. . . . .	59
7.7. Vista del módulo de análisis por lotes. Se muestra la barra de progreso, los resultados parciales y el resumen global del desempeño del modelo. . . . .	60
10.1. Portada del manual de usuario . . . . .	75
10.2. Pagina 1 manual de usuario . . . . .	76
10.3. Pagina 2 manual de usuario . . . . .	77
10.4. Pagina 3 manual de usuario . . . . .	78

10.5. Pagina 4 manual de usuario . . . . .	79
10.6. Pagina 5 manual de usuario . . . . .	80
10.7. Pagina 6 manual de usuario . . . . .	81
10.8. Pagina 7 manual de usuario . . . . .	82
10.9. Pagina 8 manual de usuario . . . . .	83
10.10 Pagina 9 manual de usuario . . . . .	84
10.11 Prototipo de Tarjeta de calibración . . . . .	85

# Índice de tablas

7.1. Resultados arquitecturas de Transfer Learning . . . . .	51
7.2. Resultados de las arquitecturas experimentales 2DCNN . . . . .	52
7.3. Resultados de la validación cruzada para el modelo de clasificación 2DCNN seleccionado	55
7.4. Resultados arquitecturas de estimación de bilirrubina sérica . . . . .	55
7.5. Resultados arquitectura de regresión seleccionada con imágenes de la base de datos NJN . . . . .	56
10.1. Valores umbral en la literatura para segmentación de piel en HSV . . . . .	68
10.2. Valores umbral en la literatura para segmentación de piel en YCrCb . . . . .	68
10.3. Estudios de clasificación de ictericia neonatal con transfer learning . . . . .	69
10.4. Hiperparámetros utilizados en los modelos evaluados . . . . .	70
10.5. Hiperparámetros utilizados en los modelos de regresión evaluados para estimación de bilirrubina sérica . . . . .	71
10.6. Descripción de los principales módulos de la interfaz desarrollada. . . . .	73
10.7. Dependencias principales del proyecto. . . . .	74



# Introducción

---

La ictericia neonatal es una condición clínica frecuente que afecta aproximadamente al 60 % de los recién nacidos a término y al 80 % de los prematuros durante la primera semana de vida [1]. Esta condición se caracteriza por una coloración amarillenta de la piel y las mucosas debido a la acumulación de bilirrubina en sangre. Aunque en la mayoría de los casos es una condición fisiológica que se resuelve espontáneamente, en ciertas situaciones puede evolucionar hacia formas patológicas que requieren intervención médica inmediata para prevenir complicaciones graves, como el kernícterus, una forma de encefalopatía bilirrubínica que puede causar daño neurológico permanente [1].

En Colombia, aunque no se dispone de estadísticas nacionales específicas recientes sobre la prevalencia general de la ictericia neonatal, estudios locales han documentado su relevancia clínica. Por ejemplo, en un estudio realizado en Tuluá, Valle del Cauca, se encontró que la ictericia fue una manifestación presente en el 38 % de los neonatos febriles sin foco aparente, lo que subraya su prevalencia en el país y su impacto en la salud neonatal [2].

La detección y el manejo oportuno de esta condición presentan desafíos significativos, especialmente en áreas rurales y en instituciones de salud con recursos limitados. Los métodos tradicionales de diagnóstico, como la evaluación visual y la medición de bilirrubina sérica, tienen limitaciones. La evaluación visual es subjetiva y puede variar según la experiencia del personal de salud y las condiciones de iluminación, mientras que la medición de bilirrubina sérica es invasiva, requiere infraestructura de laboratorio y puede no estar disponible en todos los entornos clínicos [2].

La implementación de tecnologías no invasivas para el diagnóstico de la ictericia neonatal, como los bilirrubinómetros transcutáneos, han demostrado ser útiles en la práctica clínica. Sin embargo, su alto costo limita su accesibilidad en muchas instituciones de salud en Colombia. Además, la variabilidad en los resultados obtenidos con estos dispositivos, especialmente en niveles elevados de bilirrubina, puede requerir confirmación adicional mediante métodos invasivos. En este contexto, surge la necesidad de desarrollar sistemas de diagnóstico no invasivos, accesibles y precisos para la detección temprana de la ictericia neonatal en Colombia. La aplicación de tecnologías de procesamiento digital de imágenes y algoritmos de aprendizaje profundo (deep learning) ofrecen una oportunidad para abordar este problema, permitiendo la identificación y cuantificación de la ictericia a través del análisis de imágenes digitales, reduciendo la dependencia de métodos invasivos y costosos.

Abordar este desafío es crucial para mejorar la calidad de la atención neonatal en Colombia, reducir las tasas de complicaciones asociadas a la ictericia no tratada y optimizar el uso de recursos

en el sistema de salud, por lo tanto, el presente proyecto tiene como objetivo principal la elaboración de un sistema de soporte al diagnóstico de ictericia neonatal por medio de procesamiento de imágenes con arquitecturas de deep learning, enfocado en la detección de la presencia de la enfermedad y la estimación de bilirrubina sérica, con el propósito de generar una herramienta accesible, escalable y confiable que contribuya al diagnóstico temprano y no invasivo de ictericia neonatal.

Finalmente, el presente documento se encuentra organizado por capítulos, en los cuales se presentan las distintas etapas conceptuales, metodológicas y experimentales del proyecto en cuestión. En el **Capítulo 2** se presenta el planteamiento del problema que motiva este proyecto, en el **Capítulo 3** se aborda la justificación del mismo y en el **Capítulo 4** se presentan los objetivos planteados como columna vertebral del proyecto. Posteriormente, en el **Capítulo 5** se abordan fundamentos teóricos claves y trabajos relacionados al presente proyecto. En el **Capítulo 6** se describe la metodología empleada para las distintas etapas de elaboración del sistema, desde el preprocesamiento de imágenes de ictericia neonatal, el desarrollo de modelos de clasificación y regresión, validación e implementación de una interfaz integradora. Los resultados obtenidos y su análisis se muestran en el **Capítulo 7**, seguido de las conclusiones del proyecto y proyecciones futuras en los **Capítulos 8 y 9** respectivamente. Por último, se incluyen anexos con información complementaria en el **Capítulo 10**.

# Planteamiento del Problema

---

La bilirrubina es un pigmento biliar de color amarillo que resulta de la degradación del grupo hemo de proteínas como la hemoglobina. Este compuesto es insoluble en agua y circula en la sangre en unión a la proteína albúmina que la transporta al hígado [3]. En este órgano, el pigmento es transformado en un compuesto soluble, que puede ser metabolizado en el intestino como urobilínogeno. Posteriormente, el último puede ser reabsorbido o excretado por el riñón o a través de las heces [4].

Un aumento de los niveles de bilirrubina en sangre se denomina como hiperbilirrubinemia. En neonatos es consecuencia del incremento en la destrucción de glóbulos rojos y la inmadurez fisiológica hepática para conjuguar y excretar la bilirrubina, entre otros factores. En el mundo, alrededor de 1 de cada 10 neonatos es propenso a desarrollar hiperbilirrubinemia significativa que requiera una atención médica [1]. Lo fundamental de la intervención radica en evitar la toxicidad sobre el sistema nervioso central del neonato. La neurotoxicidad puede ocasionar una encefalopatía icterica, que a la vez, puede evolucionar en parálisis cerebral y déficits sensorio-motores [5].

Una de las patologías derivadas de la hiperbilirrubinemia es la ictericia. Según la Asociación Colombiana de Dermatología y Cirugía Dermatológica (AsoColDerma), la ictericia se describe como la coloración amarillenta de la piel y las mucosas causada por el aumento en sangre de los pigmentos biliares [6]. La coloración mencionada, inicia su expresión gradual en el cuerpo del neonato con una relación directa a la concentración de bilirrubina en sangre, la cual puede variar dependiendo del tono de piel del neonato. La visualización de la patología comienza en las escleróticas de los ojos con una concentración de 2-3 mg/dL, continuando por la cara alrededor de los 4-5 mg/dL, llegando al ombligo en valores de concentración cercanos a los 15 mg/dL y alcanzando los pies alrededor de los 20 mg/dL [4].

En un estudio global, se encontró que 60% de los recién nacidos a término y el 80% de los prematuros desarrollan ictericia en la primera semana de vida, así como que el 10% de los recién nacidos alimentados con seno materno presentan ictericia durante el primer mes de vida [1]. Existen diferentes formas de diagnóstico [3] como lo son la inspección visual, donde se realiza una evaluación subjetiva de la coloración cutánea, como criterio de diagnóstico e indicativo para realizar pruebas complementarias. También se cuenta con la medición invasiva de la concentración de bilirrubina en sangre, que se basa en una punción venosa, capilar o arterial realizada al neonato con el propósito de obtener una muestra de sangre con la cual, por medio de técnicas de análisis químicas, se permita la cuantificación de la bilirrubina presente.

Un tercer método es la medición no invasiva con analizadores de bilirrubina transcutánea (BTC), que permiten obtener el resultado de forma fácil, rápida e indolora por medio de la diferencia entre densidades ópticas de la luz, en especial para longitudes de onda correspondientes a los colores azul (450 nm) y verde (550 nm). Colombia enfrenta desafíos en el uso de métodos no invasivos, con una baja disponibilidad de esta tecnología en el país y un alto costo de adquisición y mantenimiento de los mismos en instituciones de salud.

Entre las situaciones que se presentan por preferir la realización del diagnóstico de ictericia por el método invasivos tenemos: tiempos prolongados para el procesamiento de la muestra, lo cual conlleva a demoras en el diagnóstico, tratamiento y egreso del neonato; generación de dolor para el neonato y también de angustia para los padres. Adicionalmente, para la obtención de la muestra de sangre es necesaria la participación de personal entrenado para la realización del procedimiento, tecnología de soporte al mismo y un laboratorio adecuado para el análisis de la muestra; por lo que hay un gasto significativo de recursos para las instituciones. En cuanto a la inspección visual, el diagnóstico por este método es susceptible a tener un alto grado de subjetividad y una baja precisión, lo cual puede llevar a diagnósticos erróneos y un déficit en el tratamiento, o en caso contrario, un sometimiento equívoco a procedimientos para tratar una condición mal diagnosticada. En estos casos, ambos tipos de diagnóstico llevarían a un deterioro de la salud del neonato [1].

En concordancia a las complicaciones de los métodos clínicos aplicados actualmente para el diagnóstico de ictericia neonatal, se puede determinar que tanto los efectos negativos a la salud del neonato como el efecto económico en las instituciones de salud no puede seguir pasando desapercibido si se desea prestar un servicio íntegro, efectivo y con un balance costo-beneficio adecuado. Es por lo anterior que se hace presente la necesidad de generar métodos o técnicas innovadoras para el diagnóstico de la ictericia en neonatos, las cuales le permitan a la institución realizar el diagnóstico de forma no invasiva, ágil, precisa y rentable; sin dejar de salvaguardar la salud del neonato en todo momento.

Entre algunas alternativas apreciables en la literatura se encuentra el acercamiento a un diagnóstico desde el procesamiento de imágenes neonatales, sin embargo, también existen problemáticas técnicas significativas asociadas a este enfoque. La captura de imágenes neonatales puede verse afectada por variaciones en las condiciones de iluminación, lo que introduce ruido y dificulta la extracción precisa de características cromáticas relevantes [7]. Además, la relación entre las paletas de color utilizadas en los análisis y la diversidad de pigmentaciones cutáneas representa un reto para lograr un modelo generalizable [8]. El ángulo de captura, la resolución de las cámaras, y la necesidad de preprocesamiento exhaustivo, como correcciones de iluminación y segmentación precisa, son etapas críticas que influyen en la calidad y utilidad de las imágenes para los modelos de aprendizaje profundo [9]. Estos modelos de aprendizaje mencionados, a su vez presentan desafíos en el contexto de la ictericia neonatal. Entre estos se encuentran la selección de un modelo adecuado que se acomode a las características de las imágenes y al volumen de datos disponibles [7]. Este

último factor genera a su vez una dificultad recurrente por la escasez de datos clínicos etiquetados en la ictericia neonatal [8]. También, se generan complicaciones desde la adecuada selección de métricas de desempeño para evaluar la clasificación realizada de las imágenes etiquetadas de manera confiable [9].

Resolver las problemáticas técnicas relacionadas con diagnósticos alternativos de ictericia neonatal es fundamental para garantizar que un sistema enfocado desde el procesamiento de imágenes para ese fin pueda ofrecer resultados robustos y precisos, incluso en entornos clínicos con recursos limitados y que cumpla con los estándares para la implementación en estos entornos en un futuro.



# Justificación

---

La ictericia neonatal es la anomalía más común durante el periodo neonatal inmediato y constituye una de las causas de readmisión hospitalaria más frecuentes de recién nacidos. Mundialmente, un estudio dirigido por Bhutani estimó en 2013 que, en un año, 481,000 recién nacidos a término y prematuros tardíos presentan niveles séricos de bilirrubina mayores de 25 mg/dL, y de estos, 114,000 mueren y más de 63,000 sobreviven con daño neurológico de moderado a grave [10]. Más recientemente, datos del Global Burden of Disease (GBD) indican que en 2016 la ictericia fue causante de 1008 muertes por cada 100,000 nacimientos (con un intervalo de confianza del 95 %), posicionando la patología en séptimo lugar entre las causas de muerte en el periodo neonatal temprano en África, Asia y Norteamérica [11].

A su vez, la ictericia es causada por la hiperbilirrubinemia, una fisiopatología que puede ser devastadora para los recién nacidos a causa de las secuelas a largo plazo que genera, como son: parálisis cerebral, desórdenes del lenguaje, sordera y en general retraso del neurodesarrollo. Estas complicaciones son más comunes en países del tercer mundo porque se realiza un diagnóstico y tratamiento en forma tardía. A diferencia de la mayoría de los trastornos neonatales, la ictericia no se puede prevenir, por lo que el objetivo de cualquier intervención en este padecimiento es identificar y prevenir su evolución a complicaciones más graves [1]. Existen diferentes formas de diagnóstico que enfrentan limitaciones entre las cuales identificamos el costo, invasividad e inmediatez de un resultado.

El costo elevado de los dispositivos para el diagnóstico no invasivo de ictericia neonatal y las consecuencias asociadas a primar métodos de diagnóstico invasivos u observacionales, hacen necesario desarrollar un sistema alternativo que permita identificar los niveles de bilirrubina desde la no invasividad, que sea confiable y que sea accesible económicamente a las instituciones de salud, pues de esta forma se permitiría una atención y tratamiento adecuados al neonato que lo requiera. Al realizar una búsqueda en el aplicativo de consulta avanzada de registros sanitarios del INVIMA en el grupo de *médico quirúrgicos* y bajo el nombre de producto "*bilirrubinómetro*", se encuentra que existe solo un dispositivo con registro sanitario del INVIMA de la marca Dräger con registro sanitario INVIMA 2018DM-0018872 . Estos resultados evidencian la falta de oferta en el mercado de tecnologías avaladas para el diagnóstico no invasivo de ictericia y por ende, una baja disponibilidad de otras opciones para instituciones que no puedan adquirir este producto de importación.

Con el objetivo de salvaguardar las características de no invasividad y la posibilidad de realizar una medición al paciente con un resultado inmediato, se considera que el desarrollo de un sistema

de detección de signos de ictericia por medio del procesamiento de imágenes y el entrenamiento de inteligencia artificial podría permitir resolver la necesidad de las instituciones de salud y en segunda instancia de los pacientes y sus familiares. A su vez, en Colombia se carece de estudios y bases de datos autóctonas sobre la patología, por lo que el presente proyecto podría sentar las bases para el desarrollo de futuras investigaciones con datos locales de ictericia neonatal, propiciando el mejoramiento de protocolos de atención médica neonatal en el país.

Para el desarrollo de las distintas etapas del proyecto, se cuenta con el acompañamiento de profesionales en el área de neonatología que se encuentran actualmente inmersos en el ambiente hospitalario de interés. Al colaborar con ellos, este proyecto no solo responde a la necesidad de ofrecer una posible alternativa económica y precisa para la detección temprana de ictericia neonatal, sino también recibe un respaldo intelectual desde el área de la salud, el cual es necesario para desarrollar el sistema de manera aplicable al contexto clínico local. Adicionalmente, se cuenta con acceso a bases de datos con imágenes de pacientes neonatales con y sin ictericia, provenientes de proyectos internacionales, las cuales van a permitir una evaluación del sistema en cada etapa del desarrollo del mismo. Finalmente, se considera que los costos del proyecto van enfocados en el uso de software y herramientas computacionales que permitan el correcto procesamiento e identificación de imágenes en deep learning. En comparación con los costos asociados con dispositivos médicos tradicionales, el enfoque en el desarrollo de software puede resultar más económico a largo plazo y ofrece la flexibilidad de actualizar y mejorar el sistema con el tiempo.

# Objetivos

---

## 4.1. Objetivo General

Desarrollar un sistema no invasivo basado en procesamiento de imágenes para la identificación de ictericia en neonatos, capaz de generar alertas tempranas para apoyar el diagnóstico médico oportuno.

## 4.2. Objetivos Específicos

- Definir las características relevantes de las imágenes neonatales para la detección de ictericia mediante procesamiento digital de imágenes.
- Implementar una metodología de clasificación del estado de ictericia utilizando técnicas de aprendizaje automático (machine learning) y aprendizaje profundo (deep learning).
- Validar el sistema desarrollado utilizando métricas de desempeño que midan la exactitud, sensibilidad y especificidad del modelo.
- Diseñar una interfaz gráfica que permita la visualización de los resultados del diagnóstico y la generación de alertas basadas en los niveles de ictericia detectados.



# Marco de Referencia

---

## 5.1. Áreas Temáticas

- La ictericia neonatal
  - La hiperbilirrubinemia
  - Manifestación clínica y clasificación
  - Repercusiones graves en salud derivadas del tratamiento inoportuno de la patología
- Métodos existentes de diagnóstico de la ictericia neonatal
  - Inspección visual
  - Métodos invasivos
  - Métodos no invasivos
- Procesamiento digital de imágenes como herramienta para el diagnóstico de ictericia neonatal
  - Preprocesamiento y análisis de color
  - Machine learning
  - Inteligencia artificial y deep learning

## 5.2. Marco Teórico

### 5.2.1. La ictericia neonatal

#### 5.2.1.1. La hiperbilirrubinemia

La aparición de la ictericia neonatal se debe al aumento de niveles plasmáticos de bilirrubina o hiperbilirrubinemia sérica en el recién nacido, especialmente la hiperbilirrubinemia hepática no conjugada. Este tipo de hiperbilirrubinemia se relaciona directamente con el funcionamiento hepático y en el recién nacido sucede de forma fisiológica, donde la inmadurez hepática del neonato no permite la metabolización y excreción de la bilirrubina, por lo que la bilirrubina no conjugada que no recibe metabolización hepática y al acumularse en cantidades séricas considerables, hace que se observe la coloración amarillenta en mucosas y piel que caracterizan a la ictericia, en este caso del paciente neonatal [3].

La inmadurez hepática del neonato se hace visible principalmente en dos mecanismos principales dentro del proceso metabólico de la bilirrubina: la producción del compuesto y la conjugación del mismo. Inicialmente, la bilirrubina es obtenida mayormente por medio de la degradación del grupo hemo de los glóbulos rojos que se encuentran finalizando su ciclo de vida, por lo que la frecuencia de obtención del compuesto de interés se relaciona directamente con la duración del ciclo de vida del eritrocito. Para el caso del neonato, la producción de bilirrubina por kilogramo de cuerpo es del doble a comparación de lo es en etapas posteriores de vida (8.5 mg/kg/día vs 4.0 mg/kg/día) y el eritrocito tiene un ciclo de vida mucho más corto de 60 a 100 días, por lo que existe un desbalance inicial en el metabolismo de la bilirrubina del neonato por sobreproducción. Posteriormente, la bilirrubina es transportada al hígado unida a la proteína albúmina, donde al entrar en contacto con el hepatocito debe de pasar por un proceso de conjugación a una forma soluble, proceso que realiza una transferasa llamada (UGT)1A1. E, la cual en el hígado del neonato tiene una acción retardada y por ello limita la continuación de la bilirrubina en su proceso de excreción posterior en el intestino [4].

Adicionalmente a los factores mencionados previamente, la hiperbilirrubinemia neonatal puede surgir como consecuencia de concentraciones elevadas de beta-glucuronidasa en la leche materna, componente que aumenta la desconjugación y reabsorción de bilirrubina, también por incompatibilidad feto-materna de grupo sanguíneo que genera sobreproducción de bilirrubina, fracaso en la lactancia materna que aumenta la circulación enterohepática de bilirrubina en neonatos deshidratados, entre otras causas [4].

### 5.2.1.2. Manifestación clínica y clasificación

La ictericia neonatal es la principal manifestación de la hiperbilirrubinemia neonatal, caracterizándose por la aparición de coloración amarillenta de las membranas mucosas, esclera del ojo y de la piel del neonato [4]. Esta coloración, a medida que aumentan los niveles de bilirrubina sérica total, empieza su progresión en el cuerpo del neonato de manera céfalo-caudal, puesto que los neonatos presentan una falta de flujo sanguíneo capilar en las partes más distales del cuerpo al estar centrado su desarrollo principalmente en la cabeza y partes más proximales en los primeros días después de nacer, haciendo que la bilirrubina se acumule inicialmente en estos lugares y luego se disperse con el aumento de cantidad de bilirrubina en sangre, aumentando así la gravedad del estado del paciente [12].

La ictericia neonatal puede presentar una clasificación dependiendo de algunos de los factores causales a los cuales se encuentre asociada [13].

- **Ictericia fisiológica:** Ictericia neonatal causada por el aumento de los niveles de bilirrubina indirecta o no conjugada debido a que el sistema de conjugación y de transporte de bilirrubina no tienen la madurez suficiente y porque los eritrocitos del recién nacido tienen una menor

vida media. Su aparición comúnmente es considerada como benigna y puede vigilarse desde el hogar.

- **Ictericia asociada a lactancia materna:** Ictericia neonatal causada por la alimentación con leche materna, donde se ha evidenciado que los recién nacidos a término tienen entre tres y cuatro veces más posibilidad de desarrollar ictericia moderada y severa que los recién nacidos alimentados por fórmula. Se trata por medio de fototerapia.
- **Ictericia por incompatibilidad de factor Rh:** Ictericia neonatal originada por un proceso de hemólisis de los eritrocitos fetales debido a la incompatibilidad de Rh, donde la madre con Rh (-) produce anticuerpos de tipo IgG e IgM en respuesta a los glóbulos rojos Rh(+) del feto, y posteriormente, los anticuerpos IgG que logran atravesar a la placenta y llegar al torrente sanguíneo fetal marcan el eritrocito Rh(+) para ser degradado por macrófagos. Debido a esto, el recién nacido al nacer presenta un aumento en la síntesis de bilirrubina, y junto con la inmadurez del proceso de metabolismo de bilirrubina mencionado previamente, se puede presentar la patología.
- **Ictericia por incompatibilidad ABO:** Ictericia neonatal que puede aparecer si madre e hijo presentan distinto grupo sanguíneo, específicamente cuando una madre con grupo sanguíneo O tiene un hijo de grupo A o B, donde los anticuerpos de la madre interactúan con el eritrocito del neonato y producen la hemólisis de los eritrocitos del recién nacido.

### 5.2.1.3. Repercusiones graves en salud derivadas del tratamiento inoportuno de la patología

En caso de que el diagnóstico de la ictericia neonatal se realice de forma tardía, se incrementa la probabilidad de que el neonato sufra de déficits neurológicos como la encefalopatía por bilirrubina (BIE), donde la bilirrubina no conjugada presente, libre en el plasma sanguíneo del neonato, sobrepasa la capacidad de unión con la albúmina y cruza hacia las regiones cefálicas, donde actúa como un tóxico celular que altera el proceso de respiración celular y bloquea la producción de ATP, ocasionando muerte celular. Este tipo de desorden neurológico puede ocasionar distintos problemas como lo son lesiones en el sistema nervioso central, deficiencias auditivas, visuales, dentales, neuromotoras y del lenguaje; cuyos efectos, dependiendo del tiempo de inicio del tratamiento, pueden ser transitorios y mínimos o graves y permanentes [13].

La manifestación clínica de toxicidad de la bilirrubina puede manifestarse de distintas formas [13]:

- **Encefalopatía por bilirrubina transitoria:** La toxicidad de la bilirrubina en el cerebro es transitoria y reversible, donde los recién nacidos con el tiempo y tratamiento adecuado pueden detener los efectos de la toxicidad. Los síntomas de corta duración que presenta el neonato son letargo y respuestas anómalas del tronco cefálico.

- **Disfunción neurológica inducida por bilirrubina (BIND)** la toxicidad por bilirrubina genera disfunciones neurológicas menores como desórdenes de la visión, audición, función neuromotora, del habla, cognición y lenguaje. A su vez, se caracteriza por la presencia de anomalías en el tono muscular y reflejos neonatales hiperexcitables.
- **Encefalopatía por bilirrubina aguda** los síntomas se presentan en distintas etapas: temprana, intermedia y tardía. En la etapa temprana se presenta letargo, hipotonía leve, alimentación y succión deficientes, hiperreflexia y llanto ligeramente agudo. En la etapa intermedia se presenta irritabilidad, fiebre, arqueamiento de la espalda, hiperextensión de los músculos extensores, estupor moderado y llanto agudo. Por último, la etapa tardía se caracteriza por estupor profundo o coma, llanto agudo, retrocolis, opistótonos, convulsiones, anormalidad visual, entre otros síntomas.
- **Encefalopatía por bilirrubina crónica o kernicterus:** Se presenta disfunción auditiva, neuropatía auditiva, parálisis de la mirada horizontal, displasia del esmalte, síndrome extrapiramidal y toxicidad extra neural por bilirrubina que afecta células de otros órganos.

Adicional a la encefalopatía por bilirrubina, se puede presentar en el neonato no tratado distintas discapacidades del neurodesarrollo a largo plazo como la parálisis cerebral coreoatetósica, desórdenes en el espectro auditivo, retrasos generales en el desarrollo y aún se encuentra en investigación el efecto posible en aparición de asma infantil, autismo y epilepsia [4].

## 5.2.2. Métodos existentes de diagnóstico de la ictericia neonatal

### 5.2.2.1. Inspección visual

La pigmentación amarilla que se manifiesta en el cuerpo del neonato se hace visible cuando superan los 2.5mg/dL [3], por lo que se han generado mecanismos para determinar por medio de la observación del neonato la ictericia neonatal. El primer método que permitió hacer una correlación entre las zonas del cuerpo del neonato con coloración amarillenta y la cantidad de bilirrubina sérica se denomina método de Kramer, el cual desde 1969 se convirtió en una aproximación no invasiva de los niveles séricos de bilirrubina del neonato [12]. Para esto, se cuenta con una división de cinco zonas del cuerpo del bebé y una cantidad aproximada de bilirrubina sérica dependiendo de la zona en que se encuentre la coloración amarilla [13]:

1. **Zona 1:** Zona de la cabeza y cuello donde se estima que los niveles séricos de bilirrubina son menores a 5 mg/dL.
2. **Zona 2:** Zona del cuello hasta el ombligo sin contar extremidades, donde se estiman niveles de bilirrubina sérica entre los 5 y 12 mg/dL.
3. **Zona 3:** Zona del ombligo a las rodillas, donde se estiman niveles de bilirrubina sérica entre los 8 y 16 mg/dL.

4. **Zona 4:** Zona de rodillas a tobillos y de hombros a muñecas, donde se estima un nivel de bilirrubina sérica entre 10 y 18 mg/dL.
5. **Zona 5:** Zona de los pies y las manos, donde se estima un nivel de bilirrubina sérica mayor a 15 mg/dL.

Con este método de diagnóstico se utiliza la progresión cefalo-caudal de la coloración para así estimar la cantidad de bilirrubina en sangre y con ello, permitir al personal de salud tomar decisiones respecto al tratamiento del neonato y la gravedad de su condición. No obstante, la correlación que se ha encontrado entre la estimación observacional y la cantidad de bilirrubina sérica es muy baja, por lo que el método no es el más confiable ni objetivo [14].

#### 5.2.2.2. Métodos invasivos

Los métodos invasivos de diagnóstico de ictericia neonatal se centran en obtener los niveles de bilirrubina sérica para estimar así la gravedad de la condición del paciente. Estos métodos entregan mediciones de bilirrubina que son precisos y permiten así manejar adecuadamente al paciente neonatal con ictericia. En este caso, existen variedad de métodos por los cuales se obtienen los resultados de bilirrubina sérica [8]:

1. **Instrumentos multiparametro:** en estos encontramos distintos métodos de diagnóstico, como los que usan reacciones químicas de diazo o vanadato oxidasa para estimar la cantidad de bilirrubina en sangre o instrumentos analizadores de gases de sangre que estiman la bilirrubina en sangre indirectamente por cooximetría, proceso en el cual se hemoliza la sangre y se calcula así la concentración de bilirrubina equivalente.
2. **Bilirrubinómetros de mesa:** método basado en la espectrofotometría directa en que se utiliza una muestra mínima de sangre. En este método se hace uso de un suero para medir la absorbancia de bilirrubina (454 nm) y de absorbancia de hemoglobina (528 nm). La resta de ambas absorbancias genera un valor que se puede atribuir a la bilirrubina.
3. **Instrumentos portátiles POC:** este método consiste en un lector de reflectancia manual diseñado para cuantificar los niveles de bilirrubina sérica a partir de un volumen bajo de sangre que es aplicado a una tira reactiva compuesta de un separador de plasma sanguíneo acoplado a una membrana de nitrocelulosa, la cual permite por medio de la aplicación de un diodo emisor de luz a 465nm obtener la luz reflejada por el plasma sanguíneo y cuantificar la bilirrubina sérica.

#### 5.2.2.3. Métodos no invasivos

En la actualidad existen variedad de opciones que reducen la invasividad en el diagnóstico de ictericia neonatal sin perder precisión, entre los que encontramos [14]:

1. **Icterómetros:** método desarrollado en la segunda mitad del siglo XX que consta de una regla de Perspex, la cual se utilizaba para un diagnóstico temprano de la ictericia neonatal. Esta regla tiene de dos a seis escalas de amarillo organizadas en un gradiente escalonado y que para el diagnóstico de ictericia, es presionada contra la nariz, frente y el interior del labio del neonato. El color resultante del tejido presionado se compara con uno de los colores del gradiente del icterometro que corresponde a un rango de bilirrubina en sangre y así se estima la bilirrubina sérica.
2. **Bilirrubinometría transcutánea:** Este método se fundamenta principalmente en la descripción de una relación lineal entre la concentración de bilirrubina sérica y la intensidad del color amarillento de la piel del recién nacido. Para esto, se usan dispositivos electrónicos que detectan indirectamente el valor de la bilirrubina en sangre por medio de la medición de diferencia de densidades ópticas para luz en diferentes longitudes de onda.
3. **Estimación de bilirrubina basado en smartphone o cámara digital:** este método es novedoso y permite por medio de la recopilación de imágenes con la cámara de un teléfono móvil en conjunto con técnicas avanzadas de procesamiento digital de imágenes y aprendizaje automático estimar la cantidad de bilirrubina en sangre. En este caso se han avanzado en la generación de aplicaciones que integran este tipo de software de estimación de bilirrubina en donde se selecciona como parámetro de interés de una foto del neonato las coloraciones amarillas de la piel o esclera y se estima así la cantidad de bilirrubina en sangre por medio de la extracción de datos relevantes y detalles en la imagen [9].

### 5.2.3. Procesamiento digital de imágenes como herramienta para el apoyo al diagnóstico de ictericia neonatal

#### 5.2.3.1. Preprocesamiento

En [15], una imagen digital se define como un archivo resultado de la discretización de una imagen natural o sintética en elementos mínimos llamados píxeles. A su vez, esta puede ser definida como una función bidimensional  $I(x, y)$  donde  $x$  e  $y$  son coordenadas en el plano, y la amplitud  $I$  es llamada la intensidad asociada con cada píxel en escala de grises [16]. Entonces, un píxel es un elemento de imagen que en conjunto con otros píxeles, forma una imagen la cual tiene propiedades geométricas y radiométricas, como lo es el tamaño, resolución, profundidad de color y compresión, entre otras.

La adquisición de una imagen digital sigue un proceso que implica varias etapas clave. En primer lugar, la captura se realiza mediante tecnología de sensor CCD o CMOS, que convierte la escena del mundo real en una señal eléctrica. Luego, en la etapa de muestreo, se digitalizan las coordenadas espaciales  $(x, y)$ , determinando el tamaño de la imagen. Continúa la cuantización, que es donde se digitaliza la amplitud o intensidad  $I$ , definiendo el tipo de datos (entero, real) y el número de bits utilizados para representar cada píxel [16]. Finalmente, la etapa de codificación determina el tipo

de representación de la imagen.

Además, para preparar la imagen para análisis, existen técnicas de preprocesamiento como recorte y escalado para ajustar su tamaño y enfocarse en áreas de interés específicas. También se cuenta con análisis de histogramas, que proporciona información sobre la distribución de intensidades, mientras que el filtrado y la segmentación ayudan a resaltar características relevantes. Existen otro tipo de operaciones como lo son las transformaciones geométricas que permiten corregir la perspectiva o distorsión, mientras que las transformadas de frecuencia, como la transformada de Fourier, se utilizan para análisis espectral y detección de características en la imagen [16]. Estas técnicas de preprocesamiento y análisis son fundamentales para optimizar la calidad y utilidad de las imágenes digitales en una variedad de aplicaciones, en especial de imágenes médicas.

Una vez se preprocesa una imagen digital, se procede a extraer, representar y manipular la información cromática contenida en una imagen por medio del análisis del color. Esta información de color es crucial en diversas aplicaciones, como la detección de objetos, la clasificación de imágenes, la mejora de la calidad visual y la corrección de color en fotografías. Según [15], las técnicas de preprocesamiento incluyen la normalización de la iluminación, filtrado de ruido, y segmentación de regiones de interés, que en el caso del presente proyecto, serían áreas de piel afectadas por la ictericia.

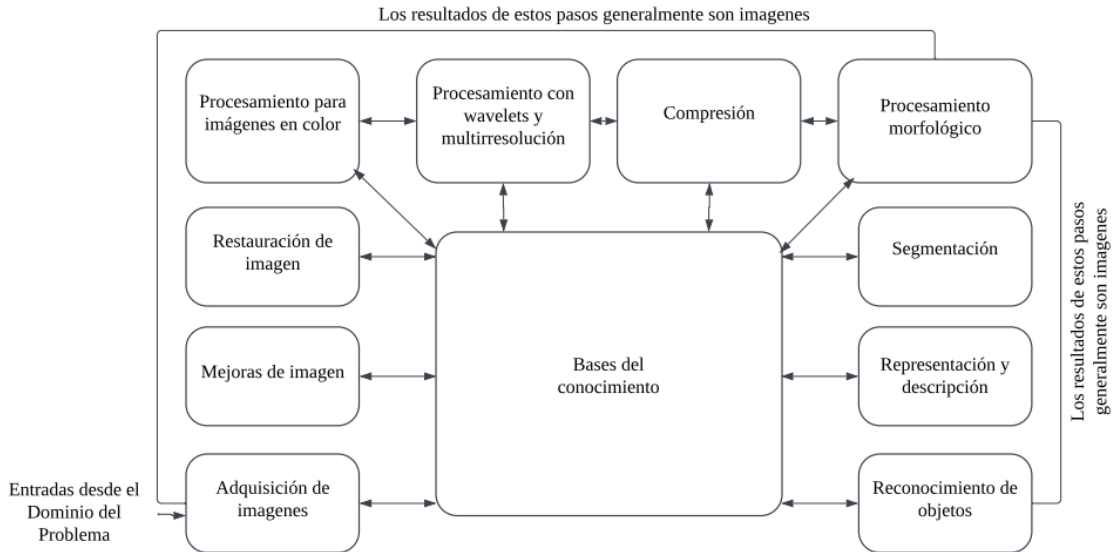


Figura 5.1: Etapas del procesamiento de imágenes

El análisis de color es particularmente relevante para el diagnóstico de ictericia neonatal, ya que la variación en la tonalidad cutánea constituye uno de los principales indicadores clínicos de la afección. En este contexto, la elección del espacio de color adecuado resulta determinante para

representar de forma consistente la información cromática presente en la piel del neonato y para mitigar los efectos de iluminación variable durante la captura de las imágenes. Todos los procesos mencionados dentro del preprocesamiento pueden observarse de manera general en la Figura 5.1. A continuación, se presentan los espacios de color más utilizados en el procesamiento digital de imágenes y su relevancia para el análisis cromático en este estudio.

### 5.2.3.2. Espacios de color

Los espacios de color constituyen modelos matemáticos que permiten representar la información cromática de una imagen mediante la organización de sus componentes en diferentes dimensiones [17]. Su elección resulta fundamental en aplicaciones donde la tonalidad y la luminancia influyen directamente en el análisis, como en el caso de la ictericia neonatal. Dado que cada modelo describe el color desde una perspectiva distinta, resulta útil visualizar cómo se estructuran conceptualmente estas representaciones antes de analizar su utilidad en este estudio.

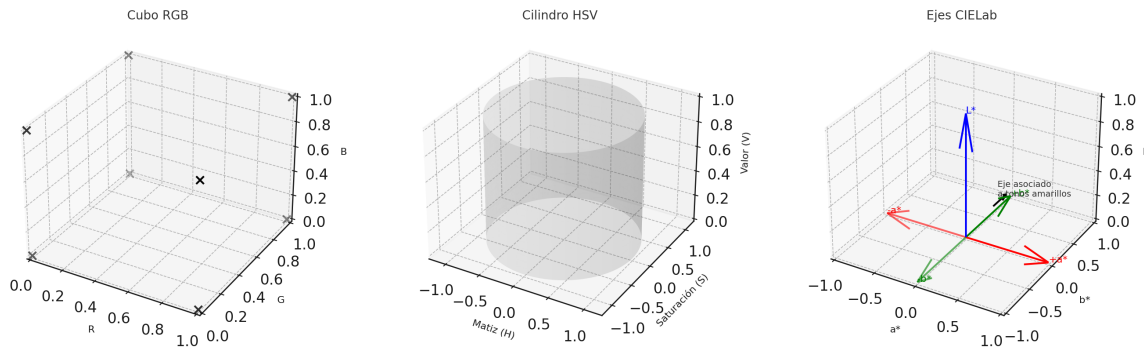


Figura 5.2: Representación conceptual de los espacios de color RGB, HSV y CIE Lab.

El modelo RGB se estructura como un cubo tridimensional basado en la combinación aditiva de los canales primarios. HSV se organiza en un cilindro que separa explícitamente el matiz, la saturación y el valor. CIE Lab se representa en un sistema perceptualmente uniforme a través de los ejes L, a y b, lo que facilita el análisis de variaciones cromáticas sutiles.

A partir de esta representación conceptual, es posible comprender cómo cada espacio enfatiza características diferentes del color. El espacio **RGB** conserva la información capturada directamente por los sensores de la cámara; **HSV** separa el matiz de las variaciones de iluminación, lo que facilita la identificación de tonalidades específicas; y **CIE Lab** proporciona un esquema perceptualmente uniforme donde los ejes de crominancia permiten cuantificar de manera más consistente los cambios asociados a tonalidades amarillentas. Estas propiedades resultan especialmente relevantes para el análisis cromático de la piel neonatal en el contexto de este estudio.

El espacio *RGB* (*Red-Green-Blue*) es la representación más ampliamente empleada en dispositivos de captura digitales. En este modelo, los colores se forman mediante la combinación aditiva de tres canales primarios. Si bien su estructura coincide con la forma en que los sensores registran la

información luminosa, su principal limitación radica en su dependencia directa de las condiciones de iluminación. A pesar de ello, el canal  $R$  puede reflejar incrementos en la tonalidad amarillenta de la piel, razón por la cual continúa siendo de utilidad en evaluaciones preliminares.

El espacio  $HSV$  (*Hue-Saturation-Value*) reorganiza la información cromática separando explícitamente el matiz, la saturación y el nivel de brillo. Esta separación permite describir el color de manera más estable ante variaciones de iluminación, especialmente a través del componente  $Hue$ , el cual cuantifica el matiz de un modo intuitivo y robusto. La capacidad del espacio HSV para aislar tonalidades específicas lo hace particularmente adecuado para segmentar regiones con predominancia amarilla, una característica clave en la detección de ictericia.

Además de los modelos RGB y HSV, el espacio de color  $YCbCr$  ha demostrado ser especialmente útil en tareas relacionadas con la segmentación de piel, debido a que desacopla la luminancia ( $Y$ ) de la crominancia ( $Cb$ ,  $Cr$ ), lo cual reduce la sensibilidad a variaciones de iluminación. Diversos estudios han mostrado que este espacio captura de manera más estable los rangos cromáticos asociados al tono cutáneo y facilita la detección de regiones de piel incluso bajo condiciones heterogéneas de captura [18]. Adicionalmente, se ha reportado que enfoques basados en YCbCr pueden integrarse en sistemas de análisis médico, pues su representación cromática favorece la delimitación de tejidos blandos y estructuras dérmicas de interés clínico [19]. En el contexto de este estudio, estas características convierten a YCbCr en un complemento valioso para el análisis de color en la piel neonatal, aportando robustez frente a variaciones de iluminación y fortaleciendo la segmentación previa al diagnóstico de ictericia.

Por último, el espacio  $CIELab$  es un modelo perceptualmente uniforme, lo que significa que diferencias similares de color corresponden a distancias comparables dentro del espacio tridimensional. Este modelo se compone de un canal de luminancia ( $L$ ) y dos canales de crominancia ( $a$  y  $b$ ). El canal  $b$ , que representa el eje azul-amarillo, es especialmente relevante para la identificación de incrementos en la tonalidad amarillenta de la piel, permitiendo un análisis más consistente frente a variaciones en las condiciones de captura. Asimismo, la uniformidad perceptual del modelo CIELab favorece una interpretación más estable de los colores en contextos clínicos donde la iluminación puede ser variable.

La elección de diferentes espacios de color aporta ventajas específicas en el análisis cromático debido a que cada uno organiza la información de forma distinta y resalta características particulares del color [20]. Esta propiedad resulta especialmente útil en aplicaciones clínicas basadas en tonalidades de la piel, como la ictericia neonatal, donde la detección depende de variaciones sutiles en los componentes amarillentos. Aunque la literatura clásica del procesamiento de imágenes no aborda este caso clínico de manera directa, los principios de representación cromática que describen estos modelos permiten adaptar sus ventajas al estudio de la piel neonatal bajo condiciones de iluminación heterogéneas.

### 5.2.3.3. Inteligencia artificial y aprendizaje automático supervisado para la ictericia neonatal

La inteligencia artificial aplicada al procesamiento de imágenes médicas constituye un área de desarrollo relevante, ya que permite extraer información clínica a partir de datos visuales que pueden pasar desapercibidos para la evaluación humana. Dentro de este campo, el aprendizaje automático (machine learning) y, de forma más específica, las redes neuronales profundas (deep learning) desempeñan un papel central debido a su capacidad para identificar patrones complejos y relaciones no lineales en imágenes digitalizadas[15]. Estas características resultan especialmente pertinentes en el análisis de la ictericia neonatal, donde leves variaciones en el tono de la piel pueden asociarse a cambios en la concentración de bilirrubina.

El aprendizaje automático supervisado se define como un enfoque en el cual cada dato de entrada posee una etiqueta o salida asociada, permitiendo entrenar modelos capaces de aprender una función de mapeo entre ambas variables [21]. Dentro de este enfoque se encuentran tanto métodos clásicos como modelos profundos. La regresión lineal es uno de los métodos fundamentales para modelar relaciones continuas entre características extraídas de la imagen y un valor numérico de interés, como la estimación de bilirrubina. Asimismo, el algoritmo de los  $k$ -vecinos más cercanos (k-NN) clasifica nuevas muestras según la similitud respecto a instancias previamente observadas, medida mediante métricas de distancia [22]. Aunque ambos métodos ofrecen interpretabilidad y simplicidad, su capacidad para resolver relaciones altamente no lineales es limitada, especialmente cuando se trabaja con imágenes de alta variabilidad cromática.

Para abordar patrones más complejos, las redes neuronales convolucionales (CNN) constituyen una de las arquitecturas más utilizadas en el procesamiento de imágenes médicas. Esta arquitectura emplea capas convolucionales capaces de detectar características de distinto nivel: bordes, texturas, gradientes de color y, en capas más profundas, patrones abstractos difíciles de definir explícitamente [23]. Estas capas se caracterizan por la aplicación de un conjunto de filtros o kernels sobre regiones de una imagen para la extracción de características por medio de la operación de **convolución**, definida de acuerdo a la Ecuación 5.1.

$$[W \circledast X](i, j) = \sum_{u=0}^{H-1} \sum_{v=0}^{W-1} w_{u,v} x_{i+u, j+v} \quad (5.1)$$

Donde:

$W$  = Kernel o filtro 2D aplicado a la imagen.

$X$  = Imagen de entrada.

$\circledast$  = Símbolo de operación de convolución 2D.

$H, W$  = Tamaño del kernel (alto y ancho).

$w_{u,v}$  = elementos del kernel.

$x_{i+u,j+v}$  = pixeles de la imagen que coinciden con el kernel.

Un ejemplo de la aplicación de la convolución en una entrada  $X$  de tamaño  $3 \times 3$  con un kernel  $W$  con tamaño  $2 \times 2$  para una salida  $Y$  se muestra en la Ecuación 5.2 y en la Figura 5.3, donde se evidencia que dentro del proceso de convolución se obtiene el resultado del producto escalar entre los pixeles de la imagen y del kernel que se desplaza sobre ella, dando como resultado lo que se conoce como mapa de características, el cual representa las características y patrones extraídos de la imagen, las cuales se convertirán en la entrada de las capas convolucionales siguientes, que aprenderán características de mayor complejidad y especificidad según la tarea propuesta para esta red.

$$Y = \begin{bmatrix} w_1 & w_2 \\ w_3 & w_4 \end{bmatrix} \otimes \begin{bmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7 & x_8 & x_9 \end{bmatrix} \quad (5.2)$$

$$= \begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_4 + w_4x_5 & w_1x_2 + w_2x_3 + w_3x_5 + w_4x_6 \\ w_1x_4 + w_2x_5 + w_3x_7 + w_4x_8 & w_1x_5 + w_2x_6 + w_3x_8 + w_4x_9 \end{bmatrix}$$

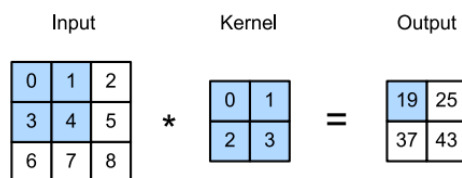


Figura 5.3: Ejemplo de aplicación de operación de convolución

Tomada de [24]

Una red convolucional neuronal, además de contener capas convolucionales para la extracción de características, cuentan también con otros componentes y operaciones de vital importancia para la recolección de las mismas, entre las cuales encontramos las siguientes:

**Función de activación:** Con el propósito de que un modelo aprenda relaciones complejas entre la entrada y salida de las neuronas que lo componen y evitar que se comporte linealmente, se aplican funciones de activación **no lineales** a la salida de la operación de convolución. Existen variedad de funciones de activación como la Sigmoide, la cual es utilizada para problemas de clasificación en la capa de salida del modelo, sin embargo, esta función presenta una saturación en 1 para valores positivos altos y 0 para valores negativos altos, una saturación que aunque es aprovechada para tareas de clasificación binaria, lleva a un problema llamado "Problema de Desvanecimiento de gradiente", el cual consiste en el descenso del gradiente a medida que aumenta la profundidad del modelo, lo cual dificulta su entrenamiento[24]. Por esto, se utilizan funciones de activación no saturantes como

la función ReLU o Unidad lineal rectificadora, que solo “apaga” valores de entrada negativos y permite el paso de valores positivos evitando el desvanecimiento del gradiente y manteniendo la no linealidad en el aprendizaje del modelo. Ejemplos adicionales de funciones de activación aplicadas en capas convolucionales son: Softmax, Elu, Hardlim, entre otras.

**Padding :** Una problemática presente en redes convolucionales con convoluciones seguidas es la reducción del tamaño de la salida de la operación, significando una pérdida de información tras la convolución [23]. Para una imagen de tamaño  $x_h \times x_w$  en una convolución con un kernel de dimensiones  $f_h \times f_w$  da como resultado una salida de dimensiones  $(x_h - f_h + 1) \times (x_w - f_w + 1)$ , debido a que el filtro se aplica únicamente a las partes "válidas" de la imagen entrada, situación a la cual se le llama *valid convolution*[24].

Para evitar que el proceso de convolución tenga pérdida de información en los mapas de características resultantes se aplica la operación llamada *Padding*, en la cual se agregan bordes de ceros a la imagen de entrada en la capa convolucional para que a la salida se mantengan las mismas dimensiones de la imagen de entrada, a esto se le llama *same convolution*. En la Figura 5.4 se muestra un ejemplo del efecto del padding en la salida de un proceso de convolución.

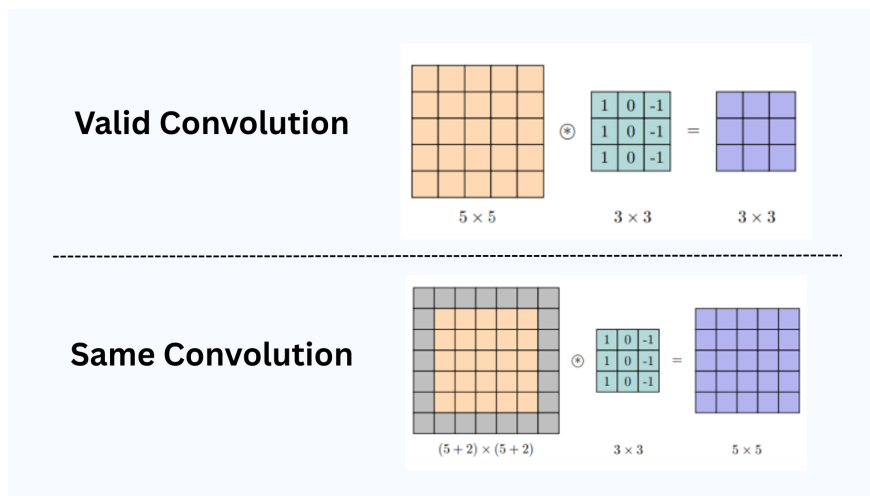


Figura 5.4: Ejemplo de aplicación de operación de Padding

Adaptada de [23]

**Pooling:** Con el propósito de reducir el tamaño de los mapas de características para optimizar la velocidad del aprendizaje pero evitando la pérdida de información importante en un modelo de clasificación con redes convolucionales, se utilizan las capas de *Pooling* [23]. Dos formas para realizar este proceso son el Max pooling y el Average pooling. Ambas utilizan un filtro con un

tamaño determinado para generar regiones en un mapa de características y luego, obtienen de cada región el valor máximo o el promedio respectivamente, para reducir así su tamaño y con ello, el costo computacional del modelo. En la Figura 5.5 se muestra el funcionamiento de las estrategias de pooling mencionadas.

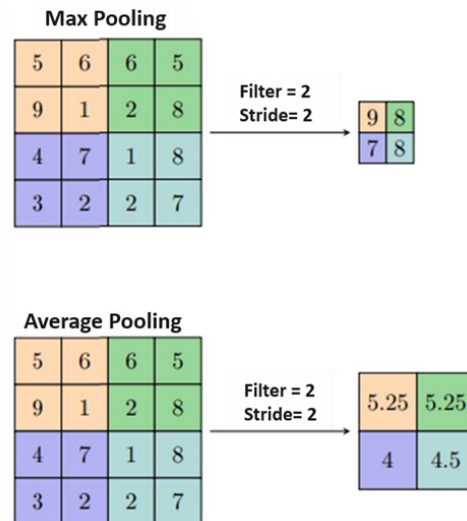


Figura 5.5: Estrategias de Pooling

Tomada de [23]

**Capa de normalización:** Como otra estrategia para solucionar el problema del desvanecimiento del gradiente, se utilizan capas de normalización en los modelos, las cuales tienen como principal función estandarizar las estadísticas de las capas del modelo [24]. Una estrategia popular de normalización es el *Batch normalization* o normalización por batches, la cual normaliza las activaciones de una red neuronal para que mantengan una media de 0 y una varianza de 1. Esto lo realiza a la salida de una capa previo a su entrada a la siguiente, dividiendo el mapa de características en lotes y calculando las varianzas y promedio del batch, para luego normalizarlos y ajustar los pesos de las activaciones de entrada a la próxima capa convolucional. Esta normalización trae como beneficio una mayor velocidad de entrenamiento y una mejor estabilidad del modelo, reduce la sensibilidad de la tasa de aprendizaje y actúa como regulador ligero.

**Capas Fully-Connected:** También llamadas capas densas, son capas donde las neuronas se encuentran conectadas a todas las neuronas de la capa anterior y a las neuronas de la capa siguiente [23]. Estas capas se encargan de la toma de decisiones en un modelo según la tarea correspondiente, como por ejemplo clasificación o regresión, evaluando las características extraídas previamente por las capas convolucionales. Estas capas reciben un vector con los valores de los mapas de características aplanados por medio de la función *Flatten*, que convierte los mapas de características en un

vector 1D y luego, empiezan a aprender patrones de la información según las entradas que reciben de las múltiples neuronas que las preceden y además integran esa información, teniendo en cuenta sesgos y los pesos de cada una de las neuronas anteriores[25]. En la Figura 5.6 se muestran las conexiones en una capa densa o fully connected. Finalmente, estas capas convierten las características interpretadas en una predicción como una clase (para modelos de clasificación) o un valor continuo (para modelos de regresión) utilizando funciones de activación acordes a ello (Sigmoid para problemas biclase, Softmax para multiclase o Lineal para regresión).

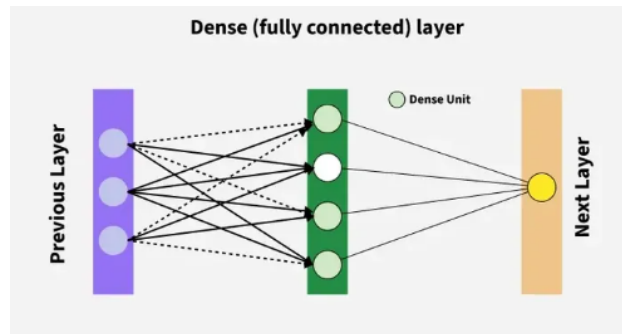


Figura 5.6: Capas densas o Fully connected

**Dropout:** El dropout es una técnica de regularización utilizada para reducir el sobreajuste [24]. Esta estrategia consiste en apagar de forma aleatoria las conexiones neuronales en las capas densas con una probabilidad  $p$ , con el propósito de evitar el aprendizaje de características complejas pero dependientes de otras neuronas, por lo que el modelo debe aprender a tener un buen desempeño aún cuando algunas neuronas aleatoriamente se encuentren apagadas. En la Figura 5.7 se observa un ejemplo del efecto que tiene la estrategia de regulación de dropout en una red neuronal.

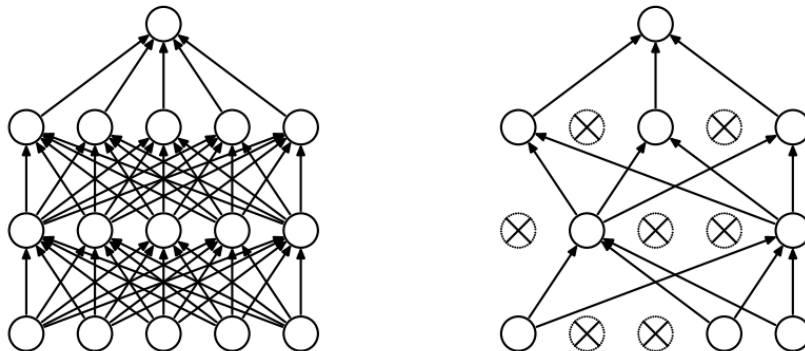


Figura 5.7: Efecto de un dropout con  $p = 0.5$  en una red neuronal

Tomada de [24]

En la Figura 5.8 se muestra un ejemplo del flujo y componentes de una red convolucional neuronal para una tarea de clasificación, dentro de la cual se encuentran los componentes explicados previamente.

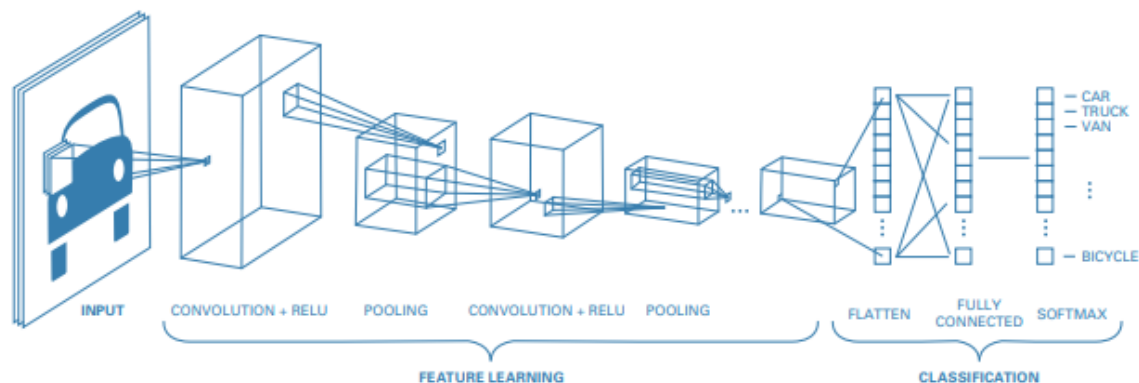


Figura 5.8: Ejemplo clásico de CNN de clasificación

Tomada de [24]

Posterior a la construcción del modelo de red convolucional viene la etapa de entrenamiento del modelo de aprendizaje profundo para desarrollar la función objetivo, ajustando iterativamente los parámetros de la red convolucional. Para este proceso, hay varios conceptos importantes, como lo son los siguientes:

**Learning Rate:** La tasa de aprendizaje o learning rate es la magnitud del paso que da un modelo en camino a su convergencia óptima [26]. En otras palabras, la tasa de aprendizaje determina la velocidad con que un modelo aprende, la duración del entrenamiento y los recursos consumidos durante este. Una tasa de aprendizaje con valores muy altos, es decir con pasos grandes, hará que el ajuste de parámetros de la red neuronal mientras el modelo aprende sea rápido, pero arriesgue que el no converja en un punto óptimo, y si el valor es muy pequeño, el modelo tardará mucho en aprender y podría atascarse.

**Función de pérdida:** la función de pérdida o loss function mide la diferencia entre las predicciones que realiza un modelo y las salidas esperadas [27]. Este valor es de vital importancia para el entrenamiento de un modelo de deep learning, puesto que la principal tarea de los modelos es optimizar sus parámetros para minimizar este valor de pérdida y acercar lo predicho a lo real. La función de pérdida utilizada para un modelo depende de la tarea específica del mismo, si es de clasificación se suele utilizar la función *Binary cross-entropy* (BCE) de forma estándar y si es de regresión las funciones Mean Squared Error (MSE) o Mean Absolute Error (MAE) son ampliamente usadas [27]; pero depende de la selección del programador.

**Optimizador:** Un optimizador es un algoritmo cuyo propósito principal es minimizar la función de pérdida [28]. Existen variedad de algoritmos con distintos abordajes matemáticos para disminuir la función de pérdida, pero el resultado de todos es modificar finalmente los parámetros del modelo para realizar esa reducción diferencia entre las predicciones y los valores reales. Ejemplos de estos optimizadores son el Stochastic Gradient Descent (SGD), Root Mean Square propagation (RMSprop) y el optimizador Adam (Adaptive Moment Optimization algorithm).

**Batch size:** El batch size o tamaño de lote, hace referencia al número de muestras de entrenamiento utilizadas antes de la optimización de parámetros [29]. En otras palabras, el conjunto de datos completo es subdividido en subgrupos o lotes con una cantidad específica de muestras que son procesadas antes de ajustar los parámetros, y a esa cantidad de muestras se le llama batch size.

**Época:** Una época es una iteración completa del modelo durante su entrenamiento donde el conjunto de datos de entrenamiento fue utilizado completamente. Un modelo necesita varias épocas o recorridos de los datos de entrenamiento para poder ajustar sus parámetros y mejorar de forma iterativa[30].

Los valores correspondientes a estos conceptos importantes para el entrenamiento de un modelo se les conocen como **hiperparámetros**, los cuales a diferencia de los parámetros de una red convolucional, requieren ser configurados para optimizar el desempeño de un modelo puesto que no son aprendidos durante el entrenamiento [31]. La literatura enfatiza que una mala configuración puede conducir a problemas como el sobreajuste (*Overfitting*) o el subajuste (*Underfitting*), por lo cual, el ajuste de hiperparámetros es un componente esencial en cualquier metodología de aprendizaje automático.

Tanto el *overfitting* y el *underfitting* representan riesgos inherentes en la construcción de modelos. El primero ocurre cuando el modelo se ajusta excesivamente a los datos de entrenamiento y pierde capacidad para generalizar, mientras que el subajuste ocurre cuando el modelo carece de la complejidad necesaria para capturar patrones esenciales. El equilibrio entre sesgo y varianza se considera fundamental para garantizar un desempeño óptimo [32]. Las estrategias de regularización, el aumento de datos y una adecuada separación de conjuntos de entrenamiento y validación son mecanismos comunes para mitigar estos efectos.

**Transfer Learning:** Una complicación común con el desarrollo de modelos de aprendizaje con CNN es el alto costo del entrenamiento desde cero, el cual requiere de grandes conjuntos de datos etiquetados y recursos computacionales extensos, requerimientos que no siempre se cumplen, especialmente en contextos médicos, donde la disponibilidad de bases de datos es limitada [33]. Una estrategia establecida como alternativa para abordar esta problemática es el *transfer learning* o aprendizaje por transferencia, el cual utiliza los parámetros aprendidos por modelos CNN entrenados en conjuntos de datos grandes y los transfiere para resolver problemas específicos aprovechando

el conocimiento previo de estos modelos, entrenando únicamente la última capa de clasificación para cumplir la tarea deseada.

Este método de aprendizaje se fundamenta en que características como las líneas curvas y rectas que construyen las imágenes son características universales en la mayoría de tareas de análisis de imágenes [33], por lo cual modelos entrenados para clasificaciones con datos no médicos pueden aprovecharse para aportar características importantes a tareas de clasificación médicas, reduciendo la necesidad de un conjunto de datos grande, el tiempo de entrenamiento y el costo de memoria.

**Pipeline metodológico para aprendizaje automático:** La literatura sobre aprendizaje automático propone un *pipeline* metodológico que organiza las etapas del desarrollo de un modelo: recolección de datos, preprocesamiento, selección del modelo, entrenamiento, validación y evaluación final [34]. En el contexto del análisis de imágenes neonatales, el preprocesamiento incluye operaciones como corrección de iluminación, normalización de color, segmentación de regiones de piel y extracción de características cromáticas relevantes. Estas etapas buscan reducir la variabilidad no clínica y mejorar la calidad de los datos que reciben los modelos.

**Threshold Tunning:** Una estrategia implementada en modelos de aprendizaje profundo para optimizar sus resultados es el *Threshold tuning* o ajuste de umbral de decisión. Esta estrategia consiste en buscar el umbral de decisión que genere una clasificación donde, según la curva precisión-recall, estas métricas de desempeño tengan un balance y el F1-Score sea alto y de esa forma, se pueda mejorar la capacidad discriminativa del modelo, especialmente en modelos donde las clases se encuentran desbalanceadas, puesto que este desbalance de datos entre las clases de interés hace que el umbral de decisión por defecto de modelos de clasificación con un valor de 0.5 no sea el umbral de decisión óptimo [35].

Para realizar el threshold tuning, inicialmente se extraen las probabilidades de las imágenes que entran al modelo respecto a la clasificación de interés y luego se evalúan distintos umbrales de decisión donde una probabilidad menor al umbral representaría la clase negativa y un valor igual o mayor al umbral representaría la clase positiva. Posteriormente, para cada valor umbral obtenido se calcula el valor de la métrica F1-Score y el umbral con mayor valor de esta métrica es aplicado al modelo de clasificación para que este obtenga mejores resultados acordes a las distribuciones de las clases que se analizan. Este procedimiento permite mejorar la capacidad de discriminación de un modelo y reducir tanto falsos negativos como positivos.

**Validación cruzada:** La validación cruzada o *cross-validation* es una técnica empleada con el propósito de evaluar la capacidad que tiene un modelo de aprender las características de interés de un conjunto de datos, en este caso características de color de piel asociadas a la ictericia neonatal. En otras palabras, la validación cruzada busca determinar si el modelo entrenado realmente obtiene

un desempeño alto porque tiene la capacidad de generalizar lo aprendido y que este no se encuentra limitado a los resultados de una única partición de los datos.

Uno de los tipos de verificación de desempeño de un modelo por medio de validación cruzada se denomina validación cruzada estratificada o de tipo *K-Fold*. Este tipo de *cross-validation* realiza una partición de la base de datos en  $k$  subconjuntos *folds* y se realiza la validación del modelo repetitivamente en  $k$  iteraciones, sin embargo, cada vez que se realiza una iteración, el modelo se entrena con  $k-1$  subconjuntos perteneciendo al conjunto de entrenamiento y se valida con el subconjunto restante. Este proceso se repite  $k$  veces hasta que todos los subconjuntos hayan participado como subconjunto de validación del modelo [36]. De esta manera, al finalizar las validaciones, se hace un promedio de los valores obtenidos en las métricas en todas las iteraciones y según el desempeño obtenido se evalúa la robustez del modelo.

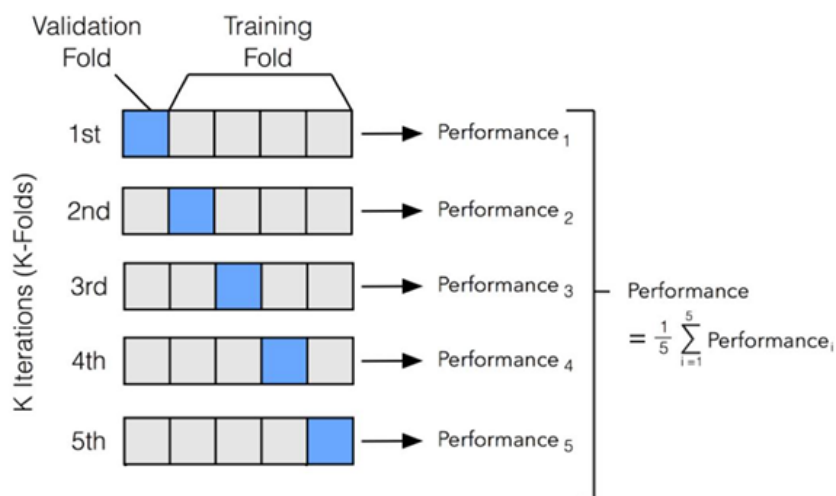


Figura 5.9: Ejemplo de Validación cruzada

Tomada de [23]

En conjunto, los fundamentos del aprendizaje automático y profundo presentados en esta subsección proporcionan el sustento teórico necesario para comprender las técnicas aplicadas en sistemas actuales de apoyo al diagnóstico basado en imágenes médicas. La integración de modelos clásicos y arquitecturas modernas, junto con metodologías rigurosas de validación y ajuste, constituyen la base conceptual sobre la cual se desarrollan aplicaciones destinadas a la estimación no invasiva de parámetros clínicos como la bilirrubina neonatal.

## 5.3. Trabajos Relacionados

La hiperbilirrubinemia es una condición que impacta a más del 60 % de los neonatos[37]. En la literatura, se han propuesto diferentes sistemas, tanto para evaluar la concentración de bilirrubina como el nivel de ictericia con diferentes enfoques invasivos y no invasivos [12]. La determinación del nivel de hiperbilirrubinemia mediante la medición de la bilirrubina sérica total (TSB) es una práctica invasiva habitual. Este método se fundamenta en la espectrofotometría y requiere una pequeña muestra de sangre obtenida con una jeringa y una aguja [38]. No obstante, la medición de la bilirrubina sérica total (TSB) requiere la intervención de profesionales médicos altamente capacitados y el uso de instalaciones adecuadas. Además, al ser un procedimiento invasivo, resulta estresante tanto para los bebés como para los padres a causa de los riesgos de infecciones, anemia y otras patologías derivadas del manejo séptico [39].

La evaluación visual se presenta como el método más sencillo y no invasivo para detectar la hiperbilirrubinemia en recién nacidos. No obstante, debido a que depende de la experticia del operador, es un proceso que puede verse afectado por el color de la piel de bebé, lo cual la hace una técnica no fiable para la detección [40]. Por otro lado, la técnica alternativa no invasiva más común es la Bilirrubinometría Transcutánea (BTC), la cual consiste en dispositivos portátiles que permiten medir la diferencia de las densidades ópticas de luz en las regiones de longitud de onda del azul y el verde (450 nm a 550 nm) [5]. Esta técnica reduce el muestreo en sangre en un 30 % pero aumenta el costo de diagnóstico y tiene limitaciones respecto a la edad gestacional, ya que solo es efectivo en bebés de más de 35 semanas y después de las primeras 24 horas de vida [40].

Con nuevas tecnologías emergentes, como cámaras digitales e inteligencia artificial, existe la posibilidad de tener sistemas más eficientes, rentables y portátiles para controlar la ictericia. La mayoría de estos métodos utilizan el procesamiento de imágenes y análisis de color para la detección de ictericia. Estos sistemas se han propuesto desde la última década y reportan un rendimiento igual o mejor en la detección de la condición. En 2021 se realizó una revisión sistemática que categoriza los sistemas de procesamiento de acuerdo al objeto estudiado: piel, esclerótica del ojo, tiras de orina y heces; siendo los últimos dos objetos inadecuados para el desarrollo de sistemas de identificación según se concluyó en la revisión [40]. Para la piel se reportaron los siguientes estudios: En [41], Rong et al propone un sistema llamado (AIB) que utiliza una tarjeta de calibración de color y un smartphone para predecir la ictericia y compara sus resultados con el BTC y TSB en recién nacidos prematuros y a término. Las imágenes de la piel capturadas se cargaron en un servidor en la nube que emplea software basado en aprendizaje de máquinas para proporcionar un índice de bilirrubina individualizado e inmediato. Los resultados reportados comprueban concordancia y correlación entre los métodos y el sistema, sin embargo, su sensibilidad y especificidad no fue lo suficientemente alta.

Juliastuti et al en [37] propusieron un sistema de identificación basado en la estadística para estimar el nivel de bilirrubina. El sistema utiliza una cámara digital para tomar imágenes de recién

nacidos y utiliza varios filtros para preprocesar y extraer características que fueron usadas en un análisis estadístico de regresión lineal multivariable, dando como resultado un nivel estimado de bilirrubina. Aunque el sistema propuesto era simple y tenía una fuerte categoría de correlación, este presentó un porcentaje de error del 17% y una precisión del 67%. En [42] se propuso un sistema que tiene como objetivo detectar ictericia en recién nacidos mediante el uso de una cámara digital. El sistema aplica diferentes técnicas de procesamiento de imágenes para extraer y analizar las características útiles de interés de las imágenes capturadas en tiempo real. Este estudio muestra avances significativos en el tema, sin embargo la población muestra es muy pequeña para verificar la eficacia del sistema.

La parte blanca del ojo humano, que es la capa exterior protectora que es opaca y fibrosa, se conoce como esclerótica. La esclerótica está formada principalmente por colágeno y algo de fibra elástica [43]. En [44] se propone una técnica que captura una imagen de la esclerótica mediante una cámara web, una linterna LED controlable y un sistema de aprendizaje de máquina. Ya que los ojos de los bebés permanecen cerrados la mayor parte del tiempo y su cabeza necesita estar en una posición fija, este método no es apropiado para esta población [45], a pesar de obtener una precisión del 90% en la detección de ictericia. Además, el sistema requiere el uso de gafas protectoras y un sistema de iluminación controlado.

Finalmente es importante mencionar al BiliCam, desarrollado por De Greef et al. [46] y Althamian [47], que emplea un smartphone para capturar imágenes, las cuales son procesadas por un algoritmo que extrae características relevantes. Estas características se utilizan como entrada para un modelo de aprendizaje automático. Este sistema también incluye una tarjeta de calibración de color que se utiliza en la frente y el esternón del recién nacido para mejorar la precisión de las mediciones.

Como se puede observar, el avance hacia sistemas de monitoreo de ictericia menos invasivos y más eficientes es evidente durante la última década. Aunque estos sistemas aún enfrentan desafíos en términos de precisión y validación clínica, representan un paso significativo hacia un cuidado neonatal más seguro y accesible. Es esencial continuar investigando y mejorando estas tecnologías para garantizar su utilidad clínica y su integración efectiva en la atención médica neonatal.

# Materiales y Métodos

---

En este capítulo se describen las distintas etapas abarcadas durante el desarrollo del proyecto, y como organización metódica de la información, se optó por dividir los métodos en cuatro etapas relacionadas a los objetivos específico del proyecto.

## 6.1. Características relevantes para el procesamiento de imágenes de ictericia neonatal

### 6.1.1. Características en la literatura

Con el propósito de identificar características de interés para el procesamiento de imágenes con relación a la ictericia neonatal, se realizó una búsqueda inicial en la literatura, específicamente en trabajos con objetivos similares al proyecto. Se identificaron variedad de características las cuales los distintos autores han utilizado como base para determinar una relación entre las imágenes de sus bases de datos y la estimación de un diagnóstico de ictericia neonatal u obtención de valores de bilirrubina sérica para el mismo fin, entre las cuales se encuentran: el promedio de los valores de los pixeles en cada canal de los espacios de color RGB, YCrCb y LAB en conjunto el gradiente de variación del color en los tres canales en RGB [46] [48], promedio de los valores de los pixeles en cada canal del espacio de color HSV [47], el promedio del valor de los pixeles en los canales B del espacio de color RGB y Cb del espacio de color YCrCb [43] [45], mapas de características concatenados de distintas agrupaciones entre de los espacios de color RGB, LAB, HSV y YCrCb [49], características de los histogramas de las imágenes en RGB, YCrCb y HSV [37], entre otras. Las distintas características encontradas fueron tomadas por los autores de sus respectivos estudios como información significativa para el análisis de la coloración amarillenta de las regiones de interés de la piel del neonato y como puntos focales para las evaluaciones posteriores de los modelos planteados en sus artículos.

A partir de estas características identificadas, se reconoció la importancia de trabajar con variedad de espacios de color para aprovechar características de la imagen que en el trabajo con un solo espacio de color no se pueden aprovechar, como la separación de luminancia y cromática en distintos canales en el espacio de color YCrCb [50], la importancia de los canales Cb y B para la detección del color amarillo [45] y los beneficios de espacios de color como HSV para la detección de texturas y de la zona de piel de la imagen neonatal [51].

### 6.1.2. Base de datos seleccionada para el proyecto

Como eje central del presente proyecto se seleccionó la base de datos de acceso abierto titulada “NeoJaundice: Neonatal Jaundice Evaluation in Demographic Images”, la cual contiene imágenes de 745 neonatos con edades gestacionales entre 30 y 42 semanas provenientes del Hospital Central de Xuzhou. El conjunto de datos se encuentra conformado en su totalidad de 2235 imágenes de tres regiones corporales específicas del neonato: cabeza, rostro y pecho. Las imágenes cuentan con una dimensión en promedio de 567x567 píxeles y con la presencia de una tarjeta de color estándar que rodea la región de piel a estudiar, con el propósito de garantizar una calibración cromática adecuada. La base de datos cuenta con etiquetas binarias basadas en el valor de bilirrubina sérica total extraído de la información otorgada por la institución hospitalaria, con un umbral en el valor de 12.9 mg/dL como punto máximo para un no tratamiento. Estas etiquetas posteriormente fueron confirmadas y validadas por un profesional en pediatría con amplia experiencia, resaltando así la confiabilidad clínica del conjunto de datos [52]. En la Figura 6.1 se muestran ejemplos de imágenes correspondientes a casos de neonatos con alta bilirrubina que requieren tratamiento (parte superior de la imagen) y neonatos con baja bilirrubina que no lo requieren (parte inferior de la imagen), al igual que la información clínica correspondiente al paciente neonatal almacenada en la base de datos.



Figura 6.1: Ejemplos de imágenes en la base de datos NeoJaundice

**Fila superior:** neonato de 5,2 días de nacido con bilirrubina sérica de 14,6 mg/dL (alta).

**Fila inferior:** neonato de 5,2 días de nacido con bilirrubina sérica de 3,9 mg/dL (baja).

6.1.3. Preprocesamiento

Teniendo en cuenta las características de interés identificadas previamente para la clasificación de ictericia neonatal y estimación de bilirrubina sérica, se tomó la decisión de dividir el preprocesamiento de las imágenes de la base de datos de interés en 3 etapas principales, las cuales se pueden apreciar en la Figura 6.2, se ejemplifican sus efectos en la Figura 6.3 y se describen a continuación:

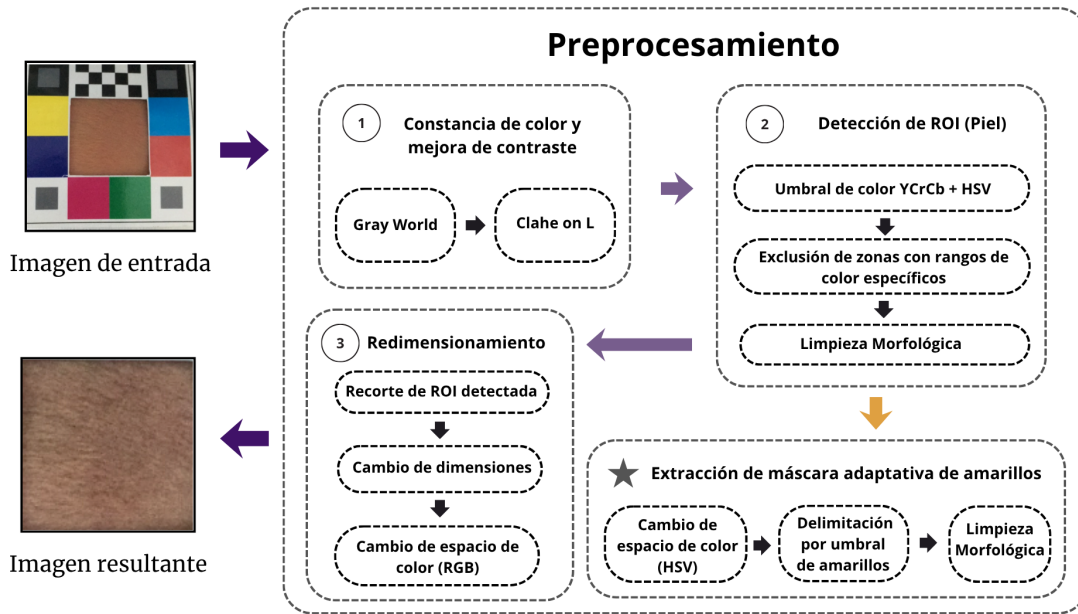


Figura 6.2: Preprocesamiento de las imágenes de la base de datos NeoJaundice

6.1.3.1. Etapa 1: Constancia de color y mejora de contraste

1. **Gray World:** Esta herramienta se fundamenta en el supuesto de que el promedio de los valores de los canales de una imagen son valores en tonalidad gris y su uso previo a la segmentación de imágenes busca permitir el reconocimiento correcto de los colores en la imagen al eliminar los efectos que puede generar la iluminación de la escena en la representación adecuada de los mismos [53]. El algoritmo hace un ajuste de los valores promedio de cada canal de color de la imagen en RGB por medio de un factor de corrección basado en la razón entre el promedio global de los 3 canales y el promedio de cada canal por separado. De esta manera, se genera una nueva imagen con mayor representación cromática al reducir los efectos de la iluminación en la tonalidad de los colores de la imagen, permitiendo una identificación de la coloración amarilla característica de la ictericia neonatal de la manera más acertada y fiel posible.
2. **Clahe on L:** Se utilizó la estrategia conocida como Contrast-Limited Adaptive Histogram Equalization (CLAHE) con el propósito de mejorar el contraste de las mismas. Para esto, se tomó el canal L de la imagen en el espacio de color LAB, correspondiente a la información

de intensidad de la luz únicamente, se subdividió la imagen en mosaicos pequeños (8x8) y se realizó la ecualización adaptativa del histograma de cada sección, para lograr una distribución uniforme de la intensidad de los píxeles en cada fragmento. Posteriormente se fusionan los mosaicos vecinos y así se genera el nuevo canal de intensidad de luz (L) con mejora del contraste hasta un límite determinado ( $clip=2$ ) para evitar amplificar ruido en la imagen [54]. Por último se juntaron los canales cromáticos A y B con el nuevo canal L, obteniendo así la imagen en LAB mejorada, la cual por último es regresada al espacio de color RGB para continuar con la siguiente etapa del preprocesamiento.

### 6.1.3.2. Etapa 2: Detección de la zona de piel de interés

El paso siguiente para el preprocesamiento de las imágenes se basó en la detección de la zona de piel de la imagen, puesto que es la región de interés para el proyecto y es necesario excluir los segmentos de la imagen que podrían generar alteraciones o errores en la identificación de características de color de la piel. Para ello, se emplearon las siguientes estrategias:

1. **Umbral de color de piel en espacios de color YCrCb y HSV:** Con el propósito de segmentar las imágenes de la base de datos para extraer la región de piel, se tomó la decisión de trabajar con los espacios de color HSV y YCrCb como punto inicial para la identificación de píxeles de piel. Se seleccionaron estos espacios de color debido a que estos permiten discriminar entre la información de color y la información de intensidad en imágenes en canales separados, lo que disminuye el efecto de las fuentes de luz aún con condiciones de iluminación irregular [55]. A partir de estos espacios de color, se buscó en la literatura distintos rangos de valores de píxeles en imágenes que se han determinado como correspondientes a píxeles de piel. En la sección de Anexos, en las tablas 10.2 y 10.1, se observan los distintos rangos de umbralización de píxeles de piel utilizados en la literatura para los espacios de color YCrCb y HSV respectivamente. Según estos rangos, se propusieron rangos umbral propios para determinar los píxeles de piel de las imágenes neonatales, para lo cual, se tomaron como condicionales que los píxeles identificados como piel debían encontrarse inmersos en ambos rangos de ambos espacios de color. De esta forma, se hizo uso de la herramienta `cv2.cvtColor()` para las transformaciones a ambos espacios de color y se empleó `cv2.inRange()` para identificar los píxeles dentro de estos rangos seleccionados y generar así una máscara binaria tanto en HSV como en YCrCb. Finalmente, se utilizó `cv2.bitwise_and()` para forjar la máscara final para la región de la piel de la imagen identificada por medio de la fusión de las máscaras obtenidas en ambos espacios de color.
2. **Exclusión de tarjeta de color:** Como paso siguiente en la segmentación de la ROI, se utilizaron como condicionales de exclusión de píxeles rangos específicos en HSV que representan los colores rojo, azul, verde, morado, blanco y negro, principalmente en intensidades de color altas para poder así extraer las secciones de la tarjeta de color que se encuentra rodeando la región de piel de interés. Para esto, se generó una máscara que acumulara los píxeles de la imagen que correspondieran dentro de los rangos de colores de la tarjeta de calibración y se

invirtió la misma con `cv2.bitwise_not()`, obteniendo así una máscara que permitiera únicamente conservar la región que no corresponda a la tarjeta de color. Finalmente, se fusionó esta máscara con la máscara de piel obtenida en el paso anterior para conservar los píxeles de piel sin la tarjeta de calibración.

3. **Limpieza morfológica:** Para mejorar la segmentación realizada previamente de la máscara de piel y garantizar un procesamiento posterior adecuado, se aplicó a la máscara binaria de piel obtenida operaciones morfológicas para eliminar el ruido de píxeles erróneamente categorizados como piel, rellenar huecos en la máscara y obtener una ROI más uniforme. Para esto, se emplearon las operaciones de apertura y cierre con un elemento estructurante elíptico de 7x7 píxeles creado con el comando `cv2.getStructuringElement(cv2.MORPH_ELLIPSE, (7,7))`. Este elemento estructurante permite la limpieza de la máscara al recorrer la imagen y posibilitar las operaciones de apertura (encargada de eliminar regiones no deseadas de píxeles aislados que no pertenecen a la región de piel) y de cierre (encargada del cierre de agujeros dentro de la región de piel para hacer la máscara una región continua para su procesamiento) [56]. Con este proceso, se obtiene una máscara de piel limpia y continua ideal para un procesamiento mucho más robusto en los modelos planteados más adelante.

#### 6.1.3.3. Etapa 3: Redimensionamiento de la imagen

1. **Recorte de la región de piel detectada:** Después de obtener la máscara para la zona de piel en la imagen, se utilizó la función `np.where()` para determinar las coordenadas de los píxeles cuya intensidad fuera mayor a 0, es decir, que correspondieran a píxeles de piel. A partir de estas coordenadas, se definió el área rectangular dentro de la cual se encontrarán incluidos todos los píxeles de interés identificando los valores máximos y mínimos en los ejes X e Y de la región segmentada. Con esta nueva área identificada, se realizó el recorte de la imagen para obtener una ROI con únicamente piel, evitando así que objetos adicionales como la tarjeta de color interfieran en el análisis posterior de la imagen para la clasificación y estimación de bilirrubina sérica.
2. **Redimensionamiento y cambio a RGB:** Como etapa final del preprocesamiento, la imagen recortada fue redimensionada a un tamaño de 224x224 y se regresó la imagen resultante de todo el proceso al espacio de color RGB.

#### 6.1.3.4. Etapa Adicional: Extracción de máscara de amarillos

Con el propósito de fortalecer las características de la coloración amarilla representativas de la ictericia neonatal en la entrada de los modelos de clasificación y regresión, se tomó la decisión de extraer una máscara adaptativa de amarillos de la imagen con base a la región de piel identificada previamente en el preprocesamiento. La máscara de amarillos se obtuvo mediante las siguientes estrategias de segmentación:

1. **Cambio a HSV y Umbral de amarillos:** Se transformó la imagen resultante de la primera etapa de preprocesamiento al espacio de color HSV con el propósito de generar una máscara

adaptativa de amarillos, la cual permita brindar información mucho más robusta de la amarillez de la piel a los modelos de clasificación y regresión. En primer lugar se separaron los canales H (matiz), S (saturación) y V (brillo) de la imagen para aislar el matiz de amarillos y posteriormente, extraer los valores de saturación y brillo de la imagen a partir de la máscara de piel obtenida previamente. Con estos valores, se generan límites adaptativos mínimos para los valores de saturación y brillo en la región de piel tomando los percentiles 35 y 30 respectivamente de los valores de estos canales. Se realiza esto con el propósito de ajustar los valores al contenido real de la imagen de piel de interés. Tomando estos valores obtenidos, se definen los rangos finales de amarillos para los tres canales para así forjar la máscara binaria de detección de amarillos correspondientes a la ictericia. Por último, se fusiona la máscara de amarillos obtenida con la máscara de piel para obtener los amarillos correspondientes únicamente a la amarillez de la piel.

2. **Limpieza morfológica:** De la misma manera que se realizó para la máscara de piel, para la máscara de amarillos se realizaron las operaciones morfológicas de apertura y cierre con un elemento estructurante de tamaño 5x5 para mejorar la continuidad de la imagen y remover el ruido de la máscara.

La máscara adaptativa de amarillos obtenida en esta etapa es finalmente concatenada con la imagen en el espacio de color RGB resultante de la etapa de redimensionamiento previamente mencionada, para así generar un cuarto canal de información para entregar a los modelos de clasificación y regresión que se describirán próximamente.

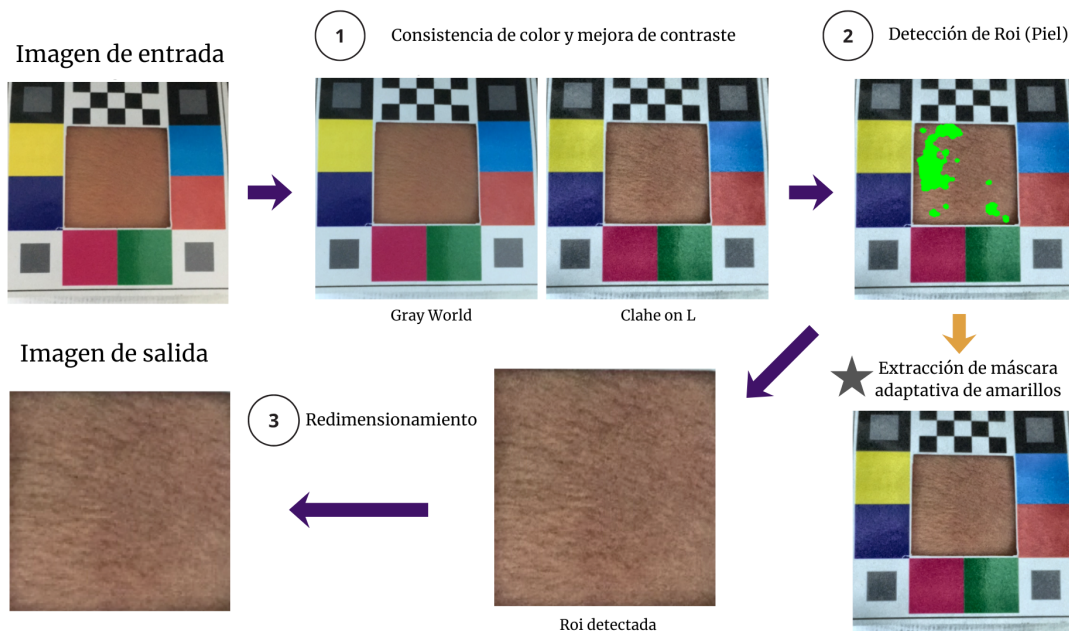


Figura 6.3: Ejemplo de los cambios de imagen en cada etapa del preprocesamiento

## 6.2. Implementación del procesamiento con deep learning

### 6.2.1. Entorno de trabajo computacional

Para la implementación computacional del proyecto, incluyendo las etapas de preprocesamiento previamente descritas y las relacionadas con el desarrollo de las arquitecturas de procesamiento de la base de datos, se utilizó **Google Colab Pro** + como el entorno principal de trabajo empleando el lenguaje de programación Python. Este entorno permitió el acceso a GPUs y TPUs, mayor capacidad de memoria en la nube y sesiones de ejecución de código prolongadas, lo cual permitió un desarrollo más eficiente de las tareas de entrenamiento de los modelos de deep learning.

### 6.2.2. Aumento de datos

Debido a la cantidad limitada de imágenes disponibles en la base de datos para el desarrollo del proyecto, se implementó un módulo de aumento de datos con el propósito de obtener modelos mucho más robustos y mitigar el sobreajuste. La estrategia se implementó exclusivamente a las imágenes resultantes del preprocesamiento sin la concatenación de la máscara de amarillos para preservar las características de imagen que se obtuvieron en ese proceso. El aumento de datos consistió en transformaciones geométricas implementadas ligeramente a las imágenes, incluyendo el reflejo con respecto al eje Y, rotación en un rango de  $\pm 15^\circ$  y escalado entre 0.9 (zoom out de 10%) y 1.1 (zoom in de 10%). Cada una de ellas cuenta con una probabilidad independiente de ser aplicada en la imagen de un 50% gracias al uso del generador de números aleatorios local `np.random.RandomState()`, el cual permite la creación de números aleatorios locales sin una semilla fija. Posteriormente, con `rand()`, se genera un número aleatorio entre 0 y 1 independiente para cada transformación, que en caso de ser menor a 0.5, aplica la transformación en la imagen que le corresponda, por lo cual en una sola imagen puede aplicarse una única transformación, más de una o ninguna.

Las tres transformaciones se aplicaron de la siguiente manera y orden:

1. **Reflexión horizontal:** Se genera un número aleatorio entre 0 y 1, y si es menor de 0.5, se realiza la reflexión con `cv2.flip()` respecto al eje vertical.
2. **Rotación:** Se genera otro número aleatorio entre el mismo rango de valores, y si cumple la condición, se genera inicialmente un valor angular aleatorio entre  $-15^\circ$  y  $+15^\circ$ , el cual, junto con las dimensiones de alto y ancho de la imagen, son utilizados para crear una matriz con el comando `cv2.getRotationMatrix2D()`, la cual cumple con la función de determina las nuevas posiciones que los píxeles de la imagen deben adoptar para cumplir con la rotación determinada en el ángulo aleatorio generado, poniendo como centro de rotación el centro de la imagen. Finalmente, el comando `cv2.warpAffine()` aplica la rotación en la imagen utilizando la matriz guía y devuelve una imagen del mismo tamaño que la original.
3. **Escalado:** Se genera un número aleatorio adicional que, al cumplir la condición previamente mencionada, permite la generación de un valor de escalado aleatorio entre 0.9 y 1.1, el cual es

utilizado junto con la posición del centro de la imagen por la función `cv2.getRotationMatrix2D()` para generar la matriz guía de la nueva posición de los píxeles, la cual es luego utilizada por `cv2.warpAffine()` para aplicar el escalado y generar la imagen final.

Con la aplicación de estas transformaciones es posible una mayor estabilidad para el entrenamiento de los modelos propuestos al brindar representaciones adicionales, lo cual permite una mayor capacidad de clasificación de las imágenes para ictericia neonatal y una estimación de bilirubina sérica más acertada.

### 6.2.3. Modelo de clasificación

#### 6.2.3.1. Modelos de transfer learning

Como propuesta inicial para el desarrollo del modelo de clasificación de imágenes de ictericia neonatal se planteó el uso de una arquitectura de transfer learning, con el propósito de aprovechar el preentrenamiento que presentan estos modelos en bases de datos de mayor magnitud y adaptarlo al problema de clasificación de interés. Para esto, se realizó una búsqueda de arquitecturas de transfer learning utilizadas en estudios con propósitos similares a este proyecto, para así escoger la más adecuada. El resultado de esa búsqueda se evidencian en la Tabla 10.3 en la sección de Anexos, donde se recopiló la información de los estudios encontrados, las arquitecturas empleadas y los resultados de desempeño obtenidos.

A partir de las arquitecturas observadas en los artículos de la Tabla 10.3, se decidió seleccionar 3 de estos para realizar la clasificación de las imágenes de la base de datos del proyecto. Las arquitecturas seleccionadas fueron las siguientes:

- **ResNet50:** Esta arquitectura con una profundidad de 50 capas y aproximadamente 25.6 millones de parámetros fue propuesta por He et al. en 2015 y fue preentrenada con la base de datos de clasificación ImageNet 2012 (1000 clases con 1.28 millones de imágenes de entrenamiento y 50 mil imágenes de validación). Esta arquitectura se basa principalmente en el aprendizaje residual, por lo cual el modelo aprende funciones residuales (la diferencia entre la salida deseada y la entrada) con ayuda de conexiones especiales entre capas llamadas *shortcut connections* en lugar de aprender directamente una función compleja [57]. La forma con que esta arquitectura entrena las redes neuronales permite reducir el problema de degradación (Al añadir más capas se incrementa el error de entrenamiento) y esto le otorgó excelentes resultados de exactitud con su dataset de entrenamiento, con exactitud de Top-1 de 74.9% y exactitud de Top-5 de 92.1% [58]; al igual que resultados considerablemente buenos en los estudios analizados en la Tabla 10.3.
- **DenseNet121:** Esta arquitectura con una profundidad de 121 capas y aproximadamente 8.1 millones de parámetros fue preentrenada a su vez con ImageNet 2012 y utiliza conexiones densas entre capas, las cuales aseguran el máximo flujo de información en la red debido a que cada una de las capas convolucionales recibe entradas adicionales de las capas anteriores

directamente y le entregan la salida a las capas convolucionales posteriores. Este patrón de conectividad permite aliviar el problema de desvanecimiento del gradiente, fortalecer la propagación de características, fomentar la reutilización de características y reducir el número de parámetros [59]. Este modelo también obtuvo unos resultados satisfactorios en clasificación para ImageNet con exactitud Top-1 de 75.0% y exactitud Top-5 de 92.3% [58], y adicionalmente ha sido trabajado con la base de datos de seleccionada como se muestra en el estudio [52] en la Tabla 10.3 con resultados satisfactorios.

- **EfficientNet B0:** Esta arquitectura es la base de una familia de arquitecturas EfficientNet caracterizadas por su método de escalado compuesto, el cual promueve la exactitud y la eficiencia de sus modelos optimizando el ancho, la profundidad y la resolución de los mismos de forma uniforme por medio del cálculo de coeficientes de escalado [60]. Esta arquitectura con aproximadamente 5.3 millones de parámetros, entrenada también con ImageNet, obtuvo resultados considerablemente buenos con una exactitud Top-1 de 77.1% y Top-5 de 93.3% [58], lo cual muestra su efectividad con una cantidad baja de parámetros a comparación de otras arquitecturas. A su vez, en el estudio [61] presentó resultados de exactitud altos y muy similares a su modelo sucesor EfficientNet B7 en la clasificación de ictericia neonatal, por lo cual se optó por emplear la arquitectura base al generar menor costo computacional.

En la Tabla 7.1 en la sección de Resultados se evidencian los resultados obtenidos para la clasificación de ictericia neonatal empleando estos 3 modelos de transfer learning.

### 6.2.3.2. Modelos de clasificación experimentales 2DCNN

Posterior al análisis de los resultados obtenidos mediante la clasificación por transfer learning, se tomó la decisión de elaborar una arquitectura propia para realizar la clasificación de las imágenes, con el propósito de obtener mejores resultados de desempeño. Para el diseño de la arquitectura propuesta se realizaron cuatro experimentos probando varias estrategias para mejorar el desempeño de los modelos, los cuales fueron los siguientes:

1) **2DCNN:** Como modelo inicial se propuso una red convolucional bidimensional para la clasificación de ictericia, que recibe como entrada la imagen de piel neonatal de tamaño 224x224 en RGB. El modelo cuenta inicialmente con dos bloques convolucionales, en los cuales se aplican convoluciones 2D con 32 filtros de tamaño 3x3 para el primer bloque y 64 filtros del mismo tamaño para el segundo bloque. Después de cada convolución se aplica la operación de Max pooling de tamaño 2x2 para reducir la resolución de los mapas de características, y después del último bloque convolucional se aplanan los mapas de características con Flatten y posteriormente se procesan en capas densas de 128 y 64 neuronas con activación ReLu y acompañadas de Dropout del 40%. En la capa de salida, la neurona con activación sigmoideal entrega el resultado de la clasificación binaria de ictericia o no ictericia.

2) **2DCNN + Máscara de amarillos:** Para este experimento, se buscó generar un modelo que le otorgara una mayor relevancia a la información cromática característica de la ictericia neo-

natal, por lo que se tomó la decisión de concatenar a la imagen de entrada en RGB la máscara de amarillos obtenida en el preprocesamiento previo realizado a la imagen. De esta manera, ahora el modelo recibe como entrada 4 canales, los tres canales normalizados de la imagen en RGB y la máscara de amarillos. En comparación al anterior experimento, a esta arquitectura se le agregó un bloque convolucional adicional con 128 filtros de dimensiones 3x3 para una mayor obtención de mapas de características y se agregó después de cada convolución en los 3 bloques la operación Batch Normalization, que reemplazó a la función Dropout en su papel regulador. Adicionalmente, las capas densas en este experimento se tomaron con 256 y 128 neuronas, manteniendo la función de activación, el valor de Dropout y la capa de salida de clasificación binaria.

3) **2DCNN + Máscara de amarillos con Threshold tuning**: En este tercer experimento se tomó la arquitectura obtenida del segundo experimento, con la misma entrada y las mismas capas, y se aplicó adicionalmente la estrategia de *Threshold tuning* con el propósito de optimizar la métrica de desempeño F1-Score del modelo y mejorar su capacidad discriminativa. Se utilizó esta estrategia en el experimento puesto que la base de datos de interés (NeoJaundice) cuenta con un desbalance entre las clases etiquetadas, lo que se observa al tener una mayor cantidad de imágenes de neonatos sin ictericia (1302 imágenes) que con ictericia (933 imágenes). Debido a esto, para posibilitar una mejor discriminación de clases, se optimizó un nuevo umbral de decisión que se adaptara a la tarea de clasificación de ictericia neonatal, las probabilidades de las clases de interés y así, mejorar la capacidad de detección del estado icterico.

4) **2DCNN + Máscara de amarillos con Grid Search y Threshold tuning**: Este último experimento realizado toma como base la arquitectura desarrollada para el segundo experimento (entrada de 4 canales y la misma cantidad de capas), sin embargo, implementa adicionalmente una búsqueda de hiperparámetros o *Grid Search* para determinar una combinación óptima para un mejor rendimiento del modelo [62]. En este experimento se evalúan distintas combinaciones de valores otorgadas manualmente para la tasa de aprendizaje ( $1x10^{-3}, 1x10^{-4}, 1x10^{-5}$ ), Dropout (0.3, 0.4, 0.5) y número de filtros ([32, 64, 128], [64, 128, 256]) en los bloques convolucionales. De esta forma, se hace una evaluación de 18 configuraciones distintas para la arquitectura, cada una con un entrenamiento rápido de 8 etapas para estimar un rendimiento relativo de las distintas combinaciones para luego seleccionar la combinación de hiperparámetros que obtenga una mayor exactitud durante su validación. Este modelo con mayor exactitud fue entrenado nuevamente y su umbral de decisión fue optimizado para maximizar el F1 score.

Cada uno de los modelos experimentales fue probado con la composición mencionada previamente en su descripción y utilizando los hiperparámetros de la Tabla 10.4 que se encuentra en la sección de Anexos. Posteriormente, se realizó el análisis de desempeño de los modelos y según los resultados obtenidos, se seleccionó el modelo con mayor exactitud como la arquitectura final de clasificación. En la sección de resultados en la Tabla 7.2 se encuentran plasmados los resultados obtenidos para cada arquitectura y en la Figura 7.3 se presenta la composición de la arquitectura final seleccionada.

### 6.2.4. Modelo de regresión

#### 6.2.4.1. Características estadísticas del conjunto de datos de bilirrubina sérica

Como primero paso en el desarrollo del modelo de regresión para la estimación de bilirrubina sérica, se analizaron los datos obtenidos de la base de datos de bilirrubina sérica en el entorno de desarrollo integrado de acceso abierto **RStudio** utilizando el lenguaje de programación R, con el cual se extrajeron medidas descriptivas y la gráfica de densidad del conjunto de datos, que se pueden visualizar en la Figura 6.4.

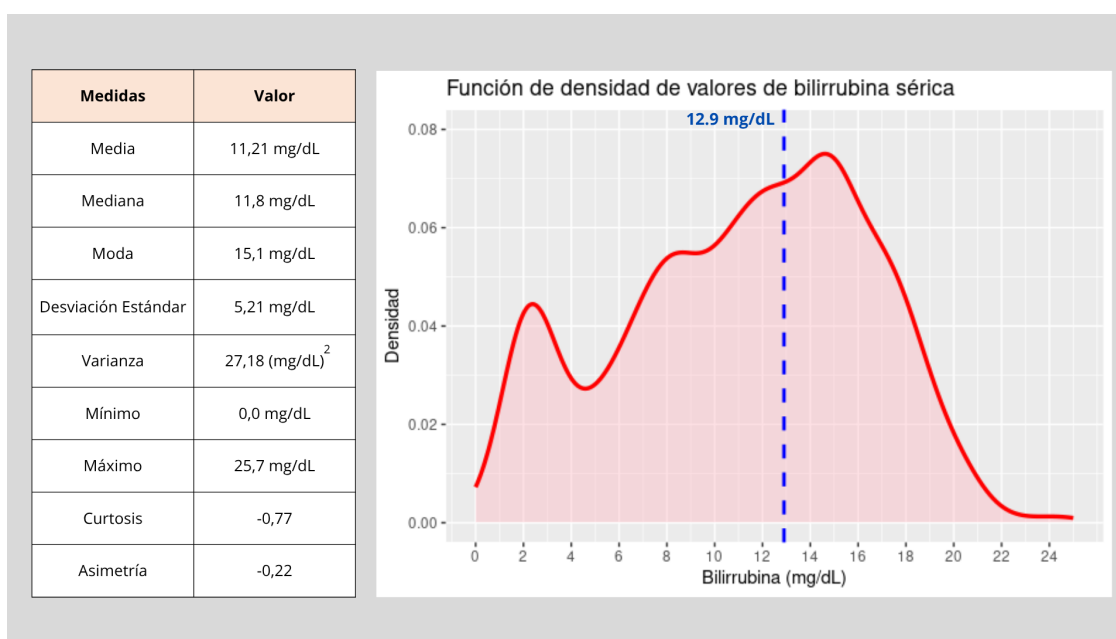


Figura 6.4: Características del conjunto de datos de bilirrubina sérica de la base de datos NeoJaundice

Los datos descriptivos en la Figura 6.4 muestran una concentración de datos en valores más altos (alrededor de la moda), sin embargo aún se evidencia una variabilidad moderada de datos de bilirrubina sérica según lo que nos muestra el valor de su desviación estándar y varianza. Con respecto a su asimetría se puede determinar que a pesar de la concentración de datos en valores altos, se encuentra una mayor dispersión de datos hacia la izquierda, lo cual lleva a los valores de media y mediana hacia la zona de valores de bilirrubina sérica normal (lado izquierdo de la línea correspondiente al valor umbral entre bilirrubina sérica normal e icterica) y por tener una curtosis negativa, se evidencia que el conjunto de datos tiene un comportamiento más aplanado que una distribución normal.

#### 6.2.4.2. Modelo experimental de regresión

Debido al acceso a los datos de bilirrubina sérica correspondientes a las imágenes de la base de datos, se tomó la decisión de implementar un modelo que permita estimar el dato de bilirrubina sérica a partir de las imágenes para generar un soporte al diagnóstico mucho más robusto, para ello se realizaron los siguientes experimentos:

1) **Arquitectura de regresión simple:** Esta arquitectura de regresión toma como entrada una imagen de 128x128 píxeles con 4 canales (R,G,B y la máscara de amarillos) y se compone principalmente de tres bloques convolucionales, los cuales se componen de convoluciones con 32, 64 y 128 filtros de tamaño 3x3, seguidos cada uno de Batch Normalization y de Max pooling 2x2. Posteriormente los mapas de características extraídos por medio de las convoluciones son aplanados con Flatten y luego pasan por capas densas de 256 y 128 neuronas seguidas cada una de un Dropout; para terminar con la capa de regresión con salida lineal para obtener el valor estimado de bilirrubina sérica. Adicionalmente, a este modelo de regresión se le agregó una etapa de validación cruzada por paciente, donde se plantea que en cada partición del dataset, los datos de un paciente queden o en el grupo de entrenamiento o el grupo de validación en los 5 folds propuestos, para luego generar un promedio de las métricas de evaluación seleccionadas para regresión.

2) **Arquitectura de regresión con Global Average Pooling:** En esta arquitectura experimental, a comparación del anterior experimento, se agrega una capa convolucional adicional, contando así con un total de cuatro capas convolucionales con 32, 64, 128 y 256 filtros con tamaños de 3x3 y con padding "same" para agregar bordes artificiales que eviten el cambio de tamaño posterior a la convolución. Las operaciones de convolución son seguidas de Batch Normalization en los cuatro bloques y Max pooling para los primeros tres, debido a que en el cuarto bloque se optó por utilizar una capa de GlobalAveragePooling2D con el propósito de reducir los valores de los mapas de características obtenidos empleando promedios globales de estos y así mitigar el sobreajuste del modelo. Posteriormente se incluyeron dos capas densas de 256 y 128 neuronas seguidas de regulación por Dropout y finalmente una capa lineal de decisión para la estimación de bilirrubina.

Cada una de las arquitecturas experimentales mencionadas fue probada con los hiperparámetros enlistados en la Tabla 10.5 que se encuentra en la sección de Anexos, y posterior a ello, se evaluó el desempeño de los mismos para la estimación de bilirrubina mediante métricas de evaluación para seleccionar la más óptima como arquitectura final de regresión. Los resultados obtenidos se encuentran en la Tabla 7.4 en la sección de Resultados y la composición de la arquitectura final seleccionada se encuentra en la Figura 7.4.

#### 6.2.5. Prueba adicional con imágenes externas

Al obtener el modelo con mejor desempeño para la estimación de bilirrubina sérica se planteó realizar una prueba adicional para determinar los alcances del modelo en imágenes con características distintas a la base de datos de entrenamiento. De esta manera se utilizaron las imágenes de la base

de datos **NJN** tomada de [63], una base de datos accesible en línea con 670 imágenes neonatales (560 normales y 200 con ictericia), las cuales se ejemplifican en la Figura 6.5. En la Tabla 7.5 en la sección de resultados se muestran los resultados obtenidos con el modelo de regresión seleccionado.

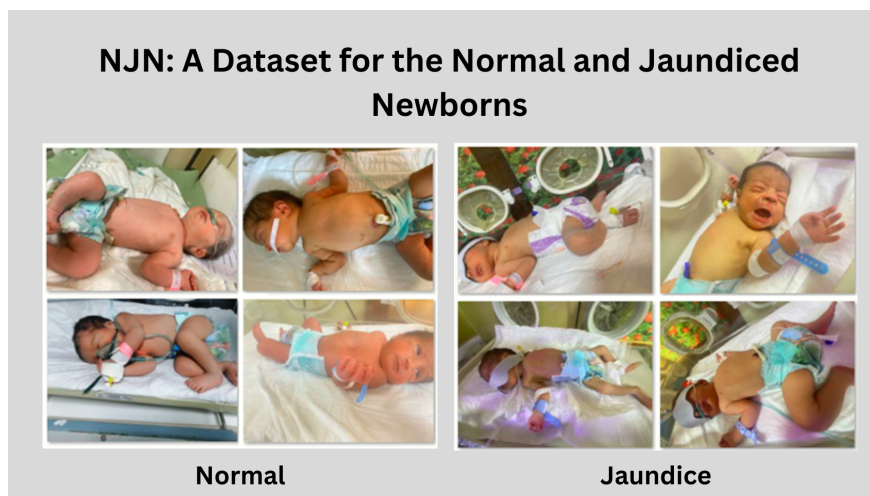


Figura 6.5: Ejemplos de las imágenes en la base de datos NJN

### 6.2.6. Modelo combinado de clasificación y regresión

Al finalizar la prueba de modelos tanto de clasificación como de regresión para la base de datos de interés, se guardaron los modelos entrenados con mejor desempeño según las métricas evaluadas (Ver sección de Resultados) y estos fueron implementados en un pipeline de inferencia conjunta, en el cual se integraron ambos modelos con las etapas de preprocesamiento y posteriormente con su etapa de validación, permitiendo así obtener de forma simultánea la clasificación de ictericia y la estimación cuantitativa de bilirrubina sérica de las imágenes neonatales.

## 6.3. Validación del sistema

Para la validación de los modelos desarrollados para la clasificación de ictericia neonatal y estimación de bilirrubina sérica se tomaron en cuenta las métricas de desempeño mencionadas en [49] y [64], las cuales fueron implementadas en código haciendo uso de la librería `scikit-learn`, específicamente de su módulo `sklearn.metrics`.

### 6.3.1. Métricas de validación para el modelo de clasificación

- **Accuracy** (Exactitud): Métrica que mide la proporción de predicciones correctas en el total de predicciones realizadas por el modelo.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (6.1)$$

- **Precision:** Métrica que mide la proporción entre predicciones correctas positivas (Neonatos con ictericia) entre todas las predicciones positivas determinadas por el modelo.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6.2)$$

- **Recall (Sensibilidad):** Métrica que mide la proporción de predicciones positivas (Neonatos con ictericia) entre las predicciones positivas reales.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6.3)$$

- **F1-score:** Métrica que es la media armónica de la Precisión y Recall (métrica única que genera balance entre las dos).

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6.4)$$

- **ROC AUC:** Métrica que hace referencia al área bajo la curva ROC (Característica Operativa del receptor). Esta métrica evalúa la capacidad diagnóstica del modelo para distinguir entre neonatos con y sin ictericia, donde un valor cercano a 1 indica un mejor desempeño en el diagnóstico.

$$\text{AUC} = \sum_{i=1}^{n-1} (FPR_{i+1} - FPR_i) \cdot \frac{TPR_{i+1} + TPR_i}{2} \quad (6.5)$$

### 6.3.2. Métricas de validación para el modelo de regresión

- **MAE (Error absoluto medio):** Métrica que mide la magnitud media de errores en un conjunto de predicciones.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6.6)$$

- **RMSE (Raíz del error cuadrático medio):** Métrica que mide el promedio de la magnitud media del error en las mismas unidades de la variable objetivo (mg/dL).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6.7)$$

- $R^2$  (Coeficiente de determinación): Métrica que indica la proporción de varianza de los datos que se predicen (estimación de bilirrubina sérica) con respecto a los valores reales (datos de bilirrubina sérica de la base de datos).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (6.8)$$

### 6.3.3. Validación cruzada

En el caso del modelo de clasificación determinado como el que obtuvo mejor rendimiento en las métricas mencionadas previamente (Ver la Tabla 7.2), se realizó un proceso de validación cruzada con una subdivisión en 10 folds y en cada iteración se calcularon las 5 métricas de desempeño definidas en esta sección, para al finalizar las iteraciones del modelo, obtener un promedio y determinar así el aprendizaje de la arquitectura de clasificación de ictericia neonatal. Adicionalmente, dentro de este *cross-validation* del modelo de clasificación se empleó una ponderación de clases con `class_weight` para solventar el desbalance de clases luego y se obtuvieron las métricas para cada uno de los fold, teniendo en cuenta el mejor threshold encontrado en cada iteración que maximizara el F1-Score (se aplicó threshold tuning para cada una de las iteraciones).

Por otra parte, se empleó adicionalmente la validación cruzada como parte de uno de los experimentos evaluados para ser seleccionados como el modelo de estimación de bilirrubina sérica, en el cual se realizó la validación cruzada con una cantidad de 5 folds sin threshold tuning para evaluar los resultados en este modelo (Ver Tabla 7.4).

## 6.4. Desarrollo de la interfaz gráfica

### 6.4.1. Características importantes para la interfaz

Como parte del presente proyecto, se desarrolló una interfaz gráfica interactiva cuyo propósito principal es articular el flujo completo de procesamiento de imágenes, inferencia y visualización de resultados, garantizando al mismo tiempo usabilidad, trazabilidad y pertinencia clínica. La interfaz fue concebida bajo los siguientes lineamientos:

- **Orientación clínica y comprensibilidad:** La presentación de los resultados se diseñó para que el usuario pueda interpretar de forma inmediata el estado del neonato con relación a la ictericia. Por tal motivo, se incluyen elementos visuales y textuales que hacen explícita la clasificación (*ictericia / no ictericia*), el nivel estimado de bilirrubina sérica (mg/dL) y una interpretación cualitativa del rango (leve, moderado, severo), con el fin de favorecer la toma de decisiones clínicas informadas sin reemplazar el criterio médico.
- **Integración directa con los modelos desarrollados:** La interfaz se diseñó como entorno de despliegue del *pipeline* propuesto en los objetivos previos, conectando de manera transparente las etapas de preprocesamiento, la extracción de características (incluyendo el canal adicional

asociado a la máscara adaptativa de amarillos) y los modelos de clasificación y regresión. Esto permite evaluar, en un contexto cercano al uso real, el desempeño conjunto de las estrategias metodológicas implementadas.

- **Modularidad y escalabilidad:** El sistema se estructuró en módulos independientes para diagnóstico individual, análisis por lotes y visualización global de resultados, lo que facilita su adaptación a distintos escenarios: validación experimental, análisis retrospectivo de bases de datos o integración futura en flujos clínicos más amplios.
- **Usabilidad y experiencia de usuario:** Se priorizó un diseño intuitivo que organizara las funcionalidades mediante navegación lateral y secciones claramente diferenciadas, minimizando la carga cognitiva del usuario. La disposición de botones, formularios y paneles de resultados busca reducir errores de uso y optimizar el tiempo requerido para ejecutar cada análisis.
- **Soporte para validación e investigación:** La interfaz no solo ofrece diagnóstico individual, sino que también permite el análisis de grandes volúmenes de imágenes y la generación automática de métricas globales de desempeño. Esta característica es fundamental para la evaluación rigurosa de los modelos y su comparación frente a otros enfoques del estado del arte.

### 6.4.2. Diseño de la interfaz

A partir de estos lineamientos, y con el propósito de definir una propuesta visual coherente con las funcionalidades descritas, se elaboró, con el apoyo de herramientas de inteligencia artificial, un **boceto preliminar de la interfaz**. Este esquema inicial permitió representar la estructura general de navegación y los componentes principales de cada módulo antes de proceder a la implementación en Streamlit. En dicho boceto se visualizaron la barra lateral de navegación, las secciones principales de visualización de resultados y los paneles informativos para la interpretación de la predicción de bilirrubina. El diseño preliminar, mostrado en la Figura 6.6, sirvió como referencia para validar la organización lógica de los elementos, la coherencia visual entre módulos y la orientación clínica del sistema antes del desarrollo final.

La interfaz fue implementada utilizando el entorno de desarrollo **Streamlit** en Python, permitiendo una integración directa con los modelos de aprendizaje profundo y los procesos de preprocesamiento definidos previamente. Su estructura modular se compuso de diferentes vistas o secciones funcionales, todas accesibles mediante una barra lateral que mantiene la estética limpia y el esquema de color azul–blanco característico de aplicaciones médicas.

#### 6.4.2.1. Vistas de la interfaz

La interfaz está compuesta por tres vistas principales, las cuales corresponden a las etapas funcionales del sistema:

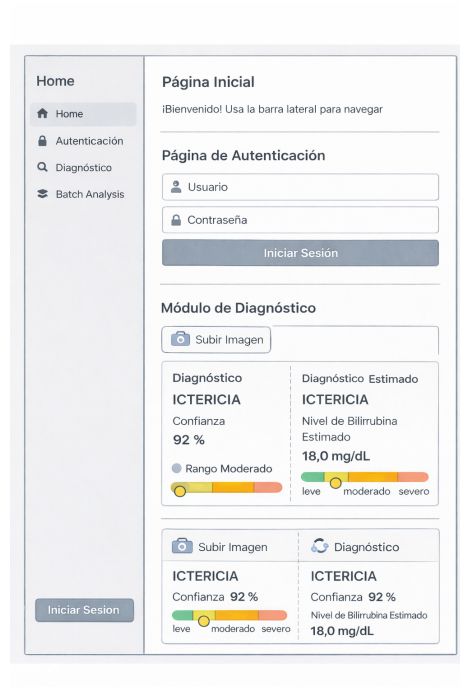


Figura 6.6: Boceto preliminar de la interfaz

1. **Página de inicio:** Presenta una descripción general del sistema, su propósito y las instrucciones de navegación. Esta sección orienta al usuario sobre el flujo de uso de la interfaz.
2. **Módulo de diagnóstico individual:** Facilita la carga de una única imagen neonatal en formato .jpg, .jpeg o .png. Tras la carga, la imagen es preprocesada (redimensionada, normalizada y filtrada con la máscara adaptativa de amarillos) y se ejecutan los modelos de clasificación y regresión para determinar el diagnóstico (*ictericia / no ictericia*) y el nivel estimado de bilirrubina sérica (mg/dL). Los resultados se presentan mediante indicadores visuales y textuales:
  - Diagnóstico predicho (etiqueta).
  - Nivel de confianza del modelo.
  - Valor estimado de bilirrubina.
  - Interpretación clínica del rango: leve, moderado o severo).
3. **Módulo de análisis por lotes:** Permite el procesamiento masivo de imágenes con fines de validación e investigación. Carga un archivo .csv con la información de referencia y procesa automáticamente las imágenes correspondientes, generando para cada una:
  - Etiqueta real y predicha.

- Nivel de confianza del modelo.
- Valor estimado de bilirrubina.
- Indicador de coincidencia .

Los resultados son almacenados en archivos con sello temporal dentro del directorio `resultados/`. Además, se generan métricas globales de desempeño (exactitud, distribución de clases y tasa de aciertos) y visualizaciones gráficas, como histogramas y gráficos de barras, para analizar el comportamiento del modelo.

#### 6.4.2.2. Funcionamiento de la interfaz

El funcionamiento de la aplicación se sustenta en una estructura modular que separa las funciones de procesamiento, inferencia y visualización, garantizando así la escalabilidad y el mantenimiento del código. Los componentes principales son los siguientes:

- **Gestión de datos y modelos:** Incluye los módulos de carga de datos (`core/database.py`) y de procesamiento de imágenes (`core/processor.py`), los cuales administran la vinculación entre las imágenes, las etiquetas y los modelos de clasificación y regresión almacenados en formato `.h5`.
- **Pipeline de preprocesamiento integrado:** Cada imagen pasa por las etapas definidas en los objetivos previos (conversión a RGB, ajuste de tamaño, normalización y generación de máscara de amarillos), asegurando que el flujo de inferencia reproduzca las mismas condiciones que el entrenamiento.
- **Ejecución de inferencia y despliegue de resultados:** Los modelos son invocados secuencialmente para emitir la clasificación binaria y la estimación numérica de bilirrubina. Los valores resultantes se transforman en información interpretable para el usuario y se muestran en paneles interactivos dentro de la interfaz.
- **Actualización y trazabilidad de resultados:** El sistema almacena cada ejecución de análisis en archivos `.csv` generados automáticamente, los cuales son posteriormente utilizados por el módulo de tablero o *dashboard* para consolidar estadísticas de desempeño a lo largo de distintas sesiones.

#### 6.4.3. Estructura del proyecto e implementación de la interfaz

La interfaz gráfica del sistema fue implementada siguiendo una estructura modular, diseñada para mantener una separación clara entre las capas de presentación, la lógica de procesamiento y los módulos de inferencia basados en aprendizaje automático. Esta organización facilita la mantenibilidad del sistema, permite su escalabilidad y favorece la incorporación de nuevas funcionalidades sin afectar la arquitectura general. En la sección de Anexos se puede observar la organización general de los directorios y archivos del sistema junto con los requerimientos operativos para desarrollar el entorno y librerías dependientes.

## 6.4.3.1. Flujo funcional del sistema

El flujo funcional de la interfaz puede resumirse de la siguiente manera:

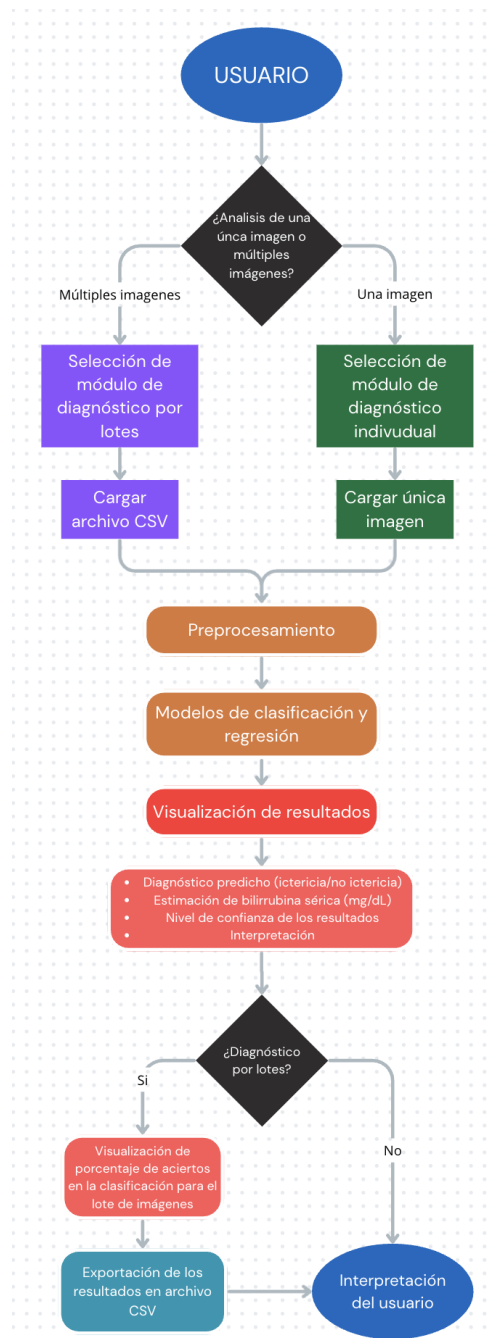


Figura 6.7: Flujo funcional del sistema de soporte al diagnóstico implementado en la interfaz desarrollada.

#### 6.4.3.2. Validaciones y pruebas

Para garantizar el correcto funcionamiento del sistema, se desarrollaron pruebas unitarias y de integración dentro del directorio `test/`. Entre ellas:

- `TEST.py`: validaciones básicas de arranque y dependencias.
- `dataset_prediccion_simulada.py`: genera datos falsos para probar el tablero sin modelos reales.
- `dataset_viewer_test.py`: verifica rutas y formatos del conjunto de datos.

Adicionalmente, se configuraron convenciones de estilo PEP8, manejo de logs, y control de errores para mantener la trazabilidad del procesamiento.

# Resultados y Discusión

---

En la presente sección se presentan los resultados obtenidos durante las etapas experimentales del proyecto para el desarrollo de los modelos de clasificación y regresión propuestos, para los cuales se realizó una división del conjunto de datos de 80/20 (80 % de los datos para entrenamiento y 20 % para validación).

## 7.1. Modelo de clasificación

### 7.1.1. Resultados de Arquitecturas de transfer learning

En la Tabla 7.1 se evidencian los resultados de exactitud y F1 score en porcentaje obtenidos para las tres arquitecturas de transfer learning propuestas. Para la implementación de estas arquitecturas se optó por descongelar 50 capas de los modelos preentrenados como estrategia de optimización de la capacidad discriminativa de los modelo aprovechando el entrenamiento previo en ImageNet y fomentando a su vez una mayor adaptación a las características para la clasificación de ictericia neonatal.

Tabla 7.1: Resultados arquitecturas de Transfer Learning

Modelo	Accuracy	F1 Score	Recall
ResNet50	59,73 %	50,82 %	50,00 %
DenseNet121	60,85 %	25,22 %	33,33 %
EfficientNet B0	37,36 %	34,89 %	38,54 %

Según los resultados de la imagen se puede observar que la arquitectura con un desempeño mejor a comparación de las otras dos es ResNet50, al tener el valor más alto de F1 Score y valores a su vez altos en Accuracy y Recall; sin embargo, para un modelo de clasificación confiable estos datos obtenidos sugieren que este modelo y los otros dos, DenseNet121 y EfficientNet B0, tienen dificultades para adaptarse a la tarea de clasificación de ictericia neonatal con el conjunto de datos propuesto, lo que se aprecia debido a que el rendimiento mostrado por los tres modelos es insuficiente en comparación con los resultados obtenidos en otros estudios con la misma temática de clasificación (Ver Tabla 10.3).

Considerando que las arquitecturas de transfer learning generan un consumo computacional de memoria alto (GPU), junto con unos tiempos de entrenamiento extensos especialmente al descongelar capas de los modelos y su desempeño se ve comprometido con bases de datos pequeñas en tareas

sustancialmente distintas y específicas por su preentrenamiento en tareas de clasificación generales con bases de datos como ImageNet, se puede justificar la creación de una arquitectura personalizada para la clasificación de ictericia neonatal diseñada para aprender las características relevantes de color específicas para la distinción de neonatos con y sin ictericia, que reduzca la carga computacional y uso de memoria, se adapte mejor al problema y otorgue resultados más satisfactorios.

### 7.1.2. Resultados de Arquitecturas 2DCNN

En la Tabla 7.2 se muestran los resultados obtenidos para los 4 experimentos evaluados para la arquitectura 2DCNN de clasificación de ictericia neonatal por medio de las métricas de Accuracy (exactitud), Precisión, F1 score, Recall y ROC AUC. Adicionalmente, en las figuras 7.1 y 7.2 se visualizan la matriz de confusión y las curvas de aprendizaje obtenidas del modelo de clasificación con mejor desempeño respectivamente.

Tabla 7.2: Resultados de las arquitecturas experimentales 2DCNN

Modelo	Accuracy	Precision	F1 Score	Recall	ROC AUC
<b>Experimento 1</b>	75,62 %	72 %	69,81 %	67,74 %	83,54 %
<b>Experimento 2</b>	78,52 %	76,16 %	73,18 %	70,43 %	85,49 %
<b>Experimento 3</b>	81,66 %	72,81 %	80,19 %	89,25 %	86,72 %
<b>Experimento 4</b>	73,83 %	64,94 %	71,94 %	80,65 %	81,07 %

- **Experimento 1:** 2DCNN.
- **Experimento 2:** 2DCNN con máscara de amarillos.
- **Experimento 3:** 2DCNN con máscara de amarillos y Threshold tuning (mejor threshold 0.5353702).
- **Experimento 4:** 2DCNN con máscara de amarillos, Grid search (  $LR = 1 \times 10^{-4}$  , Dropout = 0.4 , Número de filtros = [32,64,128]) y Threshold tuning (mejor threshold 0.260).

Los resultados observados en la Tabla 7.2 muestran que en general, los cuatro experimentos realizados muestran un desempeño alto a comparación de los resultados obtenidos para las arquitecturas de transfer learning, y además, las estrategias de mejora en los primeros tres experimentos muestran un aumento progresivo de las métricas de desempeño evaluadas, por lo cual las modificaciones empleadas fueron efectivas para un modelo con mejor rendimiento. Al comparar los resultados entre el **Experimento 1** (2DCNN base) y el **Experimento 2** (2DCNN con máscara de amarillos incorporada) se puede observar un claro aumento en todas las métricas de desempeño, lo cual confirma que la inclusión de un cuarto canal con la información cromática con la máscara de amarillos relacionada con las características de la ictericia favorece una discriminación de clases, lo cual refuerza a su vez su obtención en la etapa de preprocesamiento. En el **Experimento 3** (Adición de threshold tuning) se alcanzaron los mejores resultados entre las 4 arquitecturas, con un *Accuracy*

de 81,66 %, un *F1-Score* de 80,19 % y un *Recall* de 89,25 %, métricas que muestran que este modelo tuvo una mejor capacidad de distinción correcta entre neonatos con y sin ictericia. El **Experimento 4** (Adición de Grid Search) no mostró un incremento en las métricas de desempeño, en cambio mostró una disminución en el rendimiento del modelo de clasificación. Los resultados de este último experimento podrían deberse a que los hiperparámetros seleccionados para el modelo podrían no ser los óptimos y estarían afectando la capacidad de generalización del modelo de clasificación. Estos resultados permiten concluir que la arquitectura de clasificación 2DCNN con máscara de amarillos y Threshold tuning (Experimento 3) es la configuración que presenta un mejor desempeño en la clasificación y una mayor robustez en la discriminación de clases, razón por la cual se selecciona como la arquitectura a emplear en el proyecto para esta tarea. En la Figura 7.3 se puede observar la composición del modelo de clasificación seleccionado.

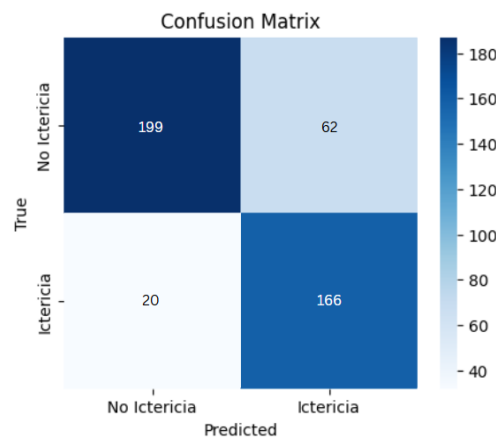


Figura 7.1: Matriz de Confusión del modelo de clasificación con threshold tuning (Experimento 3)

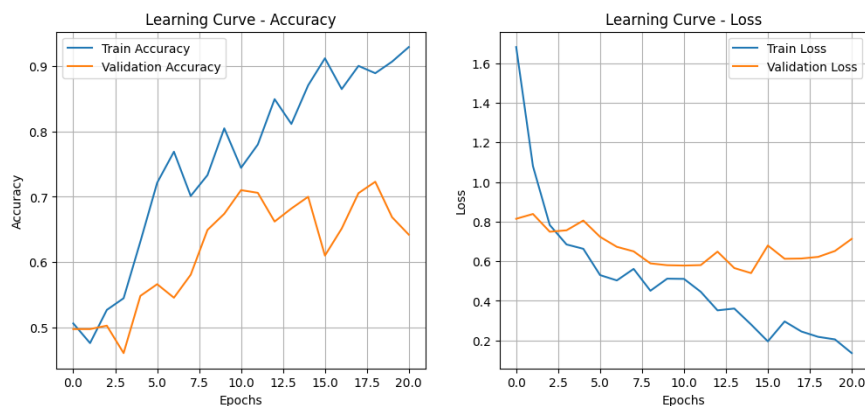


Figura 7.2: Curvas de aprendizaje del modelo de clasificación con threshold tuning (Experimento 3)

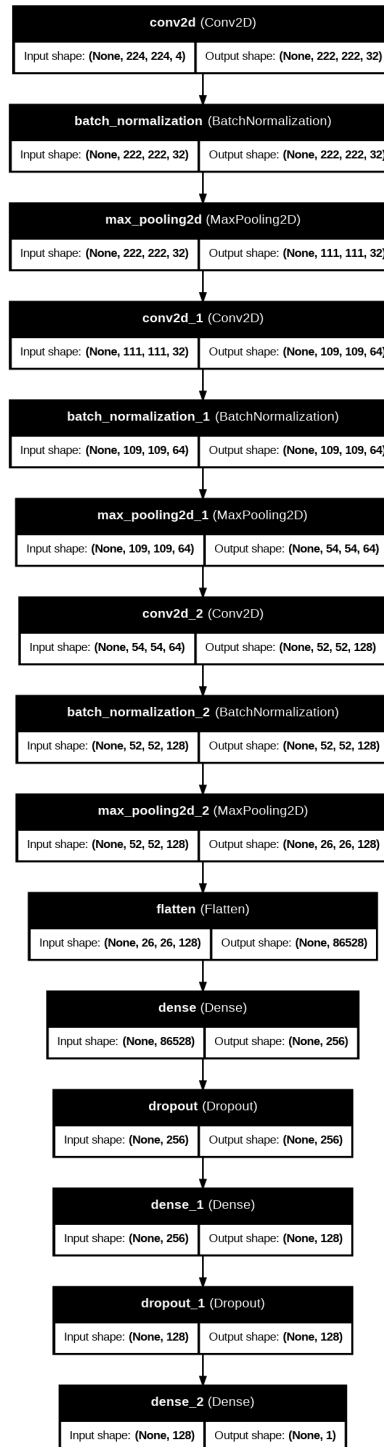


Figura 7.3: Composición de la arquitectura 2DCNN seleccionada para la clasificación

Al analizar la matriz de confusión de la arquitectura seleccionada, se puede evidenciar que el modelo tiene una capacidad de discriminación adecuada entre las clases con y sin ictericia, lo que se evidencia en la alta cantidad de verdaderos positivos (neonatos con ictericia) y verdaderos negativos (neonatos sin ictericia), sin embargo, aún se pueden evidenciar casos de clasificación errónea, principalmente de falsos positivos (neonatos sanos como ictericos), lo cual indica que deben ajustarse las estrategias de decisión del modelo. Las curvas de aprendizaje muestran a su vez que el modelo tiene potencial de mejora en su desempeño con los datos de validación, puesto que en la curva de aprendizaje (lado izquierdo) se aprecia como la exactitud tiene una curva con menor aumento y con mayores oscilaciones respecto a los datos de entrenamiento, y en la curva de Loss (lado derecho) se muestra que el descenso de la pérdida en la validación es mucho más lenta y menor que como ocurre con el entrenamiento, por lo cual se podría inferir un leve sobreajuste del modelo para los datos de validación.

### 7.1.3. Resultados de Cross-validation

En la Tabla 7.3 se muestran los resultados obtenidos del procesos de validación cruzada estratificada con 10 K-folds realizado al modelo de clasificación seleccionado. Estos resultados muestran que el modelo si tiene una capacidad alta de generalización independiente del azar, ya que los resultados promedio de las métricas obtenidas en las 10 iteraciones (K=10) se mantienen en valores altos equiparables a algunos de los resultados obtenidos en la Tabla 7.2, lo cual demuestra la robustez del modelo para la clasificación de ictericia neonatal.

Tabla 7.3: Resultados de la validación cruzada para el modelo de clasificación 2DCNN seleccionado

Modelo	Accuracy	Precision	F1 Score	Recall	ROC AUC
<b>Experimento 3</b>	73,60 ± 9,53 %	65,96 ± 9,88 %	73,36 ± 5,84 %	84,65 ± 6,22 %	80,20 ± 8,24 %

## 7.2. Modelo de Regresión

En la Tabla 7.4 se evidencian los resultados de desempeño obtenidos de los dos modelos de regresión probados experimentalmente con las métricas MAE, RMSE y  $R^2$ . En la Figura 7.5 se muestra el diagrama de correlación obtenido para la arquitectura seleccionada.

Tabla 7.4: Resultados arquitecturas de estimación de bilirrubina sérica

Modelo	MAE	RMSE	$R^2$
Regresión con validación cruzada	4,0525 mg/dL	4,9787 mg/dL	---
Regresión con <i>GlobalAveragePooling</i>	1,877 mg/dL	2,418 mg/dL	0,801

Según los resultados obtenidos para los modelos de regresión probados experimentalmente se evidencia que hay una diferencia considerable entre los desempeños de ambos modelos. Inicialmen-

te, se evidencia que el modelo de regresión con *GlobalAveragePooling* tuvo un desempeño global considerablemente mejor que el modelo de regresión con validación cruzada, lo cual se evidencia en valores de MAE y RMSE de casi la mitad de los valores obtenidos para la regresión con validación cruzada. Estos resultados muestran que el modelo de regresión con *GlobalAveragePooling* tiene una capacidad de estimación de bilirrubina sérica mucho más precisa y consistente. Esto se ve apoyado en el valor obtenido para la métrica  $R^2$  y la curva de correlación del modelo en la Figura 7.5 del modelo con *GlobalAveragePooling*, la cual muestra que las predicciones del modelo explican al 80 % el comportamiento de los datos de bilirrubina sérica de la base de datos, por lo cual las estimaciones de bilirrubina sérica tienen una variabilidad reducida al comparar con los datos reales, razón por la cual se seleccionó como el modelo adecuado para al estimación de bilirrubina en el presente proyecto. En la Figura 7.4 se puede observar la composición del modelo de regresión seleccionado.

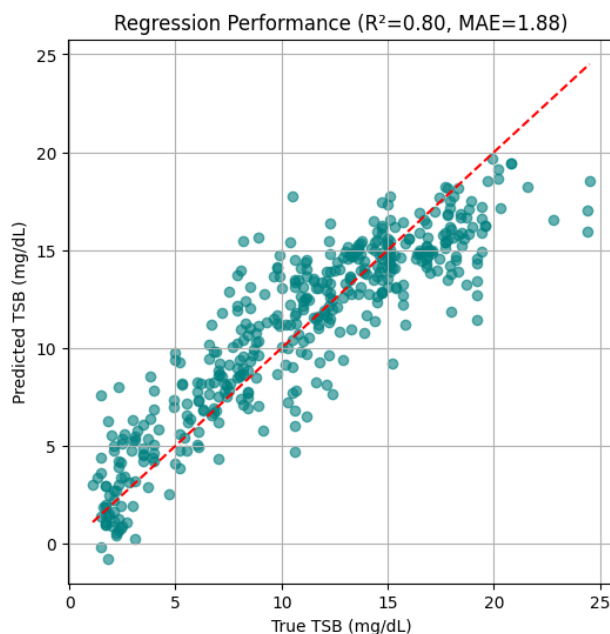


Figura 7.5: Diagrama de correlación del modelo de estimación de bilirrubina sérica

### 7.2.1. Resultados con base de datos NJN

En la Tabla 7.5 se muestran los resultados obtenidos en las métricas MAE, RMSE y  $R^2$  para la estimación de bilirrubina sérica en las imágenes de la base de datos NJN con el modelo de regresión seleccionado (con *GlobalAveragePooling*).

Tabla 7.5: Resultados arquitectura de regresión seleccionada con imágenes de la base de datos NJN

Modelo	MAE	RMSE	$R^2$
Regresión con <i>GlobalAveragePooling</i>	3,259 mg/dL	4,005 mg/dL	0,475

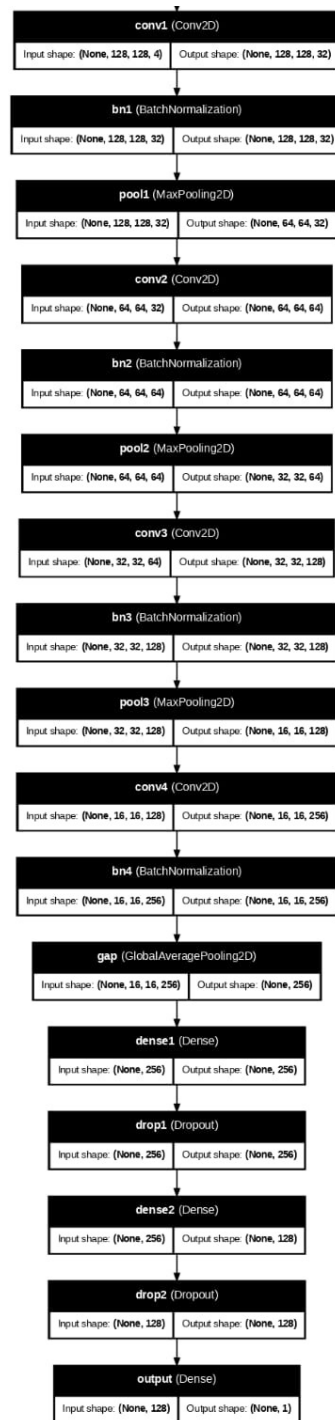


Figura 7.4: Composición de la arquitectura final de estimación de bilirrubina sérica

Los resultados obtenidos para las métricas propuestas evidencian que el modelo de regresión con GlobalAveragePooling, a pesar de sus buenos resultados en la validación con las imágenes de la base de datos NeoJaundice, con imágenes externas de la base de datos NJN obtuvo resultados poco satisfactorios, como se evidencia en los altos valores de MAE y RMSE, junto con el valor menor a 0.5 de la métrica  $R^2$ , dando a entender un bajo rendimiento y precisión del modelo con este grupo de imágenes.

Estos resultados sugieren que el modelo de regresión seleccionado aún tiene aspectos por mejorar en cuanto a la variabilidad de imágenes con las que puede trabajar para la estimación de bilirrubina, puesto que la base de datos NJN presenta altas diferencias en cuanto a las características de imagen como la falta de tarjeta de calibración, fotos de cuerpo completo del neonato y la presencia de objetos como sábanas, incubadoras, pañales, entre otros; los cuales podrían generar dificultades para el modelo desde su etapa de preprocesamiento en la identificación de la ROI de piel y posteriormente, en la estimación de bilirrubina sérica. Esto demuestra que a pesar de los buenos resultados presentados por el modelo de regresión previamente, la arquitectura tiene una alta dependencia a las características de la base de datos de entrenamiento.

### 7.3. Interfaz en streamlit

#### 7.3.1. Descripción general

Como resultado del proceso de diseño e implementación descrito en la sección metodológica, se obtuvo una interfaz funcional desarrollada en Streamlit que integra los modelos de clasificación y regresión del sistema de detección de ictericia neonatal en un solo proceso que genera varios resultados.

La herramienta permite ejecutar el flujo completo de análisis —desde la carga de imágenes, el preprocesamiento automático, la inferencia mediante redes neuronales y la visualización de resultados clínicamente interpretables— en un entorno gráfico intuitivo y de fácil uso.

El sistema final se ejecuta localmente mediante el archivo principal `app.py`, enlazando las distintas vistas modulares ubicadas en el directorio `pages/`. La interfaz conserva una estética minimalista y profesional, con predominio de tonos azul y blanco, en concordancia con entornos clínicos digitales, y se estructura mediante una barra lateral fija para la navegación entre los distintos módulos (Inicio, Diagnóstico, Análisis por Lote y Tablero de Resultados).

#### 7.3.2. Módulo de diagnóstico individual

El módulo de diagnóstico individual constituye el núcleo de la aplicación y permite el análisis de una imagen neonatal a la vez.

El usuario puede cargar un archivo con formato `.jpg`, `.jpeg` o `.png`, el cual es sometido internamente al proceso de preprocesamiento descrito en los objetivos anteriores: 1) normalización de

color e iluminación, 2) segmentación de la región de piel, 3) extracción de la máscara adaptativa de amarillos, y 4) generación del tensor RGB + canal de máscara para la inferencia.

El sistema emplea de manera subsecuente los dos modelos previamente entrenados: **(i)** el modelo de clasificación, que determina si el neonato presenta o no signos de ictericia; y **(ii)** el modelo de regresión, que estima la concentración aproximada de bilirrubina sérica en mg/dL.



Figura 7.6: Vista del módulo de diagnóstico individual en la interfaz desarrollada. Se observa la predicción del modelo, la concentración estimada de bilirrubina y la interpretación clínica correspondiente.

En la Figura 7.6 se observa la vista correspondiente al diagnóstico individual. El resultado incluye tres elementos principales:

- **Predicción del modelo:** etiqueta binaria (*ICTERICIA* / *NO ICTERICIA*).
- **Nivel estimado de bilirrubina:** valor continuo expresado en mg/dL.
- **Interpretación clínica:** clasificación cualitativa del rango (*leve*, *moderado* o *severo*), codificada por colores (verde, naranja y rojo respectivamente).

Además, la interfaz muestra una barra de confianza del modelo y un mensaje interpretativo que sugiere la acción recomendada según el rango obtenido (monitoreo, seguimiento o atención clínica inmediata). El tiempo promedio de procesamiento por imagen es de aproximadamente 2.3 segundos en entorno local (CPU Intel i7, 16 GB RAM), lo que demuestra una ejecución fluida y adecuada para su uso experimental.

### 7.3.3. Módulo de análisis por lotes

El módulo de análisis por lotes (*Batch Analysis*) permite procesar múltiples imágenes de manera automatizada, lo cual resulta de especial utilidad para la validación de desempeño de los modelos

sobre bases de datos completas o subconjuntos experimentales.

El usuario selecciona un archivo CSV con la estructura establecida (`chd_jaundice_published_2.csv`), el cual contiene la ruta de cada imagen y la etiqueta de referencia (*ictericia / no ictericia*).

Durante la ejecución, el sistema itera sobre las imágenes realizando la inferencia individual y almacenando los resultados —etiqueta predicha, probabilidad asociada, bilirrubina estimada y coincidencia con la etiqueta real— en un archivo CSV dentro del directorio `resultados/`, identificado por la fecha y hora de ejecución.

La interfaz presenta una barra de progreso y un resumen de la corrida, como se ilustra en la Figura 7.7.



Figura 7.7: Vista del módulo de análisis por lotes. Se muestra la barra de progreso, los resultados parciales y el resumen global del desempeño del modelo.

Al finalizar, se muestran las métricas agregadas del lote, incluyendo:

- Número total de imágenes procesadas,
- Porcentaje de aciertos (exactitud global),
- Distribución de clases predichas y
- Conteo de coincidencias y discrepancias entre las predicciones y las etiquetas reales.

#### 7.3.4. Visualización y trazabilidad de resultados

Todos los resultados generados por el sistema son exportados automáticamente en formato CSV y se integran posteriormente en el módulo de tablero de control (*Dashboard*). Esto permite realizar

un seguimiento histórico de las ejecuciones y visualizar el comportamiento general del modelo en distintos escenarios de validación.

El tablero consolida la información de todos los lotes procesados en el archivo `todos_lotes.csv`, mostrando la cantidad total de imágenes analizadas, el número de casos positivos y negativos de ictericia y la exactitud global alcanzada.

### 7.3.5. Análisis e interpretación final

La implementación de la interfaz permitió integrar de manera práctica los modelos desarrollados, evidenciando la posibilidad de trasladar los resultados experimentales del procesamiento digital de imágenes hacia una herramienta interactiva con aplicación potencial en entornos clínicos y de investigación.

El diseño modular y la organización del código facilitaron la depuración, el mantenimiento y la trazabilidad de los resultados, cumpliendo con los criterios de escalabilidad definidos en la metodología.

Durante las pruebas experimentales, se observó que la interfaz responde adecuadamente a distintos tamaños de imagen, mientras tengan la tarjeta de color como elemento de control cromático y calibración visual, y mantiene la estabilidad en el flujo de inferencia con tiempos de respuesta compatibles con el uso operativo.

Estos resultados confirman la viabilidad técnica de la herramienta como demostrador funcional del sistema propuesto, preservando al mismo tiempo la fidelidad visual y la reproducibilidad del análisis de color característico de la ictericia neonatal.

### 7.3.6. Acceso a la interfaz de Streamlit

La interfaz resultante del presente proyecto, la cual integra la carga de imágenes, el preprocesamiento de las mismas y el procesamiento en los modelos de clasificación y regresión se encuentra disponible en el siguiente enlace: <https://ictericiaapp-j9fappudvmc6trmrtef9s2w.streamlit.app/>. Esta versión permite mostrar el funcionamiento integrado del sistema y ejemplificar su desempeño en forma de prototipo de acuerdo a los alcances y objetivos propuestos por el proyecto.

El manual de usuario correspondiente a la interfaz gráfica se encuentra disponible para su visualización y descarga en la sección de Anexos del presente documento.

**Nota aclaratoria:** La interfaz entra en modo hibernación después de no ser utilizada por un periodo prolongado de tiempo, por lo cual para reactivarla se debe presionar el botón que aparece en pantalla y esperar de 3 a 5 minutos a que la aplicación web inicie su operación nuevamente. Este proceso se muestra detalladamente en el manual de usuario de la interfaz.



# Conclusiones

---

## Conclusiones

El desarrollo de este proyecto permitió integrar un conjunto de estrategias de procesamiento digital de imágenes y aprendizaje profundo orientadas al soporte al diagnóstico no invasivo de la ictericia neonatal, consolidando un flujo metodológico capaz de operar de manera estable con las imágenes del conjunto de datos utilizado.

La construcción del pipeline de preprocesamiento —incluyendo la normalización de color, la segmentación de la región de piel y la generación de una máscara adaptativa de amarillos— demostró ser un componente fundamental para preservar la información cromática relevante y reducir la variabilidad asociada a las condiciones de captura de las imágenes de la base de datos. No obstante, la alta dependencia del pipeline a las características de imagen de la base de datos NeoJaundice lo hizo rígido para trabajar con otro tipo de imágenes de ictericia neonatal, lo cual indica una necesidad de encontrar métodos de preprocesamiento que primen las características cromáticas y a su vez tengan mayor adaptabilidad a imágenes variables para así obtener resultados de mayor aceptabilidad en entornos menos controlados como una institución clínica.

El modelo de clasificación con máscara de amarillos y threshold tuning logró una exactitud de 81.66 %, un F1-Score de 80.19 % y un Recall de 89.25 %, lo cual evidencia que el modelo entrenado y creado desde cero tiene la capacidad de desempeñar adecuadamente la tarea de clasificación entre neonatos con y sin ictericia, superando incluso a metodologías de transfer learning adaptadas para realizar la misma clasificación. A pesar de estos resultados satisfactorios, se identificó que el modelo aún tiene dificultades en la clasificación, especialmente al generar falsos positivos, lo cual indica que existe la posibilidad y necesidad de mejorar el modelo si se busca que la herramienta sea implementada dentro de una institución clínica para reducir falsos diagnósticos o un diagnóstico no detectado, pero como primer acercamiento funcional en un ambiente controlado y con cantidad de datos limitada, los resultados fueron adecuados. Las causas de estos errores en la clasificación se pueden adjudicar al sobreajuste presente en el modelo, lo cual redujo su capacidad de generalización durante el proceso de validación, lo que hizo que el aprendizaje sesgado y limitado por el conjunto de entrenamiento generara clasificaciones erróneas en la validación, sin embargo, el modelo mantiene métricas altas en evaluaciones con validación cruzada, permitiendo reconocer así el potencial del modelo desarrollado como punto de partida para estudios más avanzados en el área de interés.

El modelo de regresión con Global Average Pooling alcanzó valores de MAE y RMS menores a 2

mg/dL y 2.5 mg/dL respectivamente, indicando que el modelo estima los valores de bilirrubina sérica con un error pequeño, haciendo que su desempeño sea aceptable para estimar este valor a manera de apoyo al diagnóstico, sin embargo, en un entorno clínico real, no contaría con la capacidad de reemplazar a métodos con una mayor exactitud como los métodos invasivos, pero como herramienta preliminar o de alerta temprana tendría validez. Adicionalmente, el coeficiente de determinación  $R^2$  indica que el modelo tiene la capacidad de explicar en un 80% la variación y comportamiento de los valores de ictericia neonatal en la base de datos, por lo cual se respalda la validez de este modelo en el contexto del proyecto.

La interfaz desarrollada permitió consolidar todos los elementos del sistema en una herramienta operativa capaz de ejecutar análisis individuales y por lotes, presentando resultados de manera visual, ordenada e interpretable. Esta implementación facilitó la evaluación práctica del funcionamiento del sistema, confirmando que los tiempos de ejecución, la estabilidad del flujo de inferencia y la gestión de imágenes son compatibles con usos experimentales y con una potencial adaptación futura a entornos clínicos.

En su conjunto, el proyecto demuestra la viabilidad de construir un sistema no invasivo basado en imágenes digitales para el apoyo en la detección temprana de ictericia neonatal. Asimismo, establece una base metodológica y técnica sólida sobre la cual pueden desarrollarse versiones posteriores con mayor rigor clínico, ampliación del conjunto de datos, optimización del desempeño de los modelos e integración con plataformas móviles o sistemas hospitalarios. Estas proyecciones abren la puerta al fortalecimiento de alternativas diagnósticas accesibles para el contexto colombiano, especialmente en escenarios donde las herramientas convencionales son limitadas o de difícil acceso.

# Trabajos futuros

---

A partir de los resultados obtenidos durante las diferentes etapas del proyecto y considerando las limitaciones observadas del sistema de procesamiento de imágenes elaborado, se identificaron varias líneas de trabajo para mejorar la robustez de la propuesta y ampliar los alcances del proyecto hacia una aplicabilidad mucho más adaptable a contextos reales en el entorno clínico.

Una línea de trabajo para el mejoramiento de los resultados del presente proyecto consiste en fortalecer la adaptabilidad del sistema ante imágenes de neonatos con características variables, con el propósito de aumentar la capacidad de aprendizaje del modelo a imágenes en condiciones distintas y robustecer su toma de decisiones. Actualmente las etapas de preprocesamiento iniciales fueron elaboradas adaptándose firmemente a la base de datos empleada en el proyecto y sus características, como, por ejemplo, la presencia de la tarjeta de color. A pesar de que estos elementos brindan resultados satisfactorios para los modelos validados, los resultados obtenidos con imágenes que presentan características y composiciones distintas no logran ser analizadas, segmentadas y procesadas adecuadamente para el fin del proyecto, lo cual limita el desempeño del sistema con datos menos controlados. Debido a esto, para favorecer un acercamiento mucho más generalizable y adaptativo a múltiples datos fotográficos relacionados con la ictericia neonatal, se pueden proponer estrategias de segmentación más robustas en las etapas de preprocesamiento de imagen, donde la extracción de características cromáticas y la detección de ROI de piel tengan mayor flexibilidad y adaptabilidad a distintos tipos de imágenes. Esto permitirá que el sistema sea utilizable con imágenes mucho más heterogéneas y, de esta manera, favorezca un poder de decisión en los modelos más fuerte por el aumento en la variabilidad de datos analizados.

Por otro lado, a pesar del acercamiento a la clasificación de ictericia neonatal y la estimación de bilirrubina sérica a partir de imágenes, en el contexto de la neonatología y la ictericia neonatal hay factores adicionales que pueden ser cruciales para el apoyo al diagnóstico y el tratamiento oportuno, como lo son los factores de riesgo individuales del neonato, los cuales podrían incorporarse como mejora al presente proyecto o a proyectos con fines similares. Los factores de riesgo, como el sexo del neonato, el color de piel y la edad gestacional, entre otros, tienen una alta relevancia en el entorno hospitalario al realizar el diagnóstico de ictericia y definir un tratamiento, por lo tanto, un acercamiento a un modelo que permita la integración de esta información clínica con las imágenes llevaría a sistemas de soporte al diagnóstico más eficientes y fomentaría un tratamiento más acorde a cada paciente. Herramientas como Bilitool [65] son ejemplo claro de la importancia del enfoque en factores de riesgo para la interpretación de niveles de bilirrubina sérica según las características del paciente y se ha validado su uso por profesionales en neonatología, por lo cual integrar estos

elementos podría significar el desarrollo de sistemas con una mayor precisión y un soporte al diagnóstico más completo y ajustado a la práctica clínica actual.

Otra línea de trabajo prometedora para proyectos de ictericia neonatal, específicamente en la región, es la construcción de un protocolo estandarizado de captura de imágenes y la recolección de fotografías de neonatos en instituciones de salud dentro del territorio nacional. Esto es importante puesto que las pocas bases de datos de acceso abierto que se encuentran en artículos en la actualidad preceden principalmente de instituciones hospitalarias en países de Asia, Europa o Norteamérica; que a pesar de su aporte hasta el momento en proyectos como el presentado, las características demográficas, fenotípicas y clínicas de sus poblaciones difieren significativamente de las autóctonas, lo cual podría afectar el desempeño de los modelos y su validez en el entorno de salud colombiano. Debido a lo anterior, la elaboración de un protocolo de captura de imágenes de neonatos en las instituciones de salud colombianas que estandarice las condiciones de iluminación, posición y uso de elementos de calibración de color permitiría el comienzo de un proceso de generación de una base de datos local con imágenes actuales de entornos clínicos del país, lo cual permitiría a mediano plazo el desarrollo de estudios nacionales de ictericia neonatal y procesamiento de imágenes, los cuales son limitados y con poca representatividad de la población local al basarse en bases de datos internacionales. A largo plazo, la continuación de crecimiento de estas bases de datos locales permitirían estudios robustos y estandarizados que podrían ser comparados y así, abrir puertas a nuevas conclusiones y avances clínicos sólidos sobre la patología.

Por último, un paso importante para el funcionamiento real del sistema es trabajar en pruebas piloto en instituciones clínicas, donde los modelos y la interfaz sean puestos a prueba por profesionales en neonatología en casos reales y se pueda así hacer una puesta en marcha del sistema en el ambiente real para el cual fue concebido. Estas pruebas permitirían medir la fiabilidad del sistema con datos tomados en tiempo real tanto de imágenes como de bilirrubina sérica y generar retroalimentación profesional sobre la utilidad del sistema, las limitaciones que presenta dentro del ambiente clínico y las posibles mejoras para la experiencia del usuario. De esta forma, con la entrada inicial al entorno clínico se podrían ampliar los alcances del proyecto y generar una adopción progresiva de la herramienta por los profesionales de interés, lo cual eventualmente podría llevar a que la herramienta alcance su propósito final para el apoyo al diagnóstico de ictericia neonatal al ser incorporado dentro de alguna institución de salud.

CAPÍTULO 10

# Anexos

---

## Anexo 1 – Rangos de umbralización de píxeles de piel utilizados en la literatura para los espacios de color YCrCb y HSV

Tabla 10.1: Valores umbral en la literatura para segmentación de piel en HSV

Ref	Valor umbral
[66][67]	$0,23 \leq S \leq 0,68$ $0^\circ \leq H \leq 50^\circ$
[68]	$0,35 < V \leq 1,0$ $0,20 \leq S \leq 0,68$ $0^\circ \leq H \leq 50^\circ$
[69]	$V \geq 40$ $0,2 \leq S \leq 0,6$ $0^\circ \leq H \leq 25^\circ$ o $335^\circ < H < 360^\circ$
[70]	Cuando $H \in (-180^\circ, 180^\circ)$ : $V \geq 40$ $H \leq (-0,4V + 75)$ $10 \leq S \leq (-H - 0,1V + 110)$ Si $H \geq 0$ , $S \leq (0,08(100 - V)H + 0,5V)$ Si $H < 0$ , $S \leq (0,5H + 35)$
[71]	Valores mínimos = [0, 58, 30] Valores máximos = [33, 255, 255]

Tabla 10.2: Valores umbral en la literatura para segmentación de piel en YCrCb

Ref	Valor umbral
[66]	$85 \leq Cb \leq 135$ $135 \leq Cr \leq 180$
[68]	$97,5 \leq Cb \leq 142,5$ $134 \leq Cr \leq 176$
[70]	$77 \leq Cb \leq 127$ $133 \leq Cr \leq 173$
[72]	$76 \leq Cb \leq 127$ $132 \leq Cr \leq 173$
[67]	$Y > 80$ $Cb > 85$ $Cr > 135,$ $Cr \leq 1,5862 Cb + 20,$ $Cr \geq 0,3448 Cb + 76,2069,$ $Cr \geq -4,5652 Cb + 234,5652,$ $Cr \leq -1,15 Cb + 301,75,$ $Cr \leq -2,2857 Cb + 432,85.$

## Anexo 2 – Estudios de clasificación con Transfer Learning

Tabla 10.3: Estudios de clasificación de ictericia neonatal con transfer learning

Ref	Objetivo del estudio	Base de datos empleada	Arquitectura de Transfer learning	Resultados (Exactitud)
[47]	Investigar la efectividad del transfer learning en el diagnóstico de ictericia neonatal utilizando distintos tipos de características de la piel, ojos y una combinación de las características de ambos.	68 imágenes de neonatos con ictericia (24 imágenes) y sin ictericia (44 imágenes) tomadas con un teléfono celular Samsung Galaxy S7 sin control de condiciones de iluminación en el Hospital Universitario King Khalid.	VGG16	Piel: 86.83 % Ojo: 79.03 % Fusion de ambos: 79.95 %
[73]	Estimar el nivel de bilirrubina en los bebés, clasificándolo en tres o cuatro categorías según el nomograma de Bhutani.	344 imágenes de neonatos con ictericia (84 imágenes) y sin ictericia (260 imágenes) tomadas con un iPhone 12 Pro Max y un anillo de luz complementario en el Hospital de Enseñanza Ibn Al-Atheer.	VGG16 ResNet50	<b>Tres Clases:</b> VGG16= 91.71 % ResNet50=95.98 % <b>Cuatro clases:</b> VGG16= 94.92 % ResNet50=94.66 %
[74]	El artículo presenta un enfoque no invasivo para la detección preventiva de la ictericia neonatal grave mediante visión artificial y aprendizaje profundo.	Dataset titulado NJN (Normal and Jaundiced Newborns) con 760 imágenes de neonatos, 560 sin ictericia y 200 con ictericia; capturadas con un iPhone 11 Pro Max en el Hospital Universitario Materno-Infantil Al-Elwiya.	MobilNet V3 EfficientNet V2 Vision Transformer	MobilNet= 64 % EfficientNet= 82 % Vision Transformer= 83 %
[61]	Diseñar e implementar un sistema inteligente de monitorización de ictericia en recién nacidos mediante el aprendizaje por transferencia de una red neuronal convolucional y Raspberry Pi.	Dataset de 145 imágenes tomadas de los hospitales de Mosul divididas en tres clases: casos normales (50 imágenes), neonatos con TSB de (3-9) mg/dL y neonatos con TSB de (10-16) mg/dL .	VGG16 VGG19 ResNet50 EfficientNet B0 EfficientNet B7	VGG16= 81.81 % VGG19= 79.95 % ResNet50= 84.09 % EfficientNet B0= 81.81 % EfficientNet B7= 81.81 %
[52]	Proponer adaptaciones de modelos básicos para la clasificación de imágenes médica en diversas tareas clínicas y examinar su rendimiento general.	2235 imágenes de neonatos del dataset NeoJaundice del hospital central de Xuzhou. (El mismo dataset utilizado en el presente proyecto)	DenseNet121 EfficientNet B4 Swin Transformer	DenseNet121= 74.2 % EfficientNet B4= 75.2 % Swin Transformer= 71.6 %

## Anexo 3 – Hiperparámetros para los modelos de clasificación

Tabla 10.4: Hiperparámetros utilizados en los modelos evaluados

Hiperparámetro	CNN Simple	CNN + Máscara	CNN + Tuning	CNN + Grid Search
Input shape	224×224×3	224×224×4	224×224×4	224×224×4
Filtros por bloque convolucional	[32, 64]	[32, 64, 128]	[32, 64, 128]	[32,64,128] / [64,128,256]
Kernel size	3×3	3×3	3×3	3×3
N. bloques Conv+Pool	2	3	3	3
Batch Normalization	No	Sí	Sí	Sí
MaxPooling	2×2	2×2	2×2	2×2
Capas densas	128 → 64	256 → 128	256 → 128	256 → 128
Dropout	0.4	0.5	0.5	[0.3, 0.4,0.5]
Activación (Conv)	ReLU	ReLU	ReLU	ReLU
Activación final	Sigmoid	Sigmoid	Sigmoid	Sigmoid
Optimizador	Adam	Adam	Adam	Adam
Learning rate (LR)	1e-4	1e-4	1e-4	[1e-3, 1e-4, 1e-5]
Función de pérdida	BCE	BCE	BCE	BCE
Épocas (entrenamiento final)*	25	25	25	25
Batch size	32	32	32	32
Balanceo de clases	Sí	Sí	Sí	Sí
Umbral de decisión	0.5	0.5	F1 óptimo	F1 óptimo

\*Nota: Las 25 épocas para el entrenamiento fueron propuestas como límite máximo, puesto que con los Callbacks un modelo podría realizar menor cantidad de épocas si no hay mejora en la métrica evaluada

## Anexo 4 – Hiperparámetros para los modelos de regresión

Tabla 10.5: Hiperparámetros utilizados en los modelos de regresión evaluados para estimación de bilirrubina sérica

Hiperparámetro	Regresión con Validación Cruzada	Regresión con GlobalAveragePooling
Input shape	128×128×4	128×128×4
Filtros por bloque convolucional	[32, 64, 128]	[32, 64, 128, 256]
Kernel size	3×3	3×3
N. Bloques Convolucionales	3	4
MaxPooling tamaño de filtro	2×2	2×2
Padding	default = "valid"	"same"
Capa de reducción	Flatten	GlobalAveragePooling2D
Capas densas	256 → 128 → 1	256 → 128 → 1
Dropout	0.5	0.5-0.3
Activación (Conv)	ReLU	ReLU
Activación final	Lineal	Lineal
Learning rate (LR)	1e-4	5e-5
Optimizador	Adam	Adam
Función de pérdida	MAE	MAE
Épocas*	20**	35
Batch size	32	32
Validación cruzada	Si	No
No. Folds (K)	5	N/A

\*Nota: El número de épocas para el entrenamiento fueron propuestas como límite máximo, puesto que con los Callbacks un modelo podría realizar menor cantidad de épocas si no hay mejora en la métrica evaluada

\*\*Nota: Para el modelo con validación cruzada, se cumple el límite máximo de 20 épocas en cada una de las iteraciones que realiza el modelo de regresión (K=5).

## Anexo 5 – Estructura de la aplicación de streamlit

A continuación, se presenta la organización general de directorios y archivos del proyecto, mientras que la Tabla 10.6 describe la función de cada uno de ellos.

```
.
assets/
core/
  database.py
  processor.py
data/
models_ml/
  preprocessing.py
pages/
  1_auth.py
  2_diagnostico.py
  3_dashboard.py
  4_analisis_por_lote.py
  5_about.py
resultados/
test/
  dataset_prediccion_simulada.py
  dataset_viewer_test.py
  TEST.py
utils/
  auth_utils.py
  layout_utils.py
.gitattributes
app.py
```

**Nota:** `app.py` es el punto de entrada principal de la aplicación. Las páginas multipágina de Streamlit se encuentran dentro del directorio `pages/`, los módulos de apoyo en `core/` y `utils/`, y los resultados exportados se almacenan en `resultados/`.

Tabla 10.6: Descripción de los principales módulos de la interfaz desarrollada.

Ruta / Módulo	Función principal	Puntos clave / Recomendaciones
<code>app.py</code>	Punto de entrada del sistema. Crea la barra lateral, maneja el estado global y enruta las páginas.	Mantener el archivo ligero; delegar la lógica en los módulos <code>core/</code> y <code>models_ml/</code> .
<code>pages/1_auth.py</code>	Página de inicio de sesión.	Depende de las funciones de <code>utils/auth_utils.py</code> . Permite activar o desactivar autenticación en modo prueba.
<code>pages/2_diagnostico.py</code>	Módulo central: carga de imágenes, preprocesamiento, inferencia y visualización de resultados.	Utiliza funciones de <code>core/processor.py</code> y <code>models_ml/preprocessing.py</code> .
<code>pages/3_dashboard.py</code>	Tablero de métricas globales.	Lee archivos CSV desde <code>resultados/</code> . Usa <code>st.cache_data</code> para optimizar el rendimiento.
<code>pages/4_analisis_por_lote.py</code>	Procesamiento en lote de múltiples imágenes.	Genera un archivo CSV por corrida con resumen estadístico.
<code>pages/5_about.py</code>	Página informativa con datos del proyecto, autores y limitaciones.	Incluye apartados de ética y privacidad.
<code>core/processor.py</code>	Orquesta el pipeline de procesamiento: lectura, normalización, segmentación, y preparación del input.	Contiene funciones puras para facilitar pruebas unitarias.
<code>core/database.py</code>	Manejo de persistencia local de resultados.	Principalmente genera archivos CSV en <code>resultados/</code> .
<code>models_ml/preprocessing.py</code>	Funciones de preprocesamiento y wrappers de modelos.	Incluye funciones <code>load_model()</code> , <code>prepare_input()</code> y <code>predict_proba()</code> .
<code>utils/auth_utils.py</code>	Validación y manejo de usuarios.	Usa variables de entorno para credenciales, evitando contraseñas en texto plano.
<code>utils/layout_utils.py</code>	Componentes de UI reutilizables.	Mejora la coherencia visual entre páginas.
<code>assets/</code>	Archivos estáticos: íconos, logos y estilos.	Evitar incluir material protegido por derechos de autor.
<code>data/</code>	Conjunto de imágenes de prueba y datos procesados temporalmente.	Asegurar anonimización en caso de datos clínicos.
<code>resultados/</code>	Almacenamiento de salidas CSV y artefactos de ejecución.	Usar nombres con fecha y hora para trazabilidad.

## Anexo 6 – Requisitos para desarrollo de la aplicación y esquema de exportación de resultados

### Requisitos para desarrollo de la aplicación

El sistema fue implementado en Python 3.11 con las librerías listadas en la Tabla 10.7. Todas las dependencias se instalaron mediante el archivo `requirements.txt`.

Tabla 10.7: Dependencias principales del proyecto.

Librería	Función principal
Streamlit	Creación de la interfaz de usuario.
OpenCV / Pillow / scikit-image	Procesamiento y manipulación de imágenes.
NumPy / Pandas	Manejo y análisis de datos.
Matplotlib / Plotly	Generación de visualizaciones y gráficos.
scikit-learn / Tensor-Flow	Carga e inferencia de modelos de aprendizaje automático.
python-dotenv	Gestión de variables de entorno ( <code>.env</code> ).

El archivo `.env` contiene las rutas y configuraciones principales del proyecto, como el umbral de riesgo y las credenciales de autenticación. El sistema se ejecuta localmente mediante el comando:

```
streamlit run app.py
```

### Esquema de exportación de resultados

Cada análisis realizado genera un archivo CSV con la siguiente estructura:

```
image_path, predicted_class, probability, risk_level, preprocess_steps,
model_version, run_id, timestamp
data/samples/img_001.jpg, ictericia, 0.82, ALTO,
{'color':'HSV','norm':'shades'}, v1.0, 20241004-170000, 2024-10-04T17:00:00-05:00
```

Estos resultados son utilizados posteriormente por el módulo de tablero (`pages/3_dashboard.py`) para consolidar las estadísticas de desempeño general.

---

## Anexo 7 – Manual de Usuario de la interfaz

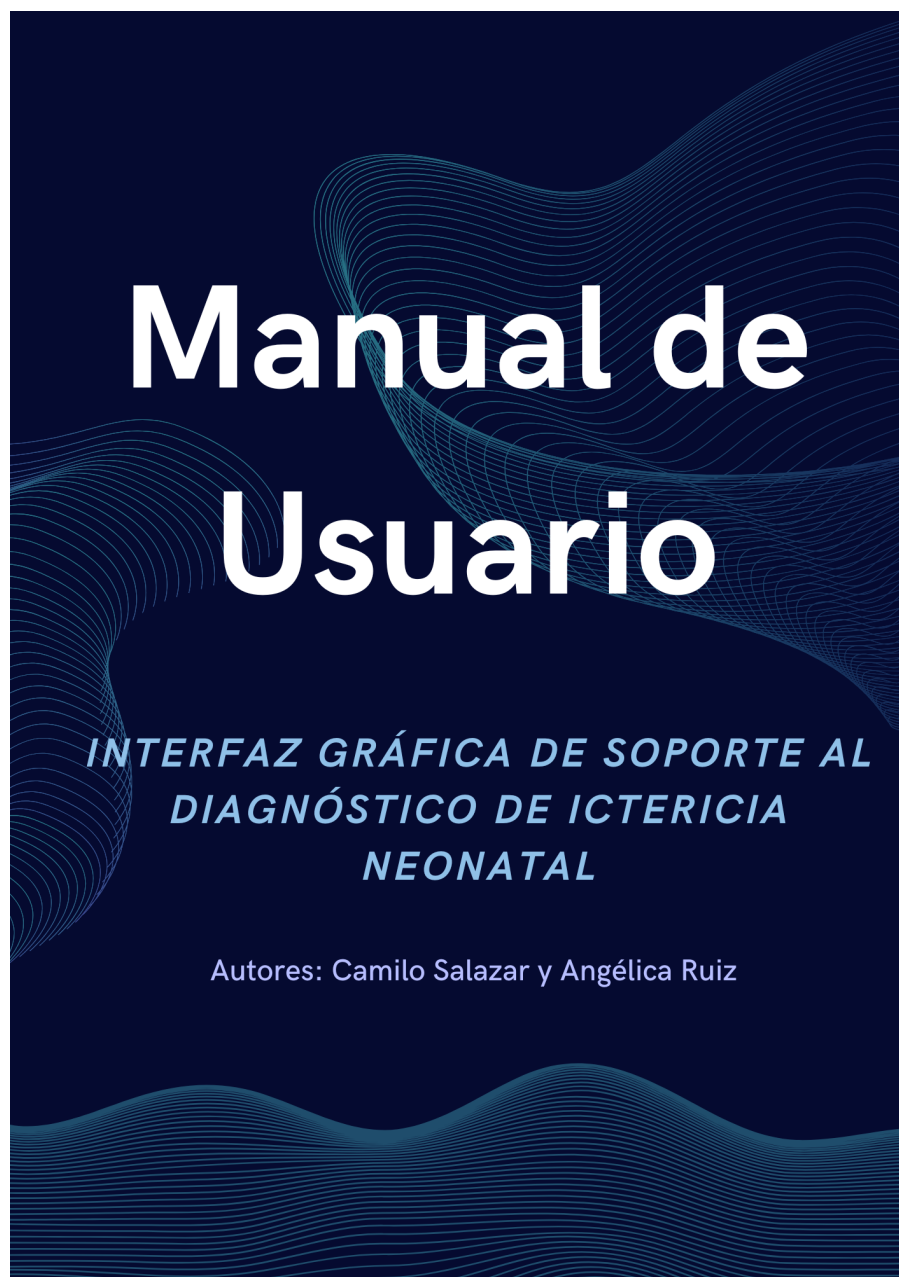


Figura 10.1: Portada del manual de usuario

[Visualizar manual de usuario aquí](#)

[Descarga aquí el manual de usuario](#)

### **Propósito de este manual:**

El propósito de este manual es ofrecer al usuario de la interfaz una guía durante el uso de la herramienta con instrucciones que le permitan comprender el funcionamiento del sistema, navegar entre sus módulos y utilizarla de manera adecuada.

### **Secciones a desarrollar en este manual:**

Este manual está estructurado en los siguientes apartados:

- 1. Contexto de desarrollo de la interfaz**
- 2. Vistas principales de la interfaz**
  - Página de inicio
  - Módulo de diagnóstico individual
  - Módulo de análisis por lotes
- 3. Flujo funcional de la interfaz**
  - Funcionamiento para diagnóstico individual
  - Funcionamiento para diagnóstico por lotes
- 4. Activación de la interfaz en modo hibernación**

Figura 10.2: Pagina 1 manual de usuario

## Contexto de desarrollo de la interfaz

Esta interfaz presentada es el resultado del diseño e implementación de un prototipo de sistema no invasivo de soporte al diagnóstico de ictericia neonatal por medio de procesamiento de imágenes. El sistema integra modelos computacionales entrenados para la identificación de la presencia de ictericia neonatal y la estimación de niveles aproximados de bilirrubina sérica, con el fin de ofrecer una visualización clara y comprensible para una valoración más fundamentada del estado de salud de un paciente.

Es necesario aclarar que este prototipo fue creado para complementar la evaluación clínica y NO debe ser tomada como única herramienta de diagnóstico de ictericia neonatal. A pesar de la capacidad del prototipo de generar estimaciones basadas en imágenes, su funcionamiento hasta el momento no cuenta con respaldo experimental ni validación en ambientes hospitalarios reales. Su propósito es de apoyo preliminar y puede ser útil para escenarios de tamizaje inicial o exploración académica, complementando el juicio profesional.

Figura 10.3: Pagina 2 manual de usuario

## Vistas principales de la interfaz

# 1

### Página de inicio

En esta sección se proporcionan descripciones generales del sistema, se menciona su propósito, las instrucciones de navegación de la interfaz y se realizan aclaraciones respecto a su alcance.

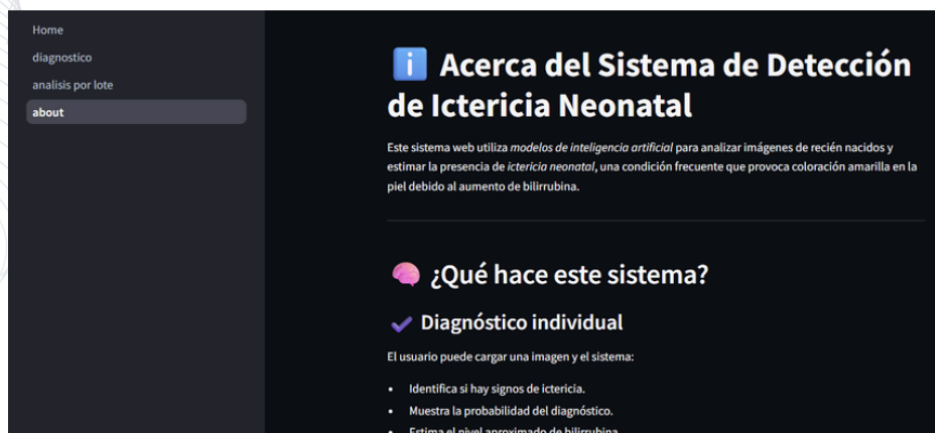


Figura 10.4: Pagina 3 manual de usuario

## Vistas principales de la interfaz

# 2

### Módulo de diagnóstico individual

En esta sección se hace un análisis de una única fotografía otorgada al sistema por parte del usuario, la cual pasará por etapas de preprocesamiento y análisis de los modelos de procesamiento para devolver al usuario un diagnóstico tentativo (ictericia/no ictericia) y el valor estimado de bilirrubina sérica en mg/dL , presentando estos resultados junto con su nivel de confianza y una interpretación clínica del rango leve, moderado o severo.

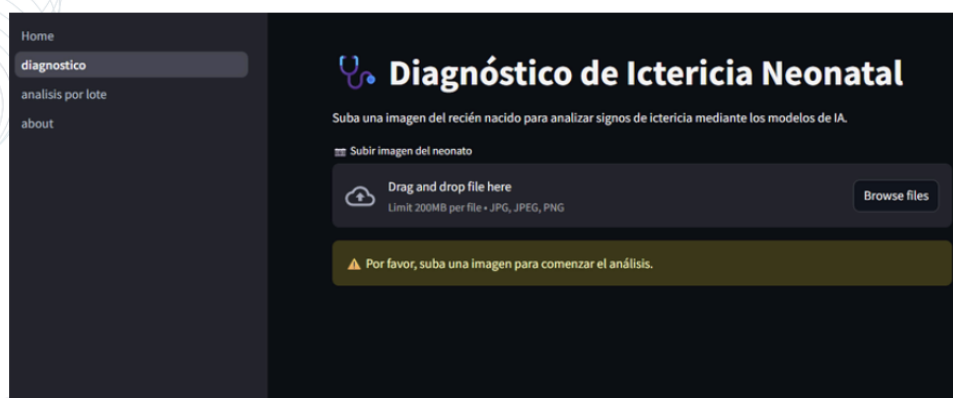


Figura 10.5: Pagina 4 manual de usuario

## Vistas principales de la interfaz

### 3 Módulo de análisis por lotes

En esta sección se hace un análisis masivo de imágenes con fines de investigación y validación de manera automatizada. El usuario carga un archivo .csv con las imágenes a analizar y para cada una se genera la estimación de bilirrubina sérica al igual que la clasificación (ictericia/no ictericia). Se presentan a su vez los niveles de confianza de los resultados y visualizaciones gráficas para una análisis global del conjunto de imágenes entregado al modelo. Este análisis tiene una índole mayormente académica y de validación del sistema, sin embargo, puede ser utilizada también para tamizaje clínico.



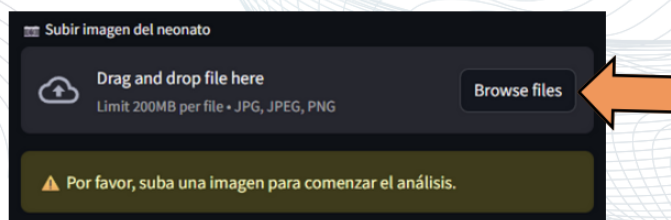
Figura 10.6: Pagina 5 manual de usuario

## Flujo funcional de la interfaz

# 1

### Funcionamiento para diagnóstico individual

**1. Cargar una imagen:** Para iniciar el procesamiento de la imagen de interés, el usuario debe cargar al sistema una imagen neonatal que desee analizar. Presionar en el botón "Browse file" para seleccionar la imagen de su dispositivo para el análisis.



**2. Resultados:** El sistema analiza la imagen y retorna la predicción del diagnóstico (ICTERICIA/NO ICTERICIA), el valor estimado de bilirrubina sérica (en mg/dL) y una interpretación clínica cualitativa con codificación de color según la gravedad de los resultados: verde para leve, naranja para moderado y rojo para severo. Finalmente, se muestran recomendaciones de acción según estos resultados.



Ejemplo para resultados severos.

Figura 10.7: Pagina 6 manual de usuario

## Flujo funcional de la interfaz

# 2

### Funcionamiento para diagnóstico por lotes

**1. Selección de archivo y de imágenes a analizar:** Por el momento para este prototipo se tiene incorporado en el sistema un archivo .csv para el análisis por lotes con las imágenes y sus datos organizados según los estándares del sistema. Para este archivo, el usuario puede utilizar los deslizadores para determinar la cantidad de imágenes a analizar y la cantidad para la cual desea visualizar su resultado.



**2. Resultados:** El sistema analiza las imágenes y retorna la predicción del diagnóstico (ICTERICIA/NO ICTERICIA), el valor estimado de bilirrubina sérica (en mg/dL) y hace un análisis de correspondencia entre la clasificación real de la imagen y la clasificación predicha, entregando el porcentaje de exactitud en la predicción para cada imagen.



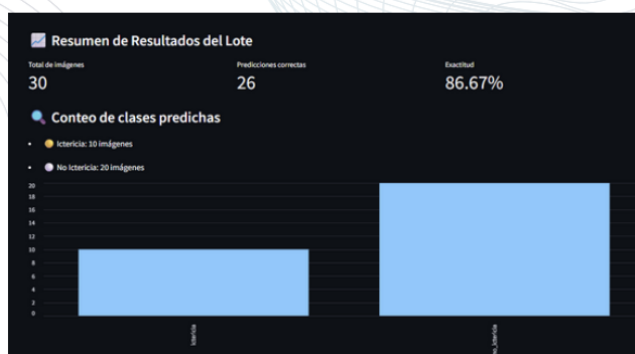
Ejemplos de visualización de resultados

## Flujo funcional de la interfaz

# 2

### Funcionamiento para diagnóstico por lotes

**3. Resumen de resultados:** el sistema además de los resultados por imágenes, entrega el porcentaje global de aciertos en la predicción que obtuvo el modelo, en conjunto con un conteo de las clases que se predijeron para las imágenes en un gráfico de barras.



Ejemplo para el análisis de 30 imágenes

**4. Guardar los resultados:** El sistema almacena los resultados obtenidos del lote en un archivo CSV. El sistema adicionalmente, tiene la opción de descargar este archivo y también, permite descargar el archivo histórico de imágenes analizadas en todos los lotes que se procesaron del archivo, con el propósito de mantener una trazabilidad de los resultados.

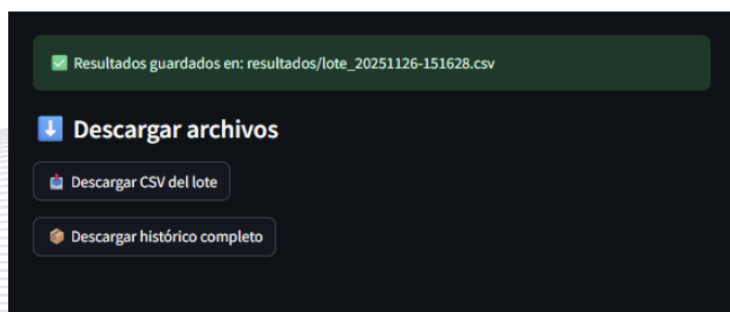


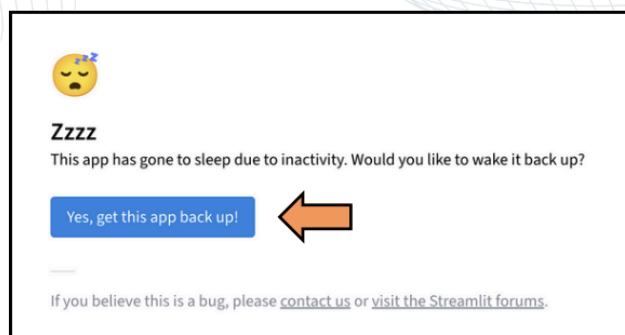
Figura 10.9: Pagina 8 manual de usuario

## Activación de la interfaz en modo hibernación

La interfaz gráfica después de no ser utilizada por un periodo prolongado de tiempo entra en un estado llamado "modo hibernación", el cual lleva a que la herramienta se desactive.

Para reactivar la interfaz se debe realizar el siguiente proceso:

- 1. Presionar el botón de activación:** al abrir la aplicación web e identificar que esta se encuentra inactiva, debe de presionarse el botón azul que contiene el texto "Yes, get this app back up!" para iniciar la reactivación de la interfaz.



- 2. Esperar la reactivación:** a la interfaz le toma entre 3 y 5 minutos reactivarse después de haberse presionado el botón, por lo que al terminar ese tiempo la herramienta iniciará su operación nuevamente.



Figura 10.10: Pagina 9 manual de usuario

## Anexo 8 – Tarjeta de Calibración

Como estrategia para una posible reproducibilidad de los resultados y pruebas futuras con nuevas imágenes, se generó una réplica de la tarjeta de calibración utilizada en las imágenes de la base de datos clave del presente proyecto. Se presenta tanto la imagen en la Figura 10.11 como el enlace para su descarga. De esta manera, en trabajos futuros, puede emplearse esta tarjeta como base para aumentar la cantidad de imágenes a procesar en el sistema desarrollado.



Figura 10.11: Prototipo de Tarjeta de calibración

[Descargar tarjeta de calibración aquí](#)



# Bibliografía

- [1] B. O. Olusanya, M. Kaplan, and T. W. R. Hansen, “Neonatal hyperbilirubinaemia: a global perspective,” *The Lancet Child Adolescent Health*, vol. 2, no. 8, p. 610–620, 2018.
- [2] A. Taipe-Paucar, A. Toaquiza-Alvarado, and G. Merchán-Coronel, “Ictericia neonatal a nivel de américa latina,” *FACSALUD-UNEMI*, vol. 6, no. 10, p. 76–84, 2022.
- [3] A. R. Guerra-Ruiz, J. Crespo, R. M. López Martínez, P. Iruzubieta, G. Casals Mercadal, M. Lalana Garcés, B. A. Lavin Gomez, and M. Morales Ruiz, “Bilirrubina: Medición y utilidad clínica en la enfermedad hepática,” *Advances in Laboratory Medicine*, vol. 2, no. 3, p. 362–372, 2021.
- [4] M. Manuals, “Hiperbilirrubinemia neonatal - pediatría - manual msd versión para profesionales,” n.d., retrieved February 17, 2024, from <https://www.msmanuals.com/es-co/professional/pediatr>
- [5] A. P. E. Vidal, M. Nannig, A. Vidal Esteban, I. Concepción, B. Fernández, M. José, and R. Martín, “Correlación entre bilirrubinometría transcutánea y bilirrubina sérica en recién nacidos 35 semanas,” *Andes Pediatrica*, vol. 95, no. 1, 2024.
- [6] Asocolderma, “Ictericia (coloración amarillenta de la piel),” retrieved February 17, 2024, from <https://asocolderma.org.co/enfermedades-de-la-piel/ictericia-coloracion-amarillenta-de-la-piel>.
- [7] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Prentice-Hall, Inc., 2006.
- [8] C. Hulzebos, L. Vitek, and C. C. Zabetta, “Diagnostic methods for neonatal hyperbilirubinemia: benefits, limitations, requirements, and novel developments,” *Pediatric Research*, vol. 90, pp. 277–283, 2021.
- [9] F. T. J. M. I. a. I. S. S. S. F. M. Miah, R. J. Tazim, “Non-invasive bilirubin level quantification and jaundice detection by sclera image processing,” in *2019 IEEE Global Humanitarian Technology Conference, GHTC 2019*, 2019.
- [10] V. Bhutani, R. Vilms, and L. Hamerman-Johnson, “Universal bilirubin screening for severe neonatal hyperbilirubinemia,” *Journal of Perinatology*, vol. 30, pp. S6–S15, 2010.
- [11] B. Olusanya, S. Teeple, and N. Kassebaum, “The contribution of neonatal jaundice to global child mortality: Findings from the gbd 2016 study,” *Pediatrics*, vol. 141, no. 2, 2018.
- [12] F. Dzulkifli, M. Mashor, and K. Khalid, “Methods for determining bilirubin level in neonatal jaundice screening and monitoring: A literature review,” in *Conference proceedings*, 2018, pp. 1–10.

- [13] C. Sierra and N. Hoyos, “Sistema para la detección de la ictericia neonatal,” escuela de Ingeniería de Antioquia. Universidad CES. Retrieved from <https://www.hgm.gov.co/loader.php?lServicio=Tools2lTipo=descargaslFuncion=descargaridFile=766>.
- [14] P. Karamzadeh, M. Fallahi, M. Kazemian, N. Taslimi, S. Nouripour, and M. Radfar, “Bilirubin induced encephalopathy,” *PubMed Central (PMC)*, retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6956966>.
- [15] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Prentice-Hall, Inc., 2006.
- [16] H. Vargas, “Introduction to digital image processing,” lección de clase, Pontificia Universidad Javeriana Cali, Abr. 8, 2024.
- [17] O. Foundation. (2025) Color spaces in opencv. Accedido: 21 de noviembre de 2025. [Online]. Available: <https://opencv.org/blog/color-spaces-in-opencv/>
- [18] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia, “Human skin detection using rgb, hsv and ycbcr color models,” *arXiv preprint arXiv:1708.02694*, 2017.
- [19] M. S. Irajy and A. Yavari, “Skin color segmentation in fuzzy ycbcr color space with the mamdani inference,” *American Journal of Scientific Research*, vol. Issue July, pp. 131–137, 2011, accedido: 22 de noviembre de 2025.
- [20] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. New York: Pearson, 2018.
- [21] GeeksforGeeks, “Types of machine learning,” <https://www.geeksforgeeks.org/machine-learning/types-of-machine-learning/>, 2025, accessed: 2025-09-01.
- [22] —, “Machine learning algorithms,” <https://www.geeksforgeeks.org/machine-learning/machine-learning-algorithms/>, 2025, accessed: 2025-09-01.
- [23] H. Vargas, “Introduction to machine learning (ml) and deep learning (ml),” lección de clase, Pontificia Universidad Javeriana Cali, Abr. 23, 2024.
- [24] K. P. Murphy, *Probabilistic Machine Learning: An introduction*. MIT Press, 2022. [Online]. Available: <http://probml.github.io/book1>
- [25] GeeksforGeeks. (2025) What is fully connected layer in deep learning? [Online]. Available: <https://www.geeksforgeeks.org/deep-learning/what-is-fully-connected-layer-in-deep-learning/>
- [26] S. Zawad and F. Yan, “Chapter 12 - hyperparameter tuning for federated learning – systems and practices,” in *Federated Learning*, L. M. Nguyen, T. N. Hoang, and P.-Y. Chen, Eds. Academic Press, 2024, pp. 219–235. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780443190377000211>

- [27] J. Terven, D. M. Córdova-Esparza, J. A. Romero-González, A. Ramírez-Pedraza, and E. A. Chávez-Urbiola, “A comprehensive survey of loss functions and metrics in deep learning,” *Artificial Intelligence Review*, vol. 58, no. 7, p. 195, 2025.
- [28] J. Pomerat, A. Segev, and R. Datta, “On neural network activation functions and optimizers in relation to polynomial regression,” in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 6183–6185.
- [29] GeeksforGeeks, “Batch size in neural network,” <https://www.geeksforgeeks.org/deep-learning/batch-size-in-neural-network/>, 2025, Última actualización: 23 Jul, 2025.
- [30] —, “Epoch in machine learning,” <https://www.geeksforgeeks.org/machine-learning/epoch-in-machine-learning/>, 2025, Última actualización: 17 Jul, 2025.
- [31] GeeksforGeeks, “Hyperparameter tuning,” <https://www.geeksforgeeks.org/machine-learning/hyperparameter-tuning/>, 2025, accessed: 2025-09-01.
- [32] —, “Underfitting and overfitting in machine learning,” <https://www.geeksforgeeks.org/machine-learning/underfitting-and-overfitting-in-machine-learning/>, 2025, accessed: 2025-09-01.
- [33] M. A. Morid, A. Borjali, and G. D. Fiol, “A scoping review of transfer learning research on medical image analysis using imagenet,” Santa Clara University, Tech. Rep., 2020.
- [34] I. Corporation, “What is a machine learning pipeline?” <https://www.ibm.com/think/topics/machine-learning-pipeline>, 2025, accessed: 2025-09-01.
- [35] Z. C. Lipton, C. Elkan, and B. Naryanaswamy, “Optimal thresholding of classifiers to maximize f1 measure,” University of California, Tech. Rep., 2014. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-662-44851-9\\_15#citeas](https://link.springer.com/chapter/10.1007/978-3-662-44851-9_15#citeas)
- [36] GeeksforGeeks. (2023) Cross validation in machine learning. Accessed: 2025-11-12. [Online]. Available: <https://www.geeksforgeeks.org/machine-learning/cross-validation-machine-learning/>
- [37] E. Juliastuti, V. Nadhira, Y. Satwika, N. Aziz, and N. Zahra, “Risk zone estimation of newborn jaundice based on skin color image analysis,” 07 2019, pp. 176–181.
- [38] A. R. Deepthi, C. Singh, and J. B. Jeeva, “A device to detect hyperbilirubin through non-invasive technique,” in *2019 Innovations in Power and Advanced Computing Technologies, i-PACT 2019*, 2019.
- [39] R. Singla and S. Singh, “A framework for detection of jaundice in newborn babies using homomorphic filtering based image processing,” in *Proceedings of the International Conference on Inventive Computation Technologies, ICICT 2016*, 2016.

- [40] F. T. Z. Khanam, A. Al-Naji, A. G. Perera, K. Gibson, and J. Chahl, "Non-contact automatic vital signs monitoring of neonates in nicu using video camera imaging," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, vol. 11, no. 2, p. 278–285, 2023.
- [41] W. Hashim, M. Alkhaled, A. Al-Naji, and I. Al-Rayahi, "A review on image processing based neonatal jaundice detection techniques," in *7th International Conference on Contemporary Information Technology and Mathematics, ICCITM 2021*, 2021, p. 213–218.
- [42] M. Nixon-Hill, R. P. Mookerjee, and T. S. Leung, "Assessment of bilirubin levels in patients with cirrhosis via forehead, sclera and lower eyelid smartphone images," *PLOS Digital Health*, vol. 2, no. 10, p. e0000357, 2023.
- [43] W. Hashim, A. Al-Naji, I. A. Al-Rayahi, and M. Oudah, "Computer vision for jaundice detection in neonates using graphic user interface," in *IOP Conference Series: Materials Science and Engineering*, vol. 1105, no. 1. IOP Publishing, 2021, p. 012076.
- [44] P. Szabo, M. Wolf, H. U. Bucher, J. C. Fauchère, D. Haensse, and R. Arlettaz, "Detection of hyperbilirubinaemia in jaundiced full-term neonates by eye or by bilirubinometer?" *European Journal of Pediatrics*, vol. 163, no. 12, p. 722–727, 2004.
- [45] A. Y. Abdulrazzak, S. L. Mohammed, A. Al-Naji, and J. Chahl, "Real-time jaundice detection in neonates based on machine learning models," *BioMedInformatics*, vol. 4, no. 1, pp. 623–637, 2024.
- [46] L. de Greef, M. Goel, M. J. Seo, E. C. Larson, J. W. Stout, J. A. Taylor, and S. N. Patel, "Bilicam: Using mobile phones to monitor newborn jaundice," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2014)*. Seattle, WA, USA: ACM, 2014, pp. 331–342.
- [47] A. Althnian, N. Almanea, and N. Aloboud, "Neonatal jaundice diagnosis using a smartphone camera based on eye, skin, and fused features with transfer learning," *Sensors*, vol. 21, no. 21, p. 7038, 2021.
- [48] M. Aydın, F. Hardalaç, B. Ural, and S. Karap, "Neonatal jaundice detection system," *Journal of Medical Systems*, vol. 40, no. 7, 2016.
- [49] F. Makhloghi, "Artificial intelligence-based non-invasive bilirubin prediction for neonatal jaundice using 1d convolutional neural network," *Scientific Reports*, vol. 15, no. 1, 2025.
- [50] W. Hashim, A. Al-Naji, I. A. Al-Rayahi, M. Alkhaled, and J. Chahl, "Neonatal jaundice detection using a computer vision system," *Designs*, vol. 5, no. 4, 2021.
- [51] V. Chernov, J. Alander, and V. Bochko, "Integer-based accurate conversion between rgb and hsv color spaces," *Computers Electrical Engineering*, vol. 46, pp. 328–337, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0045790615002827>

- [52] D. Wang, X. Wang, L. Wang, M. Li, Q. Da, X. Liu, X. Gao, J. Shen, J. He, T. Shen, Q. Duan, J. Zhao, K. Li, Y. Qiao, and S. Zhang, “A real-world dataset and benchmark for foundation model adaptation in medical image classification,” *Scientific Data*, vol. 10, no. 1, 2023.
- [53] J. Cepeda-Negrete and R. E. Sanchez-Yanez, “Gray-world assumption on perceptual color spaces,” in *Image and Video Technology. PSIVT 2013*, ser. Lecture Notes in Computer Science, R. Klette, M. Rivera, and S. Satoh, Eds. Berlin, Heidelberg: Springer, 2014, vol. 8333, pp. 451–462.
- [54] P. Musa, F. al Rafi, and M. Lamsani, “A review: Contrast-limited adaptive histogram equalization (clahe) methods to help the application of face recognition,” in *Proceedings of the International Annual Conference (IAC)*, 2019.
- [55] K. B. Shaik, P. Ganesan, V. Kalist, B. Sathish, and J. M. M. Jenitha, “Comparative study of skin color detection and segmentation in hsv and ycber color space,” *Procedia Computer Science*, vol. 57, pp. 41–48, 2015, 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050915018918>
- [56] OpenCV Development Team, “Morphological transformations,” [https://docs.opencv.org/4.x/d9/d61/tutorial\\_py\\_morphological\\_ops.html](https://docs.opencv.org/4.x/d9/d61/tutorial_py_morphological_ops.html), OpenCV, 2025.
- [57] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 12 2015. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [58] Keras Team, “Keras applications,” <https://keras.io/api/applications/>, Keras Documentation, 2025.
- [59] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” 1 2018, densenet121. [Online]. Available: <http://arxiv.org/abs/1608.06993>
- [60] M. Tan and Q. V. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” 9 2020, eFFICIENTNET. [Online]. Available: <http://arxiv.org/abs/1905.11946>
- [61] M. S. Jarjees, N. Abdulkadir, and K. Kandla, “Neonatal jaundice detection system using convolutional neural network algorithm,” 2023. [Online]. Available: [https://www.researchgate.net/publication/366839813\\_Neonatal\\_Jaundice\\_Detection\\_System\\_Using\\_Convolutional\\_Neural\\_Network\\_Algorithm](https://www.researchgate.net/publication/366839813_Neonatal_Jaundice_Detection_System_Using_Convolutional_Neural_Network_Algorithm)
- [62] S. Iqbal, A. N. Qureshi, A. Ullah, J. Li, and T. Mahmood, “Improving the robustness and quality of biomedical cnn models through adaptive hyperparameter tuning,” *Applied Sciences (Switzerland)*, vol. 12, 11 2022.
- [63] A. Y. Abdulrazzak, S. L. Mohammed, and A. Al-Naji, “Njn: A dataset for the normal and jaundiced newborns,” *BioMedInformatics*, vol. 3, no. 3, pp. 543–552, 2023. [Online]. Available: <https://www.mdpi.com/2673-7426/3/3/37>

- [64] F. O. Salami, M. Muzammel, Y. Mourchid, and A. Othmani, "Artificial intelligence non-invasive methods for neonatal jaundice detection: A review," 4 2025.
- [65] BiliTool, Inc., "Bilitool," 2024. [Online]. Available: <https://bilitool.org>
- [66] D. Dahmani, M. Cheref, and S. Larabi, "Zero-sum game theory model for segmenting skin regions," *Image and Vision Computing*, vol. 99, 7 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0262885620300573?via%3Dihub>
- [67] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia, "Human skin detection using rgb, hsv and ycbcr color models," in *Proceedings of the International Conference on Communication and Signal Processing 2016 (ICCASP 2016)*. Atlantis Press, 2016/12, pp. 324–332. [Online]. Available: <https://doi.org/10.2991/iccasp-16.2017.51>
- [68] Y. H. Chen, K. T. Hu, and S. J. Ruan, "Statistical skin color detection method without color transformation for real-time surveillance systems," *Engineering Applications of Artificial Intelligence*, vol. 25, pp. 1331–1337, 10 2012.
- [69] H. A. Alabbasi and F. Moldoveanu, *Human face detection from images, based on skin color*. IEEE, 2014. [Online]. Available: <https://ieeexplore.ieee.org/document/6982471>
- [70] F. Gasparini and R. Schettini, "Skin segmentation using multiple thresholding," in *Internet Imaging VII*, vol. 6061. SPIE, 1 2006.
- [71] S. Kamble, C. H. Muntean, and A. A. Simiscuka, "A hybrid hsv and ycrb opencv-based skin tone recognition mechanism for makeup recommender systems," in *20th International Wireless Communications and Mobile Computing Conference, IWCMC 2024*. Institute of Electrical and Electronics Engineers Inc., 2024, pp. 1224–1229. [Online]. Available: <https://ieeexplore.ieee.org/document/10592313>
- [72] A. Kaur and B. V. Kranthi, "Comparison between ycbcr color space and cielab color space for skin color segmentation," Tech. Rep., 2012. [Online]. Available: [www.ijais.org](http://www.ijais.org)
- [73] B. K. Abdulkader and M. H. Aziz, "Neonatal jaundice severity detection from skin images using deep transfer learning techniques," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, pp. 382–396, 1 2025. [Online]. Available: <https://jeemi.org/index.php/jeemi/article/view/576/203>
- [74] K. Gupta, V. Sharma, and S. S. Kathait, "Smart screening: Non-invasive detection of severe neonatal jaundice using computer vision and deep learning," *International Journal of Computer Applications*, vol. 186, pp. 35–43, 8 2024.