



Pontificia Universidad
JAVERIANA
Cali

IDENTIFICACIÓN AUTOMÁTICA DE IMÁGENES DE ECOGRAFÍAS CRANEALES CON SU RESPECTIVO PLANO ANATÓMICO

*Laura Daniela Acevedo Virgues
Luis Armando Villarreal Huertas*

*Proyecto Aplicado para optar al título de
Magíster en Ciencia de Datos*

Director
Cristian Alejandro Torres Valencia

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, JUNIO DE 2026

FICHA RESUMEN

TÍTULO DEL PROYECTO: IDENTIFICACIÓN AUTOMÁTICA DE IMÁGENES DE ECOGRAFIAS CRANEALES CON SU RESPECTIVO PLANO ANATÓMICO

1. **ÁREA DE TRABAJO:** Ciencia de Datos.
2. **TIPO DE PROYECTO (Aplicado, Innovación, Investigación):** Investigación.
3. **ESTUDIANTE(S):** Laura Daniela Acevedo Virgues y Luis Armando Villarreal Huertas.
4. **CORREO ELECTRÓNICO:** lauraacev@javerianacali.edu.co, luisvillarreal@javerianacali.edu.co.
5. **DIRECCIÓN Y TELÉFONO:** Cr. 46 #187-65 Bogotá DC, 6789877; Calle 21d #9c, Mz E casa 89 B/ Santa Mónica, Pasto, Nariño - 3166875300.
6. **DIRECTOR:** Cristian Alejandro Torres Valencia.
7. **VINCULACIÓN DEL DIRECTOR:** Profesor.
8. **CORREO ELECTRÓNICO DEL DIRECTOR:** cristian.torres@javerianacali.edu.co
9. **PALABRAS CLAVE (al menos 5):** Aprendizaje profundo, ecografías, radiología, planos anatómicos, modalidad US.
10. **FECHA DE INICIO:** 02 de junio de 2025.
11. **DURACIÓN ESTIMADA (En meses):** 12 meses.
12. **RESUMEN:**

Este proyecto se basa en el análisis de ecografías craneales fetales que son una herramienta que apoya el diagnóstico de enfermedades y anomalías durante el embarazo, pero que se ve perjudicada por la subjetividad de la interpretación del especialista. Esta clara problemática se ve reflejada por los parámetros intrínsecos de los equipos médicos donde la medición va a variar, ya que, en la modalidad de ultrasonido, es primordial hacer mediciones en el equipo para así dar un buen diagnóstico; los parámetros del equipo pueden alterar el transductor, pieza clave para la toma del estudio, sobre todo en el seguimiento del feto que está en constante desarrollo. Con un modelo de aprendizaje profundo y un conjunto de datos correctamente etiquetados, se generó un prototipo que identifique el plano anatómico de imágenes de ecografías craneales fetales, el cual se implementó en cuatro fases. En la primera fase se hizo una revisión sistemática de la información para determinar el modelo a usar y el conjunto de datos. En la segunda fase se buscó mejorar la calidad de la imagen al usar técnicas de preprocesamiento; esto permitió que a la hora de usar el método de aprendizaje profundo mejore el modelo para la detección de los planos anatómicos. En la tercera fase se aplicó el modelo de aprendizaje profundo y, por último, en la cuarta se implementó un análisis que busca evaluar la precisión del prototipo. Estas fases van de la mano con el planteamiento de los cuatro objetivos, que buscan cumplir con una estructura en el proyecto. Finalmente, se obtuvo un prototipo de aprendizaje de conjunto capaz de clasificar las imágenes fetales craneales a través de tres modelos con exactitudes mayores al 90%. Las posibles aplicaciones van dirigidas a reducir la subjetividad y variabilidad de los diagnósticos, facilitando una interpretación más precisa.

TABLA DE CONTENIDO

LISTA DE FIGURAS.....	6
LISTA DE TABLAS.....	8
1. INTRODUCCIÓN	9
2. CONTEXTUALIZACIÓN DEL PROYECTO	10
2.1. DEFINICIÓN DEL PROBLEMA	10
2.2. PLANTEAMIENTO DEL PROBLEMA	10
2.3. FORMULACIÓN DEL PROBLEMA	11
2.4. SISTEMATIZACIÓN	11
2.5. OBJETIVOS	12
2.5.1. OBJETIVO GENERAL	12
2.5.2. OBJETIVOS ESPECÍFICOS	13
2.6. MARCO DE REFERENCIA.....	13
2.6.1. MARCO TEÓRICO	13
2.6.1.1. ECOGRAFÍA	13
2.6.1.2. ECOGRAFÍA FETAL.....	13
2.6.1.3. SISTEMAS DE DIAGNÓSTICO PACS	14
2.6.2. CIENCIA DE DATOS.....	15
2.6.2.1. INTELIGENCIA ARTIFICIAL	15
2.6.2.2. MACHINE LEARNING	16
2.6.2.3. DEEP LEARNING	16
2.7. PREPROCESAMIENTO Y PROCESAMIENTO DE IMÁGENES DE ULTRASONIDO FETAL	16
2.7.1. REDUCCIÓN DE RUIDO	17
2.7.2. NORMALIZACIÓN DE INTENSIDAD	17
2.7.3. RECORTE Y ALINEACIÓN	18
2.7.4. REDES NEURONALES CONVOLUCIONALES (CNN)	18
2.7.5. OPERACIÓN DE CONVOLUCIÓN	19
2.7.6. ARQUITECTURA VGG (VGG16 / VGG19).....	20
2.7.7. RESNET (REDES NEURONALES RESIDUALES)	20
2.7.8. APRENDIZAJE PROFUNDO EN LA CLASIFICACIÓN DE PLANOS FETALES.....	22

2.8.	MÉTRICAS DE EVALUACIÓN	22
2.8.1.	MATRIZ DE CONFUSIÓN	23
2.8.2.	EXACTITUD (ACCURACY).....	23
2.8.3.	PRECISIÓN (PRECISION)	23
2.8.4.	EXHAUSTIDAD (RECALL)	23
2.8.5.	F1 SCORE.....	24
2.8.6.	CONFIANZA.....	24
2.8.7.	ENTROPÍA	24
2.9.	ANTECEDENTES	24
3.	REVISIÓN SISTEMÁTICA	27
3.1.	ESTRATEGIA DE BÚSQUEDA Y CRITERIOS DE EXCLUSIÓN	27
3.1.1.	CRITERIOS DE EXCLUSIÓN	27
3.1.2.	MÉTODO PRISMA	27
4.	ADQUISICIÓN, CONFORMACIÓN Y PREPROCESAMIENTO DE IMÁGENES	31
4.1.	JUSTIFICACIÓN	31
4.2.	ADQUISICIÓN DE DATOS	32
4.2.1.	FUENTE PRINCIPAL	32
4.2.2.	PROTOCOLO DEL ETIQUETADO	33
4.3.	ESTRATEGIAS PARA LA AMPLIACIÓN DEL DATASET.....	34
4.4.	PREPROCESAMIENTO DE IMÁGENES (PIPELINE)	36
4.4.1.	NORMALIZACIÓN.....	37
4.4.2.	REDUCCIÓN DE RUIDO	38
4.4.3.	REDIMENSIONAMIENTO.....	38
4.5.	ESTRUCTURA FINAL.....	39
4.5.1.	DIVISIÓN DEL DATASET.....	39
5.	DESARROLLO Y ENTRENAMIENTO DE MODELOS.....	44
5.1.	SELECCIÓN DE ARQUITECTURAS	44
5.2.	IMPLEMENTACIÓN DE ARQUITECTURAS	44
5.2.1.	ARQUITECTURAS BASADAS EN TRANSFER LEARNING	44
5.2.1.1.	RESNET50.....	45
5.2.1.2.	VGG16.....	49

5.2.1.3. ARQUITECTURA PERSONALIZADA	53
5.2.1.3.1. DIVERSAS ARQUITECTURAS USADAS	53
5.2.1.3.2. DISEÑO DE LA ARQUITECTURA FINAL	55
5.2.1.3.3. DIVERSOS MODELOS CON LA ARQUITECTURA	56
5.2.1.4. MODELO ESCOGIDO	57
5.3. DISCUSIÓN.....	60
6. PROTOTIPO PARA LA CLASIFICACIÓN	63
6.1. STREAMLIT.....	63
6.2. APRENDIZAJE DE CONJUNTO	63
6.3. PROTOTIPO FINAL	64
6.3.1. PREDICCIONES	67
6.3.1.1. ETIQUETA: DIVERSAS CABEZAS FETALES	67
6.3.1.2. ETIQUETA: NO ES UN CEREBRO.....	68
6.3.1.3. ETIQUETA: TRANS-CEREBELO	70
6.3.1.4. ETIQUETA: TRANS-TALÁMICO	71
6.3.1.5. ETIQUETA: TRANS-VENTRICULAR	72
7. CONCLUSIONES Y TRABAJOS FUTUROS.....	74
7.1. CONCLUSIONES	74
7.2. TRABAJOS FUTUROS.....	74
REFERENCIAS BIBLIOGRÁFICAS	76

LISTA DE FIGURAS

Figura 1: Indicaciones para el examen ecográfico en el primer trimestre [16].....	14
Figura 2: Ejemplo de modelos de aprendizaje profundo para procesamiento de imágenes fetales. Tomado de [24]	19
Figura 3: Estructura simple del VGG	20
Figura 4: Diagrama simple de bloque residual. Tomado de [26].....	21
Figura 5: Matriz de confusión. Tomado de [30].....	23
Figura 6: Flujo de selección de artículos según diagrama PRISMA.....	28
Figura 7: Planos craneales fetales: (a) trans-ventricular, (b) trans-talámico y (c) trans-cerebelar. Tomado de [38].....	31
Figura 8: Distribución de clases del dataset final.....	34
Figura 9: Undersampling de la clase mayoritaria	35
Figura 10: Primera imagen del dataset	37
Figura 11: Comparación en paralelo de imagen original y normalizada	38
Figura 12: Comparación en paralelo de imagen normalizada y filtrada.....	38
Figura 13: Comparación paralela de imagen filtrada y redimensionada.....	39
Figura 14: Estructura del dataset con la división 70/20/10.....	39
Figura 15: Distribución de clases en el conjunto de Entrenamiento	40
Figura 16: Distribución de clases en el conjunto de Validación.....	41
Figura 17: Distribución de clases en el conjunto de Prueba	41
Figura 18: Imágenes del conjunto de Entrenamiento de la etiqueta Trans-talámica	42
Figura 19: Imágenes del conjunto de Validación de la etiqueta Trans-talámica.....	43
Figura 20: Imágenes del conjunto de Prueba de la etiqueta Trans-talámica	43
Figura 21: Arquitectura del ResNet50. Tomado de [52].....	45
Figura 22: Descongelamiento de capas de ResNet50.....	46
Figura 23: Accuracy de ResNet50 con 35% de capas descongeladas	46
Figura 24: ResNet50 con 35% con 12 epochs.....	47
Figura 25: Matriz de confusión de ResNet con 35% de capas descongeladas	48
Figura 26: Arquitectura VGG16.....	50
Figura 27: Descongelamiento de capas de VGG16	50
Figura 28: Accuracy de VGG16 con 30% de capas descongeladas	51
Figura 29: VGG16 con 40% con 12 epochs	51
Figura 30: Matriz de confusión de VGG16 con 40% de capas descongeladas	52
Figura 31: Diagrama de la arquitectura base personalizada	56
Figura 32: Arquitectura personalizada con mecanismo de atención	57
Figura 33: Accuracy para la arquitectura personalizada.....	58
Figura 34: Matriz de confusión para arquitectura personalizada.....	59
Figura 35: Diagrama de bloque de aprendizaje de conjunto paralelo. Modificado de [59]	64
Figura 36: Interfaz de Streamlit del clasificador de Ultrasonidos Fetales Craneales	65

Figura 37: Mostrado de imagen cruda y preprocesada lado a lado	65
Figura 38: Predicción de cada modelo, su confianza y la entropía.....	66
Figura 39: Votación por mayoría y gráfica de probabilidades por modelo	66
Figura 40: Resumen de predicciones y diagnóstico final.....	67
Figura 41: Imagen original y preprocesada de etiqueta Diverse-Fetal-Head	68
Figura 42: Predicciones por modelos y votación de etiqueta Diverse-Fetal-Head.....	68
Figura 43: Imagen original y preprocesada de etiqueta Not-A-Brain	69
Figura 44: Predicciones por modelos y votación de etiqueta Not-A-Brain	69
Figura 45: Imagen original y preprocesada de etiqueta Trans-cerebellum.....	70
Figura 46: Predicciones por modelos y votación de etiqueta Trans-cerebellum	70
Figura 47: Imagen original y preprocesada de etiqueta Trans-thalamic.....	71
Figura 48: Predicciones por modelos y votación de etiqueta Trans-thalamic	71
Figura 49: Imagen original y preprocesada de etiqueta Trans-ventricular	72
Figura 50: Predicciones por modelos y votación de etiqueta Trans-ventricular	72

LISTA DE TABLAS

Tabla 1: Comparativa de métodos de inteligencia artificial en imágenes de ultrasonido fetal	29
Tabla 2: Etiquetas en la agrupación de conjuntos de datos	33
Tabla 3: Transformaciones de imágenes de cabeza fetal para Data Aumenting	35
Tabla 4: Reporte de clasificación de ResNet50	48
Tabla 5: Promedios de métricas de ResNet50	49
Tabla 6: Reporte de clasificación de VGG16	52
Tabla 7: Promedios de métricas de VGG16	53
Tabla 8: Arquitecturas personalizadas usadas	54
Tabla 9: Modelos con la arquitectura personalizada	56
Tabla 10: Reporte de clasificación de VGG16	59
Tabla 11: Promedios de métricas de arquitectura personalizada	60
Tabla 12: Accuracies de los modelos	60
Tabla 13: Métricas de los modelos	61

1. INTRODUCCIÓN

En el campo de las imágenes diagnósticas existen diversas modalidades para la evaluación de los pacientes, cuya selección depende del tipo de lesión o del objetivo clínico que se desea analizar. Entre estas, la ecografía se destaca por ser una técnica no invasiva que permite el seguimiento continuo de la gestante durante el embarazo, facilitando el monitoreo del crecimiento y desarrollo fetal [1]. Sin embargo, el uso de ultrasonido presenta retos importantes, especialmente en la identificación de planos anatómicos específicos, debido a la alta similitud visual entre algunos de ellos. Por ejemplo, los planos trans-ventricular y trans-talámico pueden presentar características similares, lo que dificulta su diferenciación, particularmente en operadores con menor experiencia. Esta variabilidad puede afectar la calidad del diagnóstico y la reproducibilidad de los estudios [1].

En el contexto actual de la revolución tecnológica, la medicina ha fortalecido su vínculo con la tecnología, especialmente en el uso de imágenes médicas. Un ejemplo de ello es la aplicación de la inteligencia artificial en el análisis de estudios ecográficos, donde modelos computacionales pueden asistir en la identificación de estructuras anatómicas y en la clasificación de planos estándar, contribuyendo a mejorar la precisión diagnóstica [2].

El uso de estas herramientas representa un beneficio significativo para los usuarios del sistema de salud, ya que contribuye a optimizar la calidad de la atención y a disminuir los tiempos de diagnóstico. Además, apoya al radiólogo en la interpretación de las imágenes, disminuyendo la variabilidad asociada a la subjetividad del análisis humano.

En este proyecto, se seleccionaron algoritmos y modelos de aprendizaje profundo adecuados mediante una revisión científica de la literatura. Posteriormente, se preparó un conjunto de datos amplio y correctamente etiquetado por expertos, garantizando la calidad de la información utilizada. A continuación, se utilizaron dos modelos preentrenados y se desarrolló un modelo de aprendizaje profundo adicional, todos orientados a la clasificación de imágenes ecográficas según su plano anatómico estándar. Entre los tres modelos, se obtuvieron exactitudes superiores al 90%. El desempeño del modelo fue evaluado mediante métricas de evaluación estándar, permitiendo medir su precisión y capacidad de generalización. Finalmente, se desarrolló una interfaz que permite visualizar los resultados obtenidos por el modelo de aprendizaje de conjunto, facilitando la forma en que el usuario interactúa con el sistema y la interpretación de las predicciones generadas.

2. CONTEXTUALIZACIÓN DEL PROYECTO

2.1. DEFINICIÓN DEL PROBLEMA

La ecografía craneal fetal es una herramienta diagnóstica fundamentalmente útil por ser no invasiva, de bajo costo y segura [3]. Sin embargo, su precisión diagnóstica no está establecida, debido a que es una tarea que requiere un conocimiento profundo y una habilidad inmensa del operador. Incluso para profesionales experimentados, puede ser laboriosa y susceptible a variables externas como la calibración de los equipos o la habilidad humana [4]. Además, la calidad de estas ecografías es, generalmente, baja, lo que presenta una alta variabilidad a la hora de su clasificación. Por lo tanto, se plantea el desarrollo de un prototipo que pueda superar estos desafíos: identificar los planos anatómicos estándar de ecografías craneales.

2.2. PLANTEAMIENTO DEL PROBLEMA

La ecografía es una herramienta que puede evaluar la salud y desarrollo del feto. No solo sirve para identificar la edad gestacional, sino para determinar el correcto crecimiento. A partir del primer trimestre del embarazo es posible la evaluación de las estructuras anatómicas del feto y se diagnostican diferentes anomalías y condiciones, en ellas se encuentran las cardiopatías congénitas, así como las anomalías de la pared abdominal, en extremidades y en órganos genitales, las anomalías del sistema nervioso central (como la microcefalia e hidrocefalia) y posiblemente algunas anomalías del tipo cromosómico o síndromes genéticos [5], [6]. Esta técnica presenta grandes ventajas por su seguridad y bajo costo, sin contar que al ser «no invasiva» no se genera ningún posible daño al paciente. No obstante, su eficacia diagnóstica puede verse reducida por diferentes factores, entre ellos: el movimiento del paciente, lo que podría resultar en imágenes de mala calidad; la técnica del operador a la hora de la correcta interpretación de estructuras; y factores externos como la calibración de los artefactos (que también se altera por el movimiento del paciente), uso de equipos no adecuados y el tipo de control de calidad para la adquisición de las imágenes [7].

En tanto, la correcta interpretación de la ecografía craneal depende críticamente de la determinación de los planos anatómicos estándar, y debido a la calidad de las imágenes obtenidas, puede resultar una tarea laboriosa que no depende netamente de la habilidad del profesional que la interprete, sino también del buen manejo de los factores ya presentados. Este desafío no se limita a la ecografía craneal. La baja calidad de las imágenes puede dificultar un diagnóstico fiable en otras áreas, como en la ecocardiografía [4] y en la interpretación de imágenes de fondo [8].

Si se sigue optando por la determinación manual de los planos anatómicos estándar en ecografías craneales, siendo un método sin asistencia tecnológica, persistirán los desafíos a la hora de determinar las patologías, y la dependencia a la experiencia y habilidad del operador llevará (así como se viene presentado) a la alta variabilidad en los resultados y mediciones [7]. Por consiguiente, la subjetividad inherente y la baja calidad de las imágenes seguirá dificultando la correcta interpretación clínica, lo que se traduce

como una limitación importante en la eficacia del diagnóstico, siendo, además, una tarea que consume tiempo valioso de los profesionales de la salud.

2.3. FORMULACIÓN DEL PROBLEMA

Para identificar el plano anatómico en ecografías craneales fetales se deben superar varios desafíos, entre ellos la baja calidad de las imágenes, por lo que la presente investigación se propone responder las siguientes interrogantes: en primer lugar, ¿cómo se puede implementar un prototipo basado en Aprendizaje Profundo para la identificación automática de planos anatómicos?

Para lograrlo es primordial determinar qué algoritmos y modelos de procesamiento de imágenes son los más adecuados para la clasificación de ecografías craneales. De tal manera, es crucial establecer un método para el preprocesamiento de las imágenes que resuelva los desafíos de baja calidad y asegure un análisis automatizado. A través de la resolución de esas preguntas, se explorará cómo desarrollar, entrenar y optimizar un modelo de Aprendizaje Profundo que pueda clasificar una imagen de ecografía según su plano anatómico estándar y cómo evaluar la precisión del automatizado de planos anatómicos a través de métricas estadísticas y técnicas de validación estándar.

2.4. SISTEMATIZACIÓN

La identificación de los planos anatómicos estándar en ecografías craneales fetales constituye un proceso esencial para la correcta evaluación del desarrollo neurológico y la detección temprana de anomalías congénitas. Sin embargo, en la práctica clínica actual, esta identificación se realiza de manera predominantemente manual, lo que implica una fuerte dependencia del conocimiento y experiencia del profesional de la salud, así como de la calidad de las imágenes obtenidas durante el examen ecográfico. Esta situación genera una alta variabilidad en los resultados, incrementa la subjetividad del diagnóstico y puede afectar la precisión clínica [9].

Uno de los principales factores que inciden en esta problemática es la baja calidad inherente de las imágenes ecográficas, la presencia de ruido, artefactos, sombras acústicas y distorsiones propias del proceso de adquisición. Adicionalmente, el movimiento fetal y materno durante el estudio, así como las variaciones en la configuración y calibración de los equipos médicos, dificultan la obtención de imágenes nítidas y consistentes. Estas limitaciones técnicas impactan directamente en la correcta visualización de las estructuras anatómicas craneales, lo que complica la determinación precisa de los planos estándar [10], [11].

La ausencia de herramientas automatizadas robustas que apoyen la identificación de los planos anatómicos en ecografías craneales representa otra de las principales debilidades en el contexto clínico actual. A pesar de los avances en tecnologías de procesamiento de imágenes y de inteligencia artificial aplicada a la medicina, aún persiste una brecha entre el desarrollo académico de estas herramientas y su implementación efectiva en entornos clínicos reales, especialmente en escenarios donde las imágenes

presentan baja calidad [9], [10].

En este contexto, el Aprendizaje Profundo se posiciona como una alternativa prometedora para abordar esta problemática, al ofrecer la posibilidad de desarrollar modelos capaces de aprender características relevantes directamente a partir de los datos y realizar clasificaciones automáticas con alta precisión. No obstante, su aplicación efectiva en ecografías craneales fetales implica diversos retos, entre ellos la selección adecuada de algoritmos y arquitecturas de redes neuronales, el diseño de estrategias eficientes de preprocesamiento de imágenes que permitan mitigar el ruido y mejorar la calidad visual, así como la correcta preparación y balanceo de los conjuntos de datos utilizados para el entrenamiento [10], [12], [13].

Asimismo, resulta indispensable establecer metodologías rigurosas para el entrenamiento, optimización y evaluación de los modelos propuestos, de manera que se garantice no solo un alto desempeño en términos de exactitud, sino también la confiabilidad y generalización del sistema en diferentes condiciones clínicas. La evaluación mediante métricas estadísticas adecuadas y técnicas de validación estándar se convierte, por lo tanto, en un componente esencial del proceso de investigación [10], [11], [14].

Con respecto a lo anterior, se plantean las siguientes preguntas:

¿Cómo implementar un prototipo basado en Aprendizaje Profundo que permita la identificación automática de los planos anatómicos estándar en ecografías craneales fetales, garantizando una adecuada exactitud y confiabilidad diagnóstica?

¿Qué algoritmos y arquitecturas de Aprendizaje Profundo son los más adecuados para la clasificación automática de los planos anatómicos en ecografías craneales fetales?

¿Qué técnicas de preprocesamiento de imágenes permiten mejorar la calidad de las ecografías craneales y optimizar el desempeño del modelo de clasificación?

¿Cómo preparar y organizar un conjunto de datos de ecografías craneales que sea representativo, balanceado y correctamente etiquetado para el entrenamiento del modelo?

¿Qué métricas estadísticas y métodos de validación son los más apropiados para evaluar el desempeño del prototipo desarrollado?

2.5. OBJETIVOS

2.5.1. OBJETIVO GENERAL

Implementar un prototipo basado en Aprendizaje Profundo para la identificación automática de planos anatómicos específicos en imágenes de ecografía craneal fetal, con

el fin de reducir errores de diagnóstico médico.

2.5.2. OBJETIVOS ESPECÍFICOS

- Seleccionar los algoritmos de procesamiento de imágenes y modelos de Aprendizaje Profundo más adecuados para la clasificación de imágenes médicas mediante la revisión de literatura científica, con el fin de identificar las tecnologías más prometedoras para el prototipo.
- Preparar un conjunto de datos de imágenes de ecografía craneal, asegurando que sea lo suficientemente grande y tenga correctamente etiquetados los planos anatómicos por expertos, con el fin de entrenar modelos de Aprendizaje Profundo.
- Desarrollar un modelo de Aprendizaje Profundo utilizando el conjunto de datos preparado para clasificar las imágenes de ecografías craneales fetales según su plano anatómico estándar.
- Evaluar la precisión del prototipo para detección de planos anatómicos, empleando métricas de evaluación estándar.

2.6. MARCO DE REFERENCIA

2.6.1. MARCO TEÓRICO

En el marco referencial de este proyecto se proporcionan los conceptos fundamentales para entender la importancia y las complicaciones al momento de tomar decisiones e identificación de planos anatómicos basados en Imágenes de Ecografías Craneales, haciendo un acercamiento a los fundamentos de Ecografía, Sistemas de diagnóstico PACS, Ciencia de Datos y Aprendizaje Profundo.

2.6.1.1. ECOGRAFÍA

La ecografía es una técnica de diagnóstico por imagen que utiliza ondas sonoras de alta frecuencia para producir imágenes en tiempo real del interior del cuerpo. Estas ondas rebotan en los tejidos y son registradas por un transductor, lo que permite construir una representación visual de órganos, vasos y estructuras blandas [15].

Su principal utilidad clínica es observar y evaluar estructuras internas del cuerpo, por lo que se emplea en múltiples especialidades médicas, entre ellas ginecología, obstetricia, cardiología, urología y medicina interna.

2.6.1.2. ECOGRAFÍA FETAL

La ecografía durante el embarazo es fundamental para la atención prenatal, se usa para efectuar la confirmación de la edad gestacional y la viabilidad del feto, también el estado del líquido amniótico y la placenta, con el fin de emprender procedimientos diagnósticos en caso de ser necesarios.

Dentro de los exámenes indicados en el primer trimestre de embarazo se encuentra la ecografía estándar, que para nuestro interés busca diagnosticar posibles problemas anatómicos, como los problemas cerebrales. Definitivamente, el primer trimestre de embarazo es el momento ideal para valorar el útero y el fondo del saco de Douglas, también posibilita el diagnóstico confiable de embarazos anembrionarios, la muerte del feto, embarazos ectópicos y la enfermedad trofoblástica gestacional [15].

Confirmar un embarazo intrauterino
Calcular la edad gestacional
Confirmar la actividad cardíaca
Diagnosticar/valorar un embarazo multifetal, incluida la corionicidad y amnionidad
Buscar ciertas anomalías fetales, como anencefalia
Medir la translucidez nucal fetal, cuando es parte de un programa para detección de aneuploidía
Valorar en busca de anomalías uterinas o tumoraciones pélvicas
Valorar una sospecha de embarazo extrauterino
Valorar una sospecha de enfermedad trofoblástica gestacional
Valorar la causa de hemorragia vaginal
Valorar la causa de dolor pélvico
Servir como complemento en la transferencia embrionaria, muestreo de vellosidades coriónicas y localización y extracción de dispositivo intrauterino

Figura 1: Indicaciones para el examen ecográfico en el primer trimestre [15]

Para el tercer trimestre, la ecografía está indicada para la búsqueda de anomalías fetales en un embarazo de riesgo. Sin embargo, en una exploración fetal detallada anterior, entre las semanas 12 (primer trimestre) y 14 (segundo trimestre) del feto, es de esperar que la mayoría de las anomalías no sean detectables a las limitaciones del tamizaje manual y la percepción del profesional tratante. En el caso de que una anomalía sea detectada en el primer trimestre, una ecografía detallada en el segundo trimestre ayudará en un mejor diagnóstico de los hallazgos.

De acuerdo con los porcentajes provistos en [16], el 46% de las anomalías se detectaron en embarazos que en un principio eran de bajo riesgo o no diagnosticados. Y en aquellos que habían sido clasificados como embarazos de alto riesgo, las anomalías detectadas superaron el 60%.

Si bien hay una mayor tasa en ciertos tipos de anomalías en esta etapa gestacional como, anencefalia fetal, haloprosencefalia alobular y otras condiciones, solo se detectaron un tercio de las anomalías cardíacas y ninguna anomalía relacionada a problemas cerebrales que se desarrollaron después, como microcefalia, agenesia del cuerpo calloso y anomalías cerebelosas.

2.6.1.3. SISTEMAS DE DIAGNÓSTICO PACS

Los Sistemas de Archivamiento y Comunicación de imágenes (PACS) son sistemas de gestión de imágenes diseñados específicamente para el área de la radiología, que combinan visualización de imágenes de diversas modalidades, su archivado y su distribución. Antes de que aparecieran los PACS, las imágenes solo se podían ver en estaciones de trabajo de la modalidad en la que se generaban y a través de estos se pudo definir una ruta más definida: adquisición, visualización y archivo de imágenes independientes del fabricante del equipo [17].

Un PACS consta de varios componentes. En la adquisición, se capturan los datos desde dispositivos como tomografía computarizada (CT), resonancia magnética (MRI), ultrasonido, radiografía computarizada, entre otros. En el almacenamiento, se tiene un sistema de almacenamiento electrónico a corto y largo plazo para la información de las imágenes. En la visualización, se dispone de una pantalla para revisión de las imágenes. En el de comunicación, se cuenta con redes de área local (LAN) y de área amplias (WAN) que conecta físicamente varias computadoras. Y, finalmente, en la interfaz, se conecta el PACS con otros sistemas, como el Sistema de Información Radiológica (RIS), para permitir el intercambio de información y mensajes.

Estos sistemas tienen estándares de comunicación, como el DICOM y HL7. El estándar para el formateo e intercambio de imágenes médicas y datos asociados es el DICOM (*Digital Imaging and Communications in Medicine*), que permite que diferentes fabricantes se comuniquen con los PACS. Y HL7 (*Health Level 7*) es más común en sistemas de información hospitalaria (HIS) y de radiología (RIS) para el intercambio de datos de texto electrónico. La integración de ambos es crucial para la interoperabilidad, permitiendo que la información clínica sea útil en ambos lados de la conexión [17].

La integración de un PACS con el RIS o el HIS es importante para evitar la entrada manual de datos, lo que suele provocar errores en los números de registro médico o en la ortografía de los nombres, dificultando la localización de estudios; para la recuperación de estudios previos cuando hay un nuevo examen, ahorrando tiempo al radiólogo; permite que los médicos tengan acceso rápido a las imágenes desde la propia estación de trabajo; y para la información fluya del RIS al dispositivo y que, una vez termine el estudio, regrese al RIS para su validación [18].

La implementación de PACS mejora la eficiencia, la calidad de atención, elimina potenciales pérdidas de información y se puede realizar anotaciones en las imágenes para ayudar a los clínicos a comprender los hallazgos [18].

2.6.2. CIENCIA DE DATOS

La ciencia de datos es un sector relativamente nuevo y de amplio espectro en sus bases de estadística, matemáticas, programación y que puede ser de mucha ayuda a otros sectores, proporcionando soluciones y planes de trabajo de mejora continua.

Además de esto, también lidia con una enorme cantidad de datos y de fuentes, ayudando en el desarrollo de algoritmos y métodos capaces de hacer el *Data Mining*, transformando dichos datos y generando decisiones o *insights*, identificando patrones, predicciones, extrayendo información valiosa para los diferentes sectores y negocios.

2.6.2.1. INTELIGENCIA ARTIFICIAL

La Inteligencia Artificial (IA) se define como una disciplina de la informática dedicada a la creación de sistemas capaces de ejecutar tareas que normalmente requieren capacidades cognitivas humanas, entre ellas el aprendizaje, el análisis lógico y la toma de

decisiones. A través de métodos computacionales y algoritmos avanzados, estas tecnologías pueden interpretar información, adaptarse a diferentes escenarios y resolver problemas de manera automatizada [18].

2.6.2.2. MACHINE LEARNING

El Machine Learning es un campo derivado de la Inteligencia Artificial que se centra en el diseño de algoritmos capaces de aprender mediante la experiencia obtenida de los datos. En lugar de depender de instrucciones programadas para cada situación, los modelos construyen patrones de comportamiento y ajustan su desempeño de acuerdo con la información procesada. Gracias a ello, es posible desarrollar sistemas predictivos y herramientas que automatizan procesos de análisis de datos [19].

2.6.2.3. DEEP LEARNING

El Deep Learning representa una evolución del Machine Learning y se fundamenta en el uso de redes neuronales profundas con múltiples niveles de procesamiento. Estas arquitecturas computacionales permiten analizar grandes cantidades de información y reconocer patrones complejos, simulando parcialmente ciertos mecanismos de procesamiento del cerebro humano. Su aplicación ha permitido avances significativos en áreas como el reconocimiento de imágenes, el procesamiento de señales y la automatización de tareas inteligentes [20].

2.7. PREPROCESAMIENTO Y PROCESAMIENTO DE IMÁGENES DE ULTRASONIDO FETAL

El ultrasonido fetal es una modalidad altamente dependiente del operador y susceptible a variaciones en la calidad de adquisición. Factores como el movimiento fetal, la posición de la madre, la impedancia acústica y las diferencias entre equipos provocan imágenes con bajo contraste, ruido speckle, artefactos de sombra y regiones poco definidas. Por ello, antes de aplicar modelos de aprendizaje profundo, es imprescindible realizar un conjunto de técnicas de preprocesamiento destinadas a estandarizar la imagen y mejorar la detección de patrones anatómicos. Estudios recientes coinciden en que una adecuada fase de preprocesamiento influye de manera directa en el rendimiento final de los modelos de clasificación [19].

Uno de los métodos más empleados es la reducción del ruido, que tiene como propósito suavizar la imagen sin eliminar información anatómica relevante. Este tipo de ruido es resultado de interferencias de la señal acústica y afecta la observación de las estructuras fetales, por lo cual se aplican filtros de mediana, difusión anisotrópica y suavizado adaptativo. La disminución de ruido mejora el contraste y facilita la detección de características profundas por parte de la red neuronal, ya que minimiza patrones erróneos que podrían confundirse con estructuras anatómicas reales.

Además, otro proceso fundamental es la normalización de intensidad, ya que diferentes equipos ecográficos pueden generar variaciones en brillo, contraste y

distribución tonal. Para afrontar esta variabilidad, se adoptan estrategias como normalización min–max, estandarización tipo Z-score y ecualización de histograma, las cuales favorecen la uniformidad entre imágenes provenientes de distintos estudios clínicos. La literatura demuestra que la normalización adecuada mejora la estabilidad del entrenamiento y reduce el riesgo de sobreajuste [20]. Otra técnica relevante es el recorte (cropping) y la alineación de la zona anatómica de interés. En ecografía fetal, las imágenes suelen contener zonas irrelevantes como partes maternas o regiones sin información diagnóstica. Al centrar únicamente el área del cráneo o torso fetal, el modelo aprende patrones discriminativos más fácilmente. El dataset FPUS23, por ejemplo, incluye un proceso riguroso de alineación y recorte para mejorar la clasificación de orientaciones y planos fetales [21]. Gracias a estos pasos de preprocesamiento, se obtiene una entrada limpia, uniforme y adecuada para que las redes neuronales convolucionales puedan extraer características significativas.

2.7.1. REDUCCIÓN DE RUIDO

El ruido speckle es un tipo de interferencia granular que afecta la uniformidad de la imagen. Para reducirlo se aplican filtros de mediana, filtros anisotrópicos, difusión no lineal y métodos adaptativos que mantienen bordes anatómicos esenciales. La reducción de ruido mejora la delimitación de estructuras y facilita que las redes neuronales identifiquen patrones reales y no artefactos aleatorios.

2.7.2. NORMALIZACIÓN DE INTENSIDAD

Cada ecógrafo produce imágenes con escalas tonales diferentes. Para unificar la entrada, se emplean técnicas como:

- Normalización min–max: Transforma a una escala fija, generalmente a 0 y 1, preservando las relaciones entre los valores originales.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

- Estandarización mediante Z-score: Transforma datos para tener una media de 0 y una desviación estándar de 1, indicando cuántas desviaciones estándar un valor está por encima o debajo de la media.

$$Z = \frac{X - \mu}{\sigma}$$

- Ecualización de histograma: Mejora el contraste al distribuir uniformemente las intensidades de los píxeles.

Estas técnicas permiten que los valores de los píxeles estén dentro de un rango

comparable, evitando que la red aprenda patrones basados en diferencias de intensidad que no son anatómicas [20].

2.7.3. RECORTE Y ALINEACIÓN

El recorte elimina zonas irrelevantes como bordes, áreas maternas o regiones sin información. La alineación centra la estructura fetal en la imagen, facilitando la detección automática. El dataset FPUS23 utiliza estos procesos para mejorar la clasificación de planos y orientaciones fetales [21].

2.7.4. REDES NEURONALES CONVOLUCIONALES (CNN)

Las Convolutional Neural Networks (CNN) han revolucionado el procesamiento de imágenes debido a su capacidad de adquisición de conocimiento automático de características relevantes sin la necesidad de diseñar manualmente descriptores. En el ámbito de la ecografía fetal, las CNN permiten identificar patrones anatómicos que humanamente pueden ser difíciles de distinguir debido al ruido propio del ultrasonido, tales como contornos óseos, estructuras del cerebro o líneas anatómicas internas.

El núcleo matemático de una CNN es la convolución, un proceso que aplica filtros sobre la imagen para detectar patrones. Esta operación permite que la red identifique características como bordes, formas o texturas, fundamentales para diferenciar entre planos fetales. Estudios recientes muestran que CNN estándares, incluso sin modificaciones avanzadas, han logrado buenos resultados en la clasificación de planos del cerebro fetal y otras regiones anatómicas [22].

Las capas de MaxPooling complementan este proceso al reducir la resolución del mapa de características, manteniendo los elementos más representativos y evitando redundancias. Esta reducción no solo disminuye el costo computacional, sino que también ofrece invariancia a pequeñas traslaciones o rotaciones, lo cual es especialmente útil en imágenes fetales debido al movimiento del feto y variabilidad en la manipulación del transductor.

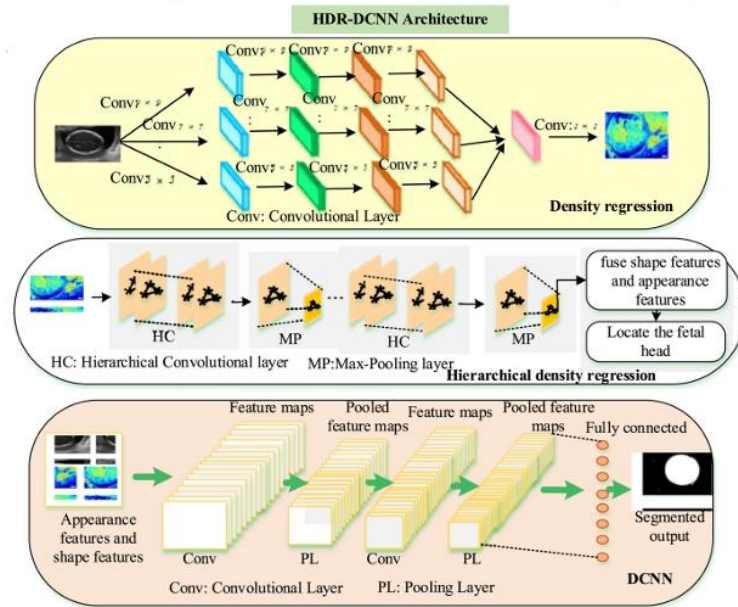


Figura 2: Ejemplo de modelos de aprendizaje profundo para procesamiento de imágenes fetales.
Tomado de [23]

Como se observa en la **Figura 2**, el flujo de procesamiento integra capas convolucionales para la extracción de rasgos jerárquicos y capas de pooling para la reducción de dimensionalidad, permitiendo fusionar información estructural y de apariencia antes de la clasificación o segmentación final.

En conjunto, la arquitectura CNN constituye la base sobre la cual se han construido modelos más complejos como VGG y ResNet, que superan considerablemente los métodos tradicionales basados en reglas o extracción manual de características.

2.7.5. OPERACIÓN DE CONVOLUCIÓN

$$Y(i, j) = \sum_{m=0}^{m-1} \sum_{n=0}^{n-1} x(i + m, j + n) \cdot K(m, n)$$

donde:

- X es la imagen de entrada,
- K es el filtro,
- Y es el mapa de características generado.

Esta operación permite que la red detecte patrones locales como bordes óseos del cráneo fetal, texturas cerebrales y diferencias sutiles entre planos correctos e incorrectos. Las CNN básicas ya han demostrado ser altamente efectivas en clasificación de planos fetales [22].

2.7.6. ARQUITECTURA VGG (VGG16 / VGG19)

La arquitectura VGG, desarrollada por el Visual Geometry Group, es una de las más influyentes en el área del Deep Learning aplicado a imágenes. Su principal característica es el uso de filtros pequeños de 3×3 distribuidos en varios bloques convolucionales, lo que permite una extracción profunda y jerárquica de características. Esta simplicidad estructural ha convertido a VGG en una arquitectura altamente estable, fácil de entrenar y capaz de generalizar bien en contextos médicos.

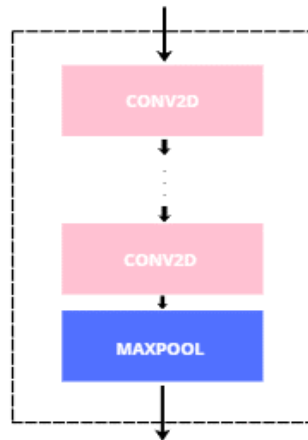


Figura 3: Estructura simple del VGG

En el campo del ultrasonido fetal, VGG16 y VGG19 han sido ampliamente utilizadas en la clasificación de planos estándares debido a su efectividad para capturar detalles texturales. Esto es relevante porque las imágenes fetales presentan cambios sutiles entre un plano correcto y uno incorrecto, por lo que los filtros pequeños ayudan a detectar pequeñas diferencias anatómicas. Investigaciones recientes demuestran que VGG, combinada con técnicas de fusión de características o clasificadores adicionales como MLP, logra mejorar significativamente la precisión en la identificación de planos fetales [22], [24].

La arquitectura se compone de bloques repetidos. Este patrón se repite cinco veces, seguido por capas totalmente conectadas y una capa Softmax para clasificación:

Bloque: Conv3 → Conv3 → MaxPool

2.7.7. RESNET (REDES NEURONALES RESIDUALES)

Las Residual Neural Networks (ResNet) representan un avance significativo respecto de las CNN tradicionales y modelos como VGG. Su innovación se basa en los bloques residuales, donde la salida de un conjunto de capas convolucionales se suma directamente a la entrada original mediante caminos de atajo (*skip connections*). Esta operación permite entrenar redes más profundas sin que se degrade el rendimiento,

problema común en modelos convencionales.

Los bloques residuales permiten que la red aprenda diferencias sutiles entre la imagen de entrada y la salida esperada, lo cual facilita la optimización y acelera la convergencia. En el ámbito del ultrasonido fetal, esta característica es particularmente relevante porque los planos fetales presentan variaciones mínimas entre categorías, y una red profunda puede captar mejor estos matices.

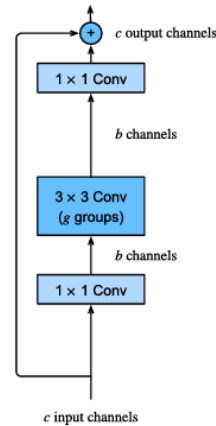


Figura 4: Diagrama simple de bloque residual. Tomado de [25]

El dataset FPUS23 demuestra que ResNet34 ofrece excelente rendimiento en la clasificación de orientaciones y planos fetales, mostrando superioridad respecto a modelos tradicionales [21]. Asimismo, se ha propuesto una arquitectura residual de 82 capas, optimizada mediante Bayesian Optimization, que mejora la clasificación de planos cerebrales y de imágenes fetales comunes con gran precisión [26].

Las Residual Neural Networks (ResNet) representan un avance significativo respecto de las CNN tradicionales y modelos como VGG. Su innovación se basa en los bloques residuales, donde la salida de un conjunto de capas convolucionales se suma directamente a la entrada original mediante caminos de atajo (*skip connections*). Esta operación permite entrenar redes más profundas sin que se degrade el rendimiento, problema común en modelos convencionales.

Los bloques residuales permiten que la red aprenda diferencias sutiles entre la imagen de entrada y la salida esperada, lo cual facilita la optimización y acelera la convergencia. En el ámbito del ultrasonido fetal, esta característica es particularmente relevante porque los planos fetales presentan variaciones mínimas entre categorías, y una red profunda puede captar mejor estos matices.

El dataset FPUS23 demuestra que ResNet34 ofrece excelente rendimiento en la clasificación de orientaciones y planos fetales, mostrando superioridad respecto a modelos tradicionales [21]. Asimismo, se ha propuesto una arquitectura residual de 82 capas, optimizada mediante Bayesian Optimization, que mejora la clasificación de planos cerebrales y de imágenes fetales comunes con gran precisión [26].

$$y = F(x, W) + x$$

donde:

- $F(x)$ es la transformación aprendida por la red,
- x es la entrada original que se suma mediante un *skip connection*.

Este proceso permite que la red aprenda la “diferencia” o residuo, lo cual facilita la optimización incluso en redes muy profundas.

2.7.8. APRENDIZAJE PROFUNDO EN LA CLASIFICACIÓN DE PLANOS FETALES

El aprendizaje profundo ha permitido automatizar tareas que antes dependían exclusivamente de la experiencia del operador, como la identificación correcta de los planos anatómicos necesarios para una evaluación fetal completa. Los modelos basados en CNN, VGG y ResNet se han aplicado exitosamente en investigaciones recientes con el fin de clasificar los planos estándares del cerebro y otras regiones fetales.

Por ejemplo, se han desarrollado sistemas basados en MLP y VGG16, los cuales combinan características profundas con clasificadores adicionales para mejorar la discriminación entre planos semejantes [22]. Otros trabajos utilizan ensembles o stacking, integrando varias redes para obtener un modelo más robusto y confiable que supera los modelos individuales [24].

Asimismo, enfoques basados en aprendizaje multitarea permiten identificar simultáneamente el plano y características anatómicas clave que ayudan a verificar la calidad de la imagen, sin involucrar segmentación explícita. El modelo propuesto por Zhao et al. demuestra que es posible detectar con precisión el plano correcto utilizando CNN y grafos de conocimiento anatómico que incorporan información contextual [27]. En conjunto, estas investigaciones demuestran que el aprendizaje profundo es altamente efectivo para reconocer patrones sutiles en las imágenes fetales, automatizar la clasificación de planos y apoyar a los especialistas durante el proceso de diagnóstico prenatal.

2.8. MÉTRICAS DE EVALUACIÓN

Las métricas de evaluación sirven para cuantificar el desempeño de un modelo de aprendizaje automático, permitiendo detectar errores, falsas predicciones, subajuste y sobreajustes. Su función principal es proporcionar una medida objetiva de qué tan cerca están las predicciones de la realidad (ground truth), lo cual es crucial para optimizar algoritmos y validar su utilidad en aplicaciones reales.

2.8.1. MATRIZ DE CONFUSIÓN

La matriz de confusión es la base para entender el funcionamiento de las métricas de evaluación. Se representa como una matriz nxn que muestra las predicciones de un clasificador en cuatro categorías: **i)** Verdadero positivo (VP), **ii)** Verdadero negativo (VN), **iii)** Falso positivo (FP) y **iv)** Falso negativo (FN). Proporciona información crucial sobre los posibles errores que comete el modelo en la clasificación de imágenes [28].

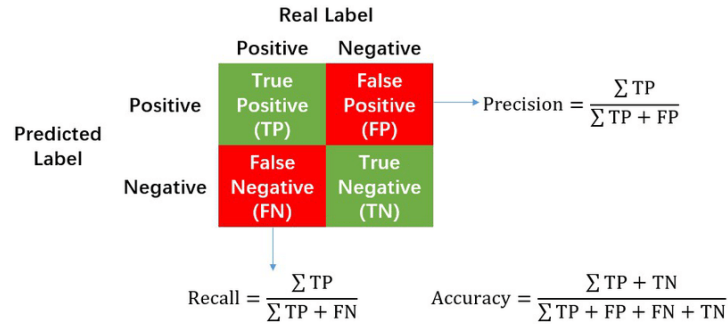


Figura 5: Matriz de confusión. Tomado de [29]

2.8.2. EXACTITUD (ACCURACY)

La exactitud mide qué tan cerca está el resultado del valor verdadero, y ofrece información sobre los errores generales en la clasificación. Matemáticamente, se define como la suma de las predicciones verdadera (VP + VN) dividida por el número total de predicciones realizadas. Es fundamental para evaluar si los modelos obtienen resultados correctos de forma global [28].

$$ACC = \frac{VP + VN}{VP + VN + FP + FN}$$

2.8.3. PRECISIÓN (PRECISION)

La precisión mide el porcentaje de predicciones positivas que fueron correctas. Se asegura que el clasificador no etiquete una muestra de clase negativa como si fuera positiva, evitando falsos positivos. Se calcula dividiendo las predicciones verdadera positiva entre la suma de predicciones positivas, tanto verdaderas como falsas [30].

$$PRE = \frac{VP}{VP + FP}$$

2.8.4. EXHAUSTIDAD (RECALL)

También se conoce como tasa de verdaderos positivos. Mide la capacidad del

modelo de detectar patrones positivos. Calcula el porcentaje de casos positivos reales que el clasificador identificó correctamente. Se obtiene dividiendo la predicción verdadera positiva entre la suma de la predicción verdadera positiva y la falsa negativa [30].

$$REC = \frac{VP}{FN + VP}$$

2.8.5. F1 SCORE

El F1 Score se define como un promedio entre la Precision y el Recall. Asigna un valor de 1 al mejor resultado posible y 0 al peor, y es utilizado para obtener un balance aceptable entre las dos métricas, siendo un valor promedio del 87% considerado como un rendimiento aceptable [28].

$$F1 = 2 \frac{PRE * REC}{PRE + REC}$$

2.8.6. CONFIANZA

La confianza es un valor, generalmente entre 0 y 1, que indica qué tan seguro está el modelo de que una predicción es correcta. Se define a través de un umbral de confianza; las predicciones por debajo de este umbral se descartan. Si se aumenta dicho umbral, la precisión suele subir, es decir, el modelo solo reporta lo que está “muy seguro”, pero el recall baja, pues se omiten casos reales de los que el modelo no estaba tan seguro [31].

2.8.7. ENTROPÍA

La entropía es una medida de disimilitud y optimización, que el contexto de las redes neuronales se utiliza principalmente mediante la entropía cruzada (*Cross-Entropy*), que mide la disimilitud, o mejor dicho la falta de semejanza, entre la distribución de probabilidad predicha por el modelo y la distribución real (etiquetas). Se usa principalmente como una función de pérdida durante el entrenamiento, pero se puede entender como una métrica más cuyo objetivo es minimizarse pues cuanto menor sea su valor, más certero se está de que esa predicción es correcta [32].

2.9. ANTECEDENTES

La ecografía es una herramienta clave en la práctica médica por su bajo costo, seguridad y accesibilidad. Sin embargo, la segmentación e interpretación automática de imágenes ecográficas sigue representando un reto técnico debido a la baja calidad de la señal respecto al ruido, presencia de artefactos de sombra y bordes anatómicos ambiguos. Estas dificultades se acentúan en imágenes craneales, donde la identificación precisa del plano anatómico es esencial para el diagnóstico neurológico fetal.

Recientemente, modelos de aprendizaje profundo han mostrado avances notables

en la segmentación automática de estructuras fetales en ecografías, como la circunferencia cefálica. Métodos como V-Net con atención supervisada profunda y Redes Convolucionales ligeras han logrado resultados de alta precisión en tareas biométricas, como la medición automática de la cabeza fetal, alcanzando coeficientes de similitud (Dice) superiores al 97% [33], [34].

Por otra parte, el modelo Segment Anything (SAM) de Meta AI ha demostrado un rendimiento prometedor para segmentar imágenes médicas de manera flexible y con mínima intervención humana. Aunque originalmente desarrollado para tareas generales de visión artificial, su capacidad para segmentar estructuras tisulares en imágenes ecográficas lo posiciona como una herramienta de alto potencial en el ámbito médico, especialmente cuando se adapta a las particularidades de las ecografías craneales [35].

El aporte de los anteriores modelos y estudios ha demostrado la fiabilidad del aprendizaje profundo para la segmentación y medición de estructuras biométricas en ecografías. Sin embargo, gran parte de estos enfoques se centra en la medición una vez que el plano anatómico había sido identificado, lo cual recae en la experiencia y habilidad manual de radiólogo.

A pesar de estos avances, la identificación automática del plano anatómico correspondiente en imágenes de ultrasonido craneal sigue siendo un área poco explorada. La correcta asociación entre imagen y plano es fundamental para el análisis clínico, ya que permite interpretar de forma contextual la información visual. Esta carencia representa una mejora para el incremento de los sistemas de aprendizaje profundo que no solo segmenten estructuras, sino que también clasifiquen automáticamente el plano anatómico de cada imagen, mejorando así la precisión diagnóstica y reduciendo la dependencia de operadores expertos.

En contraste con estos enfoques, después de la identificación manual, la aplicación de inteligencia artificial ha demostrado ser eficiente para el diagnóstico de diversas enfermedades y anomalías durante el embarazo, y en algunos casos, para la identificación de planos anatómicos. Un ejemplo se encuentra en la ecocardiografía, donde se ha demostrado la capacidad de entrenar redes convolucionales para identificar y clasificar hasta 15 cortes o planos diferentes con una seguridad cercana al 98% [4]. Aunque el estudio se centra en planos cardíacos, el éxito obtenido sugiere que técnicas similares basadas en aprendizaje profundo pueden extrapolarse a otras modalidades de ultrasonido, como las ecografías craneales fetales.

El ultrasonido fetal es una modalidad altamente dependiente del operador y susceptible a variaciones en la calidad de adquisición. Factores como el movimiento fetal, la posición de la madre, la impedancia acústica y las diferencias entre equipos provocan imágenes con bajo contraste, ruido speckle, artefactos de sombra y regiones poco definidas. Por ello, antes de aplicar modelos de aprendizaje profundo, es imprescindible realizar un conjunto de técnicas de preprocesamiento destinadas a estandarizar la imagen y mejorar la detección de patrones anatómicos. Estudios recientes coinciden en que una adecuada fase de preprocesamiento influye de manera directa en el rendimiento final de los modelos de clasificación [19].

3. REVISIÓN SISTEMÁTICA

3.1. ESTRATEGIA DE BÚSQUEDA Y CRITERIOS DE EXCLUSIÓN

Con el propósito de construir un corpus bibliográfico actualizado y relevante, se llevó a cabo una búsqueda sistemática en las siguientes bases de datos científicas de amplia cobertura: Scopus, IEEE Explore, Oxford Academic Web of Science.

La ecuación de búsqueda empleada fue:

("identificación automática" OR "clasificación automática" OR "análisis automático") AND ("ecografía" OR "ultrasonido" OR "sonografía") AND ("imágenes")

3.1.1. CRITERIOS DE EXCLUSIÓN

Con el fin de delimitar adecuadamente el conjunto de estudios incluidos y garantizar la pertinencia de estos en relación con los objetivos de la investigación, se establecieron los siguientes criterios de exclusión: Se excluyeron artículos de revisión, guías clínicas, editoriales, resúmenes de congresos y otros documentos que no correspondieran a investigaciones originales con resultados empíricos. Solo se consideraron artículos de investigación originales. Se descartaron todos los estudios publicados antes del año 2022, con el fin de asegurar la vigencia de la información y relevancia de la información recopilada.

Además, se eliminaron los artículos cuya temática principal correspondiera a áreas del conocimiento no relacionadas con el objetivo de la investigación, tales como: neurociencias, neurología, metabolismo endocrinológico, instrumentos e instrumentación, medicina experimental, física, biología reproductiva, robótica, cirugía, gastroenterología y hepatología, tecnología de laboratorio médico, óptica, ciencias de los materiales, urología y nefrología, biología a nivel molecular y bioquímica, hematología, inmunología, farmacia y farmacología, psicología, audiolología, patología del habla y lenguaje, odontología, cirugía oral y medicina, mecánica, telecomunicaciones, oncología, cardiología y sistema cardiovascular.

3.1.2. MÉTODO PRISMA

La **Figura 6** representa el flujo metodológico de selección de artículos mediante el método PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). Este diagrama resume el proceso de identificación, cribado, elegibilidad e inclusión, evidenciando el número de registros inicialmente obtenidos, los excluidos en cada etapa y el conjunto final de estudios que conformaron la base documental utilizada en esta investigación.

Este método utiliza un diagrama de flujo que resume las distintas etapas del proceso de búsqueda bibliográfica, mostrando de manera clara cuántos registros fueron

identificados en las distintas bases de datos, cuáles fueron eliminados por criterios de exclusión (duplicados, temática no pertinente, población diferente, entre otros) y el número final de estudios que fueron incluidos en el análisis

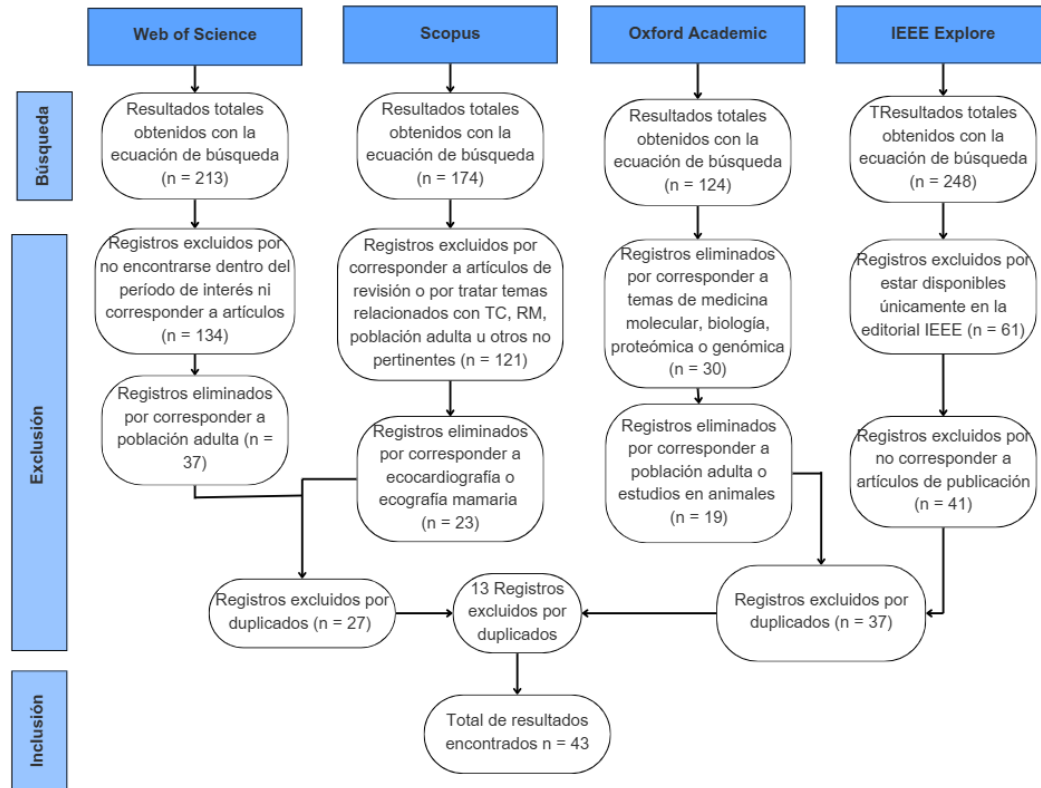


Figura 6: Flujo de selección de artículos según diagrama PRISMA

A partir de la revisión sistemática se obtienen múltiples conceptos, enfoques y resultados que permiten desarrollar una visión más completa sobre las estrategias metodológicas a implementar. Este análisis no solo facilita la comprensión de los desarrollos actuales en la aplicación de inteligencia artificial al ultrasonido fetal, sino que también orienta la definición del paso a paso que se seguirá en la tesis para abordar los objetivos propuestos. De esta manera, la información obtenida sirve como base sólida para la toma de resoluciones técnicas y metodológicas en el desarrollo del modelo planteado.

La presente revisión sistemática recopila y analiza estudios recientes enfocados en el progreso de modelos de inteligencia artificial aplicados al análisis de imágenes de ultrasonido fetal. Con el objetivo de identificar las metodologías más utilizadas, las estrategias de preprocesamiento de datos, las métricas de validación y los resultados obtenidos, se elaboró una tabla comparativa que sintetiza la información más relevante de cada trabajo, **Tabla 1**. Esta recopilación aporta una perspectiva integral del estado del arte en tareas de detección, segmentación y clasificación de planos anatómicos fetales, abarcando desde redes convolucionales clásicas como CNN, U-Net y ResNet, hasta

enfoques de mayor complejidad basados en aprendizaje multitarea, mecanismos de atención, grafos clínicos y redes generativas adversarias. La estructura de la tabla permite reconocer tendencias metodológicas y técnicas, evidenciando los principales aportes, limitaciones y logros de cada estudio. En conjunto, este análisis proporciona una guía fundamental para la definición de la metodología propuesta en el presente trabajo, al facilitar la selección de arquitecturas, técnicas de procesamiento y criterios de evaluación acordes con los avances recientes en el procesamiento automático de imágenes de ultrasonido fetal.

Tabla 1: Comparativa de métodos de inteligencia artificial en imágenes de ultrasonido fetal

Título	Método / Red usada	Preprocesamiento	Métricas de validación	Resultados cuantitativos	Resultados cualitativos	Importante
An ultrasound standard plane detection model of fetal head (2022)	Multi-task learning + grafo clínico; CNN multitarea	Normalización, anotación de 4 planos, augmentación (rotación, escala)	Accuracy, Sensibilidad, Especificidad	Mejor que baseline	Mapas de atención; robusto a ruido y rotación	Modelo interpretable con conocimiento clínico
Automated deep bottleneck residual 82-layered architecture (2023)	CNN residual 82 capas + Bayesian Optimization + Moth Flame Optimization	Resize, normalización, augmentación (rotación, traslación)	Accuracy, F1-score	$Accuracy \approx 78-79\%$	Mejor separación entre planos cerebrales	Optimización automática de hiperparámetros
Automated Detection of Common Maternal Fetal Ultrasound Planes (2022)	Transfer learning (ResNet-50 + VGG-19-GAP) + SVM	Normalización, resize, augmentación (rotación, flip, escalado)	Accuracy, Recall, Precision, F1	Superó SOTA en 6 planos	Robustez a ruido y movimiento fetal	Reduce tiempo de diagnóstico
Automatic measurement of fetal head circumference (2022)	GAC-Net: CNN encoder-decoder + GCN + atención	Normalización, supresión de ruido, realce de bordes, dataset HC18	Dice, HD, AD, DF	$DSC = 98.21\%$, $HD = 1.22 \text{ mm}$	Mejor en bordes incompletos y sombras	Robusto a ruido y estructuras similares
ECAU-Net: Efficient channel attention U-Net (2022)	U-Net + ECA (27 parámetros extra)	Dataset JSUAH-Cerebellum, blur gaussiano, flips aleatorios	Dice, Jaccard	$Dice = 91.35\%$, $Jaccard = 86.01\%$	Segmentaciones limpias en baja calidad	Optimizado para datasets pequeños
Automatic Analysis of Transverse Musculoskeletal US (2023)	MMA-Net: U-Net + VGG, multitask	1.827 imágenes, <i>split</i> 80/10/10, normalización	Dice, IoU, Accuracy, Recall	$Dice = 96.74\%$, $IoU = 94.10\%$	Preciso incluso con ruido	Reduce carga diagnóstica muscular
Automatic quality assessment for 2D fetal sonographic standard plane (2021)	CNN multitarea (FEN, CPN, RPN)	Normalización, anotación de estructuras clave	Accuracy, Precision, Recall, IoU	Evaluación en $< 1 \text{ s}$	Diferencia clara entre planos estándar/no estándar	Automatiza control de calidad

A Review of Image Processing Methods for Fetal Head Analysis (Review)	Revisión: U-Net, CNNs, contornos activos, Hough	Filtrado de ruido, normalización, realce de bordes	—	—	Analiza ventajas y limitaciones	Base teórica
Unleashing the Power of Self-Supervised Image Denoising (2024)	Revisión: U-Net, BSN, Transformers	Generación de pares ruidosos sintéticos	PSNR, SSIM	Resume performance	Evalúa robustez de métodos de denoising	Relevante para preprocesamiento en US
An Ultrasound Fetus Phantom Dataset	Dataset phantom fetal	Filtros gaussiano/mediano, normalización gris, ajuste contraste	—	—	Imágenes limpias y anotadas	Entrenamiento sin datos clínicos
FCRB U-Net: Fully Connected Residual Block U-Net	U-Net + FCRB + ECA + FRM	Augmentation: flips, recortes, ruido gaussiano	mIoU, Dice, Precision, Recall, F1	+4.16% Dice vs nnU-Net	$fps \approx 198$, tiempo real	Segmentación rápida y precisa de cerebelo
An Ultrasound Fetus Phantom Dataset With DNN Evaluations	Dataset phantom fetal evaluado con CNNs	Adquisición controlada y anotación	Accuracy, Sensibilidad, Especificidad	Benchmarks iniciales	Robustez en orientaciones y planos	Primer dataset phantom etiquetado rico
Exploiting CAMs as Prior to Generate Fetal Brain US with GANs	GAN + Class Activation Maps	Extracción de CAMs como priors	FID, SSIM, IS	Mejora calidad de imágenes	Imágenes más anatómicamente plausibles	Útil para aumentar datos
Towards Standard Plane Prediction with Domain Adaptation	CNN con adaptación adversarial	Datos de distintos centros y sondas	Accuracy, F1	Mejor en dominio externo	Generaliza en dominios distintos	Predicción robusta multi-centro

La revisión sistemática se desarrolló mediante una búsqueda estructurada y una lectura crítica de los artículos preseleccionados. Para la evaluación de calidad, se aplicó un criterio de análisis documental basado en la claridad del resumen, la definición del problema en la introducción, la solidez de la discusión y la consistencia de las conclusiones con los resultados reportados. Con ello, se aseguró que los estudios incluidos presentaran pertinencia temática, coherencia metodológica y aporte directo al objetivo del proyecto.

4. ADQUISICIÓN, CONFORMACIÓN Y PREPROCESAMIENTO DE IMÁGENES

Este capítulo detalla la metodología utilizada para crear y preparar el conjunto de datos de ecografías craneales fetales, para asegurar la calidad y el tamaño adecuado con el fin de entrenar las arquitecturas de aprendizaje profundo.

4.1. JUSTIFICACIÓN

La identificación de los tres planos anatómicos de cabeza es un requisito fundamental para el diagnóstico prenatal, ya que, a través de una imagen de ultrasonido, es posible visualizar la estructura cerebral y diagnosticar anomalías durante la gestación de manera estandarizada [36]. Contar con un dataset amplio y correctamente etiquetado es un requisito de importancia primordial para utilizar modelos de Deep Learning (DL).

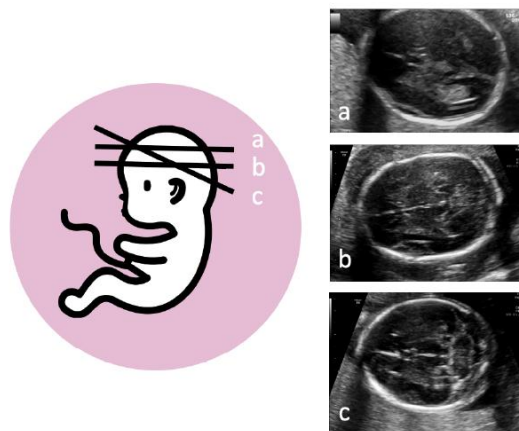


Figura 7: Planos craneales fetales: (a) trans-ventricular, (b) trans-talámico y (c) trans-cerebelar.
Tomado de [37]

Estos tres planos anatómicos, mostrados en la **Figura 7**, son:

- a. **Plano Trans-talámico (TT):** Utilizado para la estimación de la circunferencia cefálica y el diámetro biparietal, el cual permite también analizar el Cavum Septum Pellucidum (CSP) y el Tálamo (Th) [38].
- b. **Plano Trans-ventricular (TV):** Esencial para la visualización y medición de los ventrículos laterales (LV), así como la evaluación del CSP, y para la segmentación de estructuras internas [33], [38].
- c. **Plano Trans-cerebelar (TC):** Permite evaluar el cerebelo (Ce) y la Cisterna Magna (CM). Con ese plano se mide el diámetro del Ce [38], [39].

Obtener diversas imágenes de estos planos ayudará a que los modelos de inteligencia artificial puedan ser entrenados de manera adecuada, con una buena

generalización y puedan superar los desafíos inherentes a las ecografías fetales y así lograr una precisión comparable a la de los expertos. Esto implica que el conjunto de datos supere un par de desafíos, como lo son: la baja calidad de las imágenes y el ruido inherentes en estas, por lo que el dataset debe ser lo suficientemente amplio para que los modelos extraigan características a pesar de posible distorsiones, como sombras o ruido de los artefactos; el alta/baja variación intraclase, pues algunas imágenes de planos estándar pueden ser similares a los planos no estándar, con un dataset amplio se logra una excelente generalización de las estructuras; y, finalmente, la dependencia con el operador.

Los modelos de aprendizaje profundo tienen una dependencia en los datos y requieren una base de entrenamiento amplia y robusta para evitar sobreajuste y garantizar la transferibilidad del conocimiento [40]. Por lo tanto, los modelos necesitan muchas imágenes para aprender y etiquetas meticulosas, contando con el ground truth (verificación de expertos médicos) para garantizar la alta fiabilidad de las anotaciones, mediante Coeficientes de Correlación Intraclase (ICC) y el índice de Similitud Jaccard (JS).

4.2. ADQUISICIÓN DE DATOS

4.2.1. FUENTE PRINCIPAL

La fuente principal para usar se trata del dataset llamado “Large-Scale Annotation Dataset for Fetal Head Biometry in Ultrasound Images” que contiene 3.832 imágenes de ultrasonido de alta resolución, diseñadas para la investigación y diagnóstico médico. Es un recurso guiado al desarrollo y validación de algoritmos de inteligencia artificial sobre biometría fetal (es decir, medición del tamaño y crecimiento del feto, y detección de posibles complicaciones) crucial en el control del embarazo.

El conjunto de datos fue constituido por dos datasets, uno proveniente de BCNatal (Dataset A) que originalmente contenía imágenes para clasificar planos fetales. Se extrajo un subconjunto de 12.400 imágenes que mostraban tres planos de cabeza fetal: Trans-cerebelo (TC), Trans-talámico (TT) y Trans-ventricular (TV), y otro (Dataset B) que consta de 999 imágenes de cabeza fetal con sus máscaras, creada inicialmente para segmentación. Para el conjunto final, las 999 imágenes fueron re-anotadas, es decir, la etiqueta previa fue revisada y corregida para mejorar su consistencia.

Todas las imágenes fueron anonimizadas y desidentificadas previamente para mantener los estándares éticos. El conjunto está disponible como *open access* bajo la licencia CC BY y puede descargarse desde el repositorio Zenodo [41]. Se enfatiza en tres planos craneales estándar, con la adición de diversas imágenes de cabeza fetal anotadas como Diverse-Fetal-Head-Images.

Además, se cuenta con un segundo dataset del cual fue reproducido el anterior, llamado: “FETAL_PLANES_DB: Common maternal-fetal ultrasound images” [42] (Dataset A) que cuenta con cinco etiquetas, tres de las cuales son las ya mencionadas en el anterior, más la adición de una quinta rotulada como Not-A-Brain, que corresponde a las imágenes que no pertenecen a la clase principal (como las clases Abdomen (711), Femur (1.040),

Thorax (1.718), Maternal Cervix (1.626) y Other (4.213)), que son planos no cerebrales. Por lo tanto, se pretende hacer una unificación de ambos dataset para obtener uno final que posee las siguientes clases:

Tabla 2: Etiquetas en la agrupación de conjuntos de datos

Etiqueta	Número de imágenes	Dataset
Not-A-Brain	9.308	A
Trans-thalamic	1.638	A
Diverse-Fetal-Head-Images	999	B
Trans-cerebellum	714	A
Trans-ventricular	597	A
Other	143	A
Total	13.399	

Como se puede ver en la **Tabla 2**, a la hora de verificar las etiquetas de los diferentes conjuntos de datos, se agregó una sexta llamada Other, que corresponde a imágenes no correspondientes a los tres planos anatómicos cerebrales. Esta etiqueta, por ser la clase minoritaria, va a ser eliminada, quedándonos con las cinco primeras, con las que obtendremos una excelente generalización de resultados, puesto que no solo podremos clasificar una imagen en si corresponde o no a los tres planos principales, sino a si pertenece a un plano cerebral no definido (etiqueta Diverse-Fetal-Head-Images) o a planos no cerebrales (etiqueta Not-A-Brain). De esta forma, el modelo va a ser robusto para diversas aplicaciones posteriores.

4.2.2. PROTOCOLO DEL ETIQUETADO

Según [33], donde los autores crearon un modelo llamado FetSAM para la segmentación de la cabeza fetal en ecografías, el dataset original se dividió en un 60% para entrenamiento, 20% para validación y 20% para prueba. Se aplicaron “estrategias de aumento de datos exhaustivos” al conjunto de entrenamiento para expandir el conjunto original hasta 20.691 imágenes.

Las imágenes tienen una dimensión de 959x611 píxeles. La calidad de las anotaciones fue verificada a través de un proceso de validación con un médico y un tecnólogo radiológico, confirmada a través de altos ICC e índices de JS. Como el objetivo del conjunto no solo es el de clasificar, sino también de segmentar, se resaltaron estructuras clave en biometría y diagnóstico, como el cerebro, el CSP y los Ventrículos Lateral (LV).

El proceso de anotado fue estructurado y guiado por expertos, lo que asegura la transparencia y precisión de las etiquetas. Un estudiante de doctorado asistió a sesiones de enseñanza con un médico adjunto senior con 12 años de experiencia materno-fetal para mejorar la comprensión de las estructuras. Se revisaron las anotaciones iniciales

realizadas por el estudiante de doctorado, clase por clase, y el médico adjunto proporcionó la retroalimentación necesaria. Después, el estudiante volvió a revisarlas. Se evaluó la fiabilidad a través del ICC y JS.

En una segunda ronda, un tecnólogo radiológico experimentado revisó una muestra aleatoria del 20% del dataset y ofreció retroalimentación. El estudiante reajustó aún más las anotaciones.

Se volvió a evaluar la fiabilidad entre el estudiante y el radiólogo.

4.3. ESTRATEGIAS PARA LA AMPLIACIÓN DEL DATASET

La distribución de clases de las etiquetas no está balanceada. De las 13.256 imágenes del conjunto de datos final, y de las 5 etiquetas brindadas, se tiene la siguiente distribución, **Figura 8**.

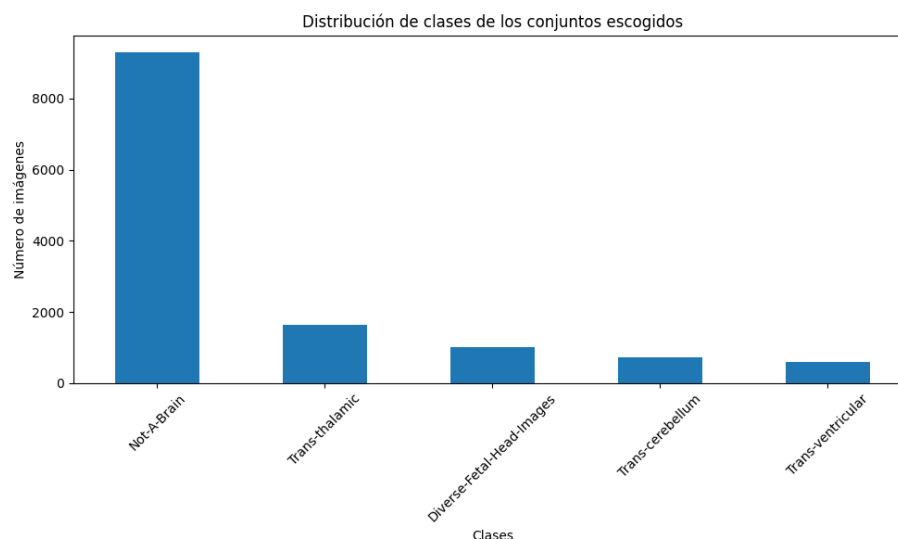


Figura 8: Distribución de clases del dataset final

De entre las etiquetas, las imágenes de planos no cerebrales son las que tiene mayor cantidad de imágenes, un total de 9.308, seguido por la del plano trans-talámico con 1.638; las imágenes de diversas cabezas fetales se encuentran de terceras con 999, seguidas por el plano trans-cerebelo con 714 y, por último, el trans-ventricular con 597. Debido a este desbalance, es necesario adoptar alguna estrategia para solucionarlo. Cuando hay clases con pocas muestras, la clasificación puede favorecer a la clase mayoritaria, lo que podría generar que una imagen fetal sea mal clasificada.

Hay muchas estrategias para el manejo de datos desbalanceados. Los modelos de clasificación pueden tomar un parámetro específico durante el entrenamiento, llamado "class_weight" para balancearlo. El problema es que no todos los algoritmos tienen estas posibilidades. Otro criterio sería reducir la clase mayoritaria, que en este caso es esencial para la clase Not-A-Brain, que es más del triple de la suma de las otras clases. Así que disminuiríamos la cantidad de imágenes a través de una estrategia de undersampling, en el que reduciríamos la cantidad de imágenes a la cantidad de la segunda clase mayoritaria,

así como se observa en la **Figura 9**.

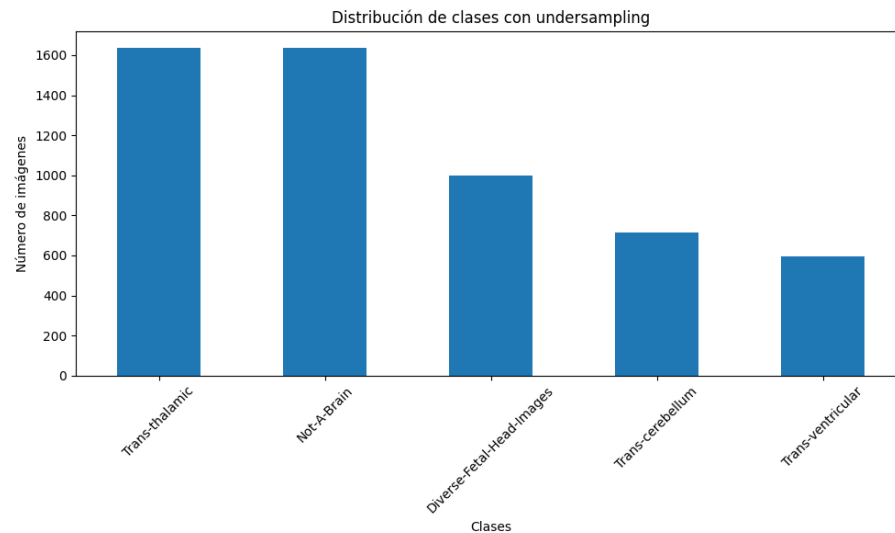


Figura 9: Undersampling de la clase mayoritaria

Después, la estrategia final sería crear “muestras sintéticas” utilizando algoritmos que diversifiquen a las clases minoritarias. Según [43], el aumento de datos es un grupo de técnicas para generar datos artificiales de alta calidad mediante la manipulación de los datos existentes, que busca combatir la cantidad limitada de datos disponibles, mejorar la generalización del modelo y empujar la distribución de la muestra hacia la verdadera distribución. Esta estrategia resuelve el problema causado por la escasez de datos y el desequilibrio, además de aumentar la diversidad durante el entrenamiento.

Se pueden separar en dos categorías: **i)** La transformación basada en valores o **ii)** la basada en estructura. Para imágenes médicas es más común la segunda transformación, donde generalmente se realizan recortes aleatorios (Random crop), translación, volteo/reflexión y cizallamiento (Shearing).

En la **Tabla 3** se muestran las transformaciones escogidas, con sus respectivos valores de probabilidad de aplicación.

Tabla 3: Transformaciones de imágenes de cabeza fetal para Data Aumenting

Parámetro	Transformación	Concepto
$rotation_range = 15$	Rotación aleatoria	Se aplica un ángulo de rotación aleatoria de hasta 15 grados.
$width_shift_range = 0.2$	Desplazamiento horizontal (Traslación)	Se desplaza la imagen horizontalmente de forma aleatoria hasta un 20% del ancho total.
$height_shift_range = 0.2$	Desplazamiento vertical (Traslación)	Se desplaza la imagen verticalmente de forma aleatoria hasta un 20% del alto total.

<i>shear_range = 0.2</i>	Cizallamiento (Shear)	Se aplica una transformación de cizallamiento.
<i>zoom_range = 0.2</i>	Zoom aleatorio	Aplica un factor de zoom aleatorio, lo que ayuda a que la red sea más robusta ante tamaños de objeto variables
<i>horizontal_flip = True</i>	Volteo horizontal (Espejo)	Se realiza un volteo de espejo a lo largo del eje vertical.
<i>fill_mode = 'reflect'</i>	Modo de relleno: Reflejo	Se especifica cómo se rellenan los píxeles recién creados después de rotaciones o desplazamientos. Esto ayuda a mantener la consistencia de la imagen
<i>brightness_range = (0.8, 1.2)</i>	Ajuste de brillo aleatorio	Se ajusta el brillo de la imagen en un rango que va del 80% al 120% del original.
<i>preprocessing_function = preprocess_input</i>	Función de preprocesamiento	El preprocesamiento de imágenes generalmente incluye operaciones como la normalización o la resta de la media de píxeles.

Estas técnicas son esenciales para ultrasonidos. Así como sucede en [44], estas técnicas aumentaron el conjunto de entrenamiento del conjunto B de 999 a 10.989 imágenes. Otras estrategias fueron filtros de desenfoque Gaussiano o la adición de ruido para simular las variaciones provocadas por los artefactos. Este proceso va guiado a la siguiente sección.

4.4. PREPROCESAMIENTO DE IMÁGENES (PIPELINE)

La fase de preprocesamiento en el conjunto de datos es crucial, puesto que para entrenar modelos de aprendizaje profundo se necesita una entrada que aborde los desafíos inherentes a las imágenes médicas, como son el ruido, el bajo contraste y la alta variabilidad [33].

Esta limpieza de las imágenes médicas se centra en mejorar su calidad. Usualmente, se usan filtros de suavizado, como el filtrado Gaussiano, para suavizar la imagen original y mitigar la intermitencia de los artefactos, como el ruido speckle, normalización y redimensionamiento respecto a la entrada de la arquitectura deseada [45].

Primera imagen

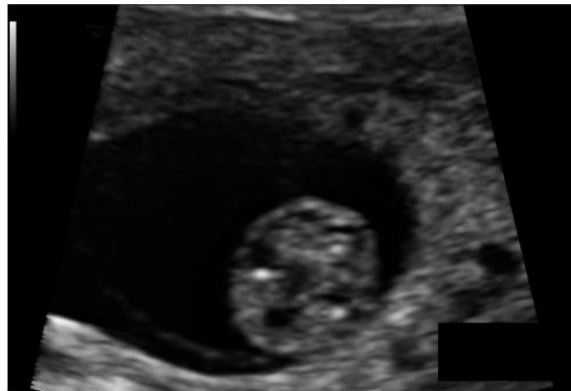


Figura 10: Primera imagen del dataset

Las imágenes del conjunto de datos tienen un tamaño de 959x661 píxeles en su mayoría, por lo que un redimensionamiento es importante, **Figura 10**. Además, pese a que a simple vista es difícil considerar si existe ruido u otros defectos, es importante tener un módulo de preprocesamiento. Generalmente, se requiere eliminar las distorsiones no deseadas y así obtener una vista clara de las características de la imagen; esto es debido a que las imágenes son borrosas y tiene variaciones en el tamaño de los píxeles. Esta fase ayuda a revelar elementos ocultos.

Los pasos más comunes son: Normalización de la resolución, por ejemplo, de 0 – 255 a 0 – 1, esencial para estandarizar intensidades y para un entrenamiento fiable. Redimensionamiento de las imágenes a una resolución estándar, por ejemplo, 512x512 píxeles o 256x256 píxeles, que simplifica la carga computacional del modelo y asegura la consistencia y preserva la información crucial y la forma original de la anatomía. Y la aplicación de algunos filtros, como aquellos que reducen los ruidos.

4.4.1. NORMALIZACIÓN

Este proceso es esencial para preparar las imágenes de ultrasonido crudas, que a menudo presentan ruido, bajo contraste y variabilidad, **Figura 11**. A través de una buena normalización se estandariza los valores de brillo y colores de los píxeles. La técnica escogida es la normalización Min-Max (o mínimo-máximo) del paquete OpenCV (*cv2*). Se utiliza para llevar los valores de intensidad de una matriz (en este caso, una imagen) a un rango deseado. Se especifica un valor mínimo *alpha* = 0 y un valor máximo *beta* = 1, y a través de la función *norm_type = cv2.NORM_MINMAX* se reescala linealmente los valores entre los límites escogidos. De esta forma aseguramos que los datos de entradas para los modelos sean consistentes. Además, garantiza la convergencia al entregar a la red entradas con distribuciones de valores adecuadas.



Figura 11: Comparación en paralelo de imagen original y normalizada

4.4.2. REDUCCIÓN DE RUIDO

Para mitigar los efectos de los artefactos en las imágenes, se aplican un filtro Gaussiano. Los ultrasonidos tienen ruido inherente, lo que resulta en una baja relación señal/ruido (SNR), que las vuelve borrosas [46]. El bajo contraste puede ocultar anatomías relevantes, como partes del cráneo, lo que complica la clasificación [44]. Por lo tanto, es un proceso vital para extraer información significativa.

El filtro Gaussiano se utiliza para suavizar la imagen original, **Figura 12**. Sin embargo, hay que tener en cuenta que según [47], este filtro es beneficioso en combinación, puesto que cuando se aplica solo, los índices de evaluación se reducen.

Se utiliza la función de la librería de OpenCV *cv2.GaussianBlur*.



Figura 12: Comparación en paralelo de imagen normalizada y filtrada

4.4.3. REDIMENSIONAMIENTO

Los diferentes modelos, como las arquitecturas VGG y ResNet (para clasificación), emplean un tamaño estándar que facilita la extracción de características profundas. Las arquitecturas para segmentación utilizan una dimensión de 244×244 píxeles. Para redes convolucionales una elección de 256×256 (así como otras como 224×224 o 512×512) facilita los procesos internos de convolución, puesto que ofrece un equilibrio entre la precisión y la eficiencia computacional y de memoria, **Figura 13**.

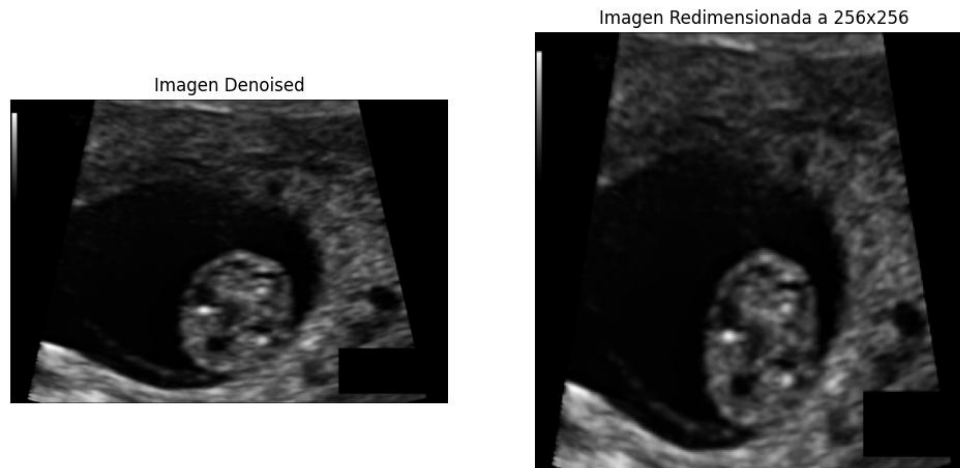


Figura 13: Comparación paralela de imagen filtrada y redimensionada

4.5. ESTRUCTURA FINAL

4.5.1. DIVISIÓN DEL DATASET

Antes de preprocesar y balancear el dataset, se debe dividirlo en los diferentes conjuntos de datos a usar. En la mayoría de las ocasiones, una división 70/30 de conjunto de entrenamiento (train) y validación (val) es suficiente, pero se harán pruebas finales con el prototipo, por lo que también se creará un conjunto de prueba (test). Como las imágenes vendrán de los dos conjuntos de datos ya establecidos, se deberá estructurar una carpeta con los diferentes conjuntos respecto a una división 70/20/10 para los conjuntos de entrenamiento, validación y prueba respectivamente, **Figura 14**.

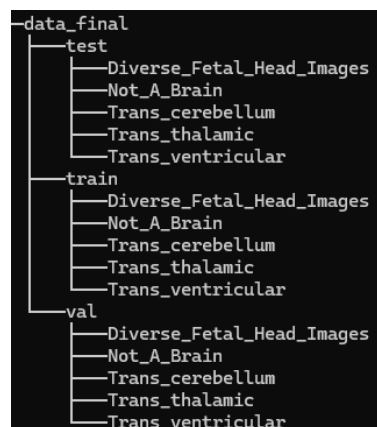


Figura 14: Estructura del dataset con la división 70/20/10

Después de esta partición, es necesario distinguir cuáles son las reglas para cada uno de los conjuntos. Según [48], las operaciones de preprocesamiento (como el centrado de media, escalado o normalización) deben ser consistentes entre todos los conjuntos. En

el caso del aumento de datos, en donde se busca introducir diversidad artificial en las muestras, se usa para ayudar al modelo a aprender características más robustas y debe usarse únicamente en el conjunto de entrenamiento [49]. Por ello, los conjuntos de validación y prueba deben ser representativos de la "población de interés" y de las circunstancias reales que el modelo encontrará en producción, por lo que no suelen estar aumentados [48].

Finalmente, debido al porcentaje de división del conjunto de datos, se obtiene un total de imágenes del conjunto de entrenamiento de 8.190, un conjunto de validación de 1.117 y uno de prueba de 559. La distribución de cada conjunto se muestra en **Figura 15**, **Figura 16** y **Figura 17**, respectivamente.

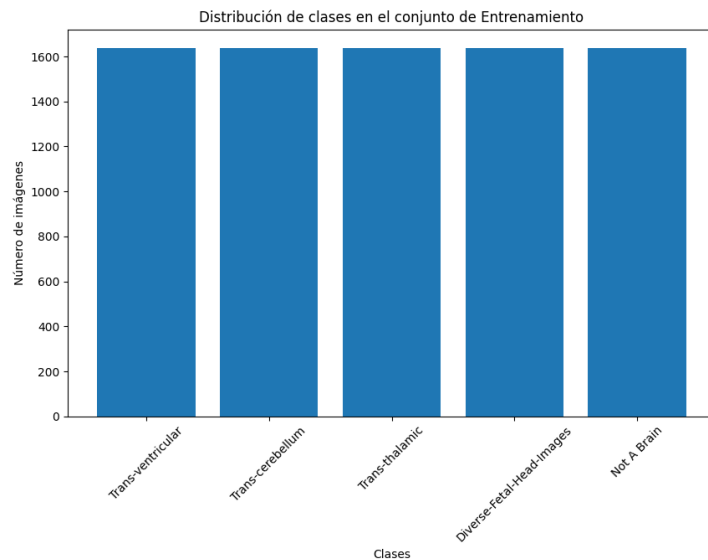


Figura 15: Distribución de clases en el conjunto de Entrenamiento

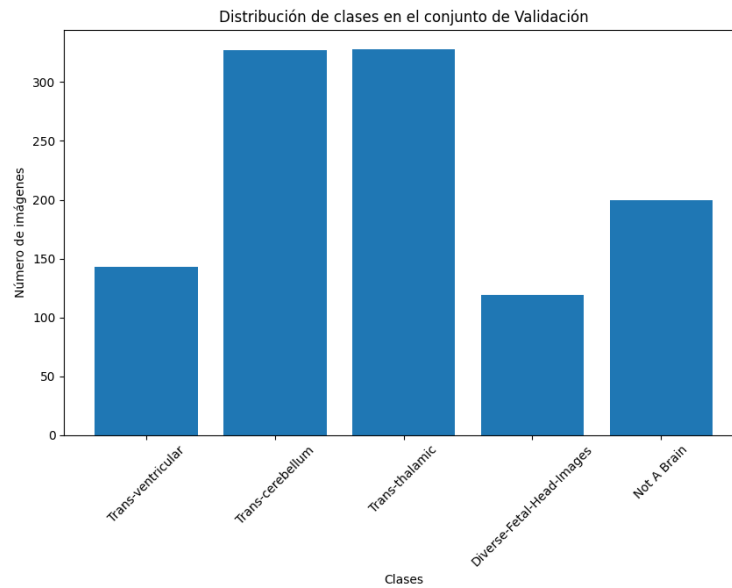


Figura 16: Distribución de clases en el conjunto de Validación

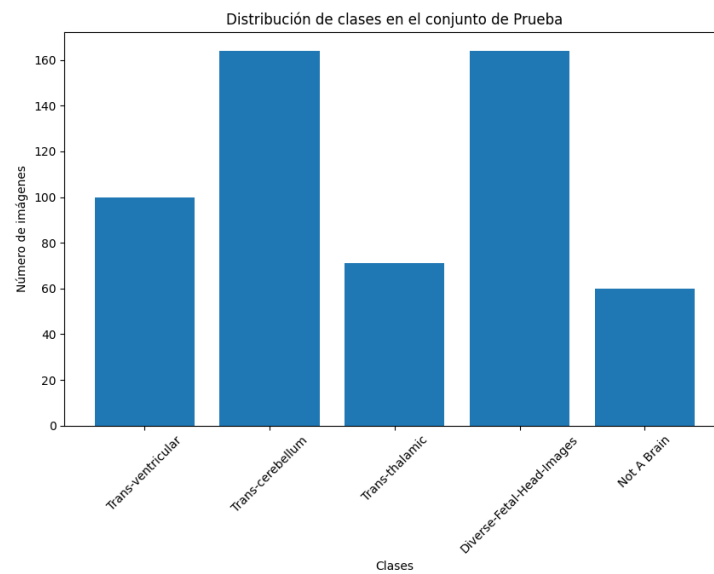


Figura 17: Distribución de clases en el conjunto de Prueba

Se muestran, además, imágenes de cada uno de los conjuntos de manera aleatoria de una de las etiquetas, **Figura 18**, **Figura 19** y **Figura 20**, en este caso la trans-talámica, en donde se podrá observar el proceso que se hizo en cada una. Todos los conjuntos fueron preprocesados, pero solo el conjunto de entrenamiento fue aumentado, evidente en las pequeñas transformaciones en algunas de ellas.

Ejemplos de imágenes preprocesadas y aumentadas: Entrenamiento - Transtalámicos

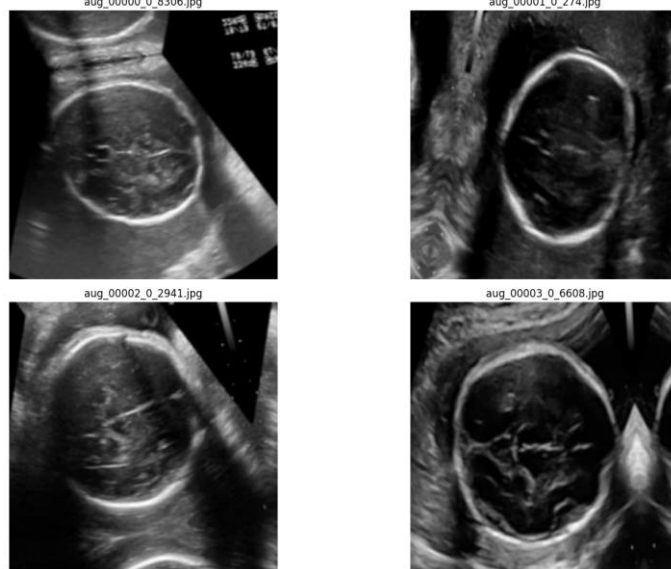


Figura 18: Imágenes del conjunto de Entrenamiento de la etiqueta Trans-talámica

Ejemplos de imágenes preprocesadas y aumentadas: Validación - Transtalámicos

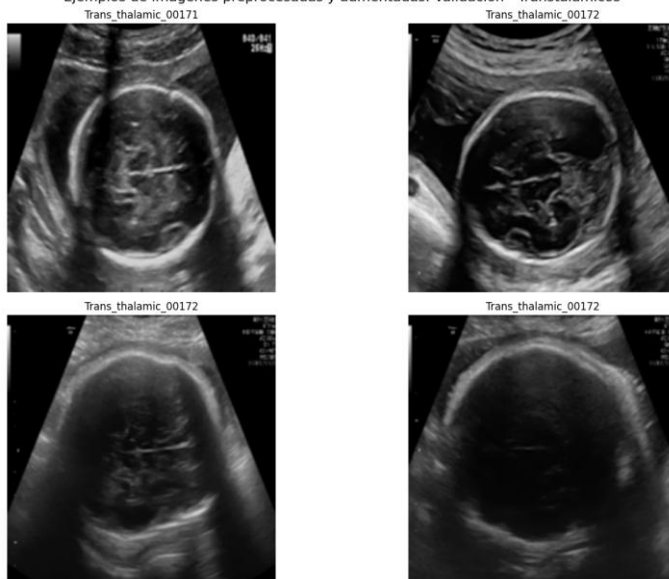


Figura 19: Imágenes del conjunto de Validación de la etiqueta Trans-talámica

Ejemplos de imágenes preprocesadas y aumentadas: Prueba - Transtalámicos

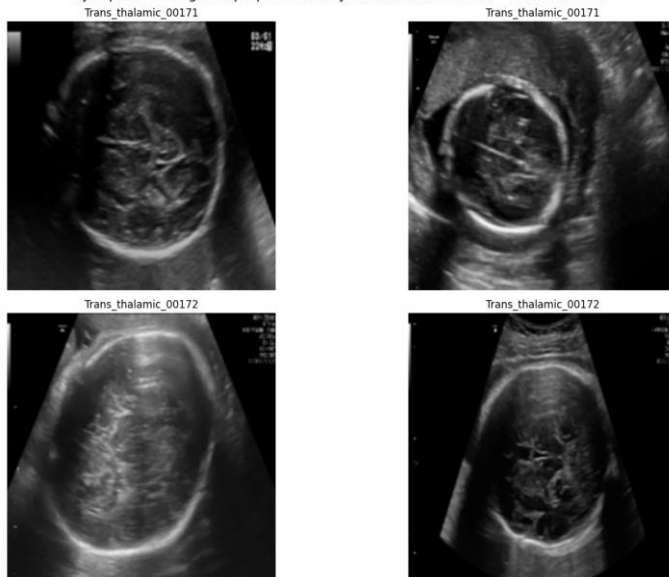


Figura 20: Imágenes del conjunto de Prueba de la etiqueta Trans-talámica

5. DESARROLLO Y ENTRENAMIENTO DE MODELOS

Este capítulo describe las arquitecturas de aprendizaje profundo seleccionadas y la metodología de entrenamiento empleada para clasificar automáticamente las ecografías craneales fetales en planos anatómicos estándar.

5.1. SELECCIÓN DE ARQUITECTURAS

Con ayuda de la revisión sistemática, se dio paso a la elección de las arquitecturas ResNet50 y VGG16 basado en la eficiencia de ambas para la extracción de características profundas y su capacidad para integrarse en aprendizaje a través de la transferencia de conocimiento (Transfer Learning).

La arquitectura VGG-Net (que podría ser VGG16 o VGG19) se selecciona por su capacidad para la extracción de atributos de imágenes mediante capas convolucionales con tamaño de kernel 3×3 [40], que permite sustraer representaciones sutiles y significativas, vitales para la tarea de clasificación. Los modelos VGG-Net son reconocidos por su rendimiento superior en la clasificación de planos de ultrasonido en comparación a otros modelos más profundos [46].

La arquitectura ResNet50 tiene una capacidad mucho más profunda y aprovecha el concepto de aprendizaje residual, que usa bloques residuales para facilitar el flujo de información, haciendo que sea una red más robusta.

Por último, la arquitectura final será una personalizada, en donde se buscará el mejor conjunto de capas para llegar a una clasificación de los planos anatómicos más acertados, explorando de esta forma técnicas que puedan ser más robustos y atractivos para la extracción de las características distintivas de las imágenes.

5.2. IMPLEMENTACIÓN DE ARQUITECTURAS

5.2.1. ARQUITECTURAS BASADAS EN TRANSFER LEARNING

La estrategia para las dos primeras arquitecturas es descongelar porcentajes de capas (también conocido como fine-tuning selectivo), apoyado en el marco del aprendizaje por transferencia. Con esta metodología, se buscará optimizar el rendimiento de los modelos y así mitigar el sobreajuste. Esta técnica permite aprovechar el conocimiento adquirido por la red preentrenada y adaptarlo al dominio de las imágenes de ultrasonido fetal [40], [50].

Las capas convolucionales iniciales de una CNN aprenden características generales, como bordes y texturas. Por eso, a menudo, se mantienen congeladas. Las capas intermedias y finales de la red aprende características más específicas, en este caso la anatomía fetal de las imágenes. El objetivo es descongelar estas capas posteriores para ajustar los pesos y que el modelo aprenda características más complejas.

Por eso, determinar qué porcentaje o número de capas se debe descongelar es esencial. Lo que faculta al modelo a usar los parámetros de entrenamiento adecuados

considerando los costes computacionales y la clasificación final.

5.2.1.1. RESNET50

La arquitectura ResNet está diseñada para construir redes más profundas sin experimentar problemas de degradación de rendimiento. La principal contribución de ResNet es la introducción del bloque residual que permite a la red mantener su precisión a medida que aumenta el número de capas, resolviendo el problema de que la precisión alcance la saturación y caiga la rápidamente en las redes muy profundas [44].

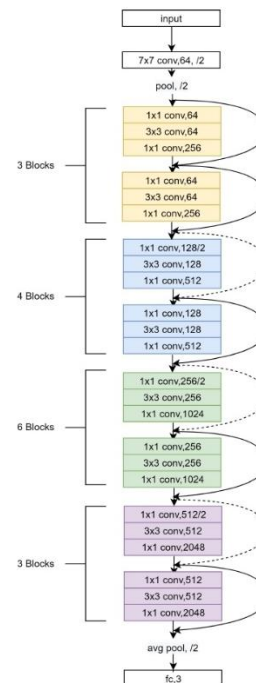


Figura 21: Arquitectura del ResNet50. Tomado de [51]

En redes convolucionales, una técnica esencial para adaptar modelos previamente entrenados a nuevas tareas es mediante el ajuste de sus parámetros, aprovechando el aprendizaje por transferencia y reduciendo las necesidades de grandes conjuntos de datos. El *fine tuning* implica congelar las capas del modelo, ajustar las capas óptimas, que podría ser entre el 20% y el 25% de las capas superiores. Finalmente, con la escogencia del mejor modelo, ajustar los hiperparámetros (como la tasa de aprendizaje, el tamaño del batch, etc.) [52].

Hemos usado una estrategia de exploración al descongelar progresivamente 5% de las capas por vez.

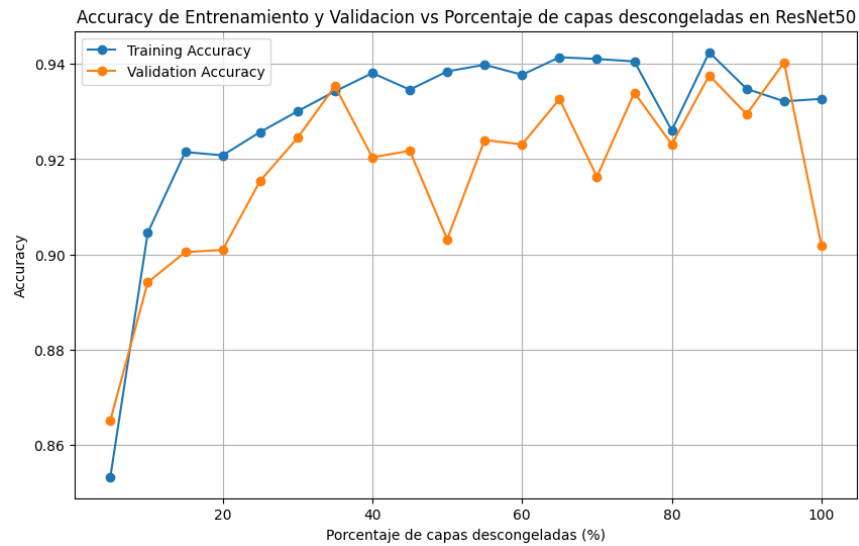


Figura 22: Descongelamiento de capas de ResNet50

La cantidad de epochs es baja (5 para estos casos), puesto que solo queremos evidenciar el comportamiento del modelo con el porcentaje de capas descongeladas. Como se observa en la imagen, **Figura 22**, la mayoría de estas iteraciones superó el 0.9 de accuracy en validación y, entre ellas, la que mejor rendimiento tuvo fue las que descongelaron el 35% de la capas, en donde se obtuvo un valor semejante en el entrenamiento y la validación de aproximadamente del 0.93.

Escogiendo un porcentaje de capas del 35%, vamos a hacer otro ajuste fino con la tasa de aprendizaje. Para el anterior ajuste, hemos usado un valor predeterminado de 0.0001. Ahora, variaremos el parámetro para ver su comportamiento.

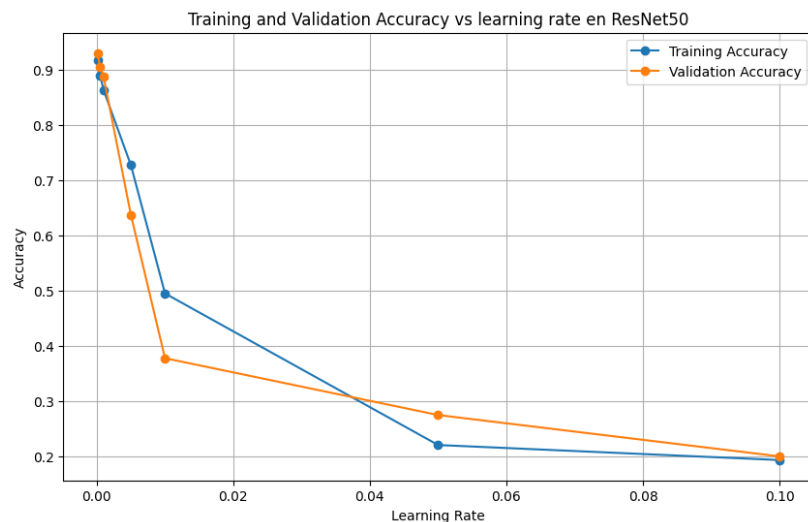


Figura 23: Accuracy de ResNet50 con 35% de capas descongeladas

La **Figura 23** muestra que entre más aumenta la tasa de aprendizaje, el accuracy del modelo fue reduciéndose poco a poco hasta llegar a valores casi de 0.1. Esto demuestra

que, para tasas de aprendizaje muy grandes, el modelo tiende a divergir. Por eso, la escogencia inicial es la adecuada.

Como última etapa, correremos el modelo con una cantidad más grande de epochs, pero se agrega dos callbacks para optimizar el entrenamiento del modelo y evitar el sobreajuste. Un EarlyStopping para monitorear una métrica (en este caso, la pérdida de validación) y detener el entrenamiento automáticamente si no hay mejorar en un número de épocas definido con una paciencia, y un ReduceLROnPlateau para reducir la tasa de aprendizaje en un factor (en este caso de la mitad).

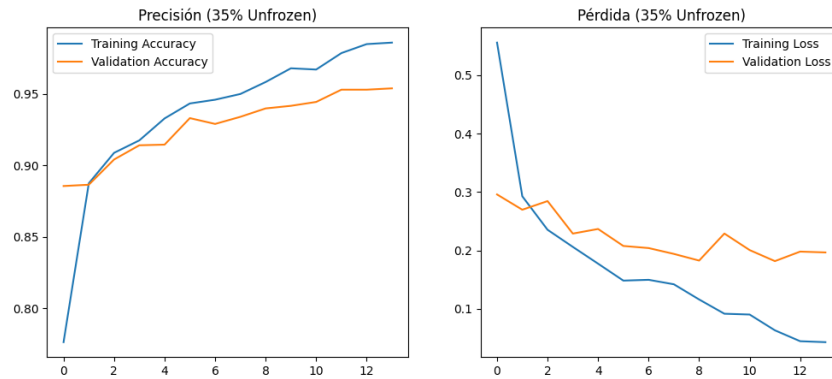


Figura 24: ResNet50 con 35% con 12 epochs

De las 100 epochs utilizadas, **Figura 24**, el entrenamiento acabó alrededor de la época 12. El accuracy para entrenamiento y validación fue aumentando lentamente hasta alcanzar valores superiores al 90%. A partir de la epoch 8, el modelo empieza a sobreajustar levemente, pero cómo el EarlyStopping se queda con el mejor resultado, el valor final de accuracy para el modelo fue del 94%.

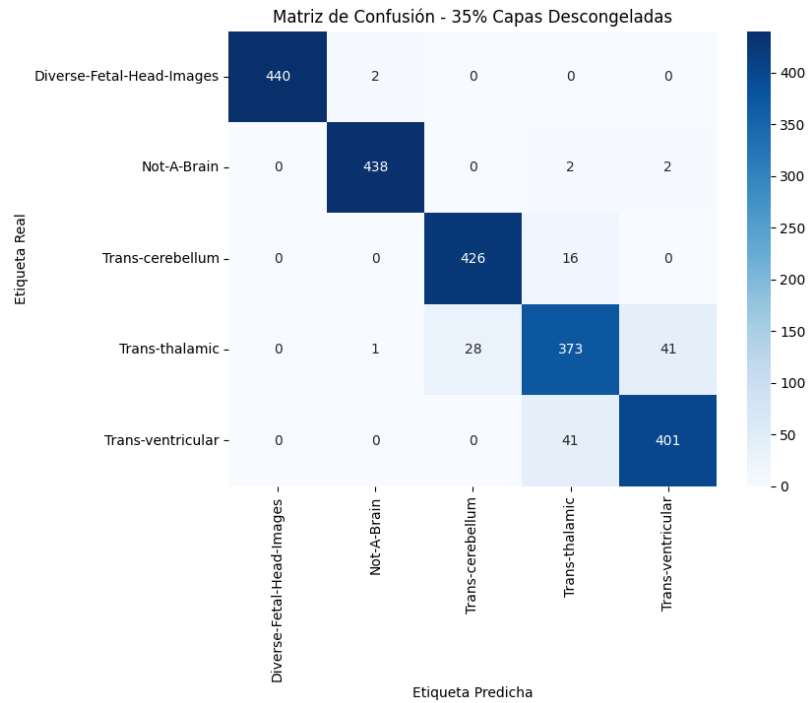


Figura 25: Matriz de confusión de ResNet con 35% de capas descongeladas

A través de la matriz de confusión, **Figura 25**, se puede ver fácilmente que las clases que no son planos anatómicos de cabeza fetal han clasificado excelentemente las imágenes, pese a que por ser clases con mucha variación, como la de Not-A-Brain, la lógica diría que tendrían más dificultades. En cambio, de los planos anatómicos principales, el talámico (thalamic) y el ventricular tuvieron una cantidad de errores mínimos. Pero un reporte de clasificación más detallado dice:

Tabla 4: Reporte de clasificación de ResNet50

Clase	Precision	Recall	F1-score
Diverse-Fetal-Head-Images	1.00	1.00	1.00
Not-A-Brain	0.99	0.99	0.99
Trans-cerebellum	0.94	0.96	0.95
Trans-thalamic	0.86	0.84	0.85

Trans-ventricular	0.90	0.91	0.91
-------------------	------	------	------

Con la **Tabla 4**, se puede aseverar que el plano con peor rendimiento fue el trans-talámico, pues obtuvo valores de precisión, Recall y F1-score inferiores al 0.9. Estas métricas de evaluación dan un panorama bastante favorable para el modelo, pues en cuanto a precisión, es capaz de predecir más del 80% de la mayoría de las clases. En el caso del Recall, el modelo tuvo una sensibilidad que se reduce para los planos trans-talámico y trans-ventricular. Y con un F1 tan alto para la mayoría de las etiquetas, se sabe que el modelo es lo suficientemente robusto para no tener muchas equivocaciones en la clasificación.

Tabla 5: Promedios de métricas de ResNet50

Métrica	Precision	Recall	F1-score
Accuracy	—	—	0.94
Macro avg	0.94	0.94	0.94
Weighted avg	0.94	0.94	0.94

Sin embargo, **Tabla 5**, el promedio de estas métricas fue de 0.94 en un accuracy general. Este resultado es bueno, puesto que tendremos una excelente clasificación de las diversas clases para los planos anatómicos principales, para los diversos y para los no cerebrales.

5.2.1.2. VGG16

La arquitectura VGG16 fue presentada por Karen Simonyan y Andrew Zisserman en el artículo "Very Deep Convolutional Networks for Large-Scale Image Recognition" [53]. Su innovación radica en la estandarización y profundización de la red mediante filtros de convolución muy pequeños (3×3). Esta arquitectura tiene 16 capas con peso (13 convolucionales y 3 totalmente conectadas), caracterizada por su uniformidad, una entrada de imagen RGB de tamaño fijo de 224×224 píxeles [54].



Figura 26: Arquitectura VGG16

Ofrece ventajas muy notables, pues aplica tres funciones de activación ReLU en lugar de una, permitiendo a la red aprender funciones más discriminativas.

Estudios sobre esta arquitectura han demostrado que supera modelos como AlexNet en tareas médicas (como por ejemplo la clasificación de melanoma) logrando una precisión del 85% [55]. Se sigue el mismo proceso de sintonización por porcentajes que en caso del ResNet. En este caso, en vez del 5% antes usado, ahora será del 10% por la cantidad inferior de capas.

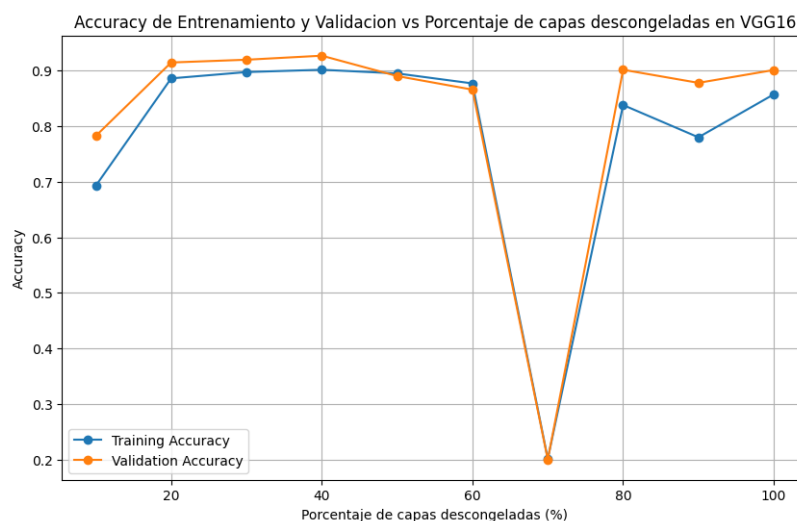


Figura 27: Descongelamiento de capas de VGG16

El proceso de descongelamiento para esta arquitectura es diferente a la del ResNet50. Como se ve en la **Figura 27**, desde un umbral cercano al 40% el modelo empieza a decaer en accuracy, además de que el accuracy en validación es en todo momento superior a su valor en entrenamiento. Hay un valle bastante distinguible en el porcentaje del 80% donde los accuracies bajan a valores de 0.2 para los conjuntos de entrenamiento y validación, pero que vuelve a retomarse en porcentajes superiores.

La escogencia del porcentaje será cerca a la anterior, del 40%, donde usaremos nuevamente una sintonización para la tasa de aprendizaje. Se podría también descongelar menos porcentaje de capas, como un 20% o un 30%, que tiene valores cercanos a 0.9 de accuracy, pero se desea un modelo un poco más robusto al enfrentarlo con más números de epochs, así que se decide por el 40% finalmente.

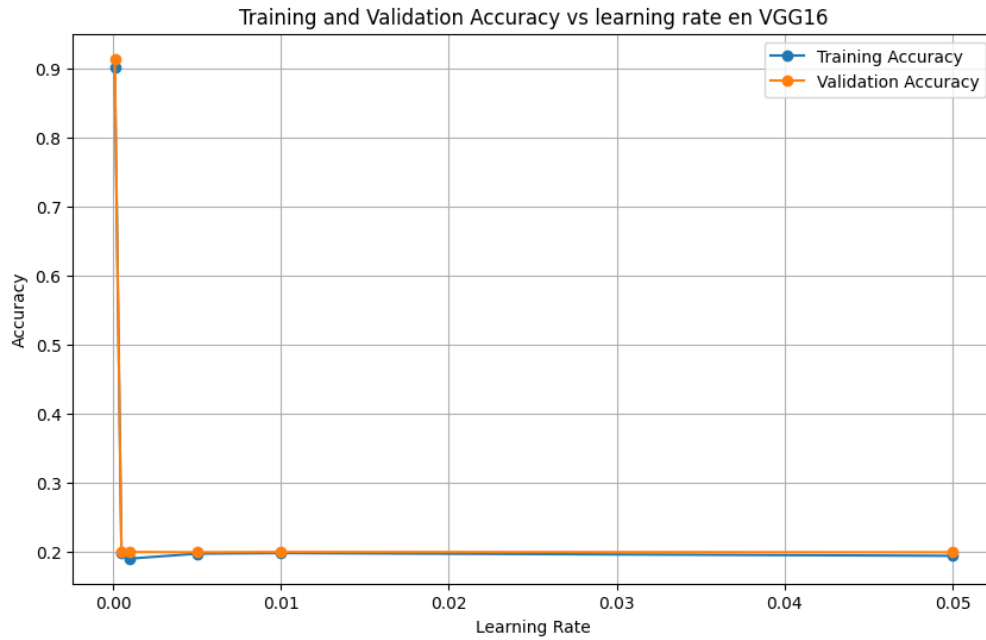


Figura 28: Accuracy de VGG16 con 30% de capas descongeladas

Así como sucedió en la anterior arquitectura, la sintonización de la tasa de aprendizaje, **Figura 28**, se redujo a un valor aproximado de 0.2 de accuracy para todos los valores superiores al ya usado, de 0.001.

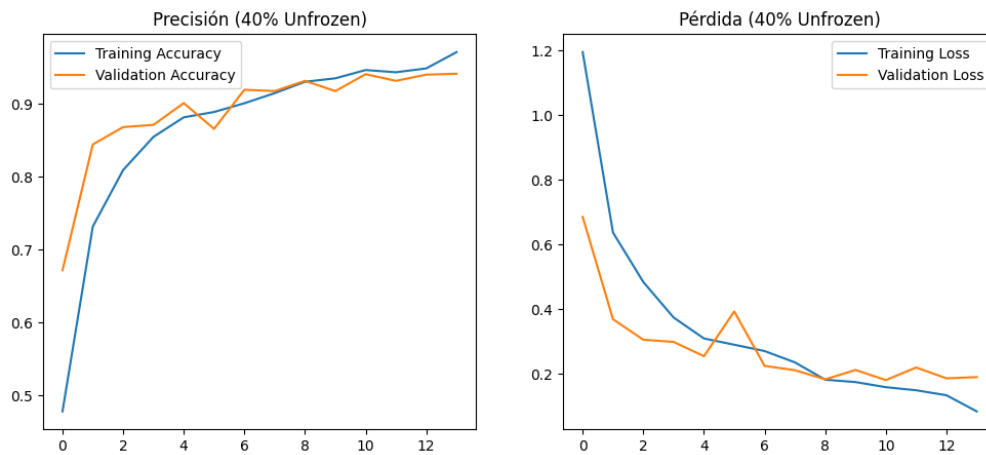


Figura 29: VGG16 con 40% con 12 epochs

Con un comportamiento semejante, **Figura 29**, cuando enfrentamos la arquitectura con varias epochs, con un máximo de 100 épocas, con el mismo EarlyStopping y ReduceLROnPlateau, llegamos a las 12 antes de que el modelo se detengan. También se nota un ligero sobreajuste al inicio, pero lentamente fue estabilizándose en las epochs 8 en adelante, con accuracy superiores al 0.9.

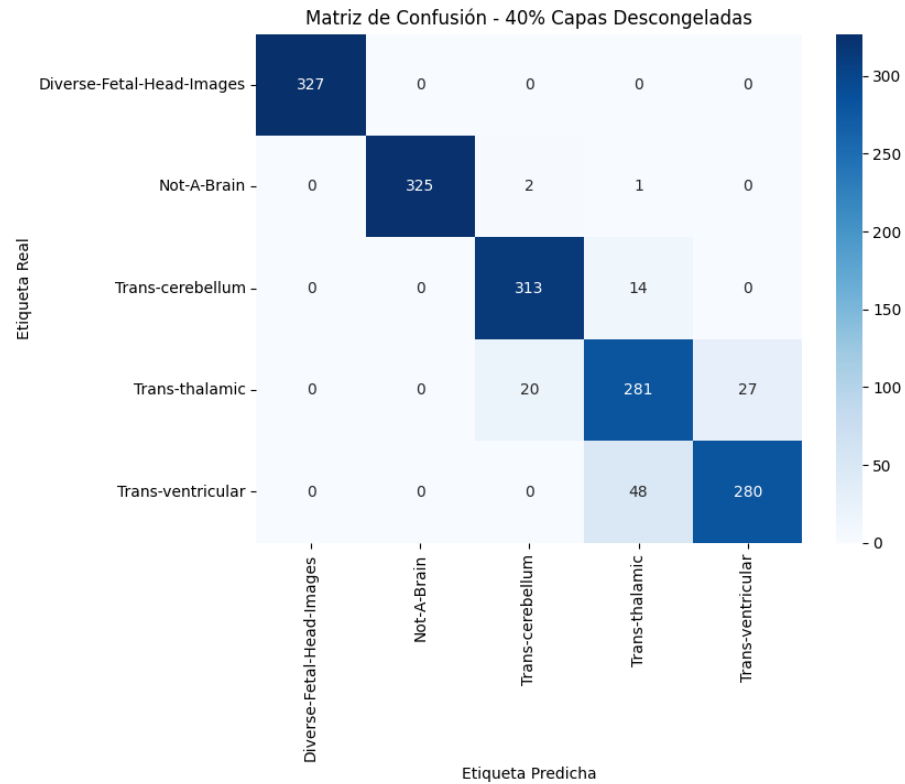


Figura 30: Matriz de confusión de VGG16 con 40% de capas descongeladas

Respecto a la matriz de confusión, **Figura 30**, la arquitectura VGG16 tuvo casi los mismos errores para los tres planos principales. Los errores de clasificación para el plano talámico y ventricular parecen semejantes con la ResNet50. Además, el plano no cerebral y los diversos fueron los mejores clasificados.

Tabla 6: Reporte de clasificación de VGG16

Clase	Precision	Recall	F1score
Diverse-Fetal-Head-Images	1.00	1.00	1.00
Not-A-Brain	1.00	0.99	1.00
Trans-cerebellum	0.93	0.96	0.95

Trans-thalamic	0.82	0.86	0.84
Trans-ventricular	0.91	0.85	0.88

El reporte de clasificación, **Tabla 6**, tiene métricas de evaluación muy semejantes al ResNet. Sin embargo, hay valores inferiores para el F1 en los planos trans-talámico y trans-ventricular, puesto que ha pedido un poco de robustez. Y asimismo, respecto al Recall, la sensibilidad también se ha reducido levemente.

Tabla 7: Promedios de métricas de VGG16

Métrica	Precision	Recall	F1-score
Accuracy	—	—	0.93
Macro avg	0.93	0.93	0.93
Weighted avg	0.93	0.93	0.93

En consecuencia, los promedios de las métricas, **Tabla 7**, solo tiene un punto porcentual menos que el ResNet50, que corresponde al 93%.

5.2.1.3. ARQUITECTURA PERSONALIZADA

Pese a que las anteriores arquitecturas tienen un excelente desempeño con las imágenes de ultrasonido fetal, para los requerimientos de estas se necesita que las arquitecturas puedan manejar las características inherentes en imágenes de ultrasonido. En muchos sentidos, explorar la creación de una red neuronal beneficiaría en la precisión, la robustez y la eficiencia de la clasificación.

5.2.1.3.1. DIVERSAS ARQUITECTURAS USADAS

Se presentan diversas arquitecturas usadas. Todas estas con un tamaño de imagen de 256x256. Para maximizar los costes de tiempo y computacionales se decidió usar una cantidad baja de epochs en el entrenamiento para solo visualizar el comportamiento de cada una de ellas.

Tabla 8: Arquitecturas personalizadas usadas

Tipo	Capas	Último Accuracy del entrenamiento	Último Accuracy de validación
Densas	[32]	0.196058	0.205128
	[64, 32]	0.197977	0.202686
	[128, 64, 32]	0.197802	0.197395
	[256, 128, 64, 32]	0.199023	0.207977
	[1024, 256, 128, 64, 32]	0.197453	0.204721
Convolucionales	[32]	0.2004	0.2002
	[64, 32]	0.2013	0.1998
	[128, 64, 32]	0.4233	0.5478
	[256, 128, 64, 32]	0.3741	0.4969
	[[128, 128], [64, 64], [32, 32]]	0.4554	0.6264
	[[128, 128, 128], [64, 64, 64], [32, 32, 32]]	0.5495	0.7053
	[[128, 128, 128, 128], [64, 64, 64, 64], [32, 32, 32, 32]]	0.5339	0.6711
Conv - BatchNorm - MaxPool	[32, 64, 128]	0.421594	0.512007
Bloques Residuales	[64, 64, 128, 128, 256, 256]	0.782138	0.540904
Bloques densos	Denso + Bloque de transición	0.653933	0.332112
Bloques Bottleneck	[256, 256, 512, 512, 1024, 1024]	0.746729	0.542939
Bloques Attention	[32, 64, 128]	0.555381	0.576313
Lightweight	[64, 64, 128, 128, 256, 256]	0.60614	0.597884
Hybrid Residual + Attention	[64, 64, 128, 128, 256, 256]	0.730508	0.442816
Simple + Bloques Residuales	[64] + [64, 64, 128, 128, 256, 256]	0.78929	0.753765
Simple + Bloques densos	[64] + Bloques densos	0.635618	0.309727
Simple + Bloques Bottleneck	[64] + [256, 256, 512, 512, 1024, 1024]	0.755276	0.3256

Se explicará cada una de estas arquitecturas presentadas en la **Tabla 8** para ahondar en el funcionamiento y comportamiento. Se presentan un nombre descriptivo de las capas

que posee el modelo, los núcleos de las diferentes capas y el último accuracy en entrenamiento y validación. Sin embargo, como se nota en los accuracies, la mayoría de las arquitecturas tienen un subajuste evidente porque no se están utilizando ni una sola estrategia de regularización.

La primera arquitectura es de capas densas, desde una capa de **32** núcleos hasta cinco capas de **[1024, 256, 128, 64, 32]** donde se obtuvieron un porcentaje cercanos a 0.2 de accuracy, que refleja un obvio de subajuste. Las segundas son convolucionales, en donde se ve un comportamiento inusual, como por ejemplo que para capas del tipo **[[128, 128, 128], [64, 64, 64], [32, 32, 32]]** se tiene un valor bastante elevado de accuracies en validación contra el de entrenamiento, con una diferencia de alrededor de 16 puntos porcentuales. Dado que todos fueron entrenados y validados con los mismos conjuntos, este comportamiento no suele ocurrir y es raro que durante la prueba de validación le modelo tenga mejor accuracy. Pese a ello, el uso de tantas capas convolucionales es computacionalmente pesado.

La tercera arquitectura con capas convolucionales **[32, 64, 128]**, un BatchNormalization y un MaxPooling2D no tuvo un desempeño adecuado. La cuarta, con el uso de bloques residuales, como la arquitectura ResNet, que permiten que la información fluya a través de la red de manera más eficiente mediante conexiones de salto, tuvo un buen comportamiento en entrenamiento, pero decayó en la validación, que podría ser un buen candidato para mejorar a través de diferentes usos de capas y de regularizaciones. La quinta, que es el componente interno de la arquitectura DenseNet, con comportamiento semejante a la ResNet, aunque los bloques se concatenan permitiendo una reutilización mucho más extensa de características [56], el cual obtuvo un sobreajuste bastante amplio.

La sexta contiene bloques cuello de botella que reducen la cantidad de computación manteniendo la capacidad de aprendizaje, utilizadas en arquitecturas como ResNet50, 101 y 152, que tuvo un buen desempeño, pero existe sobreajuste. La séptima con bloques con atención, que permiten que el modelo aprenda de las partes o de la información más importante, con subajuste y un mal accuracy. La octava, con una arquitectura ligera, que tiene casi el mismo comportamiento de la anterior. La novena, que es una red híbrida residual más el mecanismo de atención, que obtuvo un buen accuracy en entrenamiento, pero está subajustada.

La décima es la combinación de una estructura simple más bloques residuales, que es la que mejor accuracy brinda con un ligero sobreajuste, que puede ser la mejor candidata hasta el momento. La undécima que combina la estructura simple más bloques densos está sobreajustada, y por último, la combinación de estructura simple más bloques cuello de botella con la que sucede lo mismo.

5.2.1.3.2. DISEÑO DE LA ARQUITECTURA FINAL

Conociendo el comportamiento de los diferentes modelos, se escoge la arquitectura con una capa simple más bloques residuales, **Figura 31**. El bloque residual

escogido posee una primera convolución con una capa Conv2D, un BatchNormalization y una activación ReLU, una segunda convolución con una capa Conv2D y un BatchNormalization, una conexión de salto que reduciría las dimensiones espaciales a la mitad cuando el stride sea menor a 1, y, finalmente, una suma residual con activación ReLU y un Dropout que desactiva aleatoriamente el 30% de las neuronas durante el entrenamiento. El bloque de capa sencilla posee una capa Conv2D, un BatchNormalization, una activación ReLU y un MaxPooling2D. Además, las capas siguientes, que son las residuales, tendrán distintos núcleos para comprobar su comportamiento, seguidas por un GlobalAveragePooling2D, un Dropout y una capa densa respecto al número de clases del conjunto de datos.

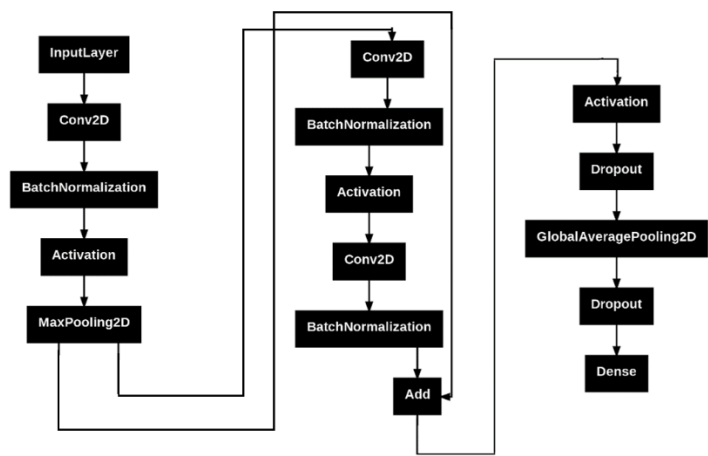


Figura 31: Diagrama de la arquitectura base personalizada

5.2.1.3.3. DIVERSOS MODELOS CON LA ARQUITECTURA

Para verificar el comportamiento de la arquitectura se plantean diversos núcleos. En este caso, lo que se pretende es brindarle a la estructura la suficiente cantidad de neuronas para que pueda aprender las características inherentes de cada clase. Por ello, los modelos usados se presentan en la **Tabla 9**:

Tabla 9: Modelos con la arquitectura personalizada

Capas bloques residuales	Último Accuracy del entrenamiento	Último Accuracy en validación
[64, 64, 128, 128, 256, 256]	0.8805	0.8974
[64, 64, 128, 128, 256, 256, 512, 512]	0.8952	0.9086
[64, 64, 64, 128, 128, 128, 256, 256, 256]	0.8804	0.9073

[64, 64, 64, 128, 128, 128, 256, 256, 256, 512, 512, 512]	0.8796	0.8307
--	--------	--------

Todos los modelos se entrenaron con los mismos callbacks y parámetros de la sección anterior, pero con 50 epochs. Además, la capa simple tiene un único núcleo de 64 para todos los casos. De entre estos, el que mejor resultado obtuvo fue el segundo modelo con un accuracy muy cercano en el entrenamiento y validación. Los siguientes, con mayor cantidad de núcleos, fueron reduciendo su rendimiento poco a poco.

5.2.1.4. MODELO ESCOGIDO

Para obtener mejor rendimiento con el modelo, se realizaron cambios a la arquitectura base personalizada, integrando bloques residuales con mecanismos de atención tipo Squeeze-and-Excitation (SE), con el fin de recalibrar los canales más relevantes de las características extraídas. Además, se utilizó una estrategia de mezcla entre Global Average Pooling y Global Max Pooling para enriquecer la representación final antes de la clasificación. Esta combinación permitió construir una arquitectura más robusta y con mayor capacidad de generalización y de esa forma aumentar un poco más la exactitud del modelo.

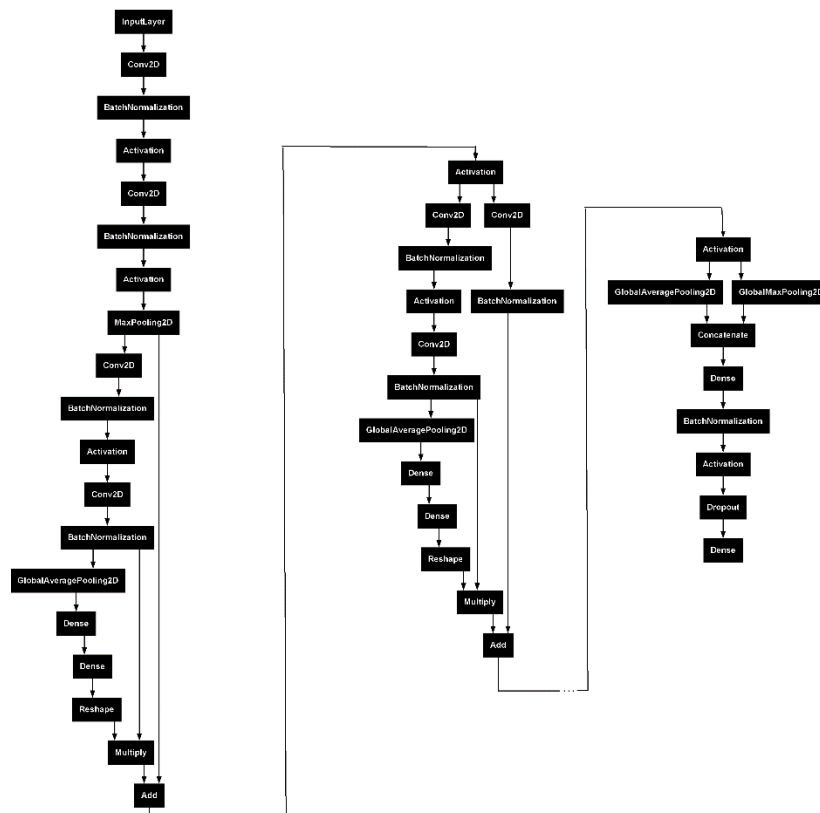


Figura 32: Arquitectura personalizada con mecanismo de atención

La **Figura 32** representa el diagrama del modelo final, en donde se implementa el bloque SE, que sirve para dar más peso a canales informativos. La sección contiene un GlobalAveragePooling, dos Dense para reducir la dimensión y generar un peso entre 0 y 1, un Reshape y Multiply para reescalar cada canal según su importancia. Después el bloque residual convolucional con atención SE, con el bloque de atención para recalibrar las características. Y, finalmente, una suma con una activación ReLu. Este modelo tuvo un comportamiento un poco mejor que a las arquitecturas ya predeterminadas como el ResNet y el VGG, pues su complejidad captura matices más finas en las imágenes médicas. Por ejemplo, no hay ese ligero sobreajuste que tuvieron los modelos de las secciones anteriores, sino que la diferencia entre accuracy de entrenamiento y validación es de menos de un punto porcentual.

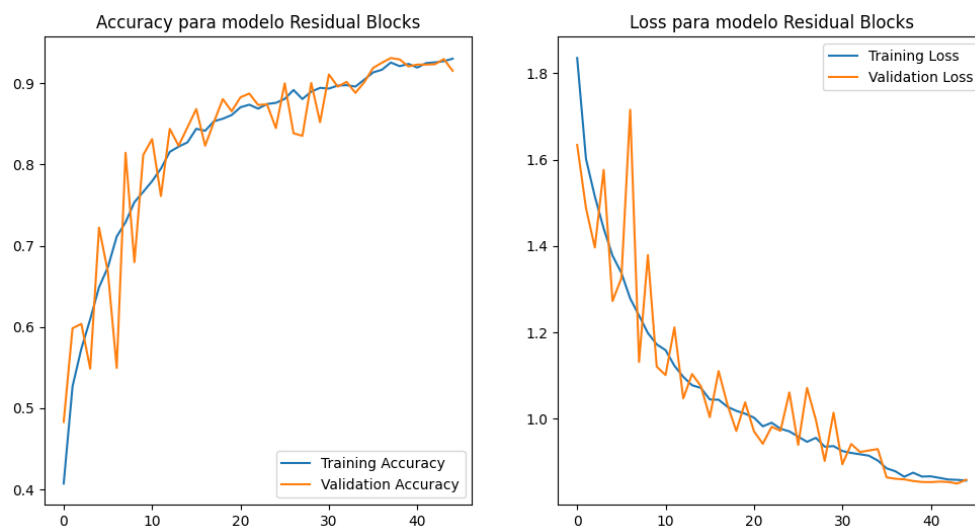


Figura 33: Accuracy para la arquitectura personalizada

La **Figura 33** muestra que el entrenamiento del modelo fue bastante errático. Desde el inicio, los valores de accuracy en entrenamiento tuvieron un aumento gradual y suave, pero en validación hubo un comportamiento zigzagueante que demuestra que el modelo necesitó alrededor de 40 epochs para estabilizarse. De allí en adelante, el modelo se estabilizó y las diferencias entre los accuracies y losses tienden a aumentar, pero debido a la cantidad de epochs escogida no hubo más evolución. En la instancia 35 hubo una reducción de la tasa de aprendizaje en un factor de 0.1, quedándonos a la final del entrenamiento con una tasa de $9.99E^{-06}$.

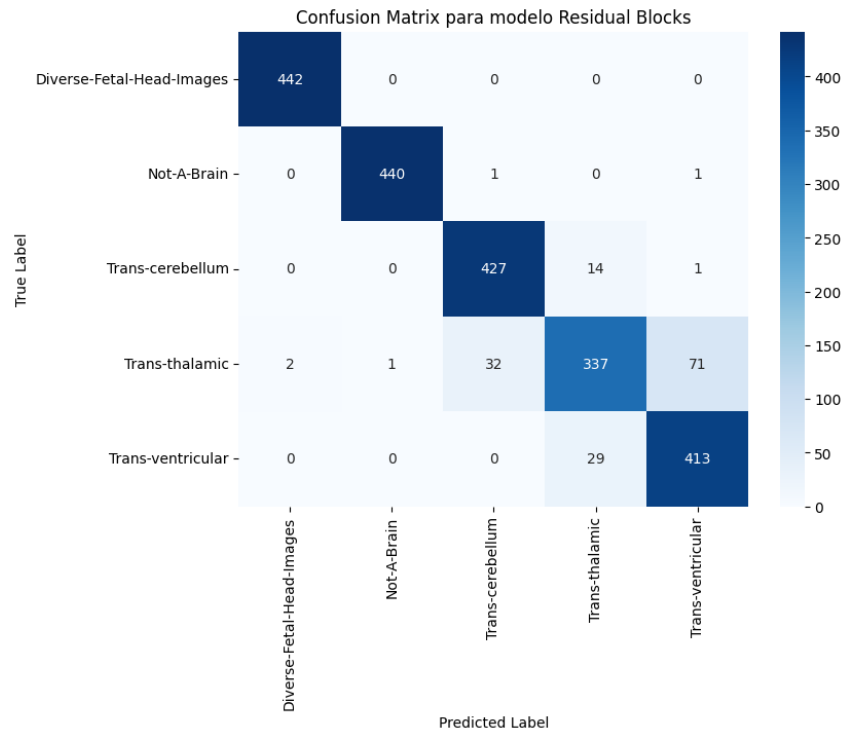


Figura 34: Matriz de confusión para arquitectura personalizada

En tanto a la clasificación de las imágenes del conjunto de validación, la **Figura 34** muestra que el comportamiento de la arquitecturas preentrenadas y la personalizada es bastante semejante. Los errores para la clase Not-A-Brain son nulos o pequeños, con solo dos mal clasificados, los de Diverse-Fetal-Head-Images no hubo errores. En cambio, cuando se clasifica los tres planos principales se tiene errores entre ellos mismos, donde los planos se pueden clasificar de manera errónea en otros casos, aunque no en el plano trans-cerebelo que es el que mejor se clasifica.

Tabla 10: Reporte de clasificación de VGG16

Clase	Precision	Recall	F1score
Diverse-Fetal-Head-Images	1.00	1.00	1.00
Not-A-Brain	1.00	1.00	1.00
Trans-cerebellum	0.93	0.97	0.95

Trans-thalamic	0.89	0.76	0.82
Trans-ventricular	0.85	0.93	0.89

El reporte de clasificación, **Tabla 10**, es muy parecido al VGG16. Hay que denotar que en el caso del plano trans-talámico la sensibilidad es mucho menor a las anteriores, y el F1 demuestra que la robustez del modelo se parece al VGG, aunque a la hora de predicciones, para las imágenes trans-talámicas, el modelo cometerá más errores o brindará porcentajes más bajo pues será capaz de clasificar menos cantidad de instancias.

Tabla 11: Promedios de métricas de arquitectura personalizada

Métrica	Precision	Recall	F1-score
Accuracy	—	—	0.93
Macro avg	0.93	0.93	0.93
Weighted avg	0.93	0.93	0.93

Asimismo, los promedios de las métricas, **Tabla 11**, es tan bueno como en los anteriores. Por lo tanto, los tres modelos escogidos tienen un rendimiento aceptable para la clasificación de las imágenes de cabeza fetal.

5.3. DISCUSIÓN

Los resultados para los modelos fueron excelentes. Para condensar de manera más eficiente estos resultados, se presenta la **Tabla 12**, en donde se puede detallar más eficientemente las diferencias entre cada uno. En este caso, los accuracies son muy similares entre los modelos, bastante adecuados para una clasificación de correcta de los planos en ultrasonidos de cabeza fetales.

Tabla 12: Accuracies de los modelos

Modelo	Accuracy
VGG16	94%
ResNet50	93%

Custom	93%
---------------	------------

Sin embargo, las métricas de Precision, Recall y F1-Score hablan un poco mejor de estas diferencias. Por eso, la **Tabla 13**, explica que la clase de Diverse-Fetal-Head es fácil de clasificar para todos los modelos. Asimismo, la clase Not-A-Brain tiene un valor cercano, excepto para el modelo ResNet50 que tiene una baja leve de 0,99. Para la clase Trans-cerebellum, el F1-Score es el mismo, puesto que el modelo puede reconocer la vista, aunque no perfectamente, pero la Precision y el Recall es muy cercanos entre todos, puesto que los modelos casi siempre aciertan la predicción de la clase y detectan la mayoría de imágenes reales.

Para la clase Trans-thalamic, la clasificación se complica. Los modelos preentrenados, ResNet50 y VGG16, confunden la vista de estas imágenes, y el modelo Custom pierde más clasificaciones con varios casos reales. El modelo Custom es el que mejor precisión detiene de los tres, pero pierden más casos reales por una sensibilidad más baja. Y, por último, para la clase Trans-ventricular, el ResNet50 es el que más equilibra precisión y recall, siendo el más robusto para su clasificación. Los modelos VGG16 y Custom tiene invertidos los valores de precisión y recall, lo que denota que el VGG16 suele acertar más, pero se le escapan más casos reales y, por lo contrario, el Custom encuentra más casos reales, pero comete más falsos positivos.

Tabla 13: Métricas de los modelos

Clase	Modelo	Precision	Recall	F1-Score
Diverse-Fetal-Head-Images	VGG16	1.00	1.00	1.00
	ResNet50	1.00	1.00	1.00
	Custom	1.00	1.00	1.00
Not-A-Brain	VGG16	1.00	0.99	1.00
	ResNet50	0.99	0.99	0.99
	Custom	1.00	1.00	1.00
Trans cerebellum	VGG16	0.93	0.96	0.95
	ResNet50	0.94	0.96	0.95
	Custom	0.93	0.97	0.95
Trans-thalamic	VGG16	0.82	0.86	0.84
	ResNet50	0.86	0.84	0.85
	Custom	0.89	0.76	0.82
Trans ventricular	VGG16	0.91	0.85	0.88
	ResNet50	0.90	0.91	0.91
	Custom	0.85	0.93	0.89

En cuanto a la validación de los modelos, se empleó una partición simple de los datos en conjuntos de entrenamiento, validación y prueba, debido al alto costo computacional que implicaría ejecutar esquemas de validación cruzada sobre arquitecturas profundas como ResNet y VGG, así como al tiempo de entrenamiento requerido por cada experimento. Esta decisión permitió mantener un proceso experimental viable dentro de los recursos disponibles, sin comprometer la separación entre ajuste y evaluación final del modelo.

6. PROTOTIPO PARA LA CLASIFICACIÓN

Este capítulo describe el prototipo realizado para la clasificación de los planos anatómicos de cabezas fetales a través de aprendizaje conjunto con los tres modelos ya establecidos.

6.1. STREAMLIT

Según la página oficial, Streamlit es un framework de código abierto en Python diseñado para crear aplicaciones web interactivas de datos y Machine Learning de forma rápida y sencilla, sin necesidad de conocimientos en front-end. Es capaz de transformar scripts sencillos en apps web en minutos, ideal para científicos de datos e ingeniero de ML. Puede integrarse con las librerías más comunes como Pandas, Matplotlib y PyTorch.

En nuestro caso, que se quiere desarrollar un prototipo (o demo) de clasificación de imágenes de cabeza fetal, Streamlit facilita el despliegue enfocándose en su trasfondo. Generalmente, se crean interfaces donde los usuarios suben imágenes para obtener predicciones en tiempo real. La metodología en esencia será a través de Aprendizaje de Conjunto (Ensemble Learning), que construye un modelo predictivo integrando múltiples modelos para mejorar el accuracy [57].

A partir de los tres modelos establecidos en las secciones anteriores, se desarrolla un prototipo con mejor accuracy predictivo, con mayor diversidad, con mayor robustez y flexibilidad.

6.2. APRENDIZAJE DE CONJUNTO

El aprendizaje de conjunto, o de ensamblado, es una técnica de aprendizaje automático que combina las predicciones de dos o más modelos base para obtener un rendimiento superior y una mejor capacidad de generalización que la que se tendría con cada modelo por separado [58]. La premisa fundamental es que los modelos individuales pueden cometer errores y estos se pueden subsanar a través de la conjunción de varios de ellos, para compensar sus debilidades y aprovechar sus fortalezas.

Se debe tener en cuenta los siguientes factores claves:

1. Los modelos individuales deben ser precisos, más allá de una tasa de error mayor al azar.
2. Los conjuntos deben ser diversos. Si cometen los mismos errores, no complementarán a la mejora del modelo principal.

Para esto, se debe proveer diversidad de modelos, en donde cada uno de ellos puede tener diferentes estructuras para que hayan aprendido de las imágenes de manera diferente (como bordes, texturas o formas complejas), logrando una extracción de características desde varios puntos de vista.

Una vez que todas las CNN generaron su propia predicción, el sistema las integra

mediante una regla de combinación para obtener el resultado final. Entre estos métodos se encuentran la votación por mayoría (o Hard Voting), que es la más intuitiva, pues simplemente se suman los votos de cada clase y la clase que haya recibido la mayor cantidad de votos se selecciona como predicción final [59]. También la votación por mayoría ponderada, donde se entiende que no todos los modelos tienen los mismos pesos predictivos y se asignan valores a cada uno, para que la influencia de las CNN más débiles no altere el resultado. Y la predicción de probabilidades (Soft Voting), en donde se promedian las probabilidades de cada clase y se selecciona la más alta [59].

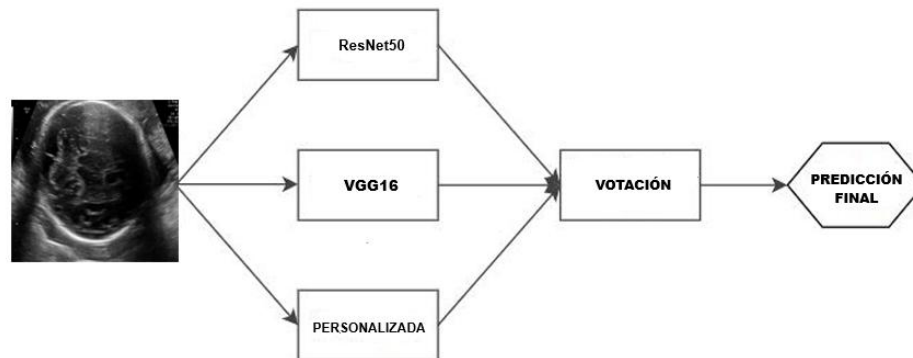


Figura 35: Diagrama de bloque de aprendizaje de conjunto paralelo. Modificado de [59]

La aplicación realizada a través de Streamlit carga una imagen de ultrasonido de cabeza fetal, usando cada uno de los modelos, permitiendo que realicen la predicción correspondiente, y, mediante una votación por mayoría, se determinará la clasificación, además de brindar un promedio de la confianza de las probabilidades, así como se muestra en la **Figura 35**.

6.3. PROTOTIPO FINAL

La interfaz final es bastante sencilla, **Figura 36**. La página principal configura el título del prototipo como: “Clasificador de Ultrasonidos Fetales Cerebrales”, a través de la función “`st.set_page_config()`”. Se muestra el título y una pequeña información de los modelos de CNN que se están usando. No hay variaciones de colores puesto que Streamlit fija mucha de estas convenciones de manera predeterminada. Tiene una barra lateral en que se informa de cada uno de los modelos y las clases que puede predecir el modelo final, a través de la función “`st.sidebar`”.



Figura 36: Interfaz de Streamlit del clasificador de Ultrasonidos Fetales Craneales

Se cargan los modelos de manera interna y se muestra un “st.success” sobre que han sido cargados correctamente. Después, se permite al usuario subir una imagen de ultrasonido de cabeza fetal a través de la función “st.file_uploader” que por defecto tiene un límite de 200 MB para cada imagen. Admite ingresar formatos comunes como PNG, JPG, JPEG y WEBP. Debido a que son imágenes de ultrasonido, se decidió prescindir del formato WEBP.



Figura 37: Mostrado de imagen cruda y preprocesada lado a lado

Se muestra en paralelo la imagen cruda y la preprocesada, **Figura 37**. Se muestran las dimensiones de cada imagen, recordando que el preprocesamiento para las imágenes sigue el siguiente pipeline:

1. Se convierte la imagen a escala de grises.
2. Se redimensiona a 256x256.

3. Se normaliza.
4. Se aplica un filtro gaussiano.
5. Se aplica el preprocesamiento del ResNet50 para mejor convergencia de los modelos. Para ello, antes de ese preprocesamiento, se devuelve la imagen a formato estándar de 8 bits y se convierte la escala de grises en un “falso RGB”, y luego se invierten los canales. (Esta inversión final es necesaria para la estructura de datos).



Figura 38: Predicción de cada modelo, su confianza y la entropía

Se realiza la predicción de cada uno de los modelos, **Figura 38**. Cada modelo es capaz de entregar la clasificación de la imagen, la confianza de su predicción y la entropía; esta última mide qué tan repartidas están las probabilidades dentro de las clases. Entre más baja la entropía, más confiable es la predicción. Como se muestra en la **Figura 38**, todos los modelos coincidieron en que la imagen es de un plano trans-cerebelo, la mayoría de ellos con confianza casi del 100% y solo la personalizada del 96%.

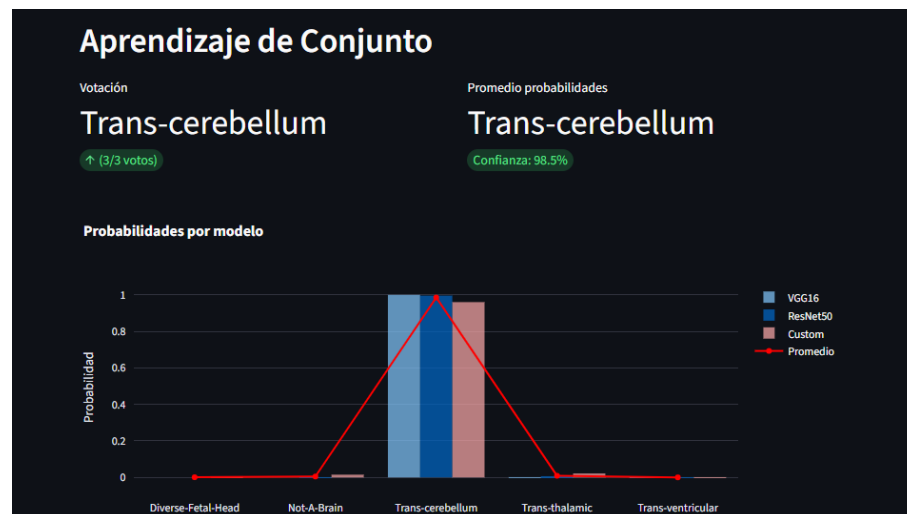


Figura 39: Votación por mayoría y gráfica de probabilidades por modelo

Se ejecuta la votación por mayoría, mostrando la confianza promedio de la clasificación, así como una gráfica de las probabilidades, **Figura 39**. Como la función “predict()” brinda las probabilidades de todas las clases en que fue entrenada la CNN, la

gráfica también muestra si el modelo piensa que la imagen pertenece a otra clase y en qué porcentaje.



Figura 40: Resumen de predicciones y diagnóstico final

Se finaliza con un resumen completo de las predicciones, mostrando una tabla con el modelo, la clase predicha, la confianza y la entropía, **Figura 40**, con un veredicto de la clase que el prototipo predice para la imagen. La baja entropía de la mayoría de modelos asegura la confianza de la predicción (que las probabilidades o probs están concentradas en una sola clase) [61]. Además, de un diagnóstico final para la imagen.

6.3.1. PREDICCIONES

Como se sabía desde la resolución de los modelos, el accuracy de todos era mayor al 90%, y, debido a sus estructuras, aprendieron de las imágenes de manera distinta, a través de valores inherentes en cada una. Sin embargo, la sensibilidad de cada modelo en los planos trans-talámicos y trans-ventriculares hacía caer en cuenta que las predicciones podrían entrar en conflicto.

Se presentan algunas predicciones de las diferentes etiquetas. Todas las imágenes son usadas del conjunto de prueba destinado para estas predicciones.

6.3.1.1. ETIQUETA: DIVERSAS CABEZAS FETALES

La clase Diverse-Fetal-Head corresponde a un conjunto de imágenes que no pertenece a los planos anatómicos axiales convencionales. Debido a ser una clase tan diversa, se esperaba que la precisión y la sensibilidad de los modelos a esta clase fuera baja, pero cada uno de ellos tuvo excelentes resultados. Por ello, la **Figura 41** y la **Figura 42** muestran la imagen original y preprocesada y el resultado de las predicciones de los modelos, respectivamente, que para los preentrenados fue el 100% y para el

personalizado fue un poco inferior, casi del 98%.

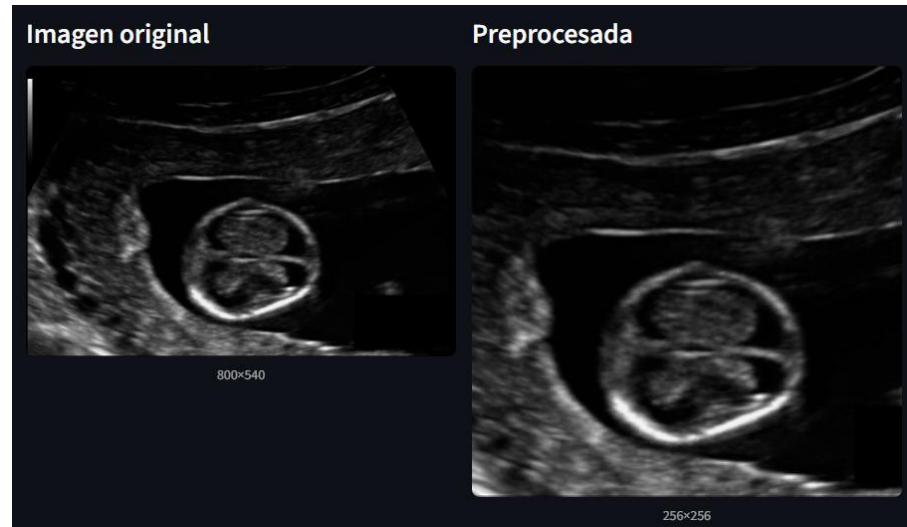


Figura 41: Imagen original y preprocesada de etiqueta *Diverse-Fetal-Head*



Figura 42: Predicciones por modelos y votación de etiqueta *Diverse-Fetal-Head*

La predicción final, **Figura 42**, fue exactamente la esperada para la imagen. Los tres modelos coincidieron en que esta corresponde a una clase *Diverse-Fetal-Head* con una confianza del 99,3%.

6.3.1.2. ETIQUETA: NO ES UN CEREBRO

La clase *Not-A-Brain* corresponde a imágenes, que como su nombre indica, no son un ultrasonido de cabeza, sino que viene de otra etiqueta respecto a planos de: abdomen fetal, femur fetal, thorax fetal y maternal cervix. Por ello, la **Figura 43** y **Figura 44** muestran

la imagen original y preprocesada y el resultado de las predicciones de los modelos, respectivamente, que para los preentrenados fue el 100% y para el personalizado fue un poco inferior, casi del 98%.



Figura 43: Imagen original y preprocesada de etiqueta Not-A-Brain



Figura 44: Predicciones por modelos y votación de etiqueta Not-A-Brain

La predicción final, **Figura 44**, fue exactamente la esperada para la imagen. Los tres modelos coincidieron en que esta corresponde a una clase Not-A-Brain con una confianza del 99,2%.

6.3.1.3. ETIQUETA: TRANS-CEREBELO

Las **Figura 45** y **Figura 46** muestran la imagen original y preprocesada y el resultado de las predicciones de los modelos, que para los modelos preentrenados fue el 100% y 99,9%, respectivamente para VGG16 y ResNet50, y para el personalizado fue un poco inferior, casi del 98,8%.

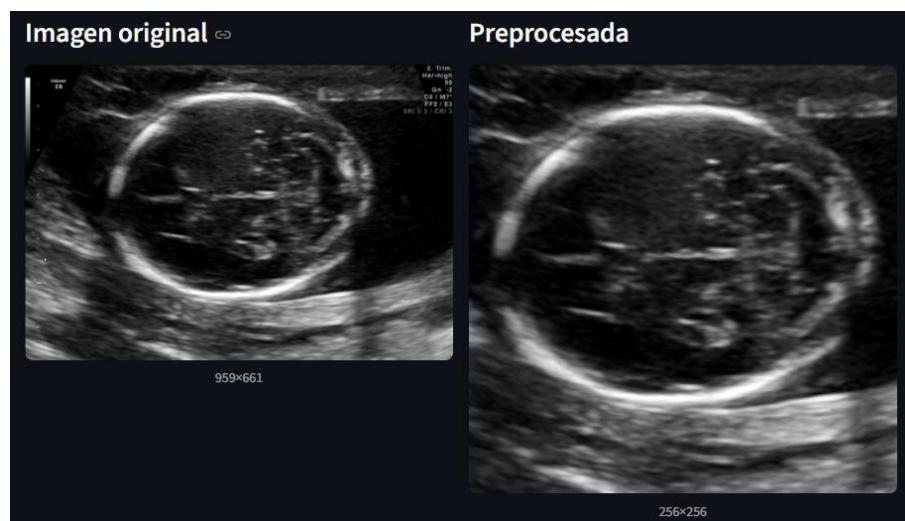


Figura 45: Imagen original y preprocesada de etiqueta Trans-cerebellum



Figura 46: Predicciones por modelos y votación de etiqueta Trans-cerebellum

La predicción final, **Figura 46**, fue exactamente la esperada para la imagen. Los tres modelos coincidieron en que esta corresponde a una clase Trans-cerebellum con una confianza del 99,6%.

6.3.1.4. ETIQUETA: TRANS-TALÁMICO

Las **Figura 47** y **Figura 48** muestran la imagen original y preprocesada y el resultado de las predicciones de los modelos, que para los modelos preentrenados fue el 99% y 93%, respectivamente para VGG16 y ResNet50, y para el personalizado fue un poco inferior, casi del 45,5%.

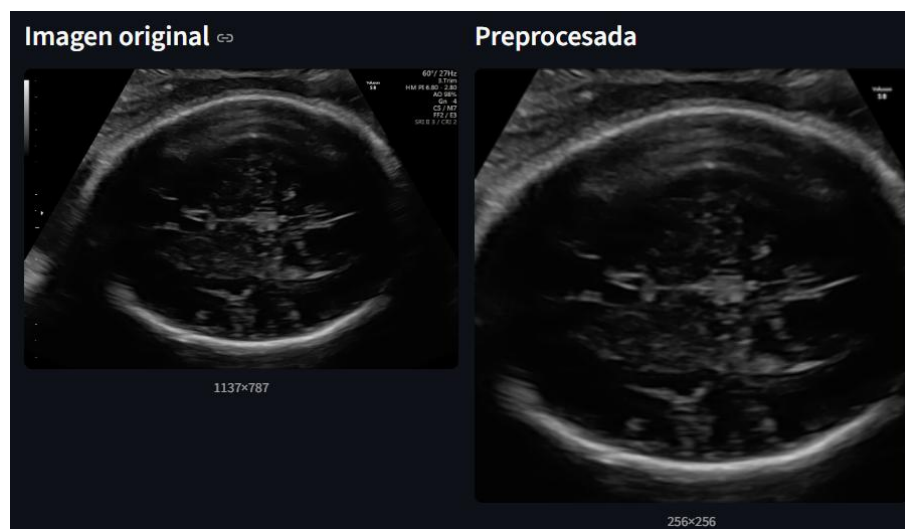


Figura 47: Imagen original y preprocesada de etiqueta Trans-thalamic



Figura 48: Predicciones por modelos y votación de etiqueta Trans-thalamic

La predicción final, **Figura 48**, fue la esperada para la imagen. En este caso, los tres modelos no pudieron decidir conjuntamente la clase, pues el modelo personalizado falló en la predicción y la certeza de la clase predicha fue bastante baja. Sin embargo, respecto al Aprendizaje de Conjunto, se llegó a un consenso con 2 votos de 3, en que la predicción

corresponde a una clase Trans-thalamic con una confianza del 78,8%.

6.3.1.5. ETIQUETA: TRANS-VENTRICULAR

Las **Figura 49** y **Figura 50** muestran la imagen original y preprocesada y el resultado de las predicciones de los modelos, que para los modelos preentrenados fue el 99,7% y 65,6%, respectivamente para VGG16 y ResNet50, y para el personalizado fue del 75,3%.

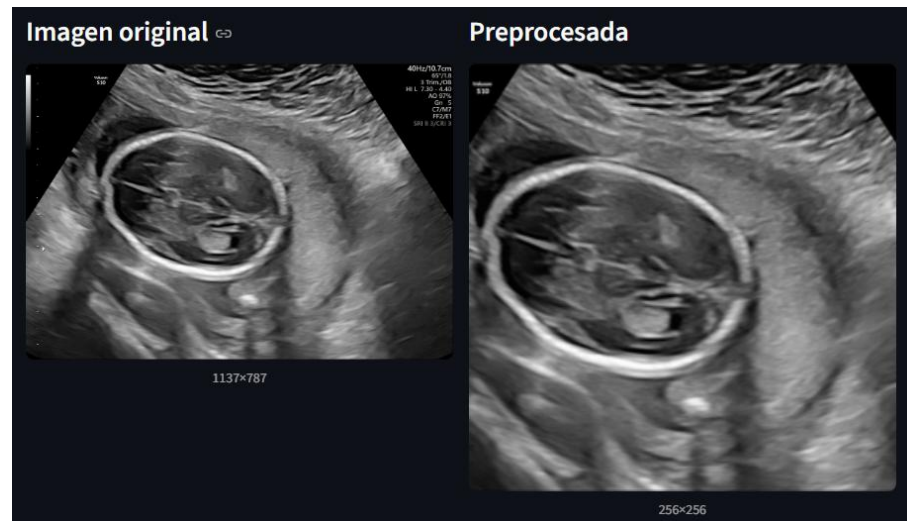


Figura 49: Imagen original y preprocesada de etiqueta Trans-ventricular



Figura 50: Predicciones por modelos y votación de etiqueta Trans-ventricular

La predicción final, **Figura 50**, fue la esperada para la imagen. En este caso, al igual que en la clase anterior, los tres modelos no pudieron decidir conjuntamente la clase, pues el modelo la ResNet50 falló en la predicción y la certeza de la clase predicha fue bastante baja. Igualmente, con ayuda del Aprendizaje de Conjunto, se llegó a un consenso con 2

votos de 3, en que la predicción corresponde a una clase Trans-Ventricular con una confianza del 69,6%.

En estas dos últimas predicciones, **Figura 48** y **Figura 50**, el Ensemble Learning ayudó al prototipo a predecir correctamente la clase de estos dos planos, trans-talámico y trans-ventricular, puesto que, por sus semejanzas, la confianza de los modelos fue reduciéndose poco a poco a otros planos, y que, si se hubiera dejado un único modelo, podría haber una mala clasificación. La votación por mayoría consiguió una confianza superior y una predicción más segura.

7. CONCLUSIONES Y TRABAJOS FUTUROS

Este capítulo describe las conclusiones respecto a los objetivos establecidos desde el inicio. Además, de los trabajos futuros guiados al potencial de crecimiento.

7.1. CONCLUSIONES

El objetivo general de este trabajo fue desarrollar un prototipo que utilizara Aprendizaje Profundo para la detección automática de planos anatómicos específicos en imágenes de ecografía craneal fetal, con el propósito de apoyar el proceso de análisis y reducir posibles errores de interpretación médica. Para ello, se realizó una revisión de la literatura científica, se preparó un conjunto de datos de ecografías craneales y se desarrollaron modelos de clasificación basados en arquitecturas de redes neuronales convolucionales, integrados posteriormente en un prototipo funcional mediante la herramienta Streamlit.

Los resultados obtenidos evidencian que las arquitecturas empleadas tienen un rendimiento sólido en la tarea de clasificación. Los modelos ResNet50, VGG16 y el modelo personalizado alcanzaron exactitudes iguales o superiores a 93%, lo que demuestra que es posible abordar el problema con técnicas de aprendizaje profundo. Estos resultados indican que las CNN son viables para la clasificación automática de los planos anatómicos fetales, especialmente cuando se dispone de un conjunto de datos correctamente etiquetado y de un proceso adecuado de entrenamiento y evaluación.

Asimismo, la estrategia de Aprendizaje de Conjunto permitió compensar algunas limitaciones individuales de los modelos, y contribuyó a mejorar la robustez del sistema final. Esta combinación de modelos ofreció una solución más estable a la sensibilidad de algunos modelos respecto a planos parecidos entre sí, lo cual resulta especialmente importante en contextos médicos donde la calidad de adquisición, la orientación del plano y las características del equipo pueden influir en el rendimiento del sistema. En ese sentido, el prototipo desarrollado no solo cumple una función demostrativa, sino que también representa una base sólida para el apoyo clínico futuro.

No obstante, es importante señalar que, aunque los resultados son prometedores, el prototipo debe interpretarse como una herramienta de apoyo y no como un sustituto del criterio médico. El desempeño alcanzado depende de las características del conjunto de datos utilizado, de la calidad de las etiquetas y de la representatividad respecto a escenarios clínicos reales. Por ello, la validación del modelo final en otros entornos, con datos procedentes de diferentes instituciones y equipos de ultrasonido, sigue siendo un aspecto necesario para considerar una posible aplicación clínica más amplia.

7.2. TRABAJOS FUTUROS

Se sugiere, para futuras líneas de trabajo, extender el conjunto de datos con imágenes provenientes de otras instituciones y equipos de adquisición diferentes, para así incrementar la variedad de casos y optimizar la capacidad de generalización del sistema; e, incluso, expandir las etiquetas de los ultrasonidos a otros planos (no únicamente al

plano axial), como el sagital y coronal. Asimismo, es relevante hacer una comprobación del rendimiento del prototipo en entornos clínicos.

Adicional a ello, se podría investigar la aplicación de modelos más modernos, como las arquitecturas más recientes o híbridas para optimizar la estabilidad y la precisión del clasificador.

Por último, se continuará perfeccionando el prototipo con el fin de transformarlo en una herramienta más completa y sólida, incluyendo su posible inclusión en equipos de ultrasonido, su análisis por lotes y medidas específicas por clase.

REFERENCIAS BIBLIOGRÁFICAS

- [1] F. Piccialli, V. Di Somma, F. Giampaolo, S. Cuomo, and G. Fortino, "A survey on deep learning in medicine: Why, how and when?," *Information Fusion*, vol. 66, pp. 111–137, Feb. 2021, doi: 10.1016/J.INFFUS.2020.09.006.
- [2] L. Zhao, K. Li, B. Pu, J. Chen, S. Li, and X. Liao, "An ultrasound standard plane detection model of fetal head based on multi-task learning and hybrid knowledge graph," *Future Generation Computer Systems*, vol. 135, pp. 234–243, Oct. 2022, doi: 10.1016/J.FUTURE.2022.04.011;PAGE:STRING:ARTICLE/CHAPTER.
- [3] V. Gómez Fonseca, P. C. Romero Lagos, and V. H. Palacios Yucci, "La inteligencia artificial en el diagnóstico de malformaciones craneofaciales," Tesis de Grado, Universidad El Bosque, 2020.
- [4] P. M. Losas, C. de C. Camarero, S. J. Loeches, A. Cortés-Beringola, I. A. González, and R. M. Aguilera, "Introducción a la ecografía pulmonar para cardiólogos," *Revista de Ecocardiografía Práctica y Otras Técnicas de Imagen Cardíaca*, vol. 8, no. 1, pp. 1–9, Apr. 2025, doi: 10.37615/retic.v8n1a2.
- [5] M. López-Azorín, M. Ybarra, and F. Cabañas, "Ecografía cerebral en neonatos," *Anales de Pediatría Continuada*, vol. 10, no. 4, pp. 228–233, Jul. 2012, doi: 10.1016/S1696-2818(12)70092-6.
- [6] İ. Ömeroğlu, H. Gölbaşı, E. Uçar, B. Sever, and H. G. Pala, "Fetal Central Nervous System Anomalies in a Tertiary Center: A Retrospective Study," *The Journal of Tepecik Education and Research Hospital*, vol. 32, no. 2, pp. 245–250, Aug. 2022, doi: 10.4274/TERH.GALENOS.2021.65872.
- [7] N. A. C. Aranguren, M. A. H. Hernández, C. M. L. Cardenal, F. Á. O. González, and P. P. Angarita, "Control de calidad en la adquisición de imágenes de tomografía cerebral y su relación con la precisión diagnóstica," Tesis de Grado, Universidad Nacional Abierta y a Distancia, 2024.
- [8] A. Rebato Ossona, "Implementación de un pipeline para el procesamiento de imágenes de fondo de ojo," Trabajo Fin de Grado, E.T.S. de Ingenieros Informáticos (UPM), Madrid, 2022.
- [9] H. R. Torres *et al.*, "A review of image processing methods for fetal head and brain analysis in ultrasound images," *Comput. Methods Programs Biomed.*, vol. 215, Mar. 2022, doi: 10.1016/J.CMPB.2022.106629.
- [10] L. Zhao, K. Li, B. Pu, J. Chen, S. Li, and X. Liao, "An ultrasound standard plane detection model of fetal head based on multi-task learning and hybrid knowledge graph," *Future Generation Computer Systems*, vol. 135, pp. 234–243, Oct. 2022, doi: 10.1016/J.FUTURE.2022.04.011.
- [11] X. Wang, W. Wang, and X. Cai, "Automatic measurement of fetal head circumference using a novel GCN-assisted deep convolutional network," *Comput. Biol. Med.*, vol. 145, p. 105515, Jun. 2022, doi: 10.1016/J.COMPBIOMED.2022.105515.

- [12] T. B. Krishna and P. Kokil, "Automated Detection of Common Maternal Fetal Ultrasound Planes Using Deep Feature Fusion," *INDICON 2022 - 2022 IEEE 19th India Council International Conference*, 2022, doi: 10.1109/INDICON56171.2022.10039879.
- [13] F. Rauf *et al.*, "Automated deep bottleneck residual 82-layered architecture with Bayesian optimization for the classification of brain and common maternal fetal ultrasound planes," *Front. Med. (Lausanne)*, vol. 10, 2023, doi: 10.3389/FMED.2023.1330218.
- [14] L. Zhao, K. Li, B. Pu, J. Chen, S. Li, and X. Liao, "An ultrasound standard plane detection model of fetal head based on multi-task learning and hybrid knowledge graph," *Future Generation Computer Systems*, vol. 135, pp. 234–243, Oct. 2022, doi: 10.1016/J.FUTURE.2022.04.011.
- [15] M. T. Pérez, E. P. De Andrés, L. H. Sánchez, M. C. Garcia, A. P. S. Romero, and E. C. Fuentes, "La física en la ecografía," *Seram*, Nov. 2018, Accessed: Jun. 15, 2026. [Online]. Available: <https://piper.espacio-seram.com/index.php/seram/article/view/1578>
- [16] F. Gary Cunningham, Kenneth J. Leveno, Jodi S. Dashe, Barbara L. Hoffman, Catherine Y. Spong, and Brian M. Casey, *Williams Obstetricia*, 26a ed. McGraw Hill Education, 2022.
- [17] "Guía de práctica clínica: Diagnóstico prenatal de los defectos congénitos. Cribado de anomalías cromosómicas," *Diagnostico Prenatal*, vol. 24, no. 2, pp. 57–72, Apr. 2013, doi: 10.1016/j.diapre.2012.06.013.
- [18] C. Creighton, "A literature review on communication between picture archiving and communication systems and radiology information systems and/or hospital information systems," *J. Digit. Imaging*, vol. 12, no. 3, pp. 138–143, 1999, doi: 10.1007/BF03168632.
- [19] J. C. Honeyman, "Information systems integration in radiology," *J. Digit. Imaging*, vol. 12, no. Suppl 1, p. 218, 1999, doi: 10.1007/BF03168810.
- [20] H. R. Torres *et al.*, "A review of image processing methods for fetal head and brain analysis in ultrasound images," *Comput. Methods Programs Biomed.*, vol. 215, Mar. 2022, doi: 10.1016/j.cmpb.2022.106629.
- [21] B. Zhang, H. Liu, H. Luo, and K. Li, "Automatic quality assessment for 2D fetal sonographic standard plane based on multitask learning," *Medicine*, vol. 100, no. 4, Jan. 2021, doi: 10.1097/MD.00000000000024427.
- [22] B. S. Prabakaran, P. Hamelmann, E. Ostrowski, and M. Shafique, "FPUS23: An Ultrasound Fetus Phantom Dataset with Deep Neural Network Evaluations for Fetus Orientations, Fetal Planes, and Anatomical Features," *IEEE Access*, vol. 11, pp. 58308–58317, Jun. 2023, doi: 10.1109/ACCESS.2023.3284315.
- [23] T. B. Krishna and P. Kokil, "Automated classification of common maternal fetal ultrasound planes using multi-layer perceptron with deep feature integration," *Biomed. Signal Process. Control*, vol. 86, p. 105283, Sep. 2023, doi: 10.1016/J.BSPC.2023.105283.

- [24] G. Dubey, S. Srivastava, A. K. Jayswal, M. Saraswat, P. Singh, and M. Memoria, "Fetal Ultrasound Segmentation and Measurements Using Appearance and Shape Prior Based Density Regression with Deep CNN and Robust Ellipse Fitting," *Journal of Imaging Informatics in Medicine*, vol. 37, no. 1, p. 247, Feb. 2024, doi: 10.1007/S10278-023-00908-8.
- [25] T. B. Krishna and P. Kokil, "Standard fetal ultrasound plane classification based on stacked ensemble of deep learning models," *Expert Syst. Appl.*, vol. 238, p. 122153, Mar. 2024, doi: 10.1016/J.ESWA.2023.122153.
- [26] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into Deep Learning," Aug. 2023.
- [27] F. Rauf *et al.*, "Automated deep bottleneck residual 82-layered architecture with Bayesian optimization for the classification of brain and common maternal fetal ultrasound planes," *Front. Med. (Lausanne)*, vol. 10, 2023, doi: 10.3389/FMED.2023.1330218.
- [28] L. Zhao, K. Li, B. Pu, J. Chen, S. Li, and X. Liao, "An ultrasound standard plane detection model of fetal head based on multi-task learning and hybrid knowledge graph," *Future Generation Computer Systems*, vol. 135, pp. 234–243, Oct. 2022, doi: 10.1016/J.FUTURE.2022.04.011.
- [29] M. México *et al.*, "Métricas de rendimiento para evaluar el aprendizaje automático en la clasificación de imágenes petroleras utilizando redes neuronales convolucionales," *Ciencia Latina Revista Científica Multidisciplinar*, vol. 6, no. 5, pp. 4624–4637, Nov. 2022, doi: 10.37811/CL_RCM.V6I5.3420.
- [30] J. Ma, Y. Ding, J. C. P. Cheng, Y. Tan, V. J. L. Gan, and J. Zhang, "Analyzing the Leading Causes of Traffic Fatalities Using XGBoost and Grid-Based Analysis: A City Management Perspective," *IEEE Access*, vol. 7, pp. 148059–148072, 2019, doi: 10.1109/ACCESS.2019.2946401.
- [31] A. León-Batallas, J. Bermeo-Paucar, J. Paredes-Quevedo, and H. Torres-Ordoñez, "Una revisión de las métricas aplicadas en el procesamiento de imágenes," *RECIMUNDO*, vol. 4, no. 3, pp. 267–273, Sep. 2020, doi: 10.26820/recimundo/4.(3).julio.2020.267-273.
- [32] J. Terven, D. M. Cordova-Esparza, J. A. Romero-González, A. Ramírez-Pedraza, and E. A. Chávez-Urbiola, "Loss Functions and Metrics in Deep Learning," *Artif. Intell. Rev.*, vol. 58, no. 7, Jul. 2023, doi: 10.1007/s10462-025-11198-7.
- [33] F. A. Mohammed, K. K. Tune, B. G. Assefa, M. Jett, and S. Muhie, "Medical Image Classifications Using Convolutional Neural Networks: A Survey of Current Methods and Statistical Modeling of the Literature," *Mach. Learn. Knowl. Extr.*, vol. 6, no. 1, pp. 699–735, Mar. 2024, doi: 10.3390/MAKE6010033/S1.
- [34] M. Alzubaidi, U. Shah, M. Agus, and M. Househ, "FetSAM: Advanced Segmentation Techniques for Fetal Head Biometrics in Ultrasound Imagery," *IEEE Open J. Eng. Med. Biol.*, vol. 5, pp. 281–295, 2024, doi: 10.1109/OJEMB.2024.3382487.
- [35] F. Chen, L. Chen, H. Han, S. Zhang, D. Zhang, and H. Liao, "The ability of Segmenting Anything Model (SAM) to segment ultrasound images," *Biosci. Trends*, vol. 17, no. 3, pp. 211–218, 2023, doi: 10.5582/BST.2023.01128,.

- [36] J. E. Figueroa Rivera, "La importancia de implementar PACS," *Revista Peruana de Radiología*, vol. 23, 2024.
- [37] A. Lasala, M. C. Fiorentino, S. Micera, A. Bandini, and S. Moccia, "Exploiting class activation mappings as prior to generate fetal brain ultrasound images with GANs," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2023, 2023, doi: 10.1109/EMBC40787.2023.10340469.
- [38] A. Lasala, M. C. Fiorentino, S. Micera, A. Bandini, and S. Moccia, "Exploiting class activation mappings as prior to generate fetal brain ultrasound images with GANs," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2023, doi: 10.1109/EMBC40787.2023.10340469.
- [39] H. R. Torres *et al.*, "A review of image processing methods for fetal head and brain analysis in ultrasound images," *Comput. Methods Programs Biomed.*, vol. 215, Mar. 2022, doi: 10.1016/j.cmpb.2022.106629.
- [40] X. Shu, Y. Gu, X. Zhang, C. Hu, and K. Cheng, "FCRB U-Net: A novel fully connected residual block U-Net for fetal cerebellum ultrasound image segmentation," *Comput. Biol. Med.*, vol. 148, p. 105693, Sep. 2022, doi: 10.1016/J.COMPBIOMED.2022.105693.
- [41] Z. Yu *et al.*, "A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 3, pp. 874–885, May 2018, doi: 10.1109/JBHI.2017.2705031.
- [42] M. Alzubaidi, M. Agus, M. Makhoul, F. Anver, K. Alyafei, and M. Househ, "Large-scale annotation dataset for fetal head biometry in ultrasound images," *Data Brief*, vol. 51, p. 109708, Dec. 2023, doi: 10.1016/J.DIB.2023.109708.
- [43] X. P. Burgos-Artizzu *et al.*, "FETAL_PLANES_DB: Common maternal-fetal ultrasound images", doi: 10.5281/ZENODO.3904280.
- [44] Z. Wang *et al.*, "A Comprehensive Survey on Data Augmentation," May 2024, doi: 10.1109/TKDE.2025.3622600.
- [45] X. Wang, W. Wang, and X. Cai, "Automatic measurement of fetal head circumference using a novel GCN-assisted deep convolutional network," *Comput. Biol. Med.*, vol. 145, Jun. 2022, doi: 10.1016/j.compbiomed.2022.105515.
- [46] B. Zhang, H. Liu, H. Luo, and K. Li, "Automatic quality assessment for 2D fetal sonographic standard plane based on multitask learning," *Medicine*, vol. 100, no. 4, Jan. 2021, doi: 10.1097/MD.0000000000024427.
- [47] T. B. Krishna and P. Kokil, "Standard fetal ultrasound plane classification based on stacked ensemble of deep learning models," *Expert Syst. Appl.*, vol. 238, Mar. 2024, doi: 10.1016/j.eswa.2023.122153.
- [48] X. Shu, F. Chang, X. Zhang, C. Shao, and X. Yang, "ECAU-Net: Efficient channel attention U-Net for fetal ultrasound cerebellum segmentation," *Biomed. Signal Process. Control*, vol. 75, p. 103528, May 2022, doi: 10.1016/J.BSPC.2022.103528.
- [49] J. Camacho, "A Set of Rules for Model Validation," *J. Chemom.*, vol. 40, no. 2, Feb. 2026, doi: 10.1002/cem.70110.

- [50] S. Korkmaz, D. Goksuluk, and E. Karaismailoglu, “fastml: Guarded Resampling Workflows for Safer Automated Machine Learning in R,” Apr. 2026.
- [51] Y. Zeng, P. H. Tsui, W. Wu, Z. Zhou, and S. Wu, “Fetal Ultrasound Image Segmentation for Automatic Head Circumference Biometry Using Deeply Supervised Attention-Gated V-Net,” *J. Digit. Imaging*, vol. 34, no. 1, pp. 134–148, Feb. 2021, doi: 10.1007/S10278-020-00410-5,.
- [52] N. A. Al-Humaidan and M. Prince, “A Classification of Arab Ethnicity Based on Face Image Using Deep Learning Approach,” *IEEE Access*, vol. 9, pp. 50755–50766, 2021, doi: 10.1109/ACCESS.2021.3069022.
- [53] A. R. Philip, “Fine-Tuning Depth Analysis: Identifying the Sweet Spot for Maximum Accuracy in CNNs,” *International Journal of Scientific Research in Computer Science and Engineering*, vol. 13, no. 3, pp. 48–63, Jun. 2025, doi: 10.26438/IJSRCSE.V13I3.703.
- [54] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Apr. 2015.
- [55] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Sep. 2014, Accessed: Dec. 07, 2025. [Online]. Available: <https://arxiv.org/pdf/1409.1556>
- [56] T. R. Reddy and J. P. J. Kumar, “VGG16 supports skin melanoma classification compared to the Alexnet algorithm,” *AIP Conf. Proc.*, vol. 3300, no. 1, Jul. 2025, doi: 10.1063/5.0277851/3356737.
- [57] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, “Dive into Deep Learning,” *Journal of the American College of Radiology*, vol. 17, no. 5, pp. 637–638, Jun. 2021, doi: 10.1016/j.jacr.2020.02.005.
- [58] Y. Akkem, B. S. Kumar, and A. Varanasi, “Streamlit Application for Advanced Ensemble Learning Methods in Crop Recommendation Systems – A Review and Implementation,” *Indian J. Sci. Technol.*, vol. 16, no. 48, pp. 4688–4702, Dec. 2023, doi: 10.17485/IJST/V16I48.2850.
- [59] I. D. Mienye and Y. Sun, “A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects,” *IEEE Access*, vol. 10, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.
- [60] D. Mishra, S. M. Tripathi, A. Chaurasia, and P. K. Chaurasia, “A Review on Ensemble Learning Methods: Machine Learning Approach,” *International Journal of Research Publication and Reviews*, vol. 6, no. 2, pp. 3795–3803, Feb. 2025, doi: 10.55248/GENGPI.6.0225.0971.
- [61] G. N. Tornetta, “Entropy methods for the confidence assessment of probabilistic classification models,” Mar. 2021.