



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado

Facultad de Ingeniería de Sistemas e Informática

Unidad de Posgrado

**Detección de displasia del desarrollo de cadera por
radiografía pélvica en infantes de 03 a 06 meses de
edad usando aprendizaje profundo**

TESIS

Para optar el Grado Académico de Magíster en Ingeniería de
Sistemas e Informática con mención en Ingeniería de Software

AUTOR

Ricardo Junior JARA GASPAR

ASESOR

Mg. Fany Yexenia SOBERO RODRÍGUEZ

Lima, Perú

2024



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Jara, R. (2024). *Detección de displasia del desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad usando aprendizaje profundo*. [Tesis de maestría, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Unidad de Posgrado]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Ricardo Junior Jara Gaspar
Tipo de documento de identidad	DNI
Número de documento de identidad	46842959
URL de ORCID	https://orcid.org/0000-0001-8914-4833
Datos de asesor	
Nombres y apellidos	Fany Yexenia Sobero Rodríguez
Tipo de documento de identidad	DNI
Número de documento de identidad	20120467
URL de ORCID	https://orcid.org/0000-0002-0323-6110
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Luis Soto Soto
Tipo de documento	DNI
Número de documento de identidad	22409532
Miembro del jurado 1	
Nombres y apellidos	Juan Gamarra Moreno
Tipo de documento	DNI
Número de documento de identidad	20039857
Miembro del jurado 2	
Nombres y apellidos	Cesar Luza Montero
Tipo de documento	DNI
Número de documento de identidad	06111988
Datos de investigación	

Línea de investigación	C.0.5.8. Inteligencia artificial
Grupo de investigación	Inteligencia Artificial - IA
Agencia de financiamiento	Sin financiamiento
Ubicación geográfica de la investigación	Edificio: Universidad Nacional Mayor de San Marcos País: Perú Departamento: Lima Provincia: Lima Distrito: Cercado de Lima Latitud: -12.055926 Longitud: -77.083469
Año o rango de años en que se realizó la investigación	2021 - 2023
URL de disciplinas OCDE	Ingeniería de sistemas y comunicaciones https://purl.org/pe-repo/ocde/ford#2.02.04



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú. Decana de América
Facultad de Ingeniería de Sistemas e Informática
Vicedecanato de Investigación y Posgrado
Unidad de Posgrado

**ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR EL GRADO ACADÉMICO
DE MAGÍSTER INGENIERÍA DE SISTEMAS E INFORMÁTICA CON MENCIÓN
EN INGENIERÍA DE SOFTWARE**

A los diecisiete (17) días del mes de octubre de 2024, siendo las 10:00 am., se reunieron en el Auditorio Alfredo Celso Alva Bravo, el Jurado de Tesis conformado por los siguientes docentes:

Dr. Luis Soto Soto (Presidente)
Mg. Juan Gamarra Moreno (Miembro)
Mg. Cesar Luza Montero (Miembro)
Mg. Fany Yexenia Sobero Rodríguez (Miembro Asesor)

Se inició la Sustentación invitando al candidato a Magíster **RICARDO JUNIOR JARA GASPAS**, para que realice la exposición oral de la tesis para optar el Grado Académico de Magíster en Ingeniería de Sistemas e Informática con mención en Ingeniería de Software, siendo la Tesis intitulada:

**“DETECCIÓN DE DISPLASIA DEL DESARROLLO DE CADERA POR RADIOGRAFÍA
PÉLVICA EN INFANTES DE 03 A 06 MESES DE EDAD USANDO APRENDIZAJE
PROFUNDO”**

Concluida la exposición, los miembros del Jurado de Tesis procedieron a formular sus preguntas que fueron absueltas por el graduando; acto seguido se procedió a la evaluación correspondiente, habiendo obtenido la siguiente calificación:

Dieciocho (18) Muy bueno

Por tanto, el presidente del Jurado, de acuerdo con el Reglamento General de Estudios de Posgrado, otorga al Bachiller **RICARDO JUNIOR JARA GASPAS** el Grado Académico de Magíster en Ingeniería de Sistemas e Informática con mención en Ingeniería de Software.

Siendo las 11:00 horas, el presidente del Jurado de Tesis, da por concluido el acto académico de Sustentación de Tesis.


Dr. Luis Soto Soto
(Presidente)


Mg. Juan Gamarra Moreno
(Miembro)


Mg. Cesar Luza Montero
(Miembro)


Mg. Fany Yexenia Sobero Rodríguez
(Miembro Asesor)

CERTIFICADO DE SIMILITUD

Yo Fany Yexenia Sobero Rodríguez en mi condición de asesor acreditado con Dictamen N° 000295-2023-UPG-VDIP-FISI/UNMSM de la tesis de investigación, cuyo título es **DETECCIÓN DE DISPLASIA DEL DESARROLLO DE CADERA POR RADIOGRAFÍA PÉLVICA EN INFANTES DE 03 A 06 MESES DE EDAD USANDO APRENDIZAJE PROFUNDO** presentado por el bachiller RICARDO JUNIOR JARA GASPAR para optar el grado de Magister en Ingeniería de Sistemas e Informática con mención en Ingeniería de Software, CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante el software de similitud textual, el documento evaluado cuenta con el porcentaje de 11% de similitud, nivel **PERMITIDO** para continuar con los trámites correspondientes y para su **publicación en el repositorio institucional.**

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del grado correspondiente.

Firma del Asesor  Firmado digitalmente por SOBERO RODRIGUEZ Fany Yexenia FAU
20148092282 soft
Motivo: Soy el autor del documento
Fecha: 13.06.2024 12:43:35 -05:00

DNI: 20120467

Nombres y apellidos del asesor: Fany Yexenia Sobero Rodríguez

DEDICATORIA

A Danna, mi amada esposa, por su constante apoyo y su continuo ímpetu de encontrar las palabras de aliento para seguir adelante en la búsqueda de ser mejor persona y profesional.

AGRADECIMIENTOS

A Dios, por brindarme la oportunidad de alcanzar un logro más en mi vida profesional.

A la Mg. Fany Sobero Rodríguez, por su dedicación y paciencia, en su labor de brindar las asesorías y correcciones oportunas para la realización del presente trabajo.

A los profesores del programa de maestría de la UNMSM, por su compromiso de guía durante cada semestre académico, con sus experiencias y conocimientos fueron de gran soporte para la elaboración del presente trabajo.

ÍNDICE GENERAL

DEDICATORIA	II
AGRADECIMIENTOS	III
ÍNDICE GENERAL	IV
LISTA DE TABLAS	VIII
LISTA DE FIGURAS	IX
RESUMEN	X
ABSTRACT	XI
CAPÍTULO 1: INTRODUCCIÓN.....	1
1.1. Situación problemática	1
1.2. Formulación del problema	3
1.2.1. <i>Problema general</i>	3
1.2.2. <i>Problemas específicos</i>	3
1.3. Justificación de la Investigación	4
1.3.1. <i>Justificación teórica</i>	4
1.3.2. <i>Justificación práctica</i>	5
1.4. Objetivos de la investigación	7
1.4.1. <i>Objetivo general</i>	7
1.4.2. <i>Objetivos específicos</i>	7
1.5. Contribuciones	7
1.6. Estructura del documento	8
CAPÍTULO 2: MARCO TEÓRICO.....	9
2.1. Antecedentes de investigación	9
2.2. Bases teóricas	16
2.2.1. <i>Displasia del desarrollo de cadera (DDC)</i>	16
2.2.2. <i>Diagnóstico de DDC</i>	17

2.2.3.	<i>Diagnóstico de DDC por radiografía</i>	17
2.2.4.	<i>Métodos clásicos de aprendizaje automático</i>	19
2.2.5.	<i>Arboles de decisión</i>	20
2.2.6.	<i>Máquinas de vectores de soporte (SVM)</i>	21
2.2.7.	<i>Aprendizaje profundo</i>	21
2.2.8.	<i>Aprendizaje supervisado</i>	23
2.2.9.	<i>Aprendizaje no supervisado</i>	23
2.2.10.	<i>Clasificación de aprendizaje supervisado</i>	24
2.2.11.	<i>Arquitecturas de aprendizaje profundo</i>	25
2.2.12.	<i>Redes generativas adversarias (GAN)</i>	27
2.2.13.	<i>Generador de una GAN</i>	28
2.2.14.	<i>Discriminador de una GAN</i>	28
2.2.15.	<i>Función objetivo de una GAN</i>	29
2.2.16.	<i>Red neuronal convolucional (CNN)</i>	29
2.2.17.	<i>Capa de convolución (Conv2D)</i>	31
2.2.18.	<i>Capa de pooling (MaxPool2D)</i>	31
2.2.19.	<i>Capa de aplanamiento (Flatten)</i>	31
2.2.20.	<i>Capa densa (Fully Connected)</i>	31
2.2.21.	<i>ResNet50</i>	32
2.2.22.	<i>Modelos híbridos</i>	33
2.2.23.	<i>Métricas de evaluación</i>	34
2.2.24.	<i>Métrica de Precisión</i>	34
2.2.25.	<i>Métrica de Sensibilidad</i>	34
2.2.26.	<i>Métrica de Especificidad</i>	35
2.2.27.	<i>Curva característica operativa del receptor</i>	35
2.2.28.	<i>Matriz de confusión</i>	36
2.3.	<i>Marcos Conceptuales</i>	37

CAPÍTULO 3: METODOLOGÍA	38
3.1. Tipo y diseño de investigación	38
3.2. Unidad de análisis	39
3.3. Población de estudio	39
3.4. Tamaño de muestra	39
3.5. Selección de muestra	39
3.6. Técnicas de recolección de datos	39
3.7. Metodología de implementación	40
CAPÍTULO 4: DETECCIÓN DE DDC	42
4.1. Aporte	42
4.2. Conjunto de datos	44
4.3. Preprocesamiento	44
4.3.1. Pasos	44
4.3.2. Implementación	46
4.3.3. Justificación y novedad	46
4.4. Aumento de datos	47
4.4.1. Implementación con transformaciones simples	47
4.4.2. Implementación con GAN	47
4.5. Entrenamiento y evaluación	50
4.5.1. Entrenamiento y evaluación de SVM	51
4.5.2. Entrenamiento y evaluación de árboles de decisión	52
4.5.3. Entrenamiento y evaluación de CNN genérica	52
4.5.4. Entrenamiento y evaluación de arquitectura híbrida	54
4.5.5. Justificación y novedad	57
4.6. Clasificación	58
4.6.1. Implementación del modelo seleccionado	58
4.6.2. Justificación y novedad	59

CAPÍTULO 5. RESULTADOS Y DISCUSIONES	61
4.7. Análisis, interpretación y discusión de resultados	61
4.7.1. Resultados con los métodos clásicos	61
4.7.2. Resultados con la CNN genérica	66
4.7.3. Resultados con el modelo híbrido propuesto	68
4.8. Pruebas de hipótesis.....	69
4.8.1. Hipótesis general	70
4.8.2. Hipótesis específicas.....	70
4.9. Presentación de resultados	71
4.9.1. Resultados con aumento de datos con transformaciones simples.....	71
4.9.2. Resultados con aumento de datos con la GAN.....	72
4.9.3. Resultados con el sistema propuesto.....	73
CONCLUSIONES Y TRABAJOS FUTUROS.....	76
RECOMENDACIONES.....	80
REFERENCIAS BIBLIOGRÁFICAS	82
ANEXOS.....	93
Anexo 1. Operacionalización de variables	93
Anexo 2. Matriz de consistencia	94

LISTA DE TABLAS

<i>Tabla 1. Especialistas en el diagnóstico de DDC por imagen radiográfica.</i>	
.....	6
<i>Tabla 2. Comparación de diferentes arquitecturas de aprendizaje profundo.</i>	
.....	26
<i>Tabla 3. Parámetros configurados para cada algoritmo.</i>	
.....	51
<i>Tabla 4. Capas de CNN genérica.</i>	
.....	53
<i>Tabla 5. Comparación de los resultados de los diferentes modelos entrenados con los datos aumentados con transformaciones simples.</i>	
.....	72
<i>Tabla 6. Comparación de los resultados de los diferentes modelos entrenados con los datos aumentados con la GAN.</i>	
.....	73

LISTA DE FIGURAS

<i>Figura 1. Comparativa de estructura acetabulares: (A) Acetábulo displásico; (B) Acetábulo en posición fisiológica</i>	<i>16</i>
<i>Figura 2. Comparativa anatómica entre una cadera normal y luxada.....</i>	<i>18</i>
<i>Figura 3. Radiografías de pacientes: (a) Con DDC; (b) Sin DDC.</i>	<i>19</i>
<i>Figura 4. Árbol de decisión para la detección de imágenes con números del dataset MNIST.</i>	<i>20</i>
<i>Figura 5. Panorama de aprendizaje profundo.</i>	<i>22</i>
<i>Figura 6. Arquitectura de una CNN</i>	<i>30</i>
<i>Figura 7. Matriz de confusión.</i>	<i>36</i>
<i>Figura 8. Método de detección de DDC.</i>	<i>42</i>
<i>Figura 9. Arquitectura propuesta de una GAN para el aumento de datos.</i>	<i>48</i>
<i>Figura 10. Arquitectura para la CNN genérica.....</i>	<i>53</i>
<i>Figura 11. Arquitectura híbrida propuesta.</i>	<i>55</i>
<i>Figura 12. Frontend de la aplicación para la detección de la DDC.....</i>	<i>59</i>
<i>Figura 13. Matrices de confusión de la Máquina de Vectores de Soporte (SVM). a) Para los datos de entrenamiento. b) Para los datos de validación.</i>	<i>62</i>
<i>Figura 14. Matrices de confusión del Árbol de decisión. a) Para los datos de entrenamiento. b) Para los datos de validación.....</i>	<i>63</i>
<i>Figura 15. Curva ROC para a) La Máquina de Vectores de Soporte, y b) El Árbol de decisión.....</i>	<i>65</i>
<i>Figura 16.a) Matriz de confusión y b) Curva ROC para la CNN genérica</i>	<i>67</i>
<i>Figura 17.a) Matriz de confusión y b) Curva ROC para híbrido propuesto</i>	<i>69</i>
<i>Figura 18. Resultados de la clasificación con el sistema propuesto.....</i>	<i>74</i>
<i>Figura 19. Probabilidades obtenidas por clase del sistema propuesto..</i>	<i>75</i>

RESUMEN

La Displasia en el Desarrollo de Cadera (DDC) es una condición que puede afectar la calidad de vida de una persona. En infantes de tres a seis meses de edad, esta afección puede diagnosticarse y tratarse de forma eficiente. Si, por otro lado, la displasia no se detecta en estas etapas de vida, es muy posible que el infante con displasia requiera de un tratamiento más agresivo en etapas posteriores de su vida. Para detectar la DDC en infantes, un profesional de la salud puede valerse de procedimientos de diagnóstico clínico como la observación, e incluso de imagenología como imágenes de rayos x de la cadera. Una desventaja suele ser la falta de profesionales de la salud diestros en el análisis de las imágenes de rayos x de caderas para el diagnóstico de la DDC, además de la falta de un consenso general sobre los grados de desviación requeridos para el diagnóstico de la displasia. En este proyecto de tesis, el objetivo es usar técnicas de aprendizaje profundo para la detección de la displasia de cadera en infantes de tres a seis meses de edad usando radiografías de rayos x de la cadera. El modelo propuesto es un modelo híbrido que combina las capas de un Red Neuronal Convolutiva entrenada con un vector de características, y un modelo con la arquitectura ResNet 50. Para comparar el modelo propuesto, se entrenaron también modelos de aprendizaje de máquina como el árbol de decisión y la máquina de vectores de soporte. De forma similar, los resultados del modelo propuesto se compararon con los resultados de una Red Neuronal Convolutiva simple. Por otro lado, para resolver el problema de escasez de datos, común en aplicaciones médicas, el modelo propuesto fue entrenado con datos generados por la combinación de transformaciones geométricas y una Red Generativa Adversaria. Los resultados finales muestran que el sistema propuesto presenta una precisión de 97.50%, una sensibilidad de 95.16% y una especificidad del 100%. Además, con el fin de exponer el modelo propuesto, se diseñó y desplegó una aplicación en la nube.

Palabras clave: Displasia de cadera, Aprendizaje profundo, Redes Generativas Adversarias.

ABSTRACT

Dysplasia in hip development is a condition that can impact a person's quality of life. In infants aged three to six months, this condition can be diagnosed and efficiently treated. However, if hip dysplasia is not detected during these early stages of life, it is highly likely that the infant with dysplasia will require more aggressive treatment in later stages of life. To detect hip dysplasia in infants, a healthcare professional may utilize clinical diagnostic procedures such as observation, and even imaging techniques like X-ray images of the hip. A common disadvantage is the lack of skilled healthcare professionals in analyzing hip X-ray images for DDH diagnosis, as well as the absence of a consensus on the degree of deviation required for the diagnosis of dysplasia, along with a lack of consensus on the required degrees of deviation for diagnosing dysplasia. In this thesis project, the objective is to use deep learning methods for the detection of hip dysplasia in children aged three to six months using X-ray hip radiographs. The proposed model is a hybrid model that combines layers of a Convolutional Neural Network trained with a feature vector, and a model with the ResNet 50 architecture. To compare the proposed model, machine learning models such as decision tree and support vector machine were also trained. Similarly, the results of the proposed model were compared with the results of a simple Convolutional Neural Network. Furthermore, to address the data scarcity issue common in medical applications, the proposed model was trained with data generated through a combination of geometric transformations and a Generative Adversarial Network. The results show that the proposed system has an accuracy of 97.50%, a sensitivity of 95.16%, and a specificity of 100%. Additionally, to showcase the proposed model, a cloud-based application was designed and deployed.

Keywords: Hip dysplasia, Deep learning, Generative Adversarial Networks.

CAPÍTULO 1: INTRODUCCIÓN

El presente capítulo tiene como objetivo proporcionar una visión general de la situación problemática que se abordará en esta tesis, así como establecer los objetivos y la relevancia de la investigación.

1.1. Situación problemática

La displasia del desarrollo de la cadera (DDC) es una alteración que conlleva a diferentes deformaciones en la estructura de la articulación en la cadera, se caracteriza por presentar variantes patológicas en el receptáculo acetabular y/o de la cabeza femoral, junto con irregularidades en los ligamentos y la cápsula articular, lo que ocasiona que la cadera sea inestable, pueda luxarse fácilmente o no se desarrolle adecuadamente (Longo et al., 2021).

En Perú, un informe con datos del 2014 de la Clínica San Juan de Dios indicó la presencia de 4,794 casos con DDC en niños de 03 meses a 14 años (Quispe, 2019), lo que sugiere que las cifras peruanas podrían alinearse con las tendencias observadas a nivel global en relación con esta condición. Por ejemplo, Brenes et., al (2020) indican prevalencias globales estimadas entre 1 y 34 por cada 1,000 neonatos. Por otro lado, el Departamento de Tecnología de la Información del Sistema Público de Salud de Brasil registró 14,584 casos de DDC en un periodo de 10 años, resultando en intervenciones quirúrgicas con un significativo impacto económico y social (Souza et al., 2021).

Los diagnósticos tardíos aumentan los riesgos asociados, como la artrosis de cadera precoz, que afecta seriamente la calidad de vida, limitando la movilidad y causando dolores constantes (Fludder et al., 2023). En caso extremos, puede provocar cojera y restricciones funcionales irreversibles, comprometiendo la independencia y el bienestar general de la persona

(Agostiniani et al., 2020). Para prevenir estas consecuencias, es crucial lograr una detección temprana y precisa, lo cual se consigue a través de exámenes clínicos meticulosos y estudios de imagenología avanzada (Saeed & Altamimi, 2021).

Los estudios médicos resaltan la necesidad de protocolos para la detección en etapas tempranas de la DDC. Agostiniani et al. (2020), enfatizan la importancia de radiografías antes de los seis meses para resultados funcionales favorables. Barrera et al. (2019), destacan las evaluaciones neonatales inmediatas, mientras Kitay et al. (2019), considera el ultrasonido viable a los seis meses, aunque prefieren las radiografías para mayor exactitud. Por otro lado, Swarup et al. (2018), abogan por exámenes clínicos y de imagenología en los primeros meses de vida, especialmente en casos con anomalías físicas o factores de riesgo. Sin embargo, estas investigaciones señalan que la dependencia de la disponibilidad especializada en análisis de imágenes médicas es un obstáculo para detecciones tempranas.

La revisión de investigaciones anteriores destaca que la identificación de la displasia de desarrollo de la cadera se basa principalmente en la habilidad y experiencia del radiólogo, lo cual puede ser subjetivo y propenso a errores. El aprendizaje profundo, y en particular las redes neuronales convolucionales se presentan como soluciones prometedoras, automatizando la identificación de la DDC en radiografías pélvicas. Además, puede reducir la dependencia de personal especializado, superando obstáculos en la detección temprana. Investigaciones recientes, incluidas las realizadas por Cheng et al. (2021), Liu et al. (2020), J. Xu et al. (2022), Den et al. (2023), y Fraiwan et al. (2022) han confirmado la efectividad del uso de técnicas de aprendizaje profundo para identificar y categorizar anomalías de la cadera en imágenes radiográficas pélvicas.

Los resultados indican que el uso del aprendizaje profundo podría constituir una herramienta significativa para potenciar el diagnóstico médico, especialmente en el contexto de la displasia de desarrollo de la cadera. Es

esencial la creación de un modelo específico de aprendizaje profundo dirigido a detectar la DDC en radiografías de pelvis, especialmente en infantes de 03 a 06 meses de edad, con el objetivo de mejorar la precisión en el diagnóstico y el tratamiento temprano de esta condición. Sin embargo, es importante resaltar que uno de los desafíos adicionales en este proceso radica en la escasez de datos adecuados y de alta calidad para el entrenamiento de estos modelos. Por consiguiente, la generación de datos sintéticos de calidad se vuelve fundamental en el desarrollo de modelos de aprendizaje profundo eficientes para la detección de la displasia de desarrollo de la cadera. Además, la implementación exitosa de esta tecnología también implica el modelado y entrenamiento de las redes neuronales convolucionales de manera adecuada.

1.2. Formulación del problema

1.2.1. Problema general

¿Es posible que un modelo basado en aprendizaje profundo contribuya a la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad?

1.2.2. Problemas específicos

Con la finalidad de abordar el problema principal de esta investigación, es necesario considerar los siguientes problemas específicos:

- ¿Es posible que un modelo basado en aprendizaje profundo contribuya con aceptable grado de Precisión en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad?
- ¿Es posible que un modelo basado en aprendizaje profundo contribuya con aceptable grado de Sensibilidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad?

- ¿Es posible que un modelo basado en aprendizaje profundo contribuya con aceptable grado de Especificidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad?
- ¿Es posible que un modelo basado en aprendizaje profundo contribuya con rendimiento óptimo en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad?

1.3. Justificación de la Investigación

1.3.1. Justificación teórica

Para la justificación teórica, se realizó una revisión de la literatura en Scopus y Web of Science, utilizando las palabras claves como “detection”, “recognition”, “detecting”, “diagnosis”, “diagnostic”, “recognition developmental dysplasia of the hip”, “detection developmental dysplasia of the hip” y “DDH”.

La mayoría de los estudios encontrados utilizan técnicas de procesamiento digital de imágenes, principalmente dirigidas a detectar diversas anomalías en la región de la cadera. Por ejemplo, Cheng et al. (2021), se enfocaron en la identificación de anomalías específicas mediante análisis de imágenes. Así mismo, Liu et al. (2020) y J. Xu et al. (2022), se centraron en la detección de malformaciones en la pelvis, haciendo uso también de enfoques basados en procesamiento de imágenes. En una línea similar, Al-Bashir et al. (2015), se destacaron por desarrollar un método de detección de DDC, aprovechando mediciones angulares obtenidas de las bordes acetabulares y el borde central extraídos de imágenes radiográficas. No obstante, es trascendental señalar que el panorama investigativo en este ámbito es limitado, lo que destaca la necesidad de ampliar la gama de enfoques y metodologías para abordar de manera más completa y precisa las diversas problemáticas relacionadas con las irregularidades en la cadera.

A partir de lo mencionado anteriormente, se resalta la importancia de llevar a cabo esta investigación, puesto que permitirá analizar en profundidad los diversos aspectos que los modelos de aprendizaje profundo aportan a la solución de este problema. Además, esta investigación busca proporcionar valiosos antecedentes que puedan servir de base para futuros estudios en esta área.

1.3.2. Justificación práctica

El empleo de un modelo de DL para la detección de la DDC en infantes a través de imágenes radiográficas emerge como una solución prometedora para afrontar la escasez de médicos especializados. Este enfoque también posibilita la expansión de la telesalud al permitir un diagnóstico oportuno (J. Xu et al., 2022). Además, se vislumbra la potencial consecuencia de disminuir la necesidad de intervenciones quirúrgicas correctivas, tal como se señala en el estudio de Husum et al. (2021). Por último, esta tecnología puede reducir los falsos positivos derivados de errores humanos, como se evidencia en la investigación de Philip et al. (2020).

La presente investigación tiene el potencial de abordar el desafío crítico que implica la escasez de profesionales médicos, especialmente en regiones remotas y de bajos recursos. Este problema se evidencia con la información de la Tabla 1, según el Colegio Médico del Perú (2022), que muestra una cifra de apenas 7,778 médicos especialistas facultados para realizar diagnósticos de DDC en infantes mediante radiografías pélvicas. Por lo general estos especialistas se encuentran en hospitales de gran envergadura en entornos urbanos, lo cual plantea dificultades considerables para que los infantes residentes en áreas rurales o ciudades de menor tamaño puedan acceder a diagnósticos profesionales y oportunos.

La incorporación de tecnologías de aprendizaje profundo en la detección de DDC puede revolucionar el paradigma actual del diagnóstico. Al proporcionar herramientas más precisas y accesibles, se pueden democratizar los servicios médicos, garantizando que ningún infante quede sin la posibilidad de un

diagnóstico temprano. Además, al reducir la carga sobre los especialistas, se pueden optimizar los recursos médicos existentes, permitiendo que estos profesionales se concentren en casos más complejos. En definitiva, la fusión de la medicina y la tecnología podría marcar el comienzo de una nueva era en la atención sanitaria, centrada en la prevención y la eficacia.

Tabla 1. Especialistas en el diagnóstico de DDC por imagen radiográfica.

Consejo Regional	Especialistas	%
CR II LIMA	4134	53,15%
CR I TRUJILLO	515	6,62%
CR V AREQUIPA	512	6,58%
CR XVIII CALLAO	319	4,10%
CR VIII CHICLAYO	301	3,87%
CR VI CUZCO	242	3,11%
CR IX ICA	222	2,85%
CR VII PIURA	200	2,57%
CR IV HUANCAYO	197	2,53%
CR XIV PUNO	151	1,94%
CR XIX CHIMBOTE	108	1,39%
CR XXXVII LIMA - PROVINCIAS	105	1,35%
CR XVII CAJAMARCA	102	1,31%
CR II IQUITOS	96	1,23%
CR XII TACNA	92	1,18%
CR X HUANUCO	77	0,99%
CR XV SAN MARTIN	58	0,75%
CR XIII PUCALLPA	57	0,73%
CR XVI AYACUCHO	52	0,67%
CR XXII APURIMAC	45	0,58%
CR XXI MOQUEGUA	39	0,50%
CR XI HUARAZ	34	0,44%
CR XXIV HUANCABELICA	32	0,41%
CR XXV AMANZONAS	24	0,31%
CR XXIII TUMBES	24	0,31%
CR XX PASCO	24	0,31%
CR XXVI MADRE DE DIOS	16	0,21%
TOTAL	7778	100,00%

Fuente. (Colegio Médico del Perú, 2022).

1.4. Objetivos de la investigación

1.4.1. *Objetivo general*

Desarrollar un modelo basado en aprendizaje profundo que contribuya a la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.

1.4.2. *Objetivos específicos*

- Desarrollar un modelo basado en aprendizaje profundo que contribuya con aceptable grado de Precisión en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.
- Desarrollar un modelo basado en aprendizaje profundo que contribuya con aceptable grado de Sensibilidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.
- Desarrollar un modelo basado en aprendizaje profundo que contribuya con aceptable grado de Especificidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.
- Desarrollar un modelo basado en aprendizaje profundo que contribuya con rendimiento óptimo en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.

1.5. Contribuciones

La tesis representa una contribución al campo del procesamiento de imágenes médicas al emplear técnicas avanzadas de aprendizaje profundo, al enfocarse en la detección temprana de la DDC en infantes de 03 a 06 meses a través de radiografías pélvicas. Mediante el desarrollo y validación de un modelo específico para este propósito, esta tesis propone un sistema con el potencial

para una detección más precisa y automatizada de la DDC, lo que podría revolucionar el enfoque de los problemas ortopédicos en infantes. Además, esta contribución no solo expande los límites de la aplicación del aprendizaje profundo en el ámbito médico, sino que también establece un valioso precedente para investigaciones futuras en el ámbito de la imagenología y la evaluación médica asistida por computadora. Cabe mencionar que se amplían los conjuntos de datos y se emplea un enfoque de modelo híbrido para lograr resultados aún más sólidos.

1.6. Estructura del documento

El resto del documento se organiza de la siguiente forma, el Capítulo 2, Marco Teórico, sienta las bases de los antecedentes y fundamentos teóricos del proyecto propuesto. El Capítulo 3, Metodología, presenta los pasos seguidos en esta tesis para el desarrollo del sistema de detección de la DDC propuesto. El aporte se describe en el Capítulo 4, mientras los resultados se exponen en el Capítulo 5. Por último, se presentan las conclusiones, trabajo futuro y recomendaciones.

CAPÍTULO 2: MARCO TEÓRICO

Este capítulo proporciona una revisión de la literatura existente relacionada con la DDC y las técnicas de aprendizaje profundo aplicadas en diagnósticos médicos. Se presentarán definiciones fundamentales, antecedentes, metodologías previamente empleadas y hallazgos significativos de investigaciones anteriores.

2.1. Antecedentes de investigación

Se han encontrado escasos estudios referentes a modelos de detección de la DDC que utilicen técnicas de aprendizaje profundo en imágenes radiográficas de cadera sobre infantes menores a seis meses de edad. Por ello, es importante la revisión de otras investigaciones que abordan otros enfoques en la detección de la DDC u otras anomalías de la cadera usando aprendizaje profundo.

Pei et al. (2023), plantean un método de detección de ocho puntos de referencias de la DDC en radiografías de pelvis, utilizando la técnica de mecanismo de atención con el fin de centrarse en elementos relevantes de la estructura de la cadera. Además, en el procesamiento de imágenes emplearon el algoritmo Non-local means, para la eliminación de ruidos en las radiografías. El método fue entrenado con un conjunto de datos de 524 radiografías de pelvis y evaluado mediante validación cruzada quíntuple. El método propuesto logró obtener un error promedio de punto a punto de 3.1350 mm, evidenciando mayor precisión de detección en comparación a U-Net, HR-Net y CE-Net.

Por su parte, Chen et al. (2022), propone un marco asistido por computadora basado en aprendizaje profundo con el objetivo de minimizar la intervención manual en el diagnóstico de la DDC de caderas Graf de tipo I y tipo II. El marco se fundamenta con el proceso de medición del método de Graf, técnica clínica basada en imágenes ecográficas para el diagnóstico de DDC. Para ello,

emplearon YOLOv3-tiny, para identificar cuatros estructuras anatómicas importantes de la imagen ecográfica, que fue fundamental con el propósito de la investigación, de detectar los planos estándar proporcionados por ecografistas y las mediciones automáticas de los ángulos de caderas. Para el entrenamiento y validación del modelo YOLOv3-tiny utilizaron 938 y 243 imágenes ecográficas de infantes menores a 06 meses, respectivamente. Sin embargo, aplicaron operaciones de aumentos de datos como escalado aleatorio, rotación aleatoria y ruido gaussiano, obteniendo finalmente un conjunto de imágenes de 19,698 para el proceso de entrenamiento. El marco propuesto fue sometido a validaciones como tasa de acuerdo de clasificación (precisión) y el coeficiente kappa de Cohen para determinar el rendimiento de clasificación. Los resultados logrados evidencia que el marco presenta mejor precisión frente a ecografistas experimentados.

De forma similar, W. Xu et al., (2022), plantean un sistema de diagnóstico para la medición y clasificación de la DDC usando técnicas de aprendizaje profundo en radiografías pélvicas pediátricas, bajo los enfoques clínicos de Tönnis y del Instituto Internacional de Displasia de Cadera (IHDI). Con ese fin, emplean Mask-RCNN para detectar los detalles específicos (features) de la imagen y segmentar la estructura anatómica relevante (la pelvis ósea). Así mismo, utilizan una red de alta resolución (HRNet) para el posicionamiento de cuatros puntos claves y la medición de parámetros radiográficos de la DDC. Para el entrenamiento del modelo consideraron 1,265 imágenes radiográficas y 133 para la validación, correspondiente a casos de infantes con edades de 06 meses a 03 años. Para la validación calcularon métricas de precisión, sensibilidad y especificidad, obteniendo valores entre 86% a 95%, 84% a 87% y 93%, respectivamente. Dichas métricas fueron comparadas con tres cirujanos ortopédicos. Del mismo, desde el punto de vista de confiabilidad y consumo de tiempo, las mediciones y clasificaciones obtenidas por el sistema propuesto fueron comparados con los cirujanos, sosteniendo que la propuesta puede medir de manera rápida y automática parámetros esenciales de la DDC.

Por otro lado, C. Li et al. (2022), proponen un método para medir los ángulos de rotación e inclinación de la cadera en imágenes radiográficas aplicando técnica de aprendizaje por transferencia y radiográficas reconstruidas digitalmente (DRR), con el propósito de evitar diagnóstico errores de la DDC. Con técnicas de DRR, generaron un total de 2,430 radiográficas pélvicas basándose en 30 imágenes de tomográficas computarizadas. Este nuevo conjunto de datos fue dividido en proporción de 2:1:1 para el entrenamiento, validación y pruebas. Así mismo, emplearon cinco redes previamente entrenadas, incluidas VGG16, Xception, VGG19, ResNet50 e InceptionV3 para desarrollar el método propuesto, los cuales fueron evaluados con el error absoluto medio para establecer su rendimiento. El modelo VGG16 obtuvo mejor rendimiento para predecir los ángulos de inclinación y rotación de la pelvis en radiografías de infantes.

Por su parte, Lin et al. (2022), aborda la detección la displasia de cadera, centrándose en las mediciones automáticas del ángulo del borde central lateral y el ángulo agudo de la articulación de la cadera usando aprendizaje profundo, a través de imágenes radiográficas. En ese sentido, utilizan Dense-U-net para segmentar el contorno femoral y los puntos clave acetabulares. Emplearon 768 imágenes radiográficas de caderas de adultos entre 20 y 72 años. Para evaluar el rendimiento del modelo propuesto aplicaron la técnica de validación cruzada de 10 veces. En cuanto, al conjunto de entrenamiento fueron sometido a operaciones de mejoras de datos incluido el giro horizontal, la rotación, cambios de brillo y recorte aleatorio. Realizaron comparaciones entre las mediciones de los ángulos por el modelo y médicos especialistas, en donde resaltan que no hubo diferencias estadísticamente significativas. La sensibilidad y precisión del diagnóstico de displasia de cadera basado en la medición automática con el modelo propuesto fue de 87,67% y 95,57%, respectivamente.

Así mismo, J. Xu et al. (2022), establece una herramienta de diagnóstico en línea basada en una red de reloj de arena multitarea que puede extraer con precisión puntos de referencia para detectar la dislocación de cadera y predecir los años de la cabeza femoral en las imágenes de rayos X.

Recopilaron 9 369 imágenes de radiografías de pelvis de pacientes entre 0,1 y 14 años. De las cuales, 7250 imágenes fueron utilizadas para entrenamiento y las 2119 imágenes restantes fueron para pruebas. Los experimentos muestran que el método propuesto alcanza un error de píxel medio de 4,64 en la detección de puntos de referencia y una precisión del 89% en la predicción de la edad. Por lo consiguiente, un modelo basado en la red de reloj de arena multitarea podría ser considerado un método confiable para detectar casos de DDC.

Por otra parte, S. Zhang et al. (2020), propone un sistema basado en aprendizaje profundo para diagnosticar la DDC mediante radiográficas pediátricas. Para ello, emplean un modelo de red neuronal convolucional ResNet-101, que es capaz de reconocer puntos claves de la estructura ósea de la cadera en las imágenes radiográficas con el fin de determinar la luxación y la medición del índice acetabular. Construyeron un conjunto de datos de 10,219 radiografías correspondiente a pacientes menores a 10 años, que fueron etiquetados, por un médico especialista, en base a los métodos clínicos de Hilgenreiner y Tönnis para clasificar con presencia o ausencia de dislocación. El conjunto de datos fue dividido en 8,000, 1,081 y 1,138 para el entrenamiento, validación y pruebas, respectivamente. En la evaluación del sistema lograron alcanzar una precisión de 99%, con sensibilidad de 95,5% y especificidad de 99.5%.

De modo similar, Q. Li et al. (2021), plantean un modelo para medir el ángulo acetabular mediante la identificación de puntos claves para apoyar al diagnóstico de DDC utilizando el algoritmo Mask R-CNN. Para el desarrollo de la investigación, se recogieron 13,228 imágenes de rayos X pélvicos anteroposteriores, de los cuales, 13,124 fueron empleados para el entrenamiento del modelo y 104 imágenes para validaciones. El modelo se verificó comparándolo con tres médicos ortopédicos. Los ángulos acetabular medios en la pelvis izquierda y derecha medidos por el modelo fue de $29,46 \pm 6,98^\circ$ y $27,92 \pm 6,56^\circ$, respectivamente, mientras que el ángulo acetabular medio medido por los médicos fue de $29,85 \pm 6,92^\circ$ y $27,75 \pm 6,45^\circ$,

respectivamente. El modelo demostró una alta consistencia en la medición del ángulo acetabular en comparación con los médicos.

En la investigación de Liu et al. (2020), exponen un método para apoyar al diagnóstico de DDC usando técnicas de aprendizaje profundo a partir de imágenes de radiografía de pelvis. El método se basa en la extracción de características morfológicas locales con redes neuronales convolucionales y la detección estructural globales con puntos de referencias deformes con Pyramid Non-local UNet (PN-UNet). Para el desarrollo del estudio, lograron recopilar 24000 imágenes de rayos X con daños en la cadera desde dislocación normal a severa, procedentes de infantes de 1 mes a 6 años. En consecuencia, del cribado estricto de un médico ortopedista, separaron 10000 imágenes, de las cuales, 8000 fueron asignados para entrenamiento y 2000 para pruebas. Los resultados experimentales muestran que el método PN-UNet logra una precisión de 90% para la detección de puntos de referencia de la DDC.

En Al-Bashir et al. (2015), proponen un algoritmo de medición automática del ángulo acetabular y del ángulo del borde central para identificar la DDC en base a imágenes radiográficas. Para las validaciones se recopilaron 16 radiografías de pelvis de infantes mayores a 06 meses de edad. La precisión de los ángulos calculados por el algoritmo fue de 85% en comparación con la medición manual realizada por el médico. Los resultados son alentadores para el desarrollo futuro de un método automático con el fin de reconocer la DDC por imágenes de rayos X.

En el estudio de Den et al. (2023), se elabora un modelo de aprendizaje profundo para detectar la DDC sin necesidad de mediciones explícitas. Para ello, aplicaron el enfoque de aprendizaje por transferencia usando You Only Look Once v5 (YOLOv5), empleando 305 imágenes radiográficas anteroposterior de cadera de pacientes menores de 12 meses de edad, constituidas por 205 imágenes normales y 100 con DDC, separando 47 imágenes para las pruebas, en donde 30 imágenes corresponde a cadera normal y 17 con DDC. Con el estudio, lograron obtener una sensibilidad y

especificidad de 94% y 96%, respectivamente, demostrando superioridad sobre al modelo detector de cajas múltiples de disparo único (SSD) para detectar la DDC.

Igualmente, Fraiwan et al. (2022) aplicaron el aprendizaje por transferencia para la detección de DDC, utilizando modelos que fueron preentrenados utilizando Imagenet. En total, fueron 13 los modelos implementados. El conjunto de datos utilizado fue de 354 radiografías de pelvis pertenecientes a sujetos menores a 07 meses de edad, comprendidas por 120 imágenes con DDC y 234 normales. Para la evaluación de los modelos se dividieron los datos en dos grupos: entrenamiento, y prueba (70/30 y 90/10), con el propósito de demostrar que suministrar más imágenes radiográficas mejora la capacidad de aprender. El modelo DarkNet53 logró alcanzar una puntuación F1, precisión, sensibilidad y especificidad de 95%, 90.6%, 100% y 94.3%, respectivamente.

De manera similar, Park et al. (2021), desarrollan un algoritmo de aprendizaje profundo basado en una nueva red neuronal convolucional para detectar la DDC en imágenes radiografías. Construyeron un conjunto de datos de 5,076 imágenes, que fueron revisados por dos radiólogos especialistas pediátricos, y etiquetadas con un total de 4,582 normal y 494 con DDC. El conjunto de datos fue dividido para las etapas de entrenamiento, validación y pruebas en 80%, 10% y 10%, respectivamente. Para la etapa de entrenamiento las imágenes fueron sometidas a cambios de tamaño a 128x128. Así mismo, aplicaron operadores de aumentos de datos con factores de 4 y 10 veces para clases normales y con DDC, correspondientemente. La evaluación del rendimiento del algoritmo fue comparada con el rendimiento de tres radiólogos de diferentes niveles de experiencia. El algoritmo logró alcanzar un 98% de sensibilidad y 98.1% de especificidad.

Por otra parte, Cheng et al. (2019), plantean identificar el uso de una red DCNN (red neuronal convolucional profunda) con el fin de detectar y localizar fracturas de cadera en imágenes radiografías pélvicas simples. Para el entrenamiento del DCNN recopilaban 29,110 radiografías de pelvis realizadas

entre 2008 y 2017. Añadieron 100 imágenes radiografías obtenidas durante el 2017, con la finalidad de evaluar la precisión, la sensibilidad, la tasa de falsos negativos y el AUC (área bajo la curva). Además, para confirmar que el modelo fuera válido, emplearon el mapeo de activación de clases ponderado por gradiente del algoritmo de visualización. El algoritmo logró una precisión del 91%, una sensibilidad del 98%, una tasa de falsos negativos del 2 % y un AUC de 0,98 para identificar fracturas de cadera. El algoritmo de visualización mostró una precisión del 95,9% para identificar lesiones. La aplicación de DCNN en imágenes radiográficas tiene el potencial de mejorar la calidad de los resultados en términos de precisión y eficiencia del diagnóstico de fractura de cadera.

En otro estudio, Cheng et al. (2021), presentan un algoritmo basado en aprendizaje profundo, llamado PelviXNet, con el fin de reconocer traumas universales de cadera por radiografía pélvica. Para el entrenamiento del modelo propuesto, recolectaron 5,024 imágenes de radiografía de pelvis registradas entre 2008 y 2019. Posteriormente, para validar el modelo, incluyeron 1888 radiografías obtenidas en el 2017 del área de urgencias mediante el muestreo consecutivo. PelviXNet logró detectar el 9 %, el 5,5 % y el 9,7 % de las fracturas de pelvis que fueron diagnosticadas erróneamente por médicos de urgencias, médicos de consulta y residentes de urgencias, respectivamente. PelviXNet obtuvo una precisión del 92,4 % en el conjunto de datos de pruebas. Los resultados evidencian posibles beneficios potenciales del algoritmo propuesto para los residentes médicos en el diagnóstico de fracturas de cadera.

Los antecedentes presentados proporcionan una visión contextual de la situación actual en relación con el uso del aprendizaje profundo en la detección de la DDC. Ahora, para comprender en profundidad los fundamentos teóricos, es esencial adentrarnos en la sección de bases teóricas. Por eso, en la siguiente sección, se exploran las teorías, conceptos y marcos conceptuales que respaldan la presente investigación y que ayudarán a construir una base sólida para analizar e interpretar los datos recopilados.

2.2. Bases teóricas

Para abordar las bases teóricas de manera organizada, primero se describirán la displasia del desarrollo de cadera, junto con las técnicas de diagnóstico relevantes, haciendo hincapié en el diagnóstico por medio de radiografías pélvicas, objeto de estudio de esta investigación. Posteriormente, se explorarán los fundamentos teóricos del aprendizaje automático, con un enfoque particular en el aprendizaje profundo.

2.2.1. Displasia del desarrollo de cadera (DDC)

La DDC, describe una anomalía en la articulación que une al hueso coxal con el fémur, manifestándose por desviaciones parcial o total de la cabeza femoral con el acetábulo, como se ilustra claramente en la Figura 1 (Shipman et al., 2006). Esta patología engloba un rango de manifestaciones que comprende tres posibles situaciones: la luxación, subluxación y la propia displasia (Gelfer & Kennedy, 2008).

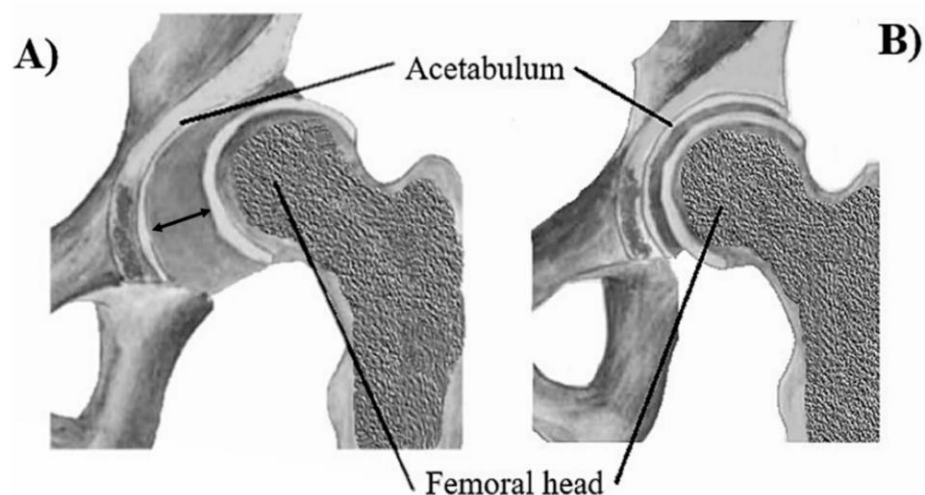


Figura 1. Comparativa de estructura acetabulares: (A) Acetábulo displásico; (B) Acetábulo en posición fisiológica

Fuente. (Harsanyi et al., 2020).

2.2.2. Diagnóstico de DDC

La DDC usualmente se detecta durante la infancia, en donde es vital el diagnóstico en una edad temprana para lograr un tratamiento óptimo. Entre los métodos de diagnóstico actuales existen el diagnóstico por examen clínico, el diagnóstico por ecografía, y el diagnóstico por radiografía.

Para un diagnóstico oportuno usualmente se aplican exámenes clínicos en los primeros meses de vida en niños con indicadores de riesgo (Swarup et al., 2018). Así mismo, las modalidades por imágenes como la ecografía y radiografías se emplean para confirmar la sospecha clínica, evaluar la gravedad y controlar la respuesta al tratamiento (Barrera et al., 2019). Así mismo, Agostiniani et al. (2020), enfatizan la importancia de radiografías antes de los seis meses para resultados funcionales favorables.

2.2.3. Diagnóstico de DDC por radiografía

Es un examen que mediante una radiografía de pelvis anteroposterior permite visualizar las alas ilíacas, los agujeros obturadores, así como también la posición de la metáfisis proximal de ambos fémures, con el fin de realizar mediciones y detectar la DDC (Mulrain et al., 2021).

Para una medición precisa, es fundamental contar con una radiografía bien tomada, que técnicamente debe capturarse con el paciente en posición decúbito supino, con las piernas extendidas de forma paralela y simétrica. Este examen ofrece mejores resultados a partir de los 3 meses de vida del bebé, ya que los núcleos de las epífisis femorales comienzan a osificarse entre los 3 y 8 meses de edad (Shaw & Segal, 2016).

En la Figura 2 se visualiza una comparación de una cadera normal (lado izquierdo) con una luxada (lado derecho). Además, incluye los aspectos radiológicos considerados con mayor frecuencia, las cuales son: la línea de Hilgenreiner, la línea de Perkins, la línea acetabular, los ángulos acetabulares y el arco de Shenton (Raimann & Aguirre, 2021).

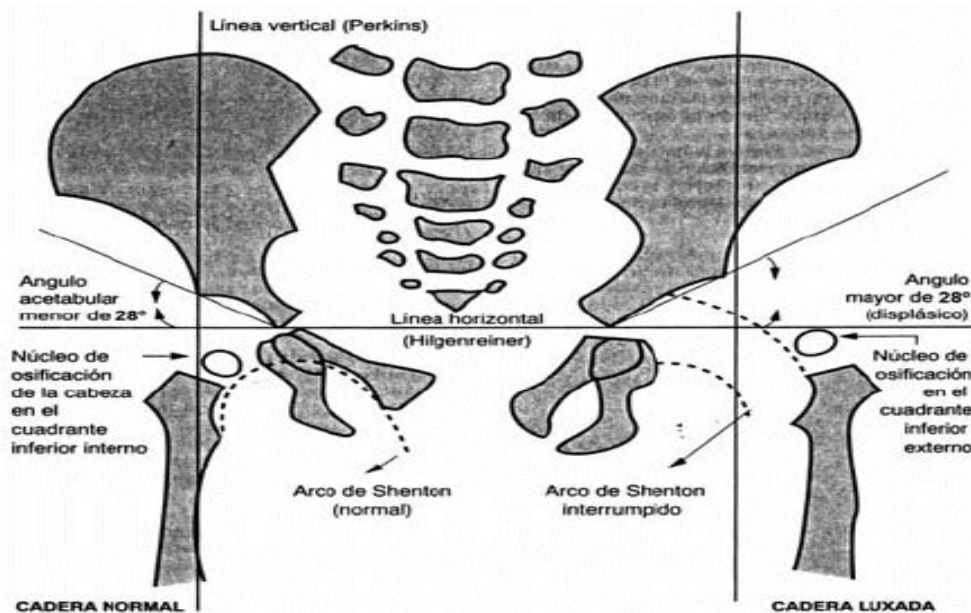


Figura 2. Comparativa anatómica entre una cadera normal y luxada
Fuente. (Mazzi, 2011).

La línea de Hilgenreiner se define como una línea horizontal que une el punto más inferior de ambos huesos ilíacos a la altura del cartílago trirradiado o en forma de Y. Mientras tanto, la línea de Perkins se traza perpendicularmente a la línea de Hilgenreiner, dividiendo el borde externo del acetábulo. Respecto a la línea acetabular, se extiende desde el borde superior del cartílago trirradiado y se desplaza de manera tangencial a lo largo del borde externo del cotilo (Atalar et al., 2020).

Los ángulos acetabulares es la formación de la línea de Perkins con la línea acetabular, los cuales, son determinantes para definir la presencia de la displasia acetabular. El ángulo cambia según la edad y género, en condiciones normales, este ángulo debe ir disminuyendo con la edad, conforme el cotilo va perdiendo oblicuidad (Barrera et al., 2019).

El arco de Shenton es una línea proyectada desde el borde inferior del cuello femoral hasta el borde superior del agujero obturador. Una ruptura en la proyección de la línea es indicativa de subluxación de la cabeza femoral del acetábulo y es altamente sugestivo de un acetábulo displásico (Rhee et al., 2011).

Otra forma de referencia en radiografías es el cuadrante de Ombredanne, que se crea donde se cruzan las líneas de Hilgenreiner y Perkins. En una cadera sana, el centro de crecimiento del hueso del muslo se encuentra en el cuadrante inferior interno. Si hay una subluxación, estará en el cuadrante inferior externo, y si la cadera está completamente desplazada, estará en la parte superior externa (Baudet et al., 2001).

La Figura 3 muestra radiografías reales de sujetos con presencia y ausencia de la DDC. El aprendizaje profundo podría ser utilizado para detectar la DDC mediante el análisis automatizado de este tipo de imágenes médicas.



Figura 3. Radiografías de pacientes: (a) Con DDC; (b) Sin DDC.

Fuente. (Fraivan et al., 2022).

2.2.4. Métodos clásicos de aprendizaje automático

A pesar de la ascendente predominancia del aprendizaje profundo en la clasificación de imágenes en años recientes, es vital reconocer la relevancia de los métodos clásicos de aprendizaje de máquina, o también conocido como aprendizaje profundo, como los árboles de decisión y las máquinas de soporte vectorial, que han sido pilares en el campo durante décadas (H. Zhang et al., 2024). Aunque estas técnicas generalmente requieren una extracción manual

y meticulosa de características, en contraste con las redes neuronales profundas, en ciertos contextos y con el preprocesamiento adecuado, pueden ser igual de efectivas y ofrecer una interpretación más transparente (Fisher et al., 2024).

2.2.5. Árboles de decisión

Los árboles de decisión se pueden usar tanto para clasificación como para regresión, en el contexto de la clasificación, un árbol de decisión divide el conjunto de datos en subconjuntos más pequeños basados en características o atributos, y asigna una etiqueta de clase a cada subconjunto final que también se conocen como hojas del árbol (Rochmawati et al., 2020). En la Figura 4, se muestra una representación esquemática de un árbol de decisión para la clasificación de imágenes.

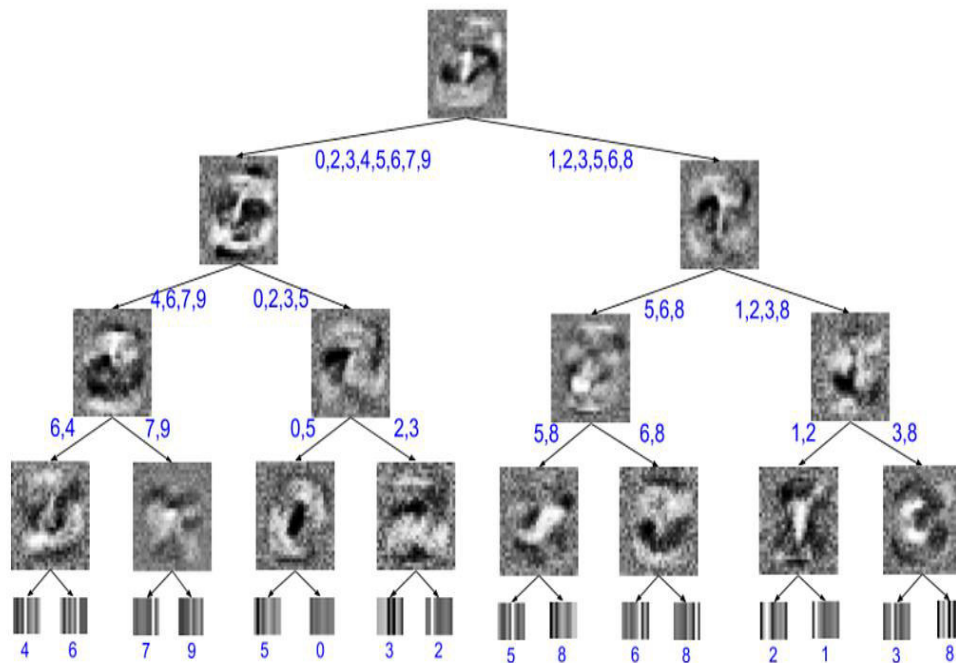


Figura 4. Árbol de decisión para la detección de imágenes con números del dataset MNIST.

Fuente. (Frosst & Hinton, 2017).

2.2.6. Máquinas de vectores de soporte (SVM)

Al igual que los árboles de decisión, las SVM son algoritmos de aprendizaje supervisado empleados para clasificar y predecir problemas tanto de clasificación como de regresión. La idea principal detrás de SVM es buscar un hiperplano óptimo que distinga de manera efectiva las clases en un espacio de características, proveyendo soluciones robustas en una variedad de contextos (Gomez-Chova et al., 2008).

2.2.7. Aprendizaje profundo

Aprendizaje profundo, conocido también como deep learning, es una especialización del aprendizaje automático, que se encuentra dentro de la rama de la inteligencia artificial y se dedica a construir modelos a partir de múltiples capas de transformaciones sucesivas, procesando grandes cantidades de datos en estructuras jerarquizadas y complejas, diferenciándose así de los enfoques más tradicionales del aprendizaje automático (Adedokun, 2019; A. Zhang et al., 2021). Estos algoritmos se distinguen por su habilidad para automatizar la extracción de características relevantes de los datos, eliminando en muchos casos la necesidad de ingeniería manual de características (Mousavi et al., 2018).

El aprendizaje profundo se distingue por utilizar modelos computacionales estructurados en múltiples capas de procesamiento, diseñados para aprender representaciones de datos en diversos niveles de abstracción, su objetivo principal es identificar estructuras complejas dentro de vastos conjuntos de datos. Para ello, utiliza algoritmos de retropropagación que ajustan los parámetros internos de la máquina, permitiendo que cada capa compute su representación en relación con la anterior. Gracias a estos métodos, en las últimas décadas, se ha logrado un avance significativo en áreas como el reconocimiento de voz, la identificación visual de objetos, la detección de objetos y otros dominios relevantes (LeCun et al., 2015).

Los avances tecnológicos como la disponibilidad de big data, computación mejorada potenciada con unidades de procesamiento gráfico ha permitido el uso de aprendizaje profundo en varios ámbitos de estudio, incluyendo imágenes médicas. Por ejemplo, en la práctica de la radiología con el objeto de detectar anomalías estructurales y clasificación de enfermedades (Lee et al., 2017).

En la Figura 5 ilustra las categorías y subcategorías predominantes del aprendizaje profundo empleadas en soluciones tecnológicas para diversas actividades médicas, particularmente en el campo de la imagenología (Yousef et al., 2022).

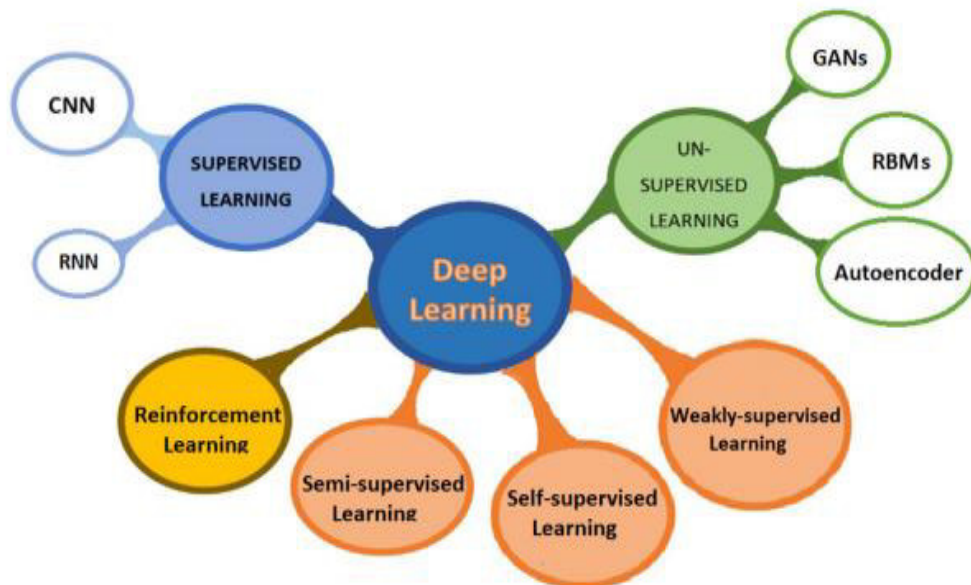


Figura 5. Panorama de aprendizaje profundo.

Fuente. (Yousef et al., 2022).

A continuación, se profundiza en las metodologías del aprendizaje supervisado y no supervisado, ambas esenciales en el desarrollo de este proyecto de tesis, como lo son la CNN para aprendizaje supervisado, y las redes generativas adversativas (GAN), para aprendizaje no supervisado. Posteriormente, se describirá con detalle la arquitectura de los modelos propuestos en el presente trabajo académico.

2.2.8. Aprendizaje supervisado

En el ámbito del aprendizaje profundo, el aprendizaje supervisado se refiere a un paradigma donde se entrena a modelos computacionales utilizando grandes conjuntos de datos que contienen ejemplos de entrada y sus correspondientes salidas deseables. Estos modelos, en particular las redes neuronales profundas, consisten en múltiples capas de nodos interconectados que pueden aprender representaciones jerárquicas de datos a partir de características de bajo nivel hasta características más abstractas. La capacidad de estas redes para modelar relaciones complejas y no lineales entre las entradas y salidas las convierte en herramientas potentes para tareas de reconocimiento de patrones, como la interpretación de imágenes médicas (J. Zhou et al., 2021).

Es relevante destacar que el aprendizaje supervisado se bifurca en dos principales categorías: clasificación y regresión (Litjens et al., 2017). En el marco de esta tesis, se abordará específicamente el aspecto de clasificación, con el objetivo de determinar la presencia o ausencia de displasia de cadera.

2.2.9. Aprendizaje no supervisado

A diferencia del aprendizaje supervisado (clasificación), donde el objetivo es predecir una salida basada en una entrada conocida, el aprendizaje no supervisado se enfoca en la exploración interna de los datos para identificar dichas relaciones (Lee et al., 2017). En el amplio espectro del aprendizaje automático, el aprendizaje no supervisado se distingue por trabajar con conjuntos de datos sin etiquetas predefinidas, orientándose a descubrir patrones, correlaciones o estructuras subyacentes que no son inmediatamente evidentes.

El aprendizaje no supervisado, y en particular las técnicas asociadas al aprendizaje profundo como las GAN, juegan un papel esencial en la generación de datos sintéticos a partir de los datos originales propuestos. Su

capacidad de descubrir patrones y estructuras ocultas brinda una herramienta potente para la comprensión y manipulación de los datos en estudio.

2.2.10. Clasificación de aprendizaje supervisado

La clasificación es uno de los pilares fundamentales en el campo del aprendizaje automático y se refiere al proceso de asignar una etiqueta o categoría a una entrada dada, basándose en ciertas características o propiedades de esa entrada (Sen et al., 2020). En términos simples, se puede pensar en la clasificación como una tarea de decisión: dada una entrada, ¿a qué grupo pertenece?

Para realizar la clasificación, es esencial tener un conjunto de datos previamente etiquetado, conocido como conjunto de entrenamiento. Existen diversas técnicas y algoritmos para la clasificación, que varían en complejidad desde métodos lineales como la regresión logística, hasta enfoques más complejos como máquinas de soporte vectorial, árboles de decisión, y, por supuesto, redes neuronales en el ámbito del aprendizaje profundo (Babaei y otros, 2024).

Una vez que un modelo ha sido entrenado usando un conjunto de datos etiquetado, puede ser probado y validado usando un conjunto de datos separado, conocido como conjunto de prueba, para evaluar su precisión y robustez (Sabharwal & Miah, 2024). Esta evaluación permite entender cuán bien el modelo ha aprendido a clasificar entradas no vistas previamente y si es capaz de generalizar su aprendizaje a nuevos datos (Fisher et al., 2024).

Dentro del ámbito médico y de la salud, el uso del aprendizaje profundo ha demostrado ser altamente prometedor en diversas aplicaciones. Por ejemplo, se ha utilizado para mejorar el diagnóstico de enfermedades, como la detección temprana de cáncer a partir de imágenes médicas, así como para predecir eventos clínicos, como la probabilidad de complicaciones después de una cirugía. Aggarwal et al. (2021) ofrecen una visión general exhaustiva de estos avances y sus implicaciones en la práctica médica. Las imágenes

médicas, en particular, se benefician del aprendizaje profundo debido a la naturaleza intrínsecamente visual de los datos y la necesidad de detectar patrones sutiles que pueden no ser evidentes para el ojo humano. Al entrenar un modelo con miles de ejemplos, es posible que el modelo aprenda a identificar indicadores y características específicas de una enfermedad o condición, incluso en casos que pueden ser difíciles para un especialista humano.

2.2.11. *Arquitecturas de aprendizaje profundo*

Las arquitecturas de aprendizaje profundo se pueden categorizar en tres clases principales, las cuales son: arquitecturas profundas generativas, discriminativas e híbridas (Deng & Yu, 2014). Por otro lado, se debe tener en consideración que para todo método de aprendizaje profundo es necesario establecer un conjunto de datos seleccionado cuidadosamente para un rendimiento óptimo (Raina et al., 2009). Para fines prácticos de este estudio, centraremos nuestra atención en dos arquitecturas específicas dentro del vasto panorama del aprendizaje profundo: las CNN y las GAN.

Las CNN se han consolidado como una herramienta poderosa en tareas de clasificación, especialmente en el ámbito de la interpretación de imágenes, y en este caso, serán implementadas para discernir entre la presencia o ausencia de displasia. Su capacidad para identificar patrones y características en imágenes las hace especialmente adecuadas para esta tarea.

Por otro lado, las GAN se destacan en la generación de datos sintéticos, replicando con precisión características de datos reales. En este contexto, las utilizaremos para generar imágenes sintéticas que complementen y enriquezcan nuestro conjunto de datos, permitiendo una formación más robusta y un análisis más detallado. Estas elecciones arquitectónicas se alinean con los objetivos de nuestra investigación y las necesidades prácticas del diagnóstico y estudio de la displasia. En la Tabla 2 presenta comparación de las arquitecturas de aprendizaje profundo y sus principales aplicaciones.

Tabla 2. Comparación de diferentes arquitecturas de aprendizaje profundo.

	Categoría	Descripción	Fortaleza	Debilidad	Principales aplicaciones
Deep AE	No supervisado	Una red neuronal entrenada para reconstruir sus entradas bajo algunas restricciones	Es capaz de aprender representación de características más complejas a través de un esquema aprendizaje no supervisado	Menos escalabilidad a datos de alta dimensión Confía en la optimización numérica y el ajuste de parámetros elevados	Aprendizaje de representaciones de características de baja dimensión a partir de la revisión de texto
DBM	No supervisado	Red neuronal profunda con una conexión no dirigida en cada capa	Proporciona una característica más robusta extracción vía entrenamiento no supervisada	Alto tiempo de cálculo debido al gran ajuste de parámetros	Modelos de conjunto
DBN	No supervisado	Consiste en una conexión dirigida en la capa inferior y una conexión no dirigida en las dos capas superiores	Potente para la extracción de funciones ocultas y útiles de los datos de audio	Difícil de entrenar debido al extenso proceso de inicialización de parámetros	Recomendación basada en audio y video
RBM	No supervisado	Bipartite, gráfico no dirigido que comprende la capa visible y la oculta	Adecuado para el aprendizaje de representación de bajo rango	No tratable como tal, la divergencia contrastiva se puede utilizar para aprender los parámetros	Recomendación basada en grupos
GAN	No supervisado / Supervisado	Red neuronal profunda que consiste discriminador y generador	Adecuado para aprendizaje unificado supervisado y no supervisado	Proceso de aprendizaje inestable dificultad en la convergencia	Recuperación de información discriminativa y generativa
CNN	Supervisado	Arquitectura interconectada inspirada en la corteza visual biológica	Potente para la extracción de características con información contextual	Requiere un alto ajuste de parametrización	Aprendizaje de representación de características de múltiples fuentes: Audio, texto, imagen, video, etc.
MLP	Supervisado	Modelado de datos con correlación simple	Transformación no lineal	Alta complejidad y convergencia lenta	Aprendizaje de representación de características a partir de contenidos de audio, video y texto

Fuente. (Da'u & Salim , 2020)

2.2.12. Redes generativas adversarias (GAN)

Dentro del ámbito del aprendizaje profundo, las GAN, emergen como una de las técnicas más destacadas del aprendizaje no supervisado. Las GAN constan de dos redes neuronales que trabajan en conjunto: una generadora, que crea datos, y una discriminadora, que evalúa la autenticidad de esos datos (Singh & Raza, 2021). A través de un proceso iterativo, la red generadora intenta producir datos cada vez más convincentes, mientras que la discriminadora busca mejorar en la identificación de los datos genuinos de los generados. Este proceso adversario conlleva a que la red generadora pueda generar datos que son notablemente similares a los reales, lo cual tiene una amplia variedad de aplicaciones, que van desde la generación de imágenes hasta la simulación de datos (Goodfellow et al., 2020).

Las imágenes médicas juegan un papel crucial en la detección, diagnóstico y seguimiento de múltiples enfermedades y condiciones. Sin embargo, obtener un volumen suficientemente grande de imágenes médicas etiquetadas es un desafío, no solo debido a las limitaciones en la recopilación de datos, sino también por cuestiones éticas, privacidad de los pacientes y la variabilidad inherente en las condiciones patológicas. En este contexto, los datos sintéticos surgen como una solución invaluable.

Uno de los principales beneficios de los datos sintéticos es la posibilidad de modelar escenarios específicos o raros. Por ejemplo, ciertas patologías o condiciones pueden no estar ampliamente representadas en un conjunto de datos original debido a su rareza. Mediante la generación sintética, es posible crear imágenes que representen estas condiciones, garantizando que los modelos entrenados sean robustos y precisos incluso en escenarios menos comunes.

Además, el uso de imágenes sintéticas puede abordar el problema de la privacidad. Al no provenir de pacientes reales, las imágenes sintéticas eliminan los riesgos asociados con la identificación del paciente, lo que permite su uso en investigaciones sin violar las normativas de privacidad.

En resumen, en el campo de las imágenes médicas, los datos sintéticos proporcionan la oportunidad de ampliar y enriquecer los conjuntos de datos de entrenamiento, aumentar la robustez y mejorar la precisión de los modelos de aprendizaje profundo, y abordar preocupaciones éticas y de privacidad. Estas ventajas hacen que los datos sintéticos sean esenciales para impulsar avances en la interpretación automática de imágenes médicas y en la asistencia diagnóstica basada en inteligencia artificial. Las GAN funcionan con dos redes: el generador y el discriminador.

2.2.13. Generador de una GAN

Es una de las principales redes neuronales de una GAN, que dado un vector de ruido aleatorio z muestreado de alguna distribución, el generador produce una muestra $G(z)$, como se observa en la Ecuación 1 (Singh & Raza, 2021).

$$G(z) = RedNeuronal(z; \theta_G) \quad (1)$$

Donde:

θ_G = Parámetros del generador.

2.2.14. Discriminador de una GAN

Es una red neuronal de una GAN, que evalúa la autenticidad de una muestra, es decir, asigna una probabilidad de que la muestra dada sea real (y no generada). En la Ecuación 2, se observa que, dado una muestra x , produce una probabilidad $D(x)$ (Y. Wang, 2020).

$$D(x) = \sigma(RedNeuronal(x; \theta_D)) \quad (2)$$

Donde:

σ = Función sigmoide

θ_D = Parámetros del discriminador.

2.2.15. *Función objetivo de una GAN*

La función objetivo, también conocida como función de pérdida, nos permite medir las equivocaciones de la red neuronal. En una GAN, el generador busca minimizar la función objetivo, mientras que el discriminador intenta maximizar. Y. Wang, (2020), define la Ecuación 3, para la función objetivo de una GAN típica.

$$\min_{\theta_G} \max_{\theta_D} = \mathbb{E}_{x \sim p_{data}} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))] \quad (3)$$

Donde:

p_{data} = Distribución de datos reales.

p_z = Distribución de donde se muestrea el ruido aleatorio z .

$D(x)$ = Estimación de la probabilidad del discriminador de que una muestra x corresponda al conjunto de datos real.

$G(z)$ = Salida del generador tomando como entrada el ruido aleatorio z .

$D(G(z))$ = Estimación del discriminador de que la salida conseguida por el generado es real o falsa.

$\mathbb{E}_x$ = Valor esperado de pertenencia de que la muestra x cuente la misma distribución que los datos reales sobre todas las muestras reales.

$\mathbb{E}_z$ = Valor esperado de la salida obtenida por el generador.

El entrenamiento involucra actualizar los parámetros θ_G y θ_D usando algún algoritmo de optimización, como el descenso de gradiente, hasta que el generador produce muestras que son indistinguibles de las reales para el discriminador.

2.2.16. *Red neuronal convolucional (CNN)*

Las CNN, son una especialización dentro del aprendizaje supervisado del ámbito del aprendizaje profundo. Están optimizadas para procesar datos estructurados en forma de matrices, como es el caso de imágenes, audios o videos. Su diseño se fundamenta en cuatro pilares: conexiones locales,

compartición de pesos, agrupamiento y la incorporación de múltiples capas (Da'u & Salim , 2020).

La profundidad inherente en las CNN permite la extracción jerárquica de características. Por ejemplo, en las primeras capas, la red podría detectar bordes simples o variaciones de color. A medida que se avanza en la profundidad de la arquitectura, las capas intermedias pueden identificar formas más complejas o patrones, y las capas más profundas, a partir de estas características acumuladas, son capaces de reconocer objetos completos (Aloysius & Geetha, 2017).

En términos estructurales, una CNN típica se compone de una serie de etapas que alternan principalmente entre capas convolucionales y de agrupación. En las capas convolucionales, las unidades se organizan en mapas de características y están conectadas a través de pequeños parches de la capa anterior, utilizando un conjunto de pesos denominado banco de filtros. Estos pesos suelen ser compartidos a lo largo de toda la capa, promoviendo la detección eficiente de características. Por su parte, las capas de agrupación reducen las dimensiones de los datos procedentes de las capas convolucionales, confiriendo a las CNN cierta invariancia ante cambios y distorsiones en las entradas (LeCun et al., 2015). En la Figura 6 proporciona una visión generalizada de la arquitectura de CNN.

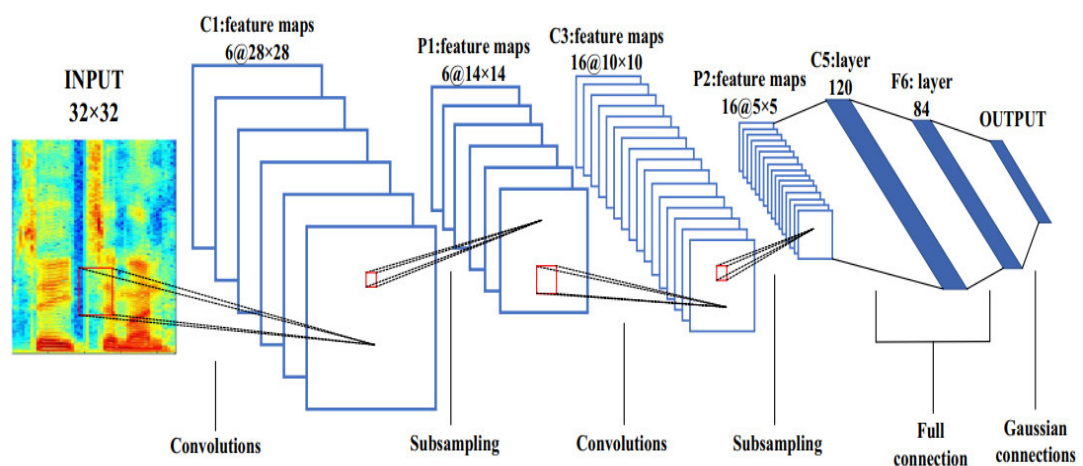


Figura 6. Arquitectura de una CNN

Fuente. (Ogunfunmi et al., 2019).

Dentro de las capas más importantes de la arquitectura destacan las siguientes: capa de convolución (Conv2D), capa de pooling (MaxPool2D), capa de aplanamiento (Flatten) y la capa densa (fully connected).

2.2.17. Capa de convolución (Conv2D)

La operación principal de una capa convolucional es la convolución, que consiste en analizar las imágenes ingresadas mediante una serie de filtros de convolución con el fin de detectar o extraer un conjunto de características (Dhillon & Verma, 2020).

2.2.18. Capa de pooling (MaxPool2D)

En pooling máximo, se trata de un procedimiento comúnmente utilizado entre dos capas de convolución. En este proceso, toma un conjunto de características generado en la salida de la capa de convolución como entrada. Su objetivo principal es reducir el tamaño de las imágenes mientras conserva sus características más importantes (Gholamalinezhad & Khosravi, 2020).

2.2.19. Capa de aplanamiento (Flatten)

Esta capa simplemente transforma un tensor tridimensional en un vector unidimensional (Sarvamangala & Kulkarni, 2022). Si un tensor tiene dimensiones (a, b, c) , el vector resultante tiene dimensiones $abc \times 1$.

2.2.20. Capa densa (Fully Connected)

En una capa totalmente conectada, cada neurona está conectada a todas las neuronas en la capa anterior y en la capa siguiente. Esta capa recibe un vector de entrada para aplicar una serie de combinación lineal y una función de activación con el objetivo de clasificar la imagen (Sarvamangala & Kulkarni, 2022).

2.2.21. *ResNet50*

ResNet, abreviatura de Redes Residuales, son un conjunto de modelos de redes neuronales profundas muy populares en el campo de la visión por computadora. Fue introducida por He K. et al. (2016), la idea principal detrás de ResNet es introducir conexiones residuales o conexiones cortas, que saltan una o más capas. Estas conexiones residuales ayudan a evitar el problema del desvanecimiento del gradiente en redes muy profundas y permiten entrenar redes con un número significativamente mayor de capas.

ResNet50 es una variante que tiene aproximadamente 50 capas. A continuación, se presenta una descripción general de la arquitectura:

- Entrada: La imagen de entrada se somete a una capa convolucional única que consta de 64 filtros, utilizando matrices de kernel de 7x7 y un paso (stride) de 2. Posteriormente, se aplica una capa de agrupación máxima (max-pooling).
- Bloques Convolucionales: Después de la entrada, la red tiene una serie de bloques convolucionales. Cada bloque tiene un cierto número de unidades residuales.
- Unidad Residual: Una unidad residual típica tiene 3 capas convolucionales. Estas capas van reduciendo el tamaño del feature map (mapa de características) mientras aumentan el número de canales. Las conexiones residuales saltan estas capas convolucionales. Si es necesario cambiar el tamaño del feature map o el número de canales, se utiliza una convolución 1x1 en la conexión residual.
- Unidades en ResNet50:
 - Primer bloque: 3 unidades con filtros de 64, 64, 256.
 - Segundo bloque: 4 unidades con filtros de 128, 128, 512.
 - Tercer bloque: 6 unidades con filtros de 256, 256, 1024.
 - Cuarto bloque: 3 unidades con filtros de 512, 512, 2048.

- Finalización: se utiliza una capa de average-pooling seguida por una capa completamente conectada que produce la salida final para la clasificación.

Gracias a las conexiones residuales, las ResNet son capaces de aprender la identidad (no hacer nada) en las capas adicionales, permitiendo a los modelos beneficiarse de capas adicionales sin el riesgo de degradación del rendimiento. Esto les permite ser más profundos y, por lo tanto, tener una mayor capacidad sin los problemas típicos asociados con las redes más profundas.

2.2.22. Modelos híbridos

Los modelos híbridos en machine learning son enfoques que combinan múltiples algoritmos o modelos diferentes para abordar de manera más efectiva tareas específicas o desafíos complejos (Salur & Aydin, 2020). Estos modelos aprovechan las fortalezas individuales de cada componente para mejorar la precisión y eficacia general del sistema. Los modelos híbridos pueden ser útiles en diversas aplicaciones, como la clasificación de datos, la predicción y la recomendación (Hashemzahi et al., 2020).

En sistemas de recomendación, por ejemplo, un modelo híbrido puede combinar enfoques basados en contenido y filtrado colaborativo. El enfoque basado en contenido se centra en las características de los elementos y sugiere elementos similares a los que un usuario ha demostrado interés (Taresh et al., 2021). El filtrado colaborativo, por otro lado, se basa en el comportamiento y preferencias de usuarios similares para realizar recomendaciones. Un modelo híbrido puede combinar estos enfoques para ofrecer recomendaciones más precisas y personalizadas, aprovechando tanto la información del contenido como la colaborativa (Celik & Inik, 2024).

2.2.23. Métricas de evaluación

La evaluación de modelos de clasificación es muy importante en el estudio del aprendizaje de máquina y la inteligencia artificial, puesto que permite validar su precisión y eficiencia, así mismo, asegurar la confiabilidad de sus predicciones en aplicaciones del mundo real. La correcta evaluación brinda una guía precisa sobre cómo mejorar y ajustar el modelo. Por otro lado, una evaluación inadecuada puede llevar a decisiones erróneas, costosas y en ocasiones peligrosas. Por ello, es imperativo emplear métricas y técnicas adecuadas para evaluar la performance de estos modelos. A continuación, mostramos las ecuaciones correspondientes a las métricas usadas en la presente tesis: Precisión, Sensibilidad y Especificidad (S. Zhou et al., 2021).

2.2.24. Métrica de Precisión

Es la proporción de identificaciones positivas que fueron efectivamente correctas, definida por la Ecuación 4:

$$Precision = \frac{VP}{VP + FP} \quad (4)$$

Donde:

VP = Verdaderos positivos.

FP = Falsos positivos.

2.2.25. Métrica de Sensibilidad

La sensibilidad, también llamada tasa de verdaderos positivos, indica la proporción de casos positivos reales que fueron correctamente identificados; es representada por la Ecuación 5.

$$Precision = \frac{VP}{VP + FN} \quad (5)$$

Donde:

VP = Verdaderos positivos

FN = Falsos negativos.

2.2.26. Métrica de Especificidad

Es la proporción de negativos reales que fueron identificados correctamente, como se observa en la Ecuación 6.

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (6)$$

Donde:

VP = Verdaderos negativos.

FP = Falsos positivos.

2.2.27. Curva característica operativa del receptor

Una curva ROC, abreviatura de la curva de características operativas del receptor, es una representación gráfica de la sensibilidad frente a $1 - \text{especificidad}$ para varios umbrales de clasificación (Bradley, 1997). El área bajo la curva ROC, es decir el área AUC, de 1 indica una clasificación perfecta, mientras que un AUC de 0.5 indica un clasificador aleatorio.

La fórmula para calcular el área bajo la curva ROC no es trivial y generalmente se basa en técnicas numéricas como la regla trapezoidal para integración. Sin embargo, el concepto principal es representada por la Ecuación 7.

$$AUC = \int_0^1 \text{Sensibilidad}(t) dt \quad (7)$$

Donde:

t = Varía a través de todos los posibles umbrales.

2.2.28. Matriz de confusión

Una matriz de confusión es una técnica empleada en el campo del aprendizaje automático para analizar el rendimiento de un algoritmo de clasificación. Su función es ofrecer una visión detallada sobre cómo el algoritmo está clasificando las instancias, permitiendo identificar posibles áreas de mejora (Ting, 2011). Supongamos que tenemos un problema de clasificación binaria, donde las clases son positivo y negativo. La matriz de confusión se compone de 4 valores, como se presenta en la Figura 7.

		Predicción	
		Negativo (N) -	Positivo (P) +
Actual	Negativo -	Verdadero Negativo (VN)	Falso Positivo (FP) Error Tipo I
	Positivo +	Falso Negativo (FN) Error Tipo II	Verdadero Positivo (VP)

Figura 7. Matriz de confusión.

Fuente. (Ting, 2011)

Donde:

- Verdaderos Positivos (VP): Número de positivos correctamente clasificados.
- Verdaderos Negativos (VN): Número de negativos correctamente clasificados.
- Falsos Positivos (FP): Representan la cantidad de casos negativos que son clasificados erróneamente como positivos.
- Falsos Negativos (FN): Corresponden a la cantidad de casos positivos que son clasificados erróneamente como negativos.

2.3. Marcos Conceptuales

- Acetábulo: Cavidad del hueso coxal que articula la cabeza del fémur.
- Cotilo: Cavidad articular de un hueso.
- DBM: Del inglés “Deep Boltzmann machines”, en otras palabras, máquina de Boltzmann restringida.
- DBN: Del inglés para “Deep Belief Network”, en otras palabras, red de creencia profunda.
- Decúbito supino: Es una posición anatómica del cuerpo humano caracterizada por una posición corporal acostada boca arriba y en un plano paralelo al suelo.
- Deep AE: Del inglés “Deep Autoencoders”, en otras palabras, codificador profundo.
- Luxación: Es una separación de dos huesos en el lugar donde se encuentran en una articulación
- MLP: Siglas en inglés de “Multilayer Perceptron”, es decir, perceptrón multicapa.
- Retropropagación: Consiste en propagar el error hacia atrás.
- RBM: Siglas en inglés de “Restricted Boltzmann Machines”, es decir, máquina de Boltzmann restringida.
- Subluxación: Los huesos de una articulación están parcialmente fuera de su posición.

CAPÍTULO 3: METODOLOGÍA

En este capítulo se detalla la metodología utilizada para abordar el problema de investigación propuesto. Este enfoque se sustenta en procesos y técnicas diseñadas para asegurar resultados fiables y coherentes, alineados con los objetivos establecidos en la investigación.

3.1. Tipo y diseño de investigación

Este estudio se basa en una investigación aplicada de naturaleza cuantitativa, cuyo objetivo es evaluar el rendimiento del modelo propuesto para detectar la DDC. Asimismo, se caracteriza por un enfoque descriptivo, ya que se emplean técnicas de aprendizaje profundo que influyen en la precisión de la detección de la DDC en infantes de 3 a 6 meses de edad.

Siguiendo la perspectiva de Hernández et al. (2014), el enfoque cuantitativo se caracteriza por su enfoque secuencial y probatorio. Este enfoque es ideal para el estudio propuesto ya que comienza con una idea clara de mejorar la detección de la DDC mediante aprendizaje profundo. Esto se traduce en objetivos e interrogantes de investigación específicos. Además, implica la formulación de hipótesis y la identificación de variables que serán analizadas mediante técnicas de evaluación. Esta metodología es esencial para evaluar objetivamente el desempeño del modelo de aprendizaje profundo propuesto.

Según Vara (2012), los estudios descriptivos buscan medir o evaluar con mayor precisión las propiedades importantes de un fenómeno. En este estudio, el diseño descriptivo es apropiado ya que busca evaluar cómo las técnicas de aprendizaje profundo impactan la precisión de la detección de la DDC en infancias de 03 a 06 meses de edad. Esto implica una medición detallada y una evaluación de las capacidades del modelo propuesto, utilizando posiblemente datos secundarios (como radiografías existentes) para este análisis.

3.2. Unidad de análisis

La unidad de análisis son imágenes radiografías de caderas.

3.3. Población de estudio

La población de estudio incluye imágenes radiográficas de cadera tomadas a infantes de entre 3 y 6 meses de edad.

3.4. Tamaño de muestra

Para este estudio, se optó por un muestreo no probabilístico, consistente en una colección de 354 imágenes radiográficas de cadera, de las cuales, 120 fueron diagnosticados con DDC, mientras que los 234 restantes no presentaron esta afección. Siguiendo las directrices de Hernández et al. (2014), la elección de los elementos de muestra en investigaciones específicas puede reflejar la población objetivo, especialmente cuando se emplean para validar hipótesis.

3.5. Selección de muestra

Cada imagen radiográfica seleccionada debe cumplir con dos criterios: primero, debe estar relacionada con la DDC o ser una imagen normal; segundo, debe corresponder a infantes menores de seis meses de edad.

3.6. Técnicas de recolección de datos

Las imágenes radiográficas de cadera utilizadas en este estudio se recolectaron mediante una revisión de la literatura. Se seleccionaron investigaciones que aborden detecciones de anomalías de la cadera mediante imágenes radiografías, cuyos conjuntos de datos correspondan a infantes y se

encuentren disponibles públicamente. Este método asegura que las imágenes sean pertinentes para nuestro análisis del modelo de aprendizaje profundo en la detección de la displasia de desarrollo de la cadera en niños.

3.7. Metodología de implementación

A través de una revisión de la literatura, se ha determinado que el aprendizaje profundo es la técnica más apropiada para abordar el problema de esta investigación, dada a su capacidad de identificar patrones en datos no estructurados, en el afán de emular el proceso de aprendizaje humano mediante el uso de redes neuronales artificiales. Para el desarrollo del presente estudio, se tomará como referencia los pasos empleados de las investigaciones propuesta por de Den et al. (2023) y Fraiwan et al. (2022):

- **Preprocesamiento de los datos:** Establecer el conjunto de datos a utilizar, para luego transformar u homogenizar las imágenes con el fin de facilitar su procesamiento.
- **Aumento de datos:** Se propone generar dos experimentos para el incremento de datos, con el fin de ayudar al modelo a generalizar mejor y a reconocer patrones en nuevas imágenes de radiografías. Para el primer experimento se pretende aplicar diversas operaciones tradicionales a las imágenes originales como volteos, rotaciones y zooms. En cuanto a segundo experimento, consta de generar imágenes sintéticas de mayor calidad usando aprendizaje profundo con GAN (Khosravi et al., 2023).
- **Entrenamiento y evaluación:** Se propone separar el conjunto de datos en un 80% para entrenamiento y un 20% para validación. Se emplearán modelos clásicos de machine learning, la SVM y árboles de decisión; y modelos de aprendizaje profundo, una CNN genérica y de la arquitectura ResNet50. Se utilizará los dos conjuntos de datos generados en el paso previo, para posteriormente evaluar los modelos con las métricas de

desempeño: precisión, sensibilidad, especificidad y la curva ROC, con fines comparativos (Lin et al., 2022; Pei et al., 2023).

- **Clasificación:** Analizar los resultados de las evaluaciones de los modelos, con el propósito de elegir el modelo con mejor rendimiento en la detección de la DDC. El modelo seleccionado, podrá ser implementado en un componente de tipo API, con el objetivo de ser consumido por otro componente de tipo cliente (móvil, web u otro). La modelo presentará como dato de entrada una imagen radiográfica para emitir una salida indicando la presencia o ausencia de la DDC.

CAPÍTULO 4: DETECCIÓN DE DDC

En este capítulo se proporciona una descripción completa del aporte y los elementos que contribuyen a resolver el problema abordado en esta investigación.

4.1. Aporte

El principal aporte de esta tesis radica en la creación de un método asistido por computadora para detectar la DDC. Este concepto, enmarcado en la rama de la medicina, se centra en el uso de un ordenador para la interpretación de imágenes de rayos x. En contextos generales, las imágenes son digitalizadas y procesadas a través de un algoritmo de aprendizaje profundo, con el objetivo de ofrecer información de apoyo en el proceso de diagnóstico de un profesional médico.

En la Figura 8, se ilustra el método propuesto, donde se aplican 4 pasos para la detección de DCC: Preprocesamiento de datos, Aumento de datos, Entrenamiento, y Clasificación.

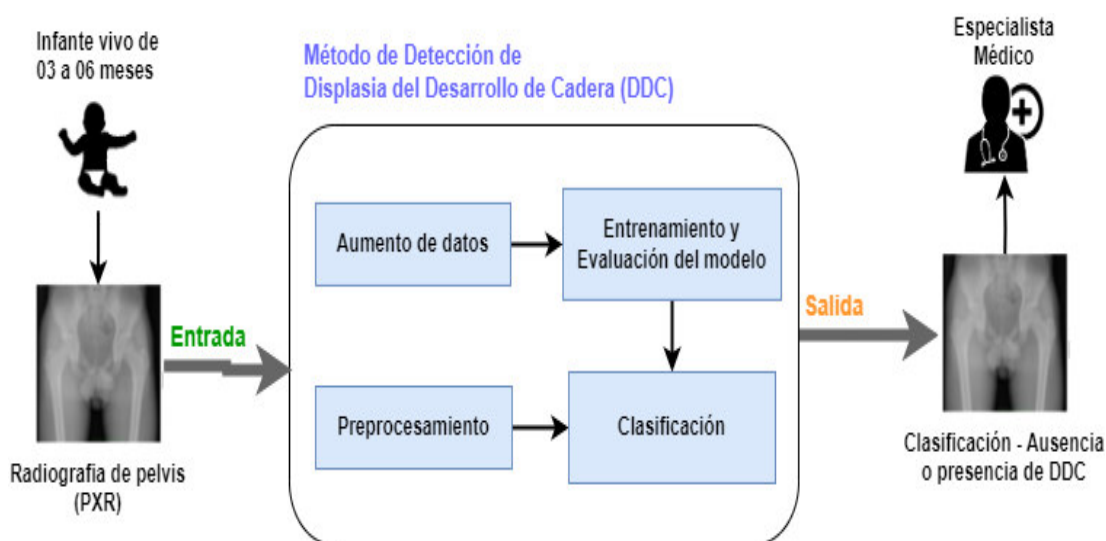


Figura 8. Método de detección de DDC.

Fuente. Elaboración Propia.

El sistema propuesto se fundamenta en un modelo híbrido que usa dos modelos: El modelo de ResNet50, y otro modelo que usa el promedio de los valores de píxeles de la imagen. Para entrenar el modelo, se emplean imágenes de radiografías de pelvis de infancias, y se realiza un aumento de datos utilizando GAN.

Este trabajo de investigación propone un método de análisis de radiografías de pelvis de bebés, tomando inspiración de investigaciones anteriores como la de Atalar et al. (2023), que analizó sonogramas de cadera utilizando aprendizaje profundo, y Pham et al. (2021), que destacó el uso de técnicas avanzadas de aprendizaje profundo en radiografías. Así como también en los trabajos de Den et al., (2023), y Fraiwan et al., (2022) que también usaron modelos de aprendizaje profundo, pero sin la necesidad de extraer información adicional de las radiografías, como el ángulo acetabular.

El componente de entrenamiento tiene como objetivo identificar los parámetros relevantes de la DDC en la imagen. Esto se logra a través del uso de un modelo híbrido de DL especializado en la clasificación de imágenes de radiografías con y sin presencia de DDC. La salida de este componente se emplea como entrada para el paso final, que corresponde al componente de clasificación. Su objetivo principal es determinar si existe o no presencia de la DDC en la imagen procesada. Este proceso secuencial garantiza una evaluación exhaustiva y precisa del estado de la cadera del infante. Adicionalmente, el componente de aumento de datos contribuye a el aseguramiento de datos suficientes para el entrenamiento del modelo para diferentes escenarios.

A continuación, se describen con más detalles cada uno de estos componentes, indicando su objetivo, pasos de implementación, y su justificación y novedad. Así mismo se detalla el conjunto de datos empleado.

4.2. Conjunto de datos

Los datos utilizados en este estudio se encuentran disponibles públicamente en el enlace <https://data.mendeley.com/datasets/jf3pv98m9g/1>, y comprenden imágenes de radiografías de cadera capturadas en vista anteroposterior (Fraïwan et al., 2022). Este conjunto de datos abarcó un total de 354 individuos, entre los cuales 120 presentaban DDC y 234 se consideraban normales. Las imágenes están disponibles en distintos tamaños que son adecuados para diversos modelos de aprendizaje profundo.

4.3. Preprocesamiento

Esta etapa marca el inicio del método propuesto, comenzando con la recepción de una radiografía de la pelvis de un infante vivo, de entre 3 y 6 meses de edad. Luego, se emplean varias técnicas para iniciar el procesamiento de la imagen. El objetivo del preprocesamiento es obtener una nueva imagen con características mejoradas para facilitar un análisis más efectivo.

4.3.1. Pasos

El preprocesamiento desempeña un papel fundamental en asegurar la calidad del modelo. En este trabajo, se resume la fase del preprocesamiento de los datos en tres pasos clave: Extracción de los valores de los píxeles, Normalización, y Escalamiento de las imágenes. Estos pasos se detallan a continuación.

- **Extracción de los valores de los píxeles:** Este paso es específico para el modelo híbrido, y más particularmente, para la red neuronal artificial (el primer modelo del sistema propuesto). Este paso extrae los valores de la intensidad de todos los píxeles, y los almacena en un vector plano. El cálculo realizado y almacenado en el vector de características es el resultado de calcular el promedio de los valores de píxeles en cada canal

de color (rojo, verde y azul) a lo largo de las dimensiones horizontal y vertical de la imagen. Este cálculo se realiza con la función `np.mean(img, axis=(0, 1))`, que devuelve un vector que representa el promedio de los valores de píxeles en cada canal de color.

En el contexto de imágenes RGB (Red, Green, Blue), cada píxel tiene tres valores asociados a los canales de color. Al tomar el promedio de estos valores para todos los píxeles en la imagen, se obtiene un resumen estadístico que refleja el tono medio de los colores presentes en la imagen. Este vector resultante se utiliza como una representación simplificada de las características visuales de la imagen.

- **Normalización:** La normalización de datos es una práctica común en el aprendizaje profundo, incluyendo el uso de redes neuronales, la normalización de datos se realiza con el objetivo de estandarizar la escala de las características de entrada, lo que puede llevar a una mejora en el rendimiento y la convergencia del modelo. En estudios realizados como el de W. Xu et al. (2022), incluye la normalización como parte de su preprocesamiento de imágenes siendo esta una referencia incluida en el presente trabajo. La normalización implica llevar las intensidades de la imagen con 3 canales de color RGB de 0 a 1.
- **Escalamiento:** ResNet50 tiene requisitos específicos sobre el tamaño de entrada, redimensionar las imágenes al tamaño requerido es común, utilizando técnicas como el reescalado o el recorte para garantizar la consistencia en el tamaño de entrada, lo que contribuye a un rendimiento más eficiente y predecible del modelo, todas las imágenes cargadas en esta tesis se ajustan uniformemente a dimensiones de 224x224 con 3 canales de color RGB. Fraiwan et al. (2022), utiliza en su trabajo esta técnica en su etapa de preprocesamiento de datos ya que es una forma general de asegurar mejor calidad en la entrada de modelos, siendo, también una referencia empleada.

4.3.2. Implementación

Los pasos de preprocesamiento se llevaron a cabo utilizando la biblioteca Numpy para la normalización de intensidad de las imágenes. Además, se emplearon los métodos provistos por TensorFlow ya que proporciona generadores de imágenes que procesan y escalan las imágenes de entrada. Esto permite cambiar el tamaño de una imagen a las dimensiones deseadas, en caso de ser necesario.

4.3.3. Justificación y novedad

De forma similar a los trabajos de Den et al., (2023), y Fraiwan et al., (2022), los datos se prepararon como la sección anterior lo comentan. La justificación de la preparación de datos mediante el ajuste uniforme de todas las imágenes a dimensiones de 224x224 con 3 canales de color RGB radica en la optimización del rendimiento del modelo y la simplificación del procesamiento. Al estandarizar las dimensiones, se logra una homogeneidad que facilita la comparación y extracción de patrones durante el entrenamiento del modelo. Esta uniformidad también contribuye a la eficiencia computacional al reducir la complejidad de las operaciones matriciales, acelerando así el proceso de entrenamiento y predicción.

La innovación en esta elección reside en la convergencia hacia un estándar que maximiza la utilidad y aplicabilidad del modelo en diversos contextos. La adopción de dimensiones específicas y canales de color RGB se alinea con las prácticas comunes en el procesamiento de imágenes, permitiendo una mayor interoperabilidad y facilitando la integración con tecnologías existentes. Además, esta estandarización posibilita la utilización de arquitecturas de modelos preentrenados en conjuntos de datos similares, aprovechando así el conocimiento previo y mejorando la eficacia del modelo en tareas específicas. En conjunto, la preparación de datos propuesta no solo optimiza el desempeño del modelo, sino que también representa una innovación pragmática al estandarizar y mejorar la versatilidad del enfoque.

4.4. Aumento de datos

El proceso de aumento de datos tiene como objetivo enriquecer el conjunto de imágenes de entrenamiento al aplicar diversas transformaciones a las imágenes originales. En este trabajo se ejecutaron dos técnicas de aumento de datos con el fin de realizar comparaciones.

4.4.1. Implementación con transformaciones simples

Para enriquecer el conjunto de datos original, se aplicó una serie de transformaciones como la rotación y el escalamiento para generar imágenes sintéticas. Estas transformaciones contribuyen a generar un dataset o conjunto de datos más variado y robusto, lo cual es esencial para mejorar el proceso de entrenamiento de nuestros modelos de aprendizaje profundo.

Se comenzó rotando las imágenes con un 70% de probabilidad, permitiendo una rotación de hasta 10 grados en ambas direcciones como sugiere Den et al., (2023). Estas transformaciones ayudan a crear un conjunto de datos más diverso y robusto, lo que es fundamental para optimizar el entrenamiento de nuestros modelos de aprendizaje profundo. Se empleó la librería Augmentor, ya que proporciona la función rotate que permite cambiar el tamaño y ángulo de una imagen a las dimensiones y ángulos deseados. Se aumentaron los datos de tal manera que hubiera 500 imágenes para cada clase.

4.4.2. Implementación con GAN

Se aplicó una técnica de aumento de datos utilizando una red Generativa Adversarial (GAN), cuya arquitectura se muestra en la Figura 9. Esta arquitectura consta de dos variantes de redes neuronales convolucionales, diseñadas para procesar y transformar entradas, en este caso, imágenes de radiografías. Se emplean dos entradas, denominadas Input (ruido) en la parte superior e Input* en la parte inferior. Cada bloque representa una operación o capa dentro de la red, con siglas que indican el tipo de operación realizada.

La parte superior representa la arquitectura del Generador, mientras que la parte inferior corresponde al Discriminador. Para la implementación de este modelo, se utilizaron las bibliotecas específicas de TensorFlow, como layers y models, para construir los modelos del Generador y el Discriminador. El entrenamiento se aceleró mediante el uso de un GPU proporcionado por Google Colab.

Como se mencionó previamente, la Figura 9 exhibe la estructura empleada, donde cada operación se representa mediante bloques. Estos bloques son convencionales en el diseño de redes neuronales destinadas a labores de visión por computadora y procesamiento de imágenes. La porción superior ilustra la arquitectura del Generador, mientras que la sección inferior de la figura corresponde al Discriminador.

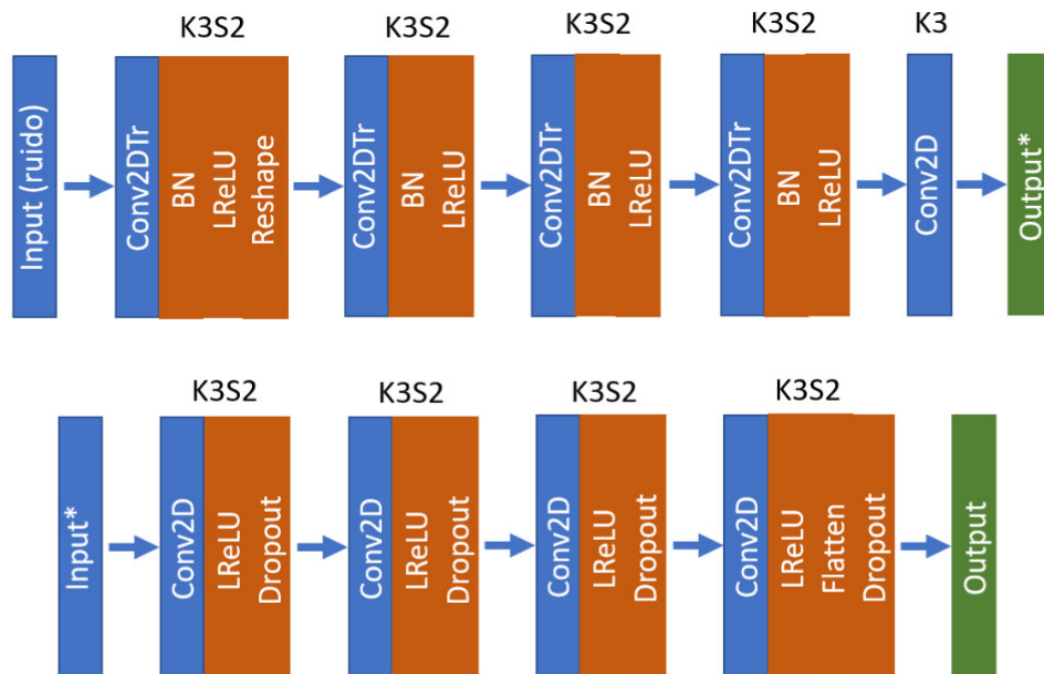


Figura 9. Arquitectura propuesta de una GAN para el aumento de datos.

Fuente. Elaboración propia.

A continuación, se describen las operaciones realizadas en la arquitectura GAN propuesta:

- **CONV2D:** Esta capa representa una convolución bidimensional que aplica una cantidad específica de filtros al input para extraer características relevantes. La designación "K3S2" indica que los filtros empleados tienen un tamaño de kernel de 3x3 y un stride de 2 (es decir, el filtro se desplaza 2 píxeles en cada paso). Mientras tanto, "K3" indica un tamaño de kernel de 3, con el stride por defecto no especificado, por lo que se asume un valor de 1.
- **BN:** La Normalización por lotes (Batch Normalization) es una técnica que normaliza las activaciones de la capa anterior. Esto contribuye a agilizar el proceso de entrenamiento y tiene un efecto de regularización que puede contrarrestar el sobreajuste.
- **ReLU:** La función de activación Rectified Linear Unit (Unidad Lineal Rectificada) introduce no linealidad en el modelo, lo que permite a la red neuronal aprender representaciones más complejas.
- **Reshape:** Esta operación cambia la forma del tensor de salida de la capa anterior sin alterar sus datos. Esta acción puede ser necesaria para ajustar las dimensiones de los datos a los requisitos de las capas subsiguientes.
- **Dropout:** La técnica de regularización en la que se apagan aleatoriamente nodos de la red durante el entrenamiento para prevenir el sobreajuste se conoce como Dropout. En Dropout, se seleccionan aleatoriamente nodos (o unidades) en la red y se desactivan temporalmente durante el entrenamiento con una cierta probabilidad. Esto obliga a la red a aprender características más robustas y reduce la dependencia de nodos específicos, lo que ayuda a prevenir el sobreajuste al hacer que la red sea más generalizable.

- **Flatten:** La operación que convierte el tensor multidimensional de salida de las capas convolucionales en un vector unidimensional es conocida como Flatten. Este paso es comúnmente utilizado antes de las capas densas en una CNN. Esta operación transforma la estructura de la salida de las capas convolucionales en un formato lineal, lo que permite que esta información sea utilizada por las capas densas para realizar la clasificación final o cualquier otra tarea de aprendizaje supervisado.

Estas técnicas se emplean en la generación de imágenes aumentadas de la siguiente forma: Las capas de CONV2D, con distintos tamaños de kernel y pasos, pueden identificar y resaltar diversas características de las imágenes, contribuyendo así a la creación de variantes únicas en las características de las imágenes sintéticas generadas. BN y ReLU se aplican de manera secuencial tras las capas de convolución para estabilizar y mejorar el proceso de aprendizaje, garantizando que las características generadas no se amplifiquen en exceso y mantengan una distribución uniforme. La capa de Dropout agrega una mayor regularización al proceso, lo que hace que la red sea menos sensible a la presencia exacta o ausencia de características específicas en las imágenes de entrada, lo que puede resultar en imágenes aumentadas más diversas. Por último, la capa de Flatten se utiliza para adecuar el tensor de salida a las capas completamente conectadas. Se aumentaron los datos de tal manera que se dispusieran de 500 imágenes para cada clase.

4.5. Entrenamiento y evaluación

El proceso de aprendizaje constituye la parte central del sistema propuesto. Aunque el énfasis del sistema se centra en entrenar un modelo híbrido de aprendizaje profundo, resulta crucial contrastar su rendimiento con los métodos clásicos de aprendizaje automático. Este análisis comparativo permitirá obtener una visión más completa sobre la precisión y efectividad de las técnicas basadas en DL en comparación con los enfoques tradicionales de

aprendizaje automático (P. Wang et al., 2021). La Tabla 3, muestra los parámetros principales y sus configuraciones para cada algoritmo.

Tabla 3. Parámetros configurados para cada algoritmo.

Método	Parámetros principales
Árbol de Decisión (AD)	- Criterio: Entropía - Mínimo de muestras: 1
Máquinas de Vectores de Soporte (SVM)	- Tipo de kernel: RBF - Parámetro C: 1.0
Sistema Propuesto (Los mismos para cada red, incluida para la CNN)	- Épocas: 50 - Tamaño del lote: 32 - Optimizador: adam - Tasa de aprendizaje: 0.001

Fuente. Elaboración propia.

4.5.1. Entrenamiento y evaluación de SVM

Para entrenar el modelo en este estudio, se eligió un clasificador de Máquinas de Vectores de Soporte (SVM) debido a su eficacia en la clasificación de datos de alta dimensión y su capacidad para manejar múltiples clases. Se utilizó un kernel radial (RBF), que es adecuado para datos donde la relación entre las clases no es lineal. El kernel RBF transforma los datos en un espacio de mayor dimensión donde es más probable que sean linealmente separables, lo que es esencial para el correcto funcionamiento de un SVM. Después de ajustar el SVM a los datos de entrenamiento, se procedió a realizar predicciones tanto en el conjunto de entrenamiento como en el conjunto de validación. Esta metodología posibilita evaluar no solamente el desempeño del modelo en los datos utilizados para el entrenamiento, sino también su capacidad para generalizar a nuevos conjuntos de datos (validación), lo cual resulta crucial para verificar la solidez del modelo.

4.5.2. Entrenamiento y evaluación de árboles de decisión

Para el entrenamiento del modelo en esta instancia, se optó por utilizar un Árbol de Decisión. Los Árboles de Decisión son conocidos por su facilidad de interpretación y su capacidad para manejar tanto datos numéricos como categóricos, lo que los hace ideales para una variedad de tareas de clasificación. Estos modelos funcionan dividiendo el espacio de los datos en regiones más pequeñas, basándose en decisiones sucesivas que se toman en cada nodo del árbol.

En el proceso, el Árbol de Decisión se entrenó con una configuración que incluye la división de los nodos basada en la pureza y la expansión del árbol hasta que todos los nodos contienen la misma clase o hasta que todos los nodos tienen menos muestras que el mínimo establecido.

Tras el proceso de entrenamiento del modelo, se llevaron a cabo predicciones tanto en el conjunto de entrenamiento como en el de validación. Este paso no solo proporciona una medida del ajuste del modelo a los datos de entrenamiento, sino que también evalúa su capacidad para generalizar a nuevos datos no vistos, lo cual es esencial en la validación de cualquier modelo de aprendizaje automático. Tanto para la máquina de vectores de soporte como para el árbol de decisión, se empleó la biblioteca Scikit Learn.

4.5.3. Entrenamiento y evaluación de CNN genérica

Una CNN se entrenó también como modelo base para la comparación del método propuesto. La primera red neuronal por entrenar es un modelo de CNN diseñado para la clasificación binaria (ver Figura 10). El modelo se entrena durante 50 épocas, utilizando los datos de entrenamiento (train_data, 80%) y los datos de validación (valid_data, 20%). El número de pasos por época se ajusta a la longitud del conjunto de datos de entrenamiento y de validación, respectivamente. Esto indica que se utilizará todo el conjunto de datos en cada época.

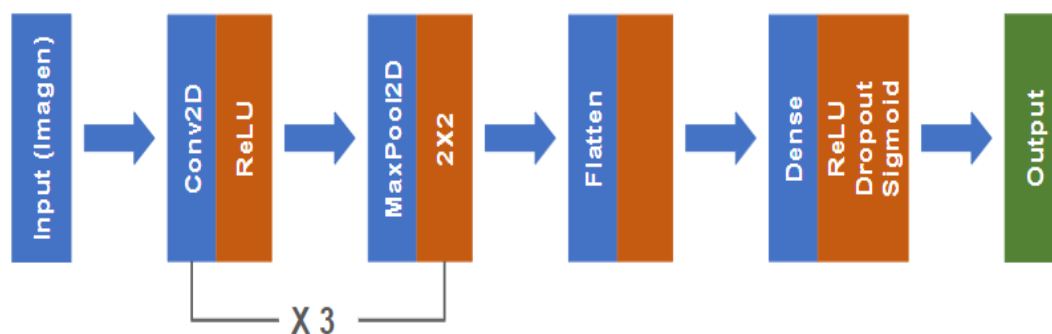


Figura 10. Arquitectura para la CNN genérica.

Fuente. Elaboración propia.

A continuación, en la Tabla 4, se detalla cada una de las capas y el proceso de entrenamiento. En primer lugar, el modelo se compila utilizando la función de pérdida Binary Cross Entropy (entropía cruzada binaria) y el optimizador Adam. Además, durante el entrenamiento, se monitorea la precisión como métrica.

Tabla 4. Capas de CNN genérica.

Capa	Descripción
Capa de Convolución (Conv2D)	- Número de filtros: 32. - Tamaño del kernel: 3x3. - Función de activación: ReLU - Tamaño de entrada: Imágenes de 224x224 píxeles en escala de grises.
Capa de Pooling (MaxPool2D)	- Tamaño del pool: 2x2.
Capa de Convolución (Conv2D)	- Número de filtros: 40. - Tamaño del kernel: 3x3. - Función de activación: ReLU.
Capa de Pooling (MaxPool2D)	- Tamaño del pool: Por defecto (2x2).
Capa de Convolución (Conv2D)	- Número de filtros: 128. - Tamaño del kernel: 3x3. - Función de activación: ReLU.

Capa de Pooling (MaxPool2D)	-	Tamaño del pool: Por defecto (2x2). Esta capa reduce a la mitad el tamaño de las características extraídas por la capa anterior.
Capa de Aplanamiento (Flatten)	-	Esta capa convierte el tensor tridimensional resultante de las capas anteriores en un vector unidimensional, preparándolo para las capas totalmente conectadas.
Capa Densa (Fully Connected)	-	Número de neuronas: 1. Función de activación: Sigmoide. Esta capa produce una salida única que representa la probabilidad de pertenencia a una de las dos clases (0 o 1).

Fuente. Elaboración propia.

En resumen, esta red neuronal consiste en una secuencia de layers o capas que ejecutan una convolución, pooling y capas densamente conectadas diseñadas para la clasificación binaria de imágenes. Utiliza la función de pérdida de Binary Cross Entropy y el optimizador Adam para el entrenamiento. La red se entrena durante 50 épocas.

4.5.4. Entrenamiento y evaluación de arquitectura híbrida

Para el sistema propuesto, se ha combinado una red neuronal convolucional (ResNet50) preentrenada con imagenet para el procesamiento de imágenes con una red neuronal densa para el manejo de características adicionales (Ver Figura 11). Al fusionar las salidas de ambas redes, se crea un modelo conjunto que puede aprender de manera más efectiva las representaciones tanto visuales como numéricas de las imágenes.

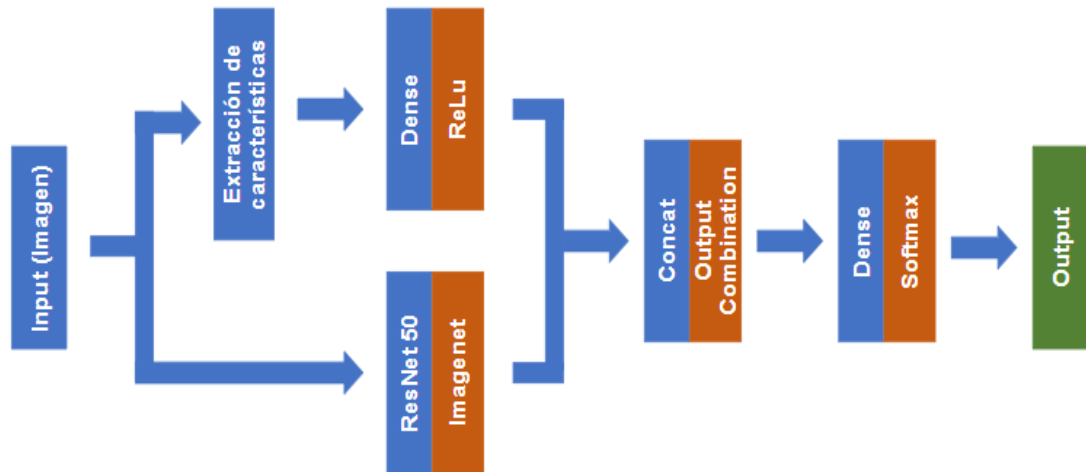


Figura 11. Arquitectura híbrida propuesta.

Fuente. Elaboración propia.

La hibridación de modelos es común en situaciones donde diferentes tipos de información deben ser considerados para realizar tareas complejas. En este contexto, la combinación de ResNet50 con una red neuronal densa agrega versatilidad al modelo, permitiéndole aprender patrones complejos y mejorar.

El proceso comienza con la carga de las imágenes desde las carpetas especificadas, donde se realiza un preprocesamiento adicional para ajustar las dimensiones de las imágenes al tamaño de entrada requerido por ResNet50 (224x224 píxeles). Además, se aplica el preprocesamiento específico de ResNet50 para normalizar las intensidades de píxeles. A continuación, se carga la arquitectura ResNet50 preentrenada, excluyendo la capa superior para permitir la extracción de características. Todas las capas de ResNet50 se congelan, asegurando que los pesos preentrenados se mantengan intactos durante el entrenamiento.

Las imágenes preprocesadas se pasan a través de ResNet50 para extraer características de alto nivel. Estas características se combinan con las características originales extraídas de las imágenes mediante capas densas adicionales. Esto permite capturar tanto las representaciones detalladas de las imágenes como las características aprendidas por ResNet50. El modelo final se compila utilizando la función de pérdida de entropía cruzada categórica

dispersa y el optimizador Adam. Luego, se procede a entrenar el modelo utilizando los datos de entrenamiento, y se evalúa su desempeño en el conjunto de prueba. La precisión en el conjunto de prueba se calcula y se presenta como resultado final.

ResNet-50 es una arquitectura de red neuronal convolucional profunda que se destaca por su capacidad para entrenar redes muy profundas con un rendimiento excepcional (He et al., 2016). Fue desarrollada por Microsoft Research y es una variante de las Redes Residuales (ResNet). La arquitectura de ResNet-50, se presenta a continuación:

- **Capa de Entrada:** La entrada de la red es una imagen de tamaño variable (generalmente 224x224 píxeles en prácticas estándar).
- **Bloques de Convolución y Pooling:** La arquitectura se compone de varias capas de convolución y max pooling organizadas en bloques.
- **Bloques Residuales:** El componente central de ResNet es el bloque residual. Consiste en dos rutas de procesamiento: la identidad directa y una ruta que pasa por capas de convolución. Luego, se combinan mediante una suma ponderada. Esta arquitectura posibilita que la red aprenda una representación residual, lo cual simplifica el entrenamiento de redes profundas sin enfrentar el problema del desvanecimiento del gradiente.
- **Capa Global Average Pooling (GAP):** Esta capa calcula el promedio de cada canal de la última capa de características, lo que resulta en un tensor de una sola dimensión.
- **Capa de Salida (Fully Connected):** Consiste en una capa densa que produce las predicciones finales. La cantidad de neuronas en esta capa depende del número de clases en el problema de clasificación.

- **Función de Activación:** Se utiliza una función de activación softmax en la capa de salida para obtener probabilidades de pertenencia a cada clase.

ResNet-50 específicamente se compone de 50 capas, incluyendo bloques residuales y capas convolucionales. Es conocida por su habilidad para aprender representaciones de características muy profundas, esto la hace especialmente efectiva para problemas de visión computacional. ResNet-50 ha demostrado un excelente rendimiento en la clasificación de imágenes en varios conjuntos de datos estándar, y es una arquitectura ampliamente utilizada en investigaciones y aplicaciones prácticas de visión por computadora.

4.5.5. Justificación y novedad

El uso de un modelo híbrido se prefirió puesto que la combinación de ResNet50, una arquitectura de redes neuronales residuales, con un modelo que utiliza características extraídas mediante el cálculo de un vector de características promediando los valores de píxeles de las imágenes de radiografía, representa un enfoque híbrido poderoso y efectivo. ResNet50 se elige por su capacidad para aprender de errores y manejar redes profundas, permitiendo capturar características complejas en las imágenes. Simultáneamente, el modelo basado en características extraídas proporciona una representación numérica resumida de las imágenes, complementando la capacidad de ResNet50. Esta combinación mejora la precisión y la eficacia en tareas médicas críticas, aprovechando tanto el aprendizaje profundo como la representación detallada de características específicas de las radiografías.

La elección de la arquitectura ResNet-50 para la clasificación de imágenes aporta una justificación sólida basada en su capacidad probada para aprender representaciones profundas de características. La novedad de esta propuesta radica en la aplicación específica de ResNet-50 para abordar problemas de clasificación en el contexto del conjunto de datos en cuestión. ResNet-50, con sus bloques residuales y capas convolucionales, ha demostrado destacarse en tareas de visión por computadora al superar desafíos asociados con la

propagación del gradiente en redes muy profundas. La aplicación de esta arquitectura no solo garantiza un aprendizaje efectivo de las características relevantes de las imágenes, sino que también potencialmente captura patrones sutiles que podrían escapar a modelos menos profundos.

La novedad en la propuesta también se destaca al optar por un enfoque híbrido que incluye una red neuronal artificial junto con ResNet-50. La combinación de estos dos modelos ofrece un enfoque integral para la clasificación, aprovechando la capacidad de ResNet-50 para aprender representaciones profundas y la flexibilidad de la red neuronal artificial para adaptarse a características específicas. La inclusión de capas densas con activación ReLU y regularización L2 en la red neuronal artificial sugiere una atención meticulosa a la capacidad de generalización y la prevención del sobreajuste.

En conjunto, la elección de ResNet-50 y la integración de una red neuronal artificial en un modelo híbrido representan una propuesta innovadora que capitaliza las fortalezas de ambas arquitecturas para abordar eficazmente el problema de clasificación de imágenes en el contexto específico del proyecto.

4.6. Clasificación

La clasificación se ejecuta con el modelo propuesto, y para lograr su implementación, se desarrolló una solución en la nube para la utilización del sistema.

4.6.1. Implementación del modelo seleccionado

Con el fin de compartir el modelo, se creó una aplicación en la nube con las siguientes capacidades:

- Cargar imagen con resolución de 224X224.
- Analizar la imagen, es decir, ejecutar la clasificación
- Visualizar la predicción y la probabilidad de la predicción

La Figura 12, muestra la pantalla de la aplicación alojada en la nube en *algorithmica.solutions*. La figura muestra cada una de las capacidades anteriormente mencionadas.

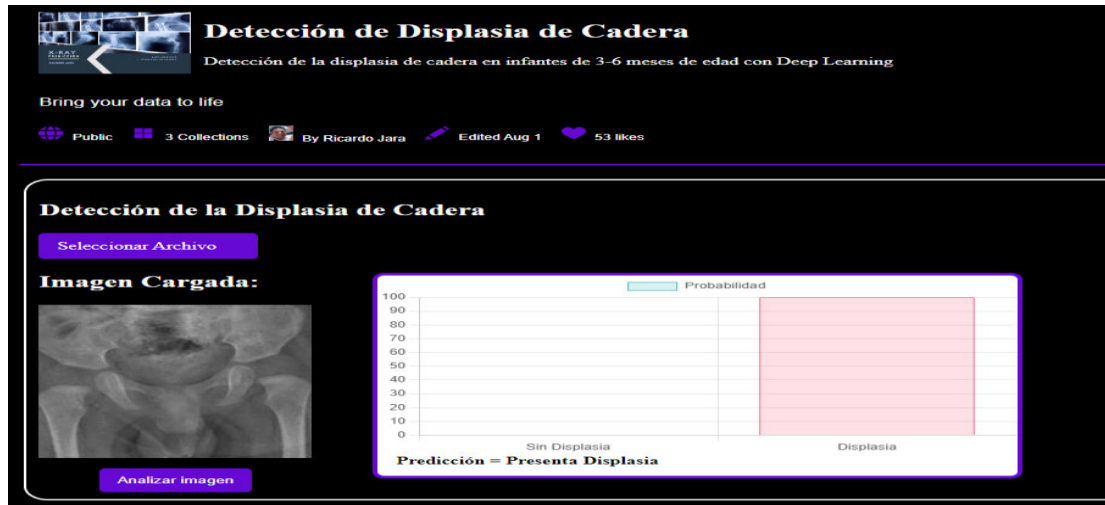


Figura 12. Frontend de la aplicación para la detección de la DDC.

Fuente. Elaboración Propia.

Los componentes para la aplicación son los siguientes:

- **AWS Lambdas:** Se crearon dos funciones Lambdas en AWS, una para el preprocesamiento de la imagen cargada, y otra para la clasificación. El modelo final entrenado se guardó en un archivo con extensión *pkl* que es usado por la aplicación.
- **MongoDB:** La base de datos no relacional MongoDB es usada para almacenar los resultados de las clasificaciones.
- **Algorithmica:** Es una plataforma para gestionar el hospedaje, el modelo y las bases de datos de la aplicación.

4.6.2. Justificación y novedad

La creación de una aplicación en la nube con las capacidades mencionadas representa una estrategia innovadora para democratizar el acceso y la utilización del modelo de clasificación desarrollado. La justificación para esta

aplicación radica en la necesidad de brindar una interfaz intuitiva y accesible que permita a usuarios con diversos niveles de experiencia interactuar con el modelo de manera efectiva. Establecer la capacidad de cargar imágenes con la resolución específica de las imágenes de entrenamiento (224x224) garantiza coherencia en el proceso de clasificación y simplifica la comparación con los resultados obtenidos durante el entrenamiento.

La capacidad para analizar la imagen en tiempo real, ejecutando la clasificación de manera inmediata, representa un paso adelante en términos de usabilidad y respuesta rápida. Esto no solo mejora la experiencia del usuario, sino que también podría ser crucial en entornos donde la eficiencia en la toma de decisiones es esencial, como en aplicaciones médicas o de diagnóstico.

La visualización de la predicción y la probabilidad proporciona una capa adicional de transparencia y comprensión para los usuarios, permitiéndoles evaluar la confianza del modelo en su clasificación. Esto es especialmente valioso en escenarios críticos donde la precisión y la confianza en la predicción son fundamentales.

La creación de esta aplicación en la nube no solo facilita el uso del modelo de clasificación, sino que también destaca la atención a la experiencia del usuario y la transparencia en el proceso de inferencia, contribuyendo así a la adopción efectiva y la confianza en la aplicación. En el siguiente capítulo, se discuten los resultados obtenidos con la metodología descrita.

CAPÍTULO 5. RESULTADOS Y DISCUSIONES

En este capítulo se describen y discuten los resultados obtenidos con la metodología expuesta en el capítulo anterior. Con fines prácticos se describe el sistema propuesto nuevamente. En esta tesis, se propone un sistema para la detección o ausencia de la DDC en radiografías pélvicas. El sistema usa un modelo híbrido de aprendizaje supervisado con ResNet 50 y una ANN densa. Para el entrenamiento, se usaron datos aumentados con transformaciones geométricas y con redes GAN.

Para contrastar los resultados obtenidos del sistema propuesto, se entrenaron distintos modelos de aprendizaje automático y de aprendizaje profundo: un Modelo de Árbol de Decisión, un Modelo de Máquina de Vectores de Soporte y una CNN genérica. Inicialmente, se analizan los resultados de cada uno de estos modelos, seguidos de la presentación de los resultados combinados, y finalmente se exponen las pruebas de hipótesis.

4.7. Análisis, interpretación y discusión de resultados

4.7.1. Resultados con los métodos clásicos

Los modelos de SVM y Árbol de decisión fueron entrenados para usarlos como referencia y comparación. En general, los métodos de ML tienen menos parámetros para sintonizar en comparación con el DL. También, las SVM pueden ser muy efectivas en espacios de alta dimensión, especialmente si hay un margen claro de separación en el espacio de características. Si sus características después de aplanar las imágenes son muy informativas, una SVM podría explotar esto de manera efectiva.

Los Árboles de decisión, por otro lado, en comparación con las redes neuronales suelen ser robustas ante valores atípicos y datos no normalizados. Los árboles de decisión son menos sensibles a los valores atípicos que las redes neuronales y no requieren que las características de entrada sean

escaladas o normalizadas. También, son más fáciles de interpretar y se pueden visualizar, lo que podría ayudar a ajustar el modelo para que funcione mejor en función del conocimiento del dominio. De forma similar a las SVM, los árboles de decisión tienen menos hiperparámetros que ajustar y, a menudo, es más rápido encontrar una buena configuración. Por el contrario, las redes neuronales tienen muchos hiperparámetros, lo que puede hacer que el proceso de ajuste sea más desafiante y requiera más tiempo. A continuación, se discuten los resultados para las SVM y los Árboles de decisión.

- **SVM:** La Figura 13 muestra los resultados de la SVM, la presencia de falsos negativos y positivos en ambas matrices sugiere que mientras el modelo es competente en la clasificación de casos normales, podría mejorar en la identificación de casos de DDC. Este desbalance en la precisión entre las clases podría ser un punto de partida para futuras investigaciones, donde se podrían explorar técnicas de balanceo de clases, ajuste de parámetros o incluso la implementación de un modelo de clasificación diferente.

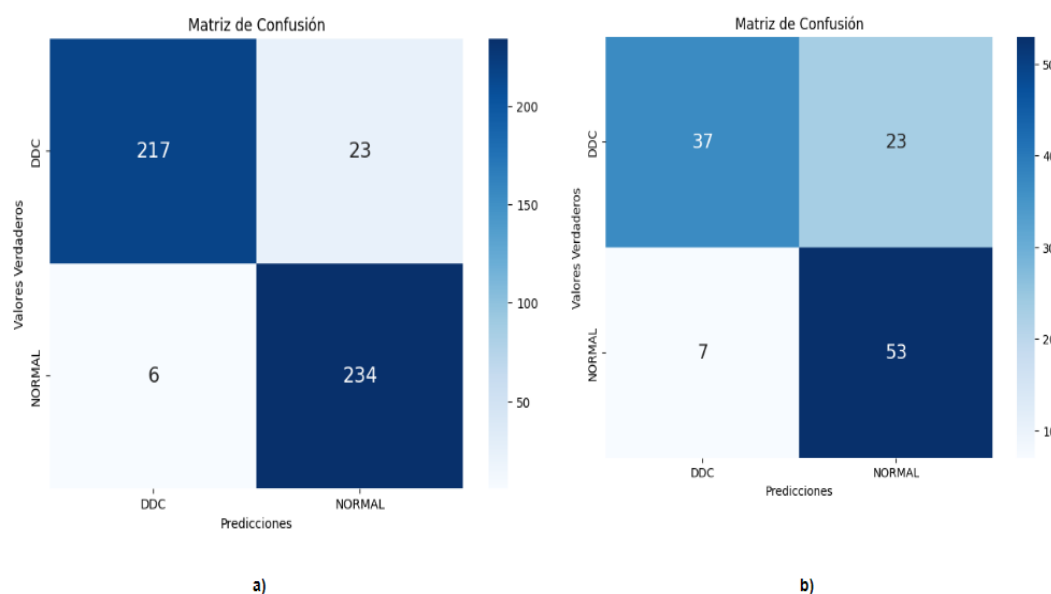


Figura 13. Matrices de confusión de la Máquina de Vectores de Soporte (SVM). a) Para los datos de entrenamiento. b) Para los datos de validación.

Fuente. Elaboración Propia.

- **Arboles de decisión:** Para analizar los resultados del Árbol de Decisión, se muestra la Figura 14. La matriz de entrenamiento (a) revela un ajuste perfecto del árbol de decisión a los datos de entrenamiento, lo que podría indicar un sobreajuste, especialmente al contrastarlo con los resultados de validación. El sobreajuste es una condición donde el modelo aprende las características de los datos de entrenamiento tan detalladamente que falla en generalizar a nuevos datos. Esto es evidenciado por la matriz de validación (b), donde el modelo ya no mantiene el rendimiento perfecto.

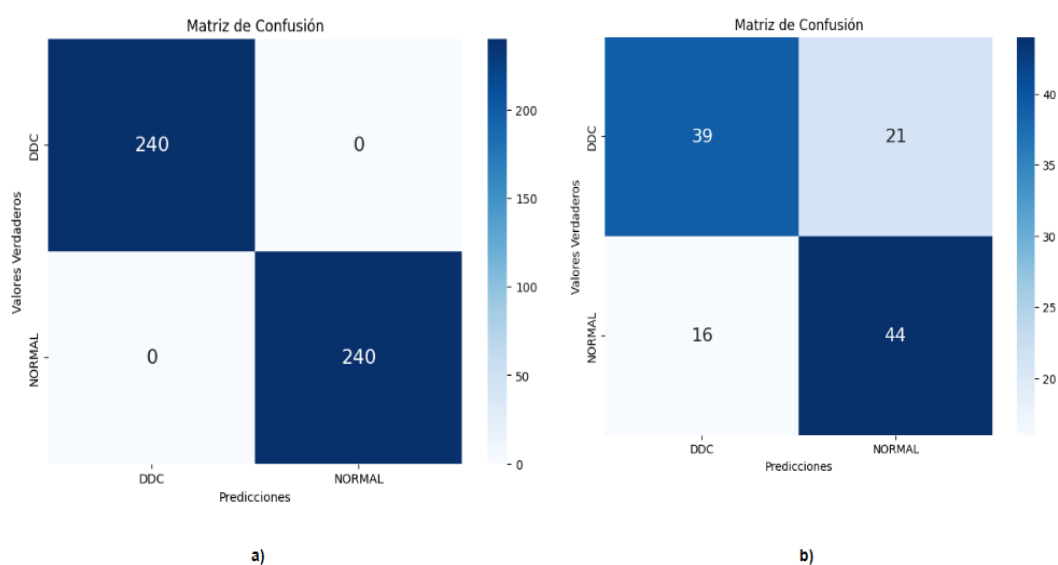


Figura 14. Matrices de confusión del Árbol de decisión. a) Para los datos de entrenamiento. b) Para los datos de validación.

Fuente. Elaboración Propia.

Continuando con la Figura 14, el número de verdaderos positivos y negativos en la matriz de validación muestra que el modelo tiene una capacidad razonable para clasificar correctamente las imágenes, pero la presencia de falsos positivos y negativos indica que hay margen de mejora. Los 21 falsos positivos pueden llevar a tratamientos o pruebas adicionales en un contexto clínico, mientras que los 16 falsos negativos podrían resultar en casos de DDC no diagnosticados, lo cual es particularmente preocupante.

Es crucial destacar que la perfecta clasificación en el conjunto de entrenamiento no se traduce en un desempeño igualmente alto en el conjunto

de validación, como lo muestra la Figura 14. Se subraya la importancia de la validación cruzada y de la selección de modelos basada en datos de validación en lugar de métricas de entrenamiento. Esto también puede implicar que el modelo necesita ser podado o que se debe ajustar la profundidad del árbol para evitar la captura de ruido estadístico o detalles irrelevantes que no se aplican al conjunto general de datos.

La eficacia de un árbol de decisión se beneficia de la simplicidad de interpretación y la rapidez de ejecución, lo que puede ser ventajoso en entornos clínicos donde las explicaciones y la rapidez son valiosas. Sin embargo, la tendencia al sobreajuste debe ser controlada mediante métodos de regularización, y las métricas de rendimiento deben ser cuidadosamente consideradas para asegurar que el modelo sea tanto preciso como generalizable.

Otra manera de evaluar los modelos es mediante el uso de la curva ROC. Esta herramienta gráfica ilustra la relación entre la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos ($1 - \text{especificidad}$) para un clasificador binario en diversos umbrales de decisión. La línea diagonal punteada representa el rendimiento de un clasificador aleatorio, y cuanto más se aleje la curva de esta línea hacia la esquina superior izquierda, mejor será el rendimiento del modelo.

En la Figura 15, La gráfica a) representa la curva ROC para el modelo SVM. Se observa que la curva se acerca bastante a la esquina superior izquierda, lo que indica una precisión de buen nivel para la clasificación. Esto se refuerza con un Área Bajo la Curva (AUC) de 0.86, lo que implica que el modelo SVM tiene un 86% de probabilidad de distinguir correctamente entre las clases positiva y negativa. La gráfica b) muestra la curva ROC para el modelo de Árbol de Decisión. Aquí, la curva está más cerca de la línea diagonal y se aleja de la esquina superior izquierda comparada con la curva SVM. El AUC es de 0.69, lo que sugiere que el Árbol de Decisión tiene un 69% de probabilidad de distinguir correctamente entre las clases, que es inferior al modelo SVM.

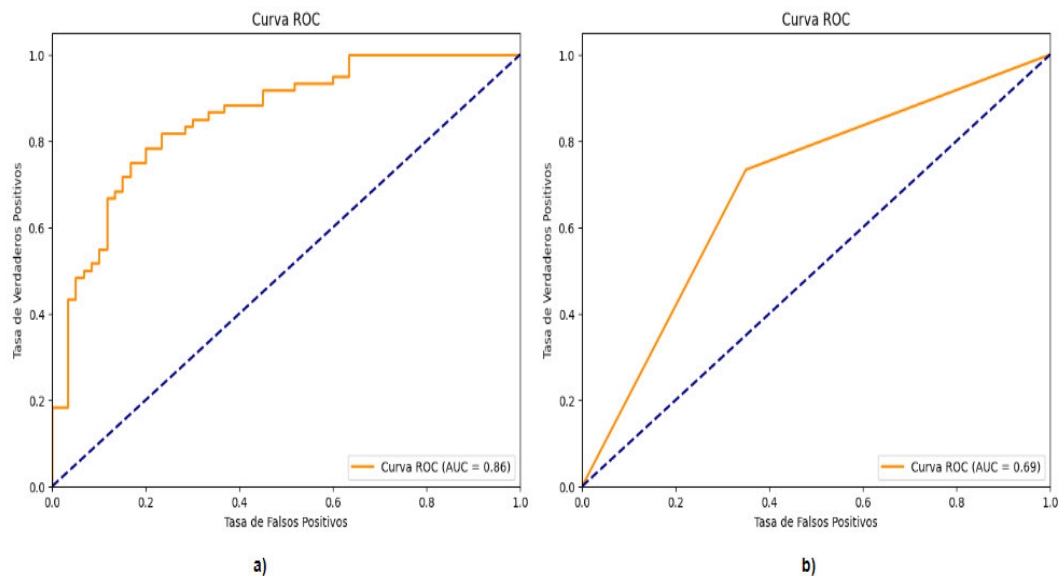


Figura 15. Curva ROC para a) La Máquina de Vectores de Soporte, y b) El Árbol de decisión.

Fuente. Elaboración Propia.

Comparando ambas curvas ROC, es evidente que el modelo SVM (gráfica a) supera al Árbol de Decisión (gráfica b) en términos de la capacidad para clasificar correctamente entre las clases positiva y negativa. El modelo SVM tiene un AUC significativamente más alto, lo que indica una mejor performance general en la clasificación. La diferencia en el AUC entre los dos modelos también puede interpretarse como una medida de la "separabilidad de clases" que cada modelo puede lograr. Un AUC de 0.86 implica que el modelo SVM es bastante eficaz en la separación de las clases, mientras que el AUC de 0.69 para el Árbol de Decisión sugiere que hay una considerable superposición entre las distribuciones de las clases positivas y negativas, lo que resulta en una mayor tasa de clasificaciones erróneas.

En el contexto de una aplicación clínica donde las ROC son utilizadas, por ejemplo, para la clasificación de imágenes médicas, el modelo con el AUC más alto (en este caso, el SVM) sería preferible debido a que ofrece una mayor confianza en la predicción de resultados verdaderamente positivos y negativos, minimizando así tanto los falsos negativos (que podrían llevar a no tratar una condición médica) como los falsos positivos (que podrían llevar a tratamientos innecesarios).

Es importante considerar que, aunque el AUC proporciona una medida útil del rendimiento general del modelo, no debe ser el único criterio para la selección del modelo. Otros factores como la interpretabilidad del modelo, el costo computacional, y la importancia de los diferentes tipos de errores (falsos positivos contra falsos negativos) también deben tenerse en cuenta al tomar una decisión en un entorno clínico.

4.7.2. Resultados con la CNN genérica

Esta red está destinada a procesar imágenes y determinar si pertenecen a una de las dos categorías: Normal o DDC. Primero, se realizan algunas importaciones necesarias, como ImageDataGenerator para generar lotes de imágenes con aumento de datos, cv2 para el procesamiento de imágenes, y otras bibliotecas esenciales como os y tensorflow. Luego, se crea el modelo de la CNN. Se utiliza un modelo secuencial (modelo) que permite apilar capas una tras otra. El proceso inicia con una capa convolucional que emplea 32 filtros de tamaño 3x3 y la función de activación ReLU. Esta capa recibe imágenes de entrada de 224 x 224 píxeles en formato RGB.

A continuación, se agregan capas de MaxPooling para reducir el tamaño de las características extraídas. Estas capas ayudan a reducir la complejidad y mejorar la eficiencia del modelo. Se repite este proceso de agregar capas convolucionales y de MaxPooling varias veces para aumentar gradualmente la capacidad del modelo para identificar características importantes en las imágenes. Después de la última capa de MaxPooling, se utiliza una capa de aplanamiento para convertir las características 2D en un vector de una sola dimensión.

Posteriormente, se añade una capa densa con 64 neuronas y la función de activación ReLU. Para evitar el sobreajuste, se emplea una técnica llamada "Dropout", que apaga aleatoriamente el 70% de las neuronas durante el entrenamiento. Finalmente, se añade una capa de salida con una sola neurona y una función de activación sigmoideal. Esta capa genera la salida final

de la red, que representa la probabilidad de pertenencia a una de las dos clases.

El modelo se configura utilizando el optimizador 'adam' y la función de pérdida 'binary_crossentropy', ambos ampliamente utilizados en problemas de clasificación binaria. Se emplea la métrica de precisión ('accuracy') para evaluar el rendimiento del modelo. El último paso implica entrenar el modelo utilizando conjuntos de datos de entrenamiento y validación. La variable batchsize se define en 32, y el número de épocas de entrenamiento se establece en 50. Durante el entrenamiento, se utiliza el conjunto de datos de validación para evaluar cómo se desempeña el modelo en datos no vistos, lo que ayuda a evitar el sobreajuste.

La Figura 16 muestra la matriz de confusión y la curva ROC resultado del modelo de CNN entrenado. Los valores dentro de la matriz de confusión indican que el modelo predijo 32 casos de DDC correctamente y 34 casos normales correctamente. Sin embargo, hubo 26 casos donde el modelo predijo DDC en lugar de Normal y 28 casos donde predijo Normal en lugar de DDC; resultados bastante deficientes.

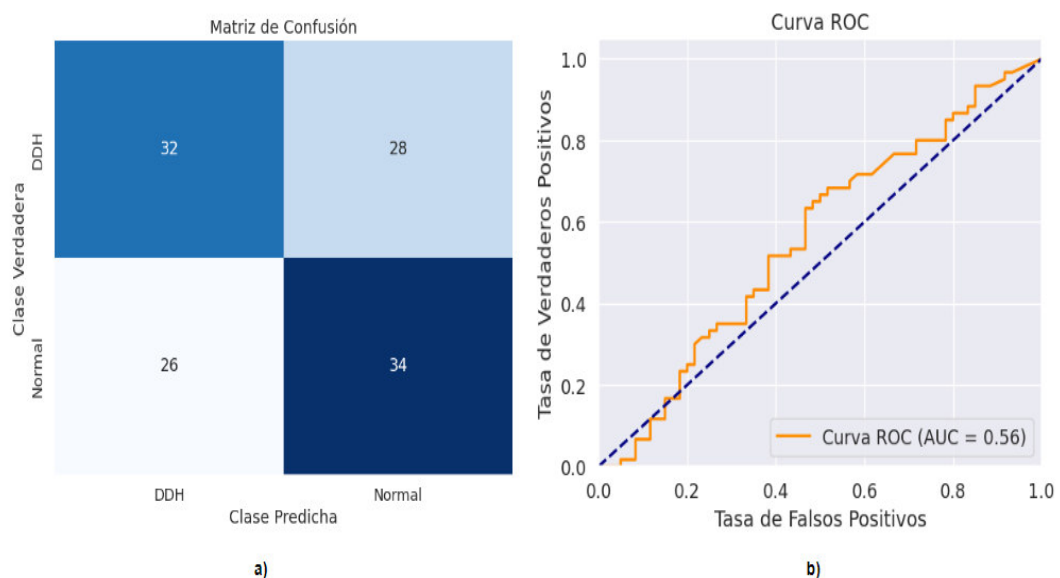


Figura 16.a) Matriz de confusión y b) Curva ROC para la CNN genérica
Fuente. Elaboración Propia.

El segundo gráfico de la Figura 16 es la curva ROC. La línea naranja muestra la tasa de verdaderos positivos frente a la tasa de falsos positivos a diferentes umbrales de decisión. La línea discontinua azul representa el rendimiento de un clasificador aleatorio. El área bajo la curva (AUC) es un indicador de cuán bueno es el modelo para distinguir entre las clases; en este caso, tiene un valor de 0.56, lo que sugiere un rendimiento apenas mejor que el azar.

4.7.3. Resultados con el modelo híbrido propuesto

El sistema propuesto es un modelo de aprendizaje automático avanzado que combina el procesamiento de imágenes con el análisis de características numéricas para clasificar con precisión las imágenes médicas en dos categorías: DDC y Normal. Mediante la integración de ResNet 50 y redes neuronales densas ANN, el modelo logra captar y comprender patrones complejos presentes tanto en los datos de imagen originales como en las características extraídas.

Para comenzar, el sistema carga imágenes desde un directorio estructurado y extrae características representativas promediando los valores de píxeles. Cada imagen se etiqueta apropiadamente según su categoría, y estas etiquetas se utilizan más adelante en el proceso de entrenamiento para guiar al modelo hacia una clasificación correcta. Las características numéricas se normalizan para proporcionar consistencia y mejorar la eficiencia del entrenamiento. Posteriormente, se procede a dividir los datos en conjuntos de entrenamiento y prueba, garantizando así una validación imparcial del modelo.

La arquitectura del modelo es una fusión de dos redes separadas. Una red convolucional con la arquitectura ResNet 50 procesa las imágenes, aprendiendo jerarquías de características visuales a través de capas que incluyen convoluciones y agrupamiento. Paralelamente, una red densa maneja las características numéricas extraídas. Ambas redes están reguladas por L2 para prevenir el sobreajuste, manteniendo así la generalización del modelo. Después de combinar las salidas de ambas redes, una capa de

clasificación final con activación 'softmax' asigna probabilidades a las categorías de clasificación. El modelo se entrena con un algoritmo de optimización 'adam' y una función de pérdida diseñada para la clasificación de múltiples categorías, con la precisión como métrica clave.

El entrenamiento del modelo se realiza durante 20 épocas, lo que permite que el sistema aprenda eficientemente de los datos disponibles. Una vez entrenado, el modelo demuestra una buena precisión en el conjunto de prueba, lo que se evidencia por una Matriz de Confusión con una alta cantidad de verdaderos positivos y mínimos errores de clasificación (ver Figura 17). Además, una Curva ROC con un AUC cercano a 1 reafirma la capacidad del modelo para distinguir entre las categorías de "DDC" y "Normal" con gran éxito.

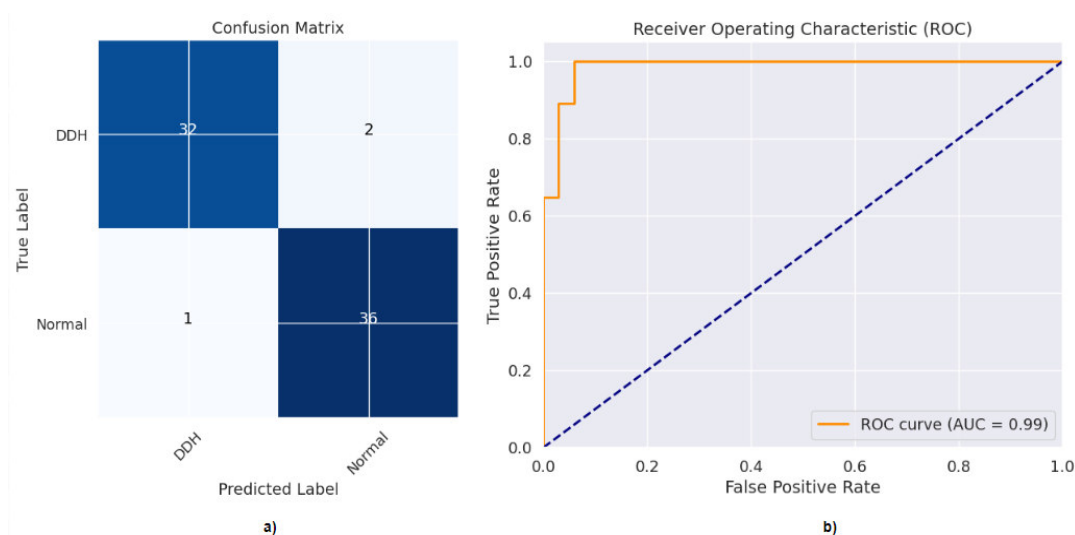


Figura 17.a) Matriz de confusión y b) Curva ROC para híbrido propuesto
Fuente. Elaboración Propia.

4.8. Pruebas de hipótesis

En esta sección, se presentan las hipótesis, así como los resultados de las pruebas efectuadas para evaluar la contribución significativa de un modelo basado en aprendizaje profundo en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.

4.8.1. Hipótesis general

Un modelo basado en aprendizaje profundo contribuye significativamente en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad. Los resultados revelaron que el sistema propuesto es útil para la clasificación de radiografías de cadera.

4.8.2. Hipótesis específicas

- **Un modelo basado en aprendizaje profundo contribuye significativamente con aceptable grado de Precisión en la detección de la DDC por radiografía pélvica en infantes de 03 a 06 meses de edad.**

En términos de precisión, el modelo basado en aprendizaje profundo demostró un rendimiento significativamente superior, con una tasa de aciertos del 97.50% en comparación con el grupo de control.

- **Un modelo basado en aprendizaje profundo contribuye significativamente con aceptable grado de Sensibilidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.**

La sensibilidad del modelo fue notablemente alta, alcanzando un 95.16%, lo que indica su capacidad para identificar con éxito casos de DDC.

- **Un modelo basado en aprendizaje profundo contribuye significativamente con aceptable grado de Especificidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.**

La especificidad del modelo también fue destacable, con una tasa de 100%, indicando su habilidad para descartar correctamente casos negativos de DDC.

- **Un modelo basado en aprendizaje profundo contribuye significativamente con rendimiento óptimo en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.**

Al evaluar el rendimiento general, se observó que el modelo logró un área bajo la curva ROC de 0.99, confirmando su capacidad para equilibrar la sensibilidad y la especificidad de manera óptima.

En conclusión, los resultados respaldan la hipótesis general de que un modelo fundamentado en el aprendizaje profundo tiene un impacto significativo en la detección de la displasia del desarrollo de cadera en infantes de 3 a 6 meses de edad, mostrando un rendimiento sobresaliente en términos de precisión, sensibilidad y especificidad.

4.9. Presentación de resultados

4.9.1. Resultados con aumento de datos con transformaciones simples

La Tabla 5 presenta una comparativa de los resultados logrados por diversos modelos en cuanto a precisión, sensibilidad y especificidad. El modelo AD (Árbol de decisión) logra una precisión respetable del 87.76%, lo que significa que acierta en la mayoría de las clasificaciones.

Sin embargo, su sensibilidad del 76.93% indica que a veces falla en la detección de casos positivos, lo que podría ser un problema en aplicaciones médicas donde la identificación de enfermedades es crítica. La especificidad del 83.28% es sólida, lo que indica que el modelo tiene una buena capacidad para identificar casos negativos. Por otro lado, el modelo SVM logra una sensibilidad significativamente mejor del 81.66%, lo que lo hace más adecuado para la detección de casos positivos. Sin embargo, su precisión y especificidad son un poco más bajas en comparación con el modelo AD.

Tabla 5. Comparación de los resultados de los diferentes modelos entrenados con los datos aumentados con transformaciones simples.

	Precisión	Sensibilidad	Especificidad
Árbol de decisión	87.76 %	76.93 %	83.28 %
SVM	83.63 %	81.66 %	89.33 %
CNN	43.55 %	45.00 %	41.67 %
Sistema Propuesto	97.50 %	95.16 %	100.00 %

Fuente. Elaboración propia.

La CNN muestra un rendimiento inferior en términos de precisión, sensibilidad y especificidad en comparación con los otros modelos. Esto podría deberse a la configuración específica de la red neuronal o a la necesidad de ajustar sus hiperparámetros. Finalmente, el Sistema Propuesto sobresale en términos de precisión, alcanzando un 97.50%. Además, muestra una sensibilidad del 95.16%, lo que indica una capacidad para detectar casos positivos, y una especificidad del 100.00%, lo que significa que no comete errores al clasificar los casos negativos. Este modelo propuesto parece ser la elección preferida cuando se busca un alto rendimiento en la detección de casos en este contexto específico.

4.9.2. Resultados con aumento de datos con la GAN

Los resultados presentados en la Tabla 6 revelan importantes hallazgos sobre el rendimiento de los diversos modelos entrenados con datos aumentados mediante la técnica GAN. El modelo AD exhibe una precisión respetable del 85.43%, lo que sugiere una capacidad sólida para clasificar con precisión. Sin embargo, su sensibilidad del 82.93% indica una propensión a errores en la detección de casos positivos. Aunque su especificidad del 85.02% es satisfactoria, estos resultados sugieren que el modelo AD podría beneficiarse de una mejora en la detección de casos positivos.

Por otro lado, el modelo SVM presenta una precisión ligeramente superior del 89.61%, destacando su habilidad para clasificar con precisión. Sin embargo,

su sensibilidad del 78.45% sugiere una menor capacidad para detectar casos positivos en comparación con el modelo AD. A pesar de ello, su alta especificidad del 89.15% indica una destacada capacidad para identificar casos negativos. Por último, el modelo CNN muestra un desempeño notablemente inferior en términos de precisión, sensibilidad y especificidad, evidenciando su limitada eficacia en la clasificación de datos aumentados.

Sin embargo, el sistema propuesto exhibe resultados excepcionales con una precisión del 98.25%, una sensibilidad del 96.55% y una especificidad del 98.39%, lo que sugiere su idoneidad para la detección precisa tanto de casos positivos como negativos, destacando su potencial para aplicaciones médicas donde la precisión es crucial. Estos resultados respaldan la efectividad del enfoque propuesto, subrayando su superioridad sobre los modelos comparativos en la tarea de clasificación.

Tabla 6. Comparación de los resultados de los diferentes modelos entrenados con los datos aumentados con la GAN.

	Precisión	Sensibilidad	Especificidad
Árbol de decisión	85.43 %	82.93 %	85.02 %
SVM	89.61 %	78.45 %	89.15 %
CNN	41.45 %	44.17 %	38.51 %
Sistema Propuesto	98.25 %	96.55 %	98.39 %

Fuente. Elaboración propia.

4.9.3. Resultados con el sistema propuesto

Los resultados obtenidos con el sistema propuesto se exponen en la Figura 18. El primer gráfico es una matriz de confusión. La matriz muestra que el modelo ha hecho predicciones muy precisas, con 32 predicciones correctas para DDC y 36 para Normal. Solo hay un pequeño número de predicciones incorrectas: 2 casos en los que se predijo Normal cuando la etiqueta verdadera era DDC, y 1 caso de lo contrario.

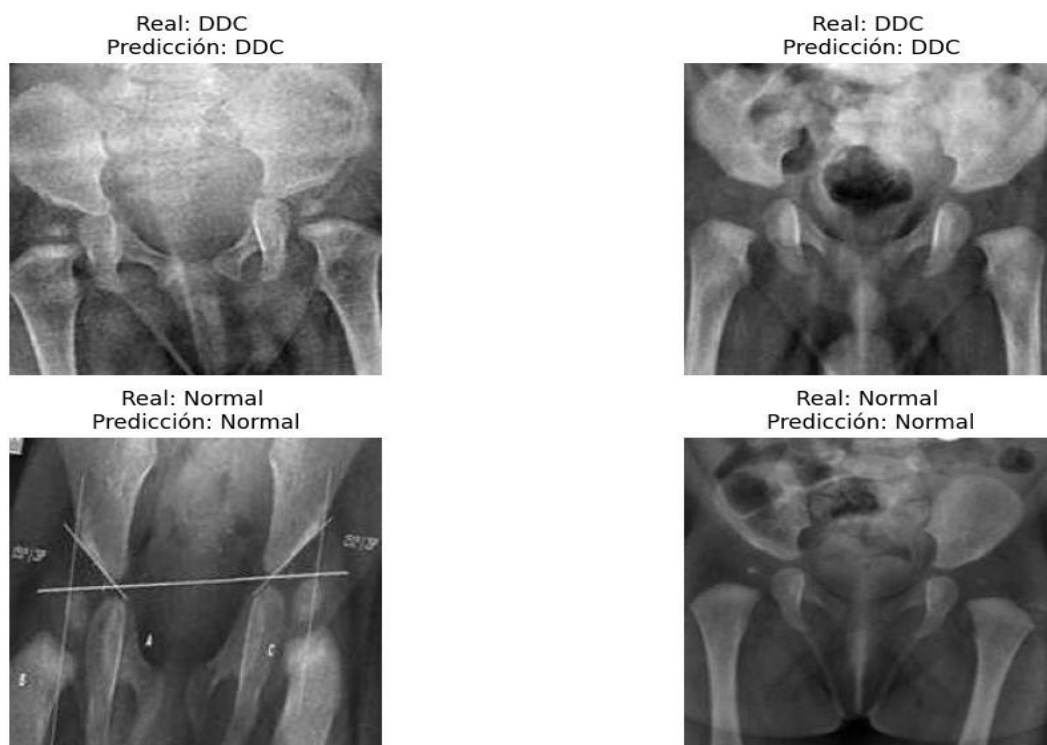


Figura 18. Resultados de la clasificación con el sistema propuesto.

Fuente. Elaboración Propia.

En las Figuras 18 y 19 se ofrecen ejemplos de los resultados alcanzados. En el primer análisis, los resultados de la Figura 18 señalan que el sistema propuesto es eficaz en diferenciar entre caderas normales y aquellas afectadas por la DDC. La alta confiabilidad en las predicciones, a menudo superior al 90%, sugiere un entrenamiento sólido del algoritmo, respaldado por un conjunto de datos robusto y representativo. Esta precisión es esencial en el contexto de diagnósticos médicos, ya que la exactitud puede tener un impacto directo en el plan de tratamiento y el pronóstico del paciente.

Sin embargo, al considerar otro aspecto importante, es importante recordar que, a pesar de la efectividad del modelo de IA, la heterogeneidad en la presentación clínica de la DDC puede plantear desafíos. Los casos con porcentajes de confianza más bajos pueden representar variaciones anatómicas o grados de severidad menos comunes en el conjunto de entrenamiento del modelo. Este escenario enfatiza la necesidad de una supervisión médica rigurosa, donde la IA actúe como una herramienta de triaje o como un segundo lector, en lugar de ser el único árbitro del diagnóstico.

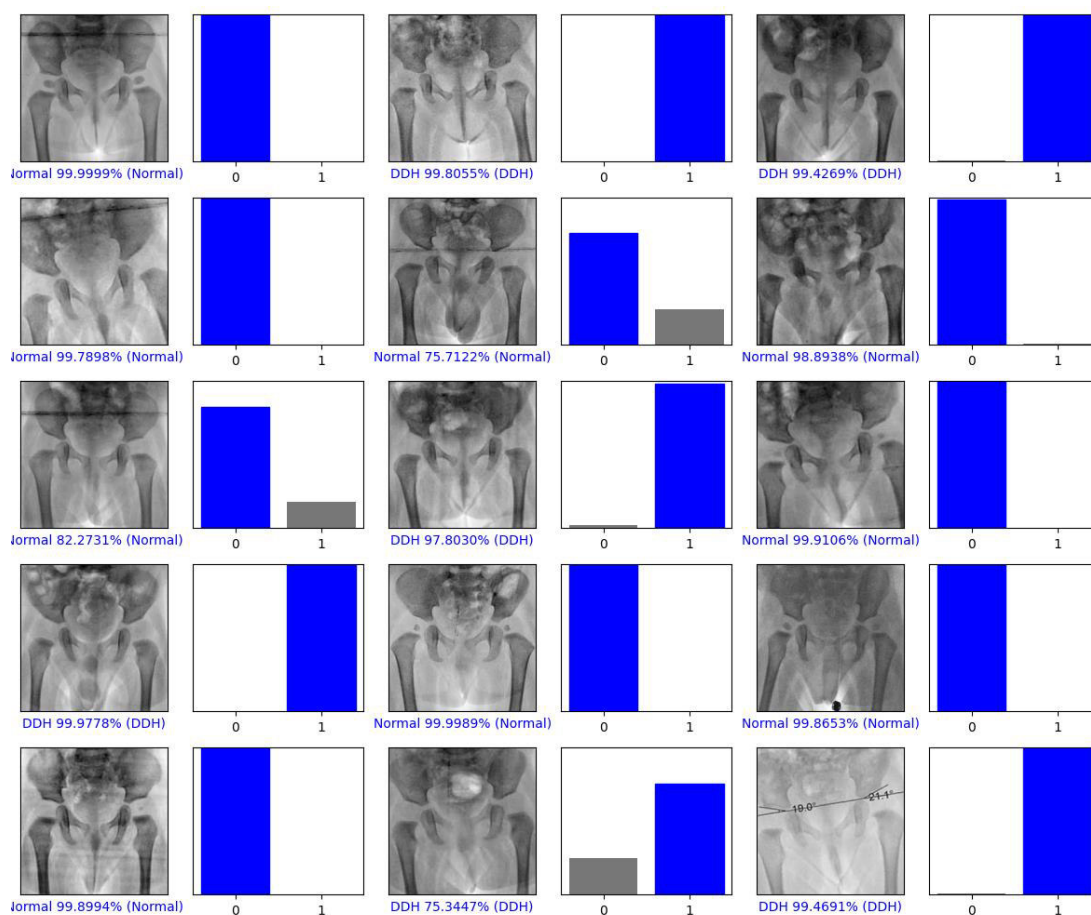


Figura 19. Probabilidades obtenidas por clase del sistema propuesto.

Fuente. Elaboración Propia.

Finalmente, en un tercer aspecto a considerar, la implementación de sistemas de IA, como los que se muestran en las imágenes, plantea cuestiones fundamentales sobre su integración en la práctica clínica. Por un lado, la adopción de esta tecnología puede aumentar la eficiencia y la capacidad para detectar condiciones en etapas tempranas, lo que puede mejorar los resultados para los pacientes. Por otro lado, la dependencia en algoritmos automáticos para diagnósticos sensibles subraya la necesidad de establecer protocolos de control de calidad, proporcionar una formación continua a los médicos de la salud en la interpretación de estos resultados y desarrollar marcos éticos que garanticen una toma de decisiones médicas responsable y centrada en el paciente.

CONCLUSIONES Y TRABAJOS FUTUROS

➤ Conclusiones generales

La presente investigación se centra en el desarrollo de un modelo basado en aprendizaje profundo para la detección de displasia de desarrollo de cadera (DDC) mediante el análisis de radiografías pélvicas de infantes con edades comprendidas entre los 3 y 6 meses.

La detección de DDC se aborda mediante el empleo de arquitecturas profundas que aprenden de imágenes radiográficas. Asimismo, se emplean redes GAN para aumentar el conjunto de datos, lo que mejora el entrenamiento de los modelos de redes neuronales al disponer de un mayor número de muestras. En este estudio, se destaca la integración de un sistema híbrido que combina una Red Neuronal Artificial (ANN) con una arquitectura CNN basada en ResNet50. Además, se aplicaron técnicas de aumento de datos tanto tradicionales como con el uso de GAN.

La utilización de ResNet50, una arquitectura avanzada de redes neuronales resulta crucial debido a su capacidad para captar características detalladas y patrones complejos, lo cual se demuestra como especialmente efectivo en contextos médicos donde la precisión en los detalles es primordial. Del mismo modo, se evidencia que el modelo de ANN contribuye significativamente a la mejora de los resultados obtenidos.

Por otro lado, el enriquecimiento del conjunto de entrenamiento mediante la generación de datos sintéticos, junto con las transformaciones geométricas y el uso de GAN, ofreció una amplia diversidad y cantidad de datos que beneficiaron de manera considerable el proceso de aprendizaje del modelo. Este enfoque se revela como especialmente valioso en el campo de la radiología, donde la obtención de grandes volúmenes de datos etiquetados suele ser un desafío constante.

Al comparar el sistema propuesto con una máquina de vectores de soporte y un árbol de decisión, se observó una mejora notable. Mientras que la CNN sin datos aumentados logró una precisión del 43.55%, nuestro sistema alcanzó un impresionante 97.5% de precisión. Esta diferencia destaca tanto la efectividad de las técnicas avanzadas de aprendizaje profundo como la importancia del aumento de datos para mejorar el rendimiento de los modelos de aprendizaje automático en tareas complejas como la clasificación de imágenes médicas.

La combinación de una ANN junto con ResNet50, además de los datos aumentados con transformaciones geométricas y GAN, resultó en una mejora significativa en la precisión y confiabilidad del modelo para clasificar radiografías con y sin DDC. Estos hallazgos son prometedores para futuras aplicaciones en diagnóstico médico asistido por computadora y sugieren posibilidades interesantes para investigaciones adicionales sobre la adaptación de estas tecnologías a otras áreas de la medicina.

Por otro lado, la creación de una aplicación en la nube representa la culminación de este proyecto, permitiendo la accesibilidad y utilización práctica del modelo por parte de usuarios con diversos niveles de experiencia. La atención a la experiencia del usuario, la transparencia en la inferencia y la capacidad de análisis en tiempo real reflejan un enfoque centrado en el usuario y la aplicación efectiva del modelo en entornos prácticos.

➤ Conclusiones específicas

OE1: Desarrollar un modelo basado en aprendizaje profundo que contribuya con aceptable grado de Precisión en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.

El modelo híbrido propuesto alcanzó un rendimiento en términos de precisión de una tasa de aciertos del 97.50% en comparación con los modelos de aprendizaje de máquina clásico y la CNN genérica.

OE2: Desarrollar un modelo basado en aprendizaje profundo que contribuya con aceptable grado de Sensibilidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.

La sensibilidad del modelo alcanzó un 95.16%, lo que indica su capacidad para identificar con éxito casos positivos de displasia de desarrollo de cadera.

OE3: Desarrollar un modelo basado en aprendizaje profundo que contribuya con aceptable grado de Especificidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.

El modelo híbrido propuesto también fue destacable en términos de especificidad. Con una tasa de 100%, se indica su habilidad para descartar correctamente casos negativos de DDC.

OE4: Desarrollar un modelo basado en aprendizaje profundo que contribuya con rendimiento óptimo en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.

Las diferentes pruebas realizadas con el modelo propuesto, además de la comparación con modelos de aprendizaje de máquina clásico y una CNN genérica, se confirmó la capacidad del modelo propuesto para equilibrar la sensibilidad y la especificidad de manera óptima.

➤ **Trabajos futuros**

En relación para futuros trabajos, existen varias direcciones prometedoras para expandir y perfeccionar los descubrimientos de este estudio. Una opción

es investigar diferentes estructuras de redes neuronales para evaluar su eficacia en la clasificación de radiografías pélvicas. Variantes como Inception, VGG y DenseNet, cada una con sus particularidades en procesamiento y aprendizaje de características, podrían ofrecer resultados valiosos y potencialmente mejorar la precisión del modelo. Asimismo, una continuación lógica de este trabajo sería entrenar modelos no solo para detectar la presencia de displasia del desarrollo de la cadera, sino también para clasificar sus distintos grados de severidad.

RECOMENDACIONES

A continuación, se presentan algunas recomendaciones pertinentes para continuar con el trabajo futuro. En el desarrollo continuo sobre la detección de la displasia del desarrollo de la cadera (DDC) mediante técnicas de aprendizaje profundo, hay diversas direcciones prometedoras que podrían enriquecer el trabajo y ofrecer avances significativos en el campo del diagnóstico médico asistido por computadora.

Primero, sería prudente expandir la investigación actual para incluir una variedad de arquitecturas de redes neuronales. Mientras que ResNet50 ha demostrado ser eficaz, explorar arquitecturas alternativas como Inception, VGG y DenseNet podría ofrecer perspectivas novedosas y potencialmente superar los resultados actuales. Cada una de estas estructuras posee atributos singulares en el procesamiento y la captación de características, aspectos que podrían resultar ventajosos para la clasificación de imágenes médicas.

Además, es crucial ampliar la base de datos empleada para el entrenamiento de los modelos. Integrar un conjunto más amplio de radiografías pélvicas provenientes de diversos grupos demográficos enriquecería la solidez y la capacidad de generalización del modelo, garantizando su eficacia en una variedad amplia de contextos clínicos. Esta diversificación es vital, ya que los conjuntos de datos limitados pueden llevar a sesgos y a una menor eficacia en condiciones reales.

La explicabilidad de los modelos de inteligencia artificial también es un aspecto que debe enfatizarse. Los modelos que pueden explicar sus decisiones en términos clínicos aumentarán la confianza y la aceptación por parte de los profesionales de la salud. La implementación de técnicas de interpretación de modelos podría ser un diferenciador clave para la adopción clínica de su sistema.

Para validar clínicamente los modelos, sería ideal realizar pruebas en colaboración con especialistas médicos. Esto no solo proporcionaría una evaluación rigurosa del trabajo, sino que también podría ayudar a refinar el modelo con base en la retroalimentación clínica.

Otra vía de investigación es el uso de técnicas de transferencia de aprendizaje, que pueden aprovechar datos de dominios no médicos para mejorar el aprendizaje de características relevantes en imágenes médicas. Esto podría ser particularmente útil para superar las limitaciones de los conjuntos de datos pequeños o desequilibrados.

La robustez del modelo frente a variaciones en la calidad de las imágenes es otro factor crítico. Realizar pruebas que simulan condiciones imperfectas ayudaría a garantizar que el modelo es fiable en una amplia gama de situaciones clínicas. La interfaz de usuario es otro componente fundamental. Desarrollar una plataforma que sea intuitiva y accesible para los médicos facilitará la adopción de esta tecnología y mejorará la experiencia del usuario final, lo cual es crucial para el éxito del proyecto.

Finalmente, es imperativo que el sistema cumpla con todas las normativas de privacidad y seguridad de datos relevantes. La protección y el manejo adecuado de los datos de los pacientes no solo es una obligación legal, sino también un imperativo ético que sustenta la integridad de su trabajo de investigación.

REFERENCIAS BIBLIOGRÁFICAS

- Adedokun, O. (2019). Deep learning – An overview. *Electrical & Electronic Systems*, 8(1). <https://doi.org/10.4172/2332-0796.1000302>
- Aggarwal, R., Sounderajah, V., Martin, G., Ting, D. S., Karthikesalingam, A., King, D., . . . Darzi, A. (2021). Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digital Medicine*, 4(1), 65. <https://doi.org/10.1038/s41746-021-00438-z>
- Agostiniani, R., Atti, G., Bonforte, S., Casini, C., Cirillo, M., De Pellegrin, M., . . . Vezzali, N. (2020). Recommendations for early diagnosis of Developmental Dysplasia of the Hip (DDH): working group intersociety consensus document. *Italian Journal of Pediatrics*, 46(150). <https://doi.org/10.1186/s13052-020-00908-2>
- Al-Bashir, A. K., Al-Abed, M., Abu Sharkh, F. M., Kordeya, M. N., & Rousan, F. M. (2015). Algorithm for automatic angles measurement and screening for Developmental Dysplasia of the Hip (DDH). *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, 6386–6389. <https://doi.org/10.1109/EMBC.2015.7319854>
- Aloysius, N., & Geetha, M. (2017). A review on deep convolutional neural networks. *International Conference on Communication and Signal Processing (ICCSP)*. <https://doi.org/10.1109/ICCSP.2017.8286426>
- Atalar, H., Gunay, C., Turanli, S., & Koktener, A. (2020). Discrepancy between ultrasonographic and radiographic findings in patients treated for developmental dysplasia of the hip; hip maturation on ultrasonography may not be consistent with radiographic hip maturation. *Journal of pediatric orthopedics. Part B*, 29(3), 228–234. <https://doi.org/10.1097/BPB.0000000000000654>
- Atalar, H., Üreten, K., Tokdemir, G., Tolunay, T., Çiçeklidağ, M., & Atik, O. Ş. (2023). The Diagnosis of Developmental Dysplasia of the Hip From Hip Ultrasonography Images With Deep Learning Methods. *Journal of*

Pediatric Orthopaedics, 43(3), e132-e137.
<https://doi.org/10.1097/BPO.0000000000002294>

- Babaei, S., Sorayaie, A., Naemi, A., Bagherzadeh, J., Pirnejad, H., & Kock, U. (2024). Survival prediction of glioblastoma patients using modern deep learning and machine learning techniques. *Scientific Reports*, 14(2371).
<https://doi.org/10.1038/s41598-024-53006-2>
- Barlow, T. (1962). Early diagnosis and treatment of congenital dislocation of the hip. *The Journal of Bone and Joint Surgery*, 44-B(2).
<https://doi.org/10.1302/0301-620X.44B2.292>
- Barrera, C., Cohen, S., Sankar, W., Ho-Fung, V., Sze, R., & Nguyen, J. (2019). Imaging of developmental dysplasia of the hip: ultrasound, radiography and magnetic resonance imaging. *Pediatr Radiol*, 49, 1652–1668.
<https://doi.org/10.1007/s00247-019-04504-3>
- Baudet, M., Revenga, C., & Bulo, M. (2001). Dysplasia in the development of the hip. *Revista de la Sociedad Andaluza de Traumatología y Ortopedia*, 21(2), 195-206. <https://www.elsevier.es/es-revista-revista-sociedad-andaluza-traumatologia-ortopedia-130-pdf-13025097>
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145-1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
- Brenes, M., Flores, A., & Meza, A. (2020). Actualización en displasia del desarrollo de la cadera. *Revista Médica Sinergia*, 5(9).
<https://doi.org/10.31434/rms.v5i9.574>
- Celik, M., & Inik, O. (2024). Development of hybrid models based on deep learning and optimized machine learning algorithms for brain tumor Multi-Classification. *Expert Systems with Applications*, 238(122159).
<https://doi.org/10.1016/j.eswa.2023.122159>
- Chen, T., Zhang, Y., Wang, B., Wang, J., Cui, L., He, J., & Cong, L. (2022). Development of a Fully Automated Graf Standard Plane and. *Diagnostics*, 12(6). <https://doi.org/10.3390/diagnostics12061423>
- Cheng, C.-T., Ho, T.-Y., Lee, T.-Y., Chang, C.-C., Chou, C.-C., Chen, C.-C., . . . Liao, C.-H. (2019). Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs.

- European Radiology*, 29(10), 5469–5477.
<https://doi.org/10.1007/s00330-019-06167-y>
- Cheng, C.-T., Wang, Y., Chen, H.-W., Hsiao, P.-M., Yeh, C.-N., Hsieh, C.-H., . . . Lu, L. (2021). A scalable physician-level deep learning algorithm detects universal trauma on pelvic radiographs. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-021-21311-3>
- Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., & Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5), 545-563. <https://doi.org/10.1111/1754-9485.13261>
- Colegio Médico del Perú. (2022). *MÉDICOS ESPECIALISTAS DEL CMP*. Retrieved July 16, 2022, from Colegio Médico del Perú: <https://www.cmp.org.pe/medicos-especialistas-del-cmp/>
- Da’u, A., & Salim, N. (2020). Recommendation system based on deep learning methods: a systematic review and new directions. *Artificial Intelligence Review*, 53, 2709–2748. <https://doi.org/10.1007/s10462-019-09744-1>
- Den, H., Ito, J., & Kokaze, A. (2023). Diagnostic accuracy of a deep learning model using YOLOv5 for detecting developmental dysplasia of the hip on radiography images. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-33860-2>
- Deng, L., & Yu, D. (2014). *Deep Learning: Methods and Applications*. Foundations and Trends in Signal Processing. <https://doi.org/10.1561/20000000039>
- Dhillon, A., & Verma, G. K. (2020). Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*, 9, 85-112. <https://doi.org/10.1007/s13748-019-00203-0>
- Fisher, T. B., Saini, G., Rekha, T. S., Krishnamurthy, J., Bhattarai, S., Callagy, G., . . . Aneja, R. (2024). Digital image analysis and machine learning-assisted prediction of neoadjuvant chemotherapy response in triple-negative breast cancer. *Breast cancer research : BCR*, 26(1), 12. <https://doi.org/10.1186/s13058-023-01752-y>

- Fludder, C. J., Keil, B. G., & Neave, M. J. (2023). Case report: Morphological changes evident after manual therapy in two cases of late-diagnosed developmental dysplasia of the hip. *Frontiers in pediatrics*, 10(1045812). <https://doi.org/10.3389/fped.2022.1045812>
- Fraiwani, M., Al-Kofahi, N., Ibranian, A., & Hanatleh, O. (2022). Detection of developmental dysplasia of the hip in X-ray images using deep transfer learning. *BMC Medical Informatics and Decision Making*, 22(216), 11. <https://doi.org/10.1186/s12911-022-01957-9>
- Frosst, N., & Hinton, G. (2017). Distilling a Neural Network Into a Soft Decision Tree. *arXiv preprint arXiv:1711.09784*. <https://doi.org/10.48550/arXiv.1711.09784>
- Gelfer, P., & Kennedy, K. (2008). Developmental dysplasia of the hip. *Journal of pediatric health care : official publication of National Association of Pediatric Nurse Associates & Practitioners*, 22(5), 318–322. <https://doi.org/10.1016/j.pedhc.2008.05.005>
- Gholamalinezhad, H., & Khosravi, H. (2020). Pooling Methods in Deep Neural Networks, a Review. *arXiv preprint arXiv:2009.07485*. <https://doi.org/10.48550/arXiv.2009.07485>
- Gomez-Chova, L., Camps-Valls, G., Munoz-Mari, J., & Calpe, J. (2008). Semisupervised Image Classification With Laplacian Support Vector Machines. *IEEE Geoscience and Remote Sensing Letters*, 5(3), 336–340. <https://doi.org/10.1109/LGRS.2008.916070>
- Goodfellow, I., Pouget-Abadie, J., Mehdi, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139–144. <https://doi.org/10.1145/3422622>
- Harsanyi, S., Zamborsky, R., Krajciova, L., Kokavec, M., & Danisovic, L. (2020). Developmental Dysplasia of the Hip: A Review of Etiopathogenesis, Risk Factors, and Genetic Aspects. *Medicina (Kaunas, Lithuania)*, 56(4), 153. <https://doi.org/10.3390/medicina56040153>
- Hashemzahi, R., Mahdavi, S. J., Kheirabadi, M., & Kamel, S. R. (2020). Detection of brain tumors from MRI images base on deep learning using

- hybrid model CNN and NADE. 1225-1232, 40(3), 1225-1232.
<https://doi.org/10.1016/j.bbe.2020.06.001>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- Hernández, R., Fernández, C., & Baptista, M. (2014). *Metodología de la investigación - 6ta edición*. McGraw-Hill.
<https://www.uncuyo.edu.ar/ices/upload/metodologia-de-la-investigacion.pdf>
- Husum, H.-C., Bjarne, M.-M., Thomsen, J., Damkjær, R., & Rahbek, O. (2021). Screening of dysplasia of the hip in Denmark. *Ugeskrift for læger*, 183(12), V05200371. <https://pubmed.ncbi.nlm.nih.gov/33829997/>
- Khosravi, B., Rouzrokh, P., Mickley, J. P., Faghani, S., Larson, A. N., Garner, H. W., . . . Wyles, C. C. (2023). Creating High Fidelity Synthetic Pelvis Radiographs Using Generative Adversarial Networks: Unlocking the Potential of Deep Learning Models Without Patient Privacy Concerns. *Journal of Arthroplasty*, 38(10), 2037-2043.
<https://doi.org/10.1016/j.arth.2022.12.013>
- Kitay, A., Widmann, R., Doyle, S., Do, H., & Green, D. (2019). Ultrasound is an Alternative to X-ray for Diagnosing Developmental Dysplasia of the Hips in 6-Month-Old Children. *HSS Journal*, 15(2), 153-158.
<https://doi.org/10.1007/s11420-018-09657-9>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*(521), 436–444. <https://doi.org/10.1038/nature14539>
- Lee, J.-G., Jun, S., Cho, Y.-W., Lee, H., Kim, G. B., Seo, J. B., & Kim, N. (2017). Deep Learning in Medical Imaging: General Overview. *Korean Journal of Radiology*, 18(4), 570-584.
<https://doi.org/10.3348/kjr.2017.18.4.570>
- Li, C., Yan, Y., Xu, H., Cao, H., Zhang, J., Sha, J., . . . Huang, L. (2022). Comparison of Transfer Learning Models in Pelvic Tilt and Rotation Measurement in Pediatric Anteroposterior Pelvic Radiographs. *Journal of Digital Imaging*, 35(6), 1506-1513. <https://doi.org/10.1007/s10278-022-00672-1>

- Li, Q., Yang, W., Xu, M., An, N., Wang, D., Wang, X., . . . Wang, J. (2021). Model construction and application for automated measurement of CE angle on pelvis orthograph based on MASK-R-CNN algorithm. *Biomedical Physics and Engineering Express*, 7(3). <https://doi.org/10.1088/2057-1976/abf483>
- Lin, Y., Yang, G., Jiang, Y., & Chang, X. (2022). Measuring lateral center edge angle and Sharp angle of hip joint on X-ray films based on deep learning for diagnosing hip dysplasia. *Chinese Journal of Medical Imaging Technology*, 38(11), 1710-1714. <https://doi.org/10.13929/j.issn.1003-3289.2022.11.026>
- Litjens, G., Kooi, T., Bejnordi, B. E., Adiyoso, A. A., Ciompi, F., Ghafoorian, M., . . . Sánchez, C. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- Liu, C., Xie, H., Zhang, S., Mao, Z., Sun, J., & Zhang, Y. (2020). Misshapen Pelvis Landmark Detection With Local-Global Feature Learning for Diagnosing Developmental Dysplasia of the Hip. *IEEE Transactions on Medical Imaging*, 39(12), 3944-3954. <https://doi.org/10.1109/TMI.2020.3008382>
- Longo, U. G., Papalia, R., De Salvatore, S., Ruzzini, L., Piergentili, I., Oggiano, L., . . . Denaro, V. (2021). Developmental Hip Dysplasia: An Epidemiological Nationwide Study in Italy from 2001 to 2016. *International Journal of Environmental Research and Public Health*, 18(12), 6589. <https://doi.org/10.3390/ijerph18126589>
- Mabee, M., Hareendranathan, A., Thompson, R., Dulai, S., & Jaremko, J. (2016). An index for diagnosing infant hip dysplasia using 3-D ultrasound: the acetabular contact angle. *Pediatric radiology*, 46(7), 1023–1031. <https://doi.org/10.1007/s00247-016-3552-8>
- Mazzi, E. (2011). Displasia del desarrollo de la cadera. *Revista de la Sociedad Boliviana de Pediatría*, 50(1), 57-64.
- Mousavi, S. S., Schukat, M., & Howley, E. (2018). Deep Reinforcement Learning: An Overview. *Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016. IntelliSys 2016. Lecture Notes in*

- Networks and Systems*, 16. https://doi.org/10.1007/978-3-319-56991-8_32
- Mulrain, J., Hennebry, J., Dicker, P., Condren, J., O'Driscoll, D., & O'Beirne, J. (2021). A normal screening ultrasound does not provide complete reassurance in infants at risk of hip dysplasia; further follow-up is required. *Irish journal of medical science*, 190(1), 233–238. <https://doi.org/10.1007/s11845-020-02296-2>
- Ogunfunmi, T., Ramachandran, R. P., Togneri, R., Zhao, Y., & Xia, X. (2019). A Primer on Deep Learning Architectures and Applications in Speech Processing. *Circuits Syst Signal Process*(3406–3432), 38. <https://doi.org/10.1007/s00034-019-01157-3>
- Ortolani, M. (1976). Congenital hip dysplasia in the light of early and very early diagnosis. *Clinical Orthopaedics and Related Research*(119), 6-10. <https://pubmed.ncbi.nlm.nih.gov/954324/>
- Palastanga, N., Field, D., & Soames, R. (2007). *Anatomía y movimiento humano : estructura y funcionamiento* (Primera ed.). Paidotribo.
- Park, H. S., Jeon, K., Cho, Y. J., Kim, S. W., Lee, S. B., Choi, G., . . . Hwang, J.-Y. (2021). Diagnostic Performance of a New Convolutional Neural Network Algorithm for Detecting Developmental Dysplasia of the Hip on Anteroposterior Radiographs. *Korean Journal of Radiology*, 22(4), 612-623. <https://doi.org/https://doi.org/10.3348/kjr.2020.0051>
- Pei, Y., Mu, L., Xu, C., Li, Q., Sen, G., Sun, B., . . . Li, X. (2023). Learning-based landmark detection in pelvis x-rays with attention mechanism: data from the osteoarthritis initiative. *Biomedical Physics and Engineering Express*, 9(2). <https://doi.org/10.1088/2057-1976/ac8ffa>
- Pham, T.-T., Le, L. H., Andersen, J., & Lou, E. (2021). Assessment of hip displacement in children with cerebral palsy using machine learning approach. *Medical & Biological Engineering & Computing*, 59, 1877–1887. <https://doi.org/10.1007/s11517-021-02416-9>
- Philip, H., Brijil, J., Nicholas, C., Herrera-Soto, J., Wudbhav, S., Schaeffer, E., . . . Aarvold, A. (2020). Even Experts Can Be Fooled: Reliability of Clinical Examination for Diagnosing Hip Dislocations in Newborns. *Journal of pediatric orthopedics*, 40(8), 408–412. <https://doi.org/10.1097/BPO.0000000000001602>

- Quispe, R. (2019, 05 25). Displasia de caderas afecta a cinco de cada cien bebés en Arequipa. *La República*. <https://larepublica.pe/archivo/869628-displasia-de-caderas-afecta-a-cinco-de-cada-cien-bebes-en-arequipa/>
- Raimann, R., & Aguirre, D. (2021). Developmental dysplasia of the hip: screening and management in the infant. *Revista Médica Clínica Las Condes*, 32(3), 263-270. <https://doi.org/10.1016/j.rmcl.2021.04.003>
- Raina, R., Madhavan, A., & Ng, A. (2009). Large-scale Deep Unsupervised Learning using Graphics Processors. *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*(873–880). <https://doi.org/10.1145/1553374.1553486>
- Rhee, P., Woodcock, J., Clohisy, J., Millis, M., Sucato, D., Beaulé, P., . . . Sierra, R. (2011). The Shenton line in the diagnosis of acetabular dysplasia in the skeletally mature patient. *The Journal of bone and joint surgery. American volume*, 93(2), 35–39. <https://doi.org/10.2106/JBJS.J.01717>
- Rochmawati, N., Hidayati, H. B., Yamasari, Y., Yustanti, W., Rakhmawati, L., Tjahyaningtjas, H. P., & Anistyasari, Y. (2020). Covid Symptom Severity Using Decision Tree. *2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE)*, 1-5. <https://doi.org/10.1109/ICVEE50212.2020.9243246>
- Sabharwal, R., & Miah, S. J. (2024). Evaluating teachers' effectiveness in classrooms: an ML-based assessment portfolio. *Social Network Analysis and Mining* , 14(1), 28 . <https://doi.org/10.1007/s13278-023-01195-5>
- Saeed, A.-N., & Altamimi, A. (2021). DDH Diagnosis: What do we Know so Far? *Bentham Science Publishers*, 14(1), 469 - 477. <https://doi.org/10.2174/1874944502114010469>
- Salur, M. U., & Aydin, I. (2020). A Novel Hybrid Deep Learning Model for Sentiment Classification. *IEEE Access*, 8, 58080-58093. <https://doi.org/10.1109/ACCESS.2020.2982538>
- Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolutional neural networks in medical image understanding: a survey. *Evolutionary Intelligence*, 15(1), 1-22. <https://doi.org/10.1007/s12065-020-00540-3>

- Sen, P. C., Hajra, M., & Ghosh, M. (2020). Supervised classification algorithms in machine learning: A survey and review. In *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph. Springer Singapore*, 937, 99-111. https://doi.org/10.1007/978-981-13-7403-6_11
- Shaw, B., & Segal, L. (2016). Evaluation and Referral for Developmental Dysplasia of the Hip in Infants. *Pediatrics*, 138(6). <https://doi.org/10.1542/peds.2016-3107>
- Shipman, S., Helfand, M., Moyer, V., & Yawn, B. (2006). Screening for developmental dysplasia of the hip: a systematic literature review for the US Preventive Services Task Force. *Pediatrics*, 117(3), e557–e576. <https://doi.org/10.1542/peds.2005-1597>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(60), 2196-1115. <https://doi.org/10.1186/s40537-019-0197-0>
- Singh, N. K., & Raza, K. (2021). Medical Image Generation Using Generative Adversarial Networks: A Review. In R. Patgiri, A. Biswas, & P. Roy, *Health Informatics: A Computational Perspective in Healthcare* (Vol. 932, pp. 77-96). Springer Singapore. https://doi.org/https://doi.org/10.1007/978-981-15-9735-0_5
- Souza, B. G., Vasconcelos, B., Pujoni, H., Nogueira, M., Oliveira, V., & Chaoubah, A. (2021). Epidemiology and costs of surgical treatment of developmental dysplasia of hip in the Brazilian Public Health System in a decade. *Einstein (São Paulo)*, 19(eGS5625). https://doi.org/10.31744/einstein_journal/2021GS5625
- Swarup, I., Penny, C., & Dodwell, E. (2018). Developmental dysplasia of the hip: An update on diagnosis and management from birth to 6 months. *Lippincott Williams and Wilkins*, 30(1), 84-92. <https://doi.org/10.1097/mop.0000000000000574>
- Taresh, M. M., Zhu, N., Ali, T. A., Alghaili, M., Hameed, A. S., & Mutar, M. L. (2021). KL-MOB: automated COVID-19 recognition using a novel approach based on image enhancement and a modified MobileNet CNN. *PeerJ Computer Science*, 7, 1-23. <https://doi.org/10.7717/PEERJ-CS.694>

- Ting, K. M. (2011). Confusion Matrix. En C. Sammut, *Encyclopedia of Machine Learning* (pág. 209). Springer.
https://doi.org/https://doi.org/10.1007/978-0-387-30164-8_157
- Vara, A. A. (2012). *Desde la Idea hasta la sustentación: siete pasos para una tesis exitosa. Un método efectivo para las ciencias empresariales*. Universidad de San Martín de Porres.
<https://doi.org/10.13140/RG.2.1.2238.4080>
- Wang, P., Fan, E., & Wang, P. (2021). Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognition Letters*, 141, 61-67.
<https://doi.org/10.1016/j.patrec.2020.07.042>
- Wang, Y. (2020). A Mathematical Introduction to Generative Adversarial Nets (GAN). *arXiv preprint arXiv:2009.00169*.
<https://doi.org/10.48550/arXiv.2009.00169>
- Xu, J., Xie, H., Tan, Q., Wu, H., Liu, C., Zhang, S., . . . Zhang, Y. (2022). Multi-task hourglass network for online automatic diagnosis of developmental dysplasia of the hip. *World Wide Web*. <https://doi.org/10.1007/s11280-022-01051-0>
- Xu, W., Shu, L., Gong, P., Huang, C., Xu, J., Zhao, J., . . . Yu, G. (2022). A Deep-Learning Aided Diagnostic System in Assessing Developmental Dysplasia of the Hip on Pediatric Pelvic Radiographs. *Frontiers in Pediatrics*, 9. <https://doi.org/10.3389/fped.2021.785480>
- Yousef, R., Gupta, G., Yousef, N., & Khari, M. (2022). A holistic overview of deep learning approach in medical imaging. *Multimedia Systems*(28), 881–914. <https://doi.org/10.1007/s00530-021-00884-5>
- Yu, J., Yu, S., Erdal, B., Demirer, M., Gupta, V., Bigelow, M., . . . White, R. (2020). Detection and localisation of hip fractures on anteroposterior radiographs with artificial intelligence: proof of concept. *Clinical Radiology*, 75(3), 237.e1-237.e9.
<https://doi.org/10.1016/j.crad.2019.10.022>
- Zhang, A., Lipton, Z., Li, M., & Smola, A. (2021). *Dive into Deep Learning*. *arXiv preprint arXiv:2106.11342*.
<https://doi.org/10.48550/arXiv.2106.11342>

- Zhang, H. W., Huang, D. L., Wang, Y. R., Zhong, H. S., & Pang, H. W. (2024). CT radiomics based on different machine learning models for classifying gross tumor volume and normal liver tissue in hepatocellular carcinoma. *Cancer imaging : the official publication of the International Cancer Imaging Society*, 24(1), 20. <https://doi.org/10.1186/s40644-024-00652-4>
- Zhang, S., Sun, J., Liu, C.-B., Fang, J.-H., Xie, H.-T., & Ning, B. (2020). Clinical application of artificial intelligence-assisted diagnosis using anteroposterior pelvic radiographs in children with developmental dysplasia of the hip. *Bone and Joint Journal*, 102(11), 1574-1581. <https://doi.org/10.1302/0301-620X.102B11.BJJ-2020-0712.R2>
- Zhou, J., Gandomi, A., Chen, F., & Holzinger, A. (2021). Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics. *Electronics*, 10(5), 593. <https://doi.org/10.3390/electronics10050593>
- Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Van Ginneken, B., Madabhushi, A., . . . Summers, R. M. (2021). A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises. *Proceedings of the IEEE*, 109(5), 820-838. <https://doi.org/10.1109/JPROC.2021.3054390>

ANEXOS

Anexo 1. Operacionalización de variables

Variables	Tipo de variable	Definición conceptual	Dimensión	Indicador	Definición de indicador
Modelo de detección usando aprendizaje profundo	Independiente	Es un subconjunto del aprendizaje automático, perteneciente a la rama de la inteligencia artificial, que busca modelar a un alto nivel de abstracción en los datos (Adedokun, 2019).	Precisión	Casos clasificados correctamente por el modelo.	Porcentaje de precisión indica la efectividad del modelo en su tarea.
			Sensibilidad	Casos con DDC detectados correctamente por el modelo.	Porcentaje de sensibilidad indica que el modelo detecta la mayoría de los casos con DDC.
			Especificidad	Casos Normales detectados correctamente por el modelo.	Porcentaje de especificidad indica que el modelo evita falsos positivos.
			Rendimiento	Tiempo en dar resultado	Cantidad de tiempo que tarda el modelo en analizar una radiografía y dar un resultado, cuanto menor sea el tiempo, más eficiente es el modelo en términos de rapidez.
Detección de DDC por radiográfica pélvica en infantes de 03 a 06 meses de edad	Dependiente	La displasia del desarrollo de la cadera (DDC), es una deformación en la articulación coxofemoral, en donde presenta desviaciones parcial o total de la cabeza femoral con el acetábulo (Shipman et al., 2006).	-	-	-

Anexo 2. Matriz de consistencia

PROBLEMA	OBJETIVOS	HIPÓTESIS
Problema general	Objetivo general	Hipótesis general
¿Es posible que un modelo basado en aprendizaje profundo contribuya a la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad?	Desarrollar un modelo basado en aprendizaje profundo que contribuya a la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.	Un modelo basado en aprendizaje profundo contribuye significativamente en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.
Problemas específicos	Objetivos específicos	Hipótesis específicas
¿Es posible que un modelo basado en aprendizaje profundo contribuya con aceptable grado de Precisión en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad?	Desarrollar un modelo basado en aprendizaje profundo que contribuya con aceptable grado de Precisión en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.	Un modelo basado en aprendizaje profundo contribuye significativamente con aceptable grado de Precisión en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.
¿Es posible que un modelo basado en aprendizaje profundo contribuya con aceptable grado de Sensibilidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad?	Desarrollar un modelo basado en aprendizaje profundo que contribuya con aceptable grado de Sensibilidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.	Un modelo basado en aprendizaje profundo contribuye significativamente con aceptable grado de Sensibilidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.
¿Es posible que un modelo basado en aprendizaje profundo contribuya con aceptable grado de Especificidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad?	Desarrollar un modelo basado en aprendizaje profundo que contribuya con aceptable grado de Especificidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.	Un modelo basado en aprendizaje profundo contribuye significativamente con aceptable grado de Especificidad en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.
¿Es posible que un modelo basado en aprendizaje profundo contribuya con rendimiento óptimo en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad?	Desarrollar un modelo basado en aprendizaje profundo que contribuya con rendimiento óptimo en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.	Un modelo basado en aprendizaje profundo contribuye significativamente con rendimiento óptimo en la detección de displasia de desarrollo de cadera por radiografía pélvica en infantes de 03 a 06 meses de edad.