

Utilización de técnicas de aprendizaje profundo para la detección temprana de displasia de cadera en infantes mediante radiografías pélvicas

Use of deep learning techniques for early detection of hip dysplasia in infants through pelvic radiographs

Ricardo Junior Jara Gaspar

<https://orcid.org/0000-0001-8914-4833>

ricardo.jara2@unmsm.edu.pe

Universidad Nacional Mayor de San Marcos,
Lima, Perú

RECIBIDO: 04/11/2023 - ACEPTADO: 25/11/2023 - PUBLICADO: 30/12/2023

RESUMEN

La clasificación exacta y temprana de la Displasia del Desarrollo de la Cadera (DDC) en pacientes pediátricos es imperativa para un tratamiento eficaz, reduciendo así las complicaciones a largo plazo. Las metodologías convencionales de diagnóstico manual muestran limitaciones, principalmente por la variabilidad interobservador y la necesidad de pericia especializada, llevando a diagnósticos inconsistentes o retrasados. Este artículo aborda esta problemática mediante la implementación de técnicas avanzadas de aprendizaje profundo para analizar radiografías pélvicas de infantes de 3 a 6 meses, utilizando ResNet50 para superar los desafíos del desvanecimiento del gradiente y proporcionar evaluaciones precisas. Adicionalmente, enfrentamos las cuestiones de privacidad, escasez y desequilibrio de los datos médicos aplicando Redes Generativas Antagónicas (GANs) para generar datos sintéticos de alta calidad, enriqueciendo así nuestro conjunto de entrenamiento y fortaleciendo la robustez de nuestro modelo. La integración del sistema propuesto con ResNet50 y técnicas de GANs fue clave para alcanzar una precisión del 97.3% en la clasificación de radiografías pélvicas con presencia o ausencia de DDC.

Palabras clave: Displasia del desarrollo de cadera, aprendizaje profundo, imagen médica para diagnóstico, redes generativas antagónicas, clasificación de imágenes.

ABSTRACT

Accurate and early classification of Developmental Hip Dysplasia (DHD) in pediatric patients is imperative for effective treatment, thereby reducing long-term complications. Conventional manual diagnosis methodologies show limitations, mainly due to interobserver variability and the need for specialized expertise, leading to inconsistent or delayed diagnoses. This article addresses this issue by implementing advanced deep learning techniques to analyze pelvic radiographs of 3- to 6-month-old infants, using ResNet50 to overcome the challenges of gradient fading and provide accurate assessments. Additionally, we address issues of privacy, sparsity, and imbalance of medical data by applying Generative Adversarial Networks (GANs) to generate high-quality synthetic data, thus enriching our training set and strengthening the robustness of our model. The integration of the proposed system with ResNet50 and GANs techniques was key to achieving an accuracy of 97.3% for the classification of pelvic radiographs with DHD.

Keywords: Hip Developmental Dysplasia, Deep Learning, Diagnostic Medical Imaging, Generative Adversarial Networks, Image Classification.

I. INTRODUCCIÓN

La Displasia del Desarrollo de la Cadera (DDC) representa una de las anomalías musculoesqueléticas más prevalentes en la población pediátrica, la cual puede conducir a problemas graves de movilidad y calidad de vida si no se diagnostica y trata de manera oportuna. Esta alteración se manifiesta principalmente en variaciones patológicas del receptáculo acetabular y de la cabeza femoral, además de anomalías en ligamentos y la cápsula articular (Longo et al., 2021). La habilidad para detectar con precisión esta condición en sus etapas iniciales es fundamental para un tratamiento exitoso, y para prevenir secuelas a largo plazo que pueden variar desde el dolor hasta la artritis y la necesidad de cirugía reconstructiva en la edad adulta.

Tradicionalmente, la detección de la DDC ha dependido en gran medida de la pericia de profesionales de la salud para interpretar las señales sutiles en pruebas clínicas y radiográficas (Al-Naser & Altamimi, 2021). El diagnóstico clínico, realizado en las primeras semanas de vida, incluye maniobras como la de Ortolani y Barlow, que buscan inestabilidad de la articulación, y observaciones como la limitación de la abducción de las caderas y el signo de Galeazzi (Baudet et al., 2001). Aunque estas técnicas son informativas, la habilidad para reconocer señales de alerta de DDC es altamente subjetiva y puede variar entre los clínicos. Por ejemplo, la interpretación de una abducción de cadera menor de 60° puede ser inconsistente, lo que eleva el riesgo de diagnósticos desiguales y, a menudo, retrasados (Raimann & Aguirre, 2021). Para complementar las pruebas clínicas, los profesionales de la salud recurren a la imagenología como las radiografías pélvicas.

Aunque las radiografías son un medio de diagnóstico valioso a partir de los 3 meses de edad, su eficacia está condicionada por la calidad de la imagen y la técnica de captura. Debe obtenerse con el infante en decúbito supino y las piernas paralelas para visualizar adecuadamente las estructuras de la cadera. Sin embargo, la interpretación de estas imágenes no está exenta de variabilidad interobservador. La evaluación de la simetría de los agujeros obturadores y la posición de la metafisis proximal de los fémures pueden ser interpretadas de manera diferente entre radiólogos, lo que conlleva una posible falta de consenso en la evaluación de la presencia y severidad de la DDC (Shaw et al., 2016). Esta subjetividad y variabilidad subrayan la necesidad de métodos diagnósticos más objetivos y reproducibles, que puedan ofrecer resultados

consistentes independientemente del observador. La aplicación de tecnologías avanzadas como el aprendizaje profundo promete una estandarización en la interpretación de estas pruebas, lo que podría representar una mejora en la precisión del diagnóstico de la DDC.

Las tecnologías emergentes, como el aprendizaje profundo, están allanando el camino hacia una mayor estandarización en la interpretación de las pruebas radiográficas, potencialmente transformando el paradigma actual en el diagnóstico de la DDC. La aplicación de algoritmos de aprendizaje profundo ya ha demostrado su eficacia en aumentar la precisión diagnóstica y en reducir la subjetividad que naturalmente acompaña a la interpretación humana de las imágenes médicas. En estudios específicos, como se reporta en la literatura (Atalar et al., 2023), se han comparado diversas arquitecturas de redes neuronales, incluyendo VGG-16, ResNet-101, MobileNetV2 y GoogLeNet, logrando también resultados prometedores con modelos preentrenados como VGG-19. Estas arquitecturas son ampliamente reconocidas en el campo de la visión por computadora y han mostrado su potencial en el análisis de imágenes médicas. De manera similar, los hallazgos de Park et al. (2021), subrayan una alta consistencia y efectividad en el diagnóstico de la DDC usando estos modelos, superando a los diagnósticos clínicos y simplificando el actual proceso de referencia y detección. No obstante, el principal reto de estos algoritmos radica en la necesidad de volúmenes de datos médicos etiquetados y balanceados, un recurso que es a menudo escaso.

Para superar la limitación de los datos, se ha recurrido a la estrategia de transferencia de aprendizaje, que permite capitalizar el conocimiento derivado de modelos previamente entrenados en otros dominios (McEvoy et al., 2021). Por ejemplo, en Fraiwan et al. (2022), se explora el uso de DarkNet53 aplicando transferencia de aprendizaje, donde se adapta un modelo preentrenado en un conjunto de datos amplio y general para que funcione con los específicos y más limitados datos de radiografías pélvicas. De manera similar, el sistema YOLOv5, fue adaptado para identificar la DDC utilizando radiografías pélvicas (Den et al., 2023), beneficiándose significativamente del aprendizaje por transferencia. La transferencia de conocimiento es una herramienta útil por su capacidad de iniciar el proceso de aprendizaje con una base de conocimientos preexistentes. Sin embargo, la transferencia de aprendizaje no está exenta de desafíos.

Uno de los problemas de la transferencia de conocimiento es la divergencia entre las características de

los datos en los que el modelo fue preentrenados y las características de los nuevos datos médicos que se desean analizar. Esta discrepancia puede conducir a lo que se conoce como transferencia negativa (Agarwal et al., 2021), donde el conocimiento previo interfiere y reduce el desempeño del modelo en la nueva tarea. Además, puede haber una sobre dependencia en los datos originales que puede no ser completamente relevante para las nuevas aplicaciones, especialmente en un campo tan especializado como el diagnóstico médico, donde las diferencias en modalidades de imagen, contrastes y patologías específicas son críticas. Por tanto, aunque el aprendizaje por transferencia facilita el proceso, sigue siendo fundamental contar con un ajuste fino y una validación exhaustiva para asegurar que el modelo se generalice bien al dominio específico de interés.

Para sortear los retos asociados con la insuficiencia de datos y como alternativa al uso de transferencia de conocimiento, nuestro método se centra en la generación de imágenes sintéticas mediante redes generativas antagónicas (GANs) (Goodfellow et al., 2020). Esta técnica nos permite enriquecer y expandir nuestro conjunto de datos para el entrenamiento, proporcionando una mayor variedad y volumen de ejemplos clínicos sin violar la confidencialidad ni la privacidad de los datos de los pacientes. Además, las GANs contribuyen a mejorar la robustez del modelo, ya que generan imágenes variadas que pueden ayudar al sistema a aprender a reconocer patrones y anomalías con mayor eficacia, incluso en escenarios donde los datos reales son escasos o incompletos.

Tener un conjunto de datos enriquecido y aumentado usando GANs facilita la integración de una arquitectura de red neuronal avanzada como ResNet (He et al., 2016). Esta arquitectura tiene beneficios notorios. Por ejemplo, ResNet se destaca por su capacidad de administrar efectivamente redes neuronales de gran profundidad, posibilita un entrenamiento eficiente y previene el problema de desvanecimiento de gradiente, un desafío común en modelos profundos. La combinación de ResNet y GANs en nuestra propuesta representa un paso significativo hacia un diagnóstico más preciso y accesible de la DDC, teniendo el potencial de mejorar las capacidades diagnósticas de los profesionales de la salud y contribuyendo a la calidad del cuidado pediátrico.

El resto del artículo se organiza de la siguiente manera. En la sección 2 se presenta los métodos de investigación para la detección de DDC. La sección 3 describe la metodología propuesta para desarrollar

el modelo de aprendizaje profundo que identifique la presencia o ausencia de la DDC. En la sección 4 se presentan los resultados obtenidos, seguidos de una discusión sobre las implicancias de estos resultados. Finalmente, la sección 5 concluye el artículo, resumiendo los hallazgos principales y proponiendo pasos futuros para la investigación en esta área.

II. MARCO TEÓRICO

El aprendizaje profundo, o Deep Learning (DL), se ha convertido en un pilar crucial en el campo de la inteligencia artificial, proporcionando un enfoque para construir modelos que simulan altos niveles de abstracción en la representación de datos (Mousavi et al., 2018). Estos modelos son especialmente valiosos por su habilidad para aprender de manera automática las características relevantes de grandes cantidades de datos, prescindiendo de la necesidad de intervención manual en la ingeniería de características, una ventaja significativa sobre métodos tradicionales de aprendizaje automático (Zhang et al., 2021). En la clasificación, un proceso esencial en este dominio, un conjunto de entrenamiento etiquetado enseña al modelo cómo categorizar datos basándose en sus características, mientras que los conjuntos de prueba validan la precisión y la habilidad del modelo para generalizar a nuevos datos (Sen et al., 2020).

La aplicación de DL en el ámbito médico ha mostrado resultados prometedores, con una amplia gama de aplicaciones que van desde el diagnóstico hasta la predicción de eventos clínicos, apalancándose en la capacidad del aprendizaje profundo para identificar patrones complejos en los datos visuales que pueden pasar desapercibidos para el ojo humano (Aggarwal et al., 2021). Entrenando modelos con miles de ejemplos de imágenes médicas, el DL puede revelar indicadores y características específicas de condiciones médicas, brindando herramientas valiosas para los especialistas en la toma de decisiones clínicas.

2.1. Redes Generativas Antagónicas

Las Redes GANs, son una arquitectura innovadora que incorpora dos redes neuronales, el generador y el discriminador, que se entrenan simultáneamente en un marco adversativo. Inspiradas por los fundamentos de la teoría de juegos, estas redes se enfrentan en un entorno de optimización diferencial, donde cada una tiene un objetivo claramente opuesto (Singh & Raza, 2021). El discriminador se entrena para mejorar su habilidad de distinguir entre datos reales y sintéticos, maximizando así la

precisión en la asignación de categorías. En contraparte, el generador se enfoca en la producción de datos que imiten de forma convincente la distribución auténtica de las muestras reales, con el fin de engañar al discriminador (Goodfellow et al., 2020).

2.2. ResNet50

La ResNet50, una variante específica de las Redes Residuales se ha establecido como un pilar dentro del ámbito de la visión computacional debido a su enfoque innovador en el manejo de redes neuronales profundas. Introducida por He et al. en 2016 (He et al., 2016), su diseño arquitectónico se orienta a resolver el problema del desvanecimiento y la explosión del gradiente impedimentos notorios que obstaculizan la convergencia y el rendimiento durante el entrenamiento de modelos profundos. La ResNet50, a través de sus distintivas conexiones residuales, permite que las señales de error se propaguen efectivamente a través de capas más profundas, posibilitando así el entrenamiento de modelos de alta profundidad sin comprometer la eficacia y la precisión en tareas de clasificación de imágenes.

III. METODOLOGIA

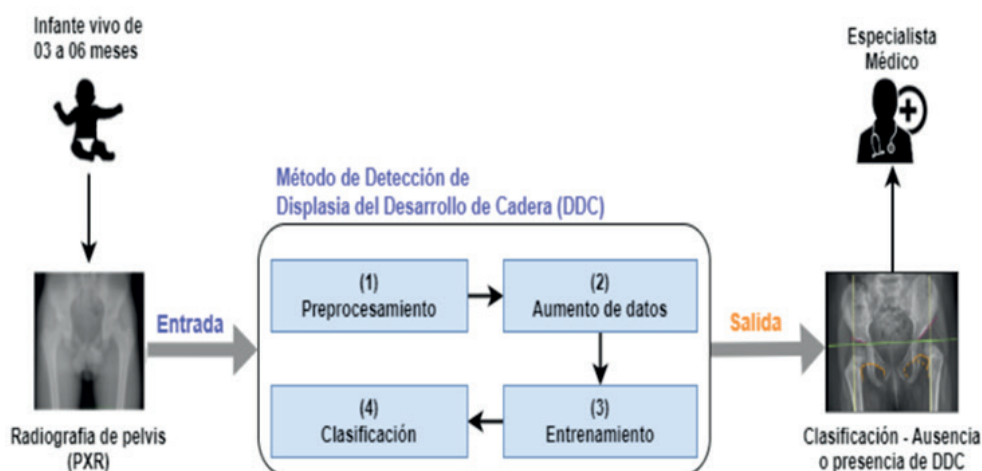
Para abordar el desafío de identificar la DDC en infantes de tres a seis meses mediante el análisis de imágenes radiográficas pélvicas, se ha concebido una metodología integral que incorpora técnicas de aprendizaje profundo. El proceso está diseñado

para mejorar la precisión y confiabilidad del diagnóstico, sirviendo de apoyo a la evaluación realizada por especialistas médicos. La Figura 1 ilustra el flujo de trabajo de este enfoque metodológico. La entrada del proceso se inicia con la adquisición de radiografías de pelvis de infantes en el rango de edad de interés, seguido por el preprocesamiento y el aumento de datos para mejorar la calidad y cantidad del conjunto de entrenamiento. Después del entrenamiento del modelo, la clasificación final determina la presencia o ausencia de DDC, con la verificación de un especialista médico como paso final.

3.1. Preprocesamiento

Este componente representa el primer paso en la metodología propuesta. En esta fase, se inicia con la recepción de una radiografía de la pelvis proveniente de un infante en vivo, cuya edad oscila entre los tres y seis meses. El conjunto de datos utilizado consiste en imágenes de radiografías de cadera tomadas en vista anteroposterior. Estas imágenes fueron recopiladas localmente en dos hospitales: el Hospital Universitario Rey Abdullah en la Universidad de Ciencia y Tecnología de Jordania, en Irbid, Jordania; y el Hospital Especializado Alsafa en Jarash, Jordania. El conjunto de datos empleado en el artículo de investigación relevante incluyó un total de 354 sujetos, de los cuales 120 presentaban displasia del desarrollo de la cadera y 234 eran considerados normales. Las imágenes están disponibles en diferentes tamaños, adecuados para diversos

Figura 1
Metodología propuesta para la detección de la DDC.



Fuente: Elaboración propia.

modelos de aprendizaje profundo y fueron tomadas de Fraiwan et al. (2022).

Durante el preprocesamiento, se hizo uso de TensorFlow para optimizar y preparar los datos para el entrenamiento de modelos. Inicialmente, se ejecutó la extracción de datos de los directorios pertinentes. Para facilitar este proceso y mejorar la eficiencia, se optó por utilizar la librería ImageDatagenerator de TensorFlow. Este enfoque permitió instanciar la librería tanto para la etapa de entrenamiento como para la de validación, minimizando así los errores potenciales al cargar imágenes directamente en entornos como Keras o TensorFlow. Como resultado, se obtuvieron dos conjuntos distintos de datos: uno para entrenamiento y otro para validación, ambos configurados con imágenes de dimensiones normalizadas a 224x224.

Dado que las imágenes eran en blanco y negro y pertenecían a dos clases distintas, se configuró el modo de clasificación en binario y se especificó que las imágenes eran monocromáticas. Además, se normalizaron las intensidades de las imágenes para que los valores de píxeles oscilaran entre 0 y 1. Esta normalización es crucial para facilitar el aprendizaje del modelo al proporcionar una escala de intensidad consistente, además de promover la estabilidad numérica durante el entrenamiento de los modelos.

Por último, se aplicó un filtro Gaussiano de 3x3 a las imágenes. El filtrado Gaussiano ayuda a reducir el ruido y las irregularidades en las imágenes, lo que es esencial para mejorar la precisión del modelo al enfocarse en las características relevantes. Estas estrategias de preprocesamiento fueron esenciales para asegurar datos limpios y estructurados para las etapas de entrenamiento y validación.

3.2. Aumento de datos:

Para abordar el desafío de tener un conjunto de datos limitado y desbalanceado, que es común en aplicaciones médicas, se emplean GANs para ampliar el conjunto de datos existente. Esta técnica permite la generación de imágenes radiográficas adicionales, realistas pero sintéticas, que pueden mejorar la robustez del modelo al proporcionar una variación más amplia de casos de entrenamiento.

Una vez que el conjunto de datos fue preparado, se realizó un aumento de datos. El conjunto de datos utilizado en el estudio se compone de 354 sujetos, con una división que no refleja un balance perfecto: 120 sujetos con diagnóstico de DDC y 234 sujetos clasificados como normales. Esta distribución indica

una mayor proporción de casos normales en comparación con los casos patológicos, lo cual podría inclinar el aprendizaje del modelo hacia una mejor detección de sujetos normales mientras potencialmente reduce su sensibilidad a detectar DDC.

Para mejorar el equilibrio del conjunto de datos y, por consiguiente, la capacidad del modelo para generalizar y detectar DDC, fue beneficioso utilizar GANs para generar más imágenes sintéticas de casos con DDC. Idealmente, el número de datos generados debería apuntar a igualar la cantidad de sujetos normales, por lo que se deberían generar 116 casos sintéticos de sujetos sin patología y 230 casos sintéticos de DDC. Esto equilibra el conjunto de datos a 350 sujetos en cada categoría, permitiendo que el modelo aprenda con la misma representación de ambas clases y mejore su precisión diagnóstica en la detección de DDC. Para el aumento de datos, se propone el uso de una red GAN con una arquitectura como la mostrada en la Figura 2.

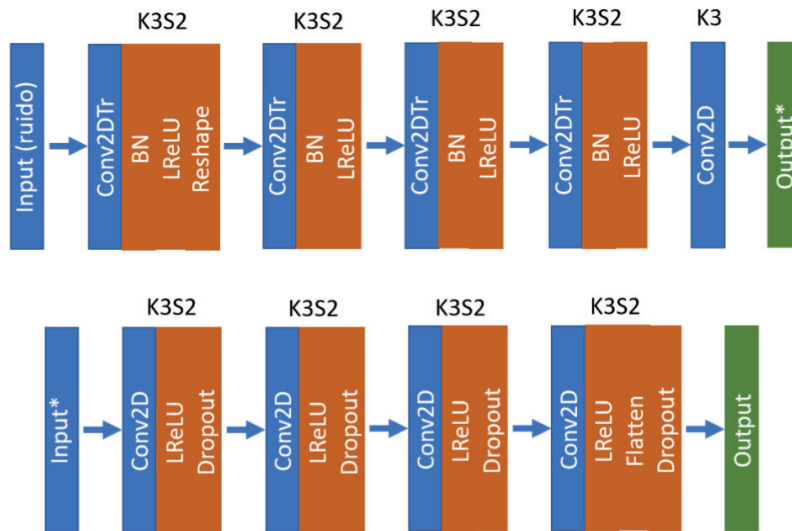
La Figura 2 muestra dos arquitecturas de red neuronal convolucional diferentes, las cuales se utilizan para procesar y transformar una entrada, en este caso, son imágenes, indicado por Input (ruido) en la parte superior e Input* en la parte inferior. Cada bloque representa una operación o una capa dentro de la red, y las siglas indican el tipo de operación que se lleva a cabo. Estos bloques son típicos en el diseño de redes neuronales para tareas de visión por computadora y procesamiento de imágenes. La parte superior representa la arquitectura del Generador, mientras que la parte inferior de la figura representa al Discriminador.

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

- **CONV2D:** Capa de convolución bidimensional que aplica un número de filtros al input para extraer características. El término "K3S2" indica que los filtros utilizados tienen un tamaño de kernel de 3x3 y un stride de 2 (el filtro se mueve 2 píxeles cada vez). "K3" indica un tamaño de kernel de 3 con el stride no especificado por defecto de 1.
- **BN:** Normalización por lotes (Batch Normalization). Esta técnica normaliza las activaciones de la capa anterior, lo que ayuda a acelerar el entrenamiento y tiene un efecto regularizador, lo que puede ayudar a combatir el sobreajuste.
- **ReLU:** Función de activación Rectified Linear Unit, que introduce no linealidad en el modelo,

Figura 2

Arquitectura propuesta de una GAN para el aumento de datos.



Fuente: Elaboración propia.

permitiendo a la red aprender representaciones más complejas.

- **Reshape:** Cambia la forma del tensor de salida de la capa anterior sin cambiar sus datos. Esto puede ser necesario para hacer que las dimensiones de los datos coincidan con las requeridas por las capas subsiguientes.
- **Dropout:** Técnica de regularización en la que se "apagan" aleatoriamente nodos de la red durante el entrenamiento para prevenir el sobreajuste.
- **Flatten:** Esta operación convierte el tensor multidimensional de salida de las capas convolucionales en un vector 1D. Es un paso común antes de las capas densas en una CNN.

Estas operaciones se utilizan en la generación de imágenes aumentadas de la siguiente manera. Las capas de CONV2D con diferentes tamaños de kernel y strides pueden capturar y enfatizar diferentes características de las imágenes, lo que puede resultar en variantes únicas en las características de las imágenes sintéticas generadas. BN y ReLU se utilizan consecutivamente después de las capas de convolución para estabilizar y mejorar el aprendizaje, asegurando que las características generadas no escalen demasiado y mantengan una distribución consistente.

La capa de Dropout añade más regularización al proceso, haciendo que la red sea menos sensible

a la exacta presencia y ausencia de características específicas en las imágenes de entrada, lo que puede resultar en imágenes aumentadas más variadas. Finalmente, la capa de Flatten se usa para preparar el tensor de salida para capas completamente conectadas, aunque en este contexto podría estar relacionada con la preparación de la salida para la evaluación o quizás para un proceso de aumento que implica transformaciones que son más adecuadas para vectores que para tensores 3D. Una vez que se tiene un conjunto de datos aumentado, se procede al entrenamiento de la ResNet50.

3.3. Entrenamiento:

En esta etapa, se llevó a cabo un proceso de entrenamiento sistemático destinado a desarrollar modelos de aprendizaje automático para la identificación de la DDC mediante el uso de imágenes. Se estableció un entorno controlado inicialmente, fijando la semilla de aleatoriedad para garantizar la reproducibilidad de los resultados, un paso esencial en la validación científica y la comparativa objetiva de diferentes modelos.

Se construyó un modelo de red neuronal convolucional como base de referencia, compuesto por capas que extraen características relevantes de las imágenes y reducen la dimensionalidad para optimizar el rendimiento y evitar el sobreajuste. Este modelo incluyó una capa de aplanamiento de datos

y una capa densa para la clasificación, utilizando una función de activación adecuada para problemas binarios, como discernir entre imágenes normales y defectuosas.

Para el manejo eficiente de los datos, se utilizó un generador de imágenes que facilita la carga y el ajuste de las imágenes en lotes adecuados, incluyendo la normalización de los valores de los píxeles. Posteriormente, se implementó un segundo modelo utilizando la arquitectura ResNet50 preentrenados, ajustando capas adicionales sin modificar los pesos preexistentes para aprovechar los patrones ya aprendidos en clasificaciones de imágenes generales.

Este segundo modelo, compilado y entrenado con el mismo conjunto de datos ampliados y durante el mismo número de épocas, permite un ajuste fino de los parámetros con base en la retroalimentación del entrenamiento. En resumen, el proceso descrito esboza una metodología rigurosa que aplica técnicas de procesamiento de imágenes y aprendizaje transferido, con el fin último de establecer un modelo confiable y eficiente en la clasificación precisa de imágenes.

3.4. Clasificación:

La clasificación se realizó con los modelos CNN y ResNet50 entrenados previamente. La salida del flujo de trabajo es una radiografía pélvica anotada, donde se ha identificado y clasificado la presencia o ausencia de DDC.

IV. RESULTADOS Y DISCUSIÓN

Para obtener los resultados, se entrenó una CNN como línea base en este estudio. La CNN tiene una estructura relativamente sencilla pero efectiva, diseñada para la clasificación binaria de imágenes. La arquitectura comienza con la inicialización del estado aleatorio para asegurar la reproducibilidad de los resultados. A continuación, se describe la estructura de la red:

- **Capa Convolutiva:** La primera capa es una capa convolutiva con 40 filtros y un kernel de tamaño 3x3. La función de activación utilizada es 'ReLU' (Rectified Linear Unit). Esta capa está diseñada para extraer características de la imagen de entrada, que tiene un tamaño esperado de 224 X 224 píxeles con 1 canal (grayscale).
- **Capa de Agrupación Máxima (MaxPooling):** Seguida por una capa de MaxPooling con

un tamaño de pool de 2x2. Esta capa reduce a la mitad el número de características, disminuyendo así la complejidad computacional y el sobreajuste.

- **Capas Convolucionales y de MaxPooling Adicionales:** Se repite este patrón de una capa convolutiva seguida de una capa de MaxPooling dos veces más. Cada capa convolutiva tiene también 40 filtros de tamaño 3x3 y utiliza la función de activación 'ReLU'.
- **Aplanamiento (Flatten):** La capa Flatten convierte los mapas de características 2D resultantes en un vector 1D, preparando los datos para la capa densa.
- **Capa Densa:** Finalmente, una capa densa con 1 unidad y función de activación sigmoidea se utiliza para la clasificación binaria.

Los parámetros de entrenamiento para la CNN de línea base como para la ResNet50 del sistema propuesto se muestran en la Tabla 1.

Para evaluar el rendimiento de los modelos se calcularon las métricas de exactitud, sensibilidad y precisión. Para ello, se establecieron los resultados de la muestra en cuatro clases: Verdadero positivo (TP), falso positivo (FP), verdadero negativo (TN) y falso negativo (FN).

La exactitud mide la proporción de predicciones correctas, tanto positivas como negativas, en relación con el total de predicciones realizadas. Se calcula como la suma de TP y TN dividida por el total de casos. La exactitud es útil como una medida general de rendimiento, pero puede ser engañosa en conjuntos de datos desequilibrados donde una clase es mucho más frecuente que la otra. En nuestro caso, al balancear los datos con GANs, la exactitud sigue siendo una métrica útil.

La sensibilidad mide la proporción de casos positivos reales que fueron correctamente identificados por el modelo. Se calcula como el número de TP dividido por la suma de TP y FN. La sensibilidad es crucial en situaciones donde no detectar un caso positivo, como una enfermedad, es crítico.

La precisión se refiere a la proporción de predicciones positivas que fueron realmente correctas. Se calcula como el número de TP dividido por la suma de TP y FP. La precisión es especialmente importante en situaciones donde los falsos positivos son problemáticos o costosos.

Además de las métricas antes mencionadas, para la evaluación se graficó la curva ROC. La curva

ROC es una herramienta gráfica utilizada para evaluar el rendimiento de un modelo de clasificación. Representa la relación entre la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos (1 - especificidad) para diferentes umbrales de decisión. El área bajo la curva ROC (AUC-ROC) es una medida de la capacidad del modelo para distinguir entre las clases. Un área bajo la curva cercana a 1 indica un rendimiento muy bueno, mientras que

un área cercana a 0.5 sugiere un rendimiento no mejor que el azar.

Los resultados obtenidos en el estudio muestran una distinción clara en el rendimiento entre la red CNN básica y el modelo ResNet50 más avanzado. La Curva ROC, mostrada en la Figura 3, y las métricas de desempeño, presentadas en la Tabla 2, reflejan una superioridad significativa de ResNet50

Tabla 1
Parámetros de entrenamiento para la CNN línea base y la ResNet50.

Parámetro	CNN	ResNet50
Función de Pérdida	BinaryCrossentropy	BinaryCrossentropy
Optimizador	Adam	Adam
Métricas	Accuracy	Accuracy
Épocas	500	500

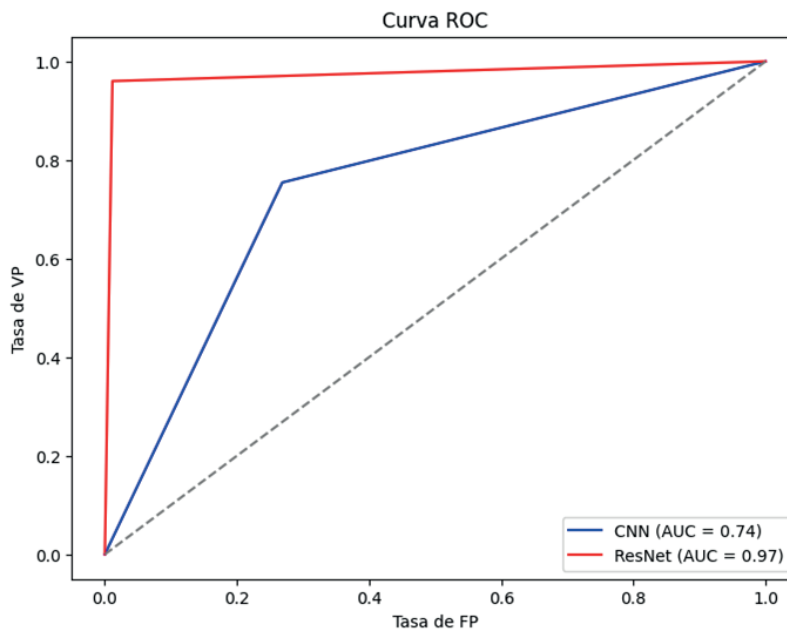
Fuente: Elaboración propia.

Tabla 2
Resultados de la CNN y ResNet50.

Métrica	Valor CNN	Valor ResNet50
Exactitud	74.29%	97.43%
Sensibilidad	73.74%	98.82%
Precisión	75.43%	96.00%

Fuente: Elaboración propia.

Figura 3
Curva ROC de la CNN de línea base y modelo propuesto.



Fuente: Elaboración propia.

en la tarea de clasificación de imágenes para la detección de la DDC. Con una exactitud que alcanza el 97.43% y una sensibilidad del 98.82%, ResNet50 demuestra una capacidad notoria para identificar correctamente los casos de DDC, lo que sugiere que la profundidad y la sofisticación de la arquitectura contribuyen de manera importante a la precisión del diagnóstico.

La CNN tradicional, con una exactitud de 74.29% y una sensibilidad de 73.74%, queda rezagada en comparación con el modelo ResNet50. Aunque la CNN puede ser más rápida y menos exigente en términos de recursos computacionales, la compensación en la precisión del modelo puede ser significativa, lo cual es crítico en aplicaciones médicas donde las decisiones basadas en el modelo pueden tener consecuencias directas en la salud de los pacientes. Además, la precisión de la CNN se registró en 75.43%, mientras que la de ResNet50 fue del 96.00%. Esta métrica es especialmente relevante, ya que una alta precisión implica una menor cantidad de falsos positivos, un aspecto crucial en el contexto médico donde los diagnósticos erróneos pueden llevar a tratamientos innecesarios o a la omisión de cuidados necesarios. La precisión superior de ResNet50 indica que es mucho menos probable que identifique incorrectamente una condición normal como DDC.

La Curva ROC subraya estos hallazgos, mostrando que ResNet50 tiene una capacidad mucho mejor para diferenciar entre las clases positivas y negativas, con un Área Bajo la Curva (AUC) de 0.97, comparado con un AUC de 0.74 para la CNN. Un AUC más cercano a 1 indica un modelo de mayor calidad, y en este caso, sugiere que ResNet50 tiene una capacidad de discriminación muy alta, lo que es esencial para un diagnóstico fiable. Estos resultados respaldan la adopción de arquitecturas de aprendizaje profundo más complejas, como ResNet50, en la clasificación de imágenes médicas para detectar DDC. La combinación de alta exactitud, sensibilidad y precisión hace que ResNet50 sea una opción prometedora para futuras aplicaciones clínicas y subraya el potencial del aprendizaje profundo para mejorar los procesos de diagnóstico médico.

V. CONCLUSIONES Y TRABAJO FUTURO

La detección de DDC es posible usando arquitecturas profundas que aprendan a partir de imágenes de radiografías pélvicas. También, el uso de redes GANs para el aumento de datos contribuye al mejor entrenamiento de modelos de redes neuronales.

En el artículo se destaca la contribución de la integración del sistema con ResNet50 y las técnicas de GANs para lograr una alta precisión en la clasificación de radiografías pélvicas, enfocándose específicamente en la identificación de la displasia del desarrollo de la cadera. La aplicación de ResNet50, una arquitectura avanzada de redes neuronales fue esencial por su capacidad para aprender características detalladas y patrones complejos, ResNet50 demostró ser muy efectivo en contextos médicos donde la precisión en los detalles es fundamental.

El uso de GANs también contribuyó a los resultados obtenidos. Mediante la generación de datos sintéticos que enriquecieron el conjunto de entrenamiento, las GANs proporcionaron una diversidad y volumen de datos que mejoraron significativamente el aprendizaje del modelo. Este enfoque fue particularmente valioso en la radiología, donde a menudo es difícil obtener grandes cantidades de datos etiquetados.

La comparación entre el sistema propuesto y una CNN que no utilizaba datos aumentados reveló una mejora notable. Mientras que la CNN sin aumento de datos alcanzó una precisión del 74.3%, el sistema que integraba ResNet50 y GANs logró una precisión del 97.3%. Esta diferencia no solo resalta la efectividad de las técnicas avanzadas de aprendizaje profundo, sino también 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 incorporación de ResNet50 y GANs en el sistema propuesto resultó en una mejora significativa en la precisión y confiabilidad del modelo para la clasificación de imágenes médicas. Estos resultados son prometedores para futuras aplicaciones en diagnóstico médico asistido por computadora y sugieren un camino interesante para investigaciones adicionales en la adaptación de estas tecnologías a otras áreas de la medicina. En cuanto al trabajo futuro, hay varias vías prometedoras que se podrían explorar para extender y mejorar los hallazgos de este artículo. Una posibilidad es probar diferentes arquitecturas de redes neuronales para evaluar su eficacia en la clasificación de radiografías pélvicas. Diversas arquitecturas como Inception, VGG y DenseNet, cada una con sus propias peculiaridades en términos de procesamiento y aprendizaje de características, podrían proporcionar hallazgos valiosos y posiblemente mejorar la precisión del modelo. Además, una extensión natural de este trabajo sería el entrenamiento de modelos no solo para identificar la presencia de la

displasia del desarrollo de la cadera, sino también para clasificar sus diferentes niveles de severidad.

REFERENCIAS BIBLIOGRAFICAS

- [1] Agarwal, N., Sondhi, A., Chopra, K., & Singh, G. (2021). Transfer learning: Survey and classification. In *Smart Innovations in Communication and Computational Sciences: Proceedings of ICSICCS 2020* (pp. 145-155). Springer Singapore. https://doi.org/10.1007/978-981-15-5345-5_13
- [2] Aggarwal, R., Sounderajah, V., Martin, G., Ting, D. S., Karthikesalingam, A., King, D., Ashrafian, H., & 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>
- [3] Al-Naser, S., & Altamimi, A. A. (2021). DDH Diagnosis: What do we Know so Far? *The Open Public Health Journal*, 14, 469-477. <https://doi.org/10.2174/1874944502114010469>
- [4] 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>
- [5] Baudet, M., Revenga, C., & Buló, 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>
- [6] 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>
- [7] Fraiwan, M., Al-Kofahi, N., Ibnian, 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>
- [8] Goodfellow, I., Pouget-Abadie, J., Mehdi, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144. <https://doi.org/10.1145/3422622>
- [9] 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>
- [10] Longo, U. G., Papalia, R., De Salvatore, S., Ruzzini, L., Piergentili, I., Oggiano, L., Costici, P. F., & 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). <https://doi.org/10.3390/ijerph18126589>
- [11] McEvoy, F., Proschowsky, H., Müller, A., Moorman, L., Bender-Koch, J., Svalastoga, E., Frellsen, J., & Nielsen, D. (2021). of hip joints in pelvis radiographs and the classification of their hip dysplasia status. *Vet Radiol Ultrasound*, 62(4), 387-393. <https://doi.org/10.1111/vru.12968>
- [12] 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
- [13] Park, H. S., Jeon, K., Cho, Y. J., Kim, S. W., Lee, S. B., Choi, G., Lee, S., Choi, Y. H., Cheon, J.-E., Kim, W. S., Ryu, Y. J., & 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>
- [14] 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.rmclc.2021.04.003>
- [15] 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

- [16] Shaw, B. A., Segal, L. S., & ORTHOPAEDICS, S. O. (2016). Evaluation and Referral for Developmental Dysplasia of the Hip in Infants. *Pediatrics*, 138(6), e20163107. <https://doi.org/10.1542/peds.2016-3107>
- [17] 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
- [18] 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>

Conflictos de Interés

El autor declara no tener conflictos de interés.