

Diagnóstico radiográfico de displasia del desarrollo de la cadera infantil asistido por inteligencia artificial: estudio retrospectivo

Melina I. Luna, Román Bertolotti, Gerardo M. Ferrer, Miguel C. González, Antonio E. Pedro Maraude

Servicio de Ortopedia y Traumatología, Hospital de Niños de la Santísima Trinidad de Córdoba, Córdoba, Argentina

RESUMEN

Introducción: La displasia del desarrollo de la cadera (DDC) es la alteración musculoesquelética pediátrica más frecuente y su diagnóstico radiográfico puede verse afectado por variaciones técnicas e interpretativas. La inteligencia artificial (IA), particularmente mediante plataformas AutoML, ofrece herramientas para estandarizar la evaluación de imágenes y asistir en la clasificación diagnóstica. **Objetivos:** Desarrollar y evaluar un modelo de IA para diferenciar radiografías panorámicas de pelvis de niños normales o con DDC utilizando la plataforma Landing Lens. **Materiales y Métodos:** Se realizó un estudio retrospectivo de 328 radiografías de pelvis de pacientes de entre 1 mes y 14 años. Cuatro ortopedistas infantiles con distintos niveles de experiencia realizaron el etiquetado independiente de las imágenes. El diagnóstico final se definió mediante mayoría simple y consenso en casos discordantes. Las radiografías se dividieron en conjuntos de entrenamiento (70%), validación (20%) y prueba (10%). El rendimiento del modelo se evaluó mediante exactitud, sensibilidad, especificidad, F1 score y coeficientes de concordancia. **Resultados:** El modelo clasificó correctamente 320 imágenes (exactitud 97,6%, sensibilidad 92,8%, especificidad 98%, F1 score 86,6%). La concordancia interobservador fue casi perfecta (kappa de Fleiss 0,86), y el acuerdo entre el modelo y los especialistas aumentó con la experiencia clínica (kappa de Cohen 0,85 con el observador senior). **Conclusiones:** Estos resultados respaldan el uso de la IA como herramienta complementaria para estandarizar el diagnóstico radiográfico de la DDC y reducir la variabilidad interobservador. Se requieren validaciones externas y estudios multicéntricos antes de su implementación clínica rutinaria.

Palabras clave: Displasia del desarrollo de la cadera; inteligencia artificial; AutoML; radiografía de pelvis; niños; aprendizaje profundo.

Nivel de Evidencia: III

Artificial Intelligence–Assisted Radiographic Diagnosis of Developmental Dysplasia of the Hip in Children: A Retrospective Study

ABSTRACT

Introduction: Developmental dysplasia of the hip (DDH) is the most common pediatric musculoskeletal disorder, and its radiographic diagnosis may be affected by technical and interpretive variability. Artificial intelligence (AI), particularly through automated machine learning (AutoML) platforms, offers tools to standardize image assessment and assist in diagnostic classification. **Objectives:** To develop and evaluate an AI model for distinguishing between normal pediatric pelvic radiographs and those showing DDH using the Landing Lens platform. **Materials and Methods:** A retrospective study included 328 pelvic radiographs from patients aged 1 month to 14 years. Four pediatric orthopedic surgeons with varying levels of experience independently labeled the images. The final diagnosis was established by a simple majority, with consensus reached in cases of disagreement. The radiographs were divided into training (70%), validation (20%), and test (10%) sets. Model performance was assessed using accuracy, recall, specificity, F1 score, and agreement coefficients. **Results:** The model correctly classified 320 images (accuracy, 97.6%; recall, 92.8%; specificity, 98%; F1 score, 86.6%). Interobserver agreement was almost perfect (Fleiss' kappa, 0.86), and agreement between the model and specialists increased with clinical experience (Cohen's kappa, 0.85 for the senior observer). **Conclusions:** These findings support the use of AI as an adjunctive tool to standardize the radiographic diagnosis of DDH and reduce interobserver variability. External validation and multicenter studies are required before routine clinical implementation.

Keywords: Developmental dysplasia of the hip; artificial intelligence; AutoML; pelvic radiography; children; deep learning.

Level of Evidence: III

Recibido el 26-12-2025. Aceptado luego de la evaluación el 7-7-2026 • Dra. MELINA I. LUNA • meli.itati4@gmail.com

 <https://orcid.org/0000-0002-2466-9492>

Cómo citar este artículo: Luna MI, Bertolotti R, Ferrer GM, González MC, Pedro Maraude AE. Diagnóstico radiográfico de displasia del desarrollo de la cadera infantil asistido por inteligencia artificial: estudio retrospectivo. *Rev Asoc Argent Ortop Traumatol* 2026;91(5):420-426. <https://doi.org/10.15417/issn.1852-7434.2026.91.5.2281>

INTRODUCCIÓN

La displasia del desarrollo de la cadera (DDC) se debe a una relación anormal entre la cabeza femoral y el acetábulo de la articulación coxofemoral inmadura, abarca un espectro que va desde una displasia acetabular leve hasta la subluxación o luxación completa. Es el trastorno musculoesquelético pediátrico más frecuente, tiene una incidencia estimada del 0,1 al 11%, predomina en el sexo femenino y en la cadera izquierda.¹⁻¹⁰ La radiografía panorámica de pelvis es el estudio de primera línea para el diagnóstico de la DDC en niños >4-6 meses.^{1,3-6,8,9,11,12}

La inteligencia artificial (IA) ha emergido como una herramienta de creciente aplicación en medicina, particularmente en el análisis de imágenes. A través de técnicas de visión artificial y aprendizaje automático, es posible entrenar algoritmos capaces de identificar patrones complejos y asistir en la clasificación diagnóstica y, de esta manera, contribuir a reducir la variabilidad interobservador.^{13,14}

En los últimos años, diversos estudios han aplicado modelos de aprendizaje profundo para el diagnóstico radiográfico de la DDC, y han obtenido resultados alentadores en términos de exactitud, sensibilidad y especificidad.^{3,5-7,9,11,12,15-17} Dentro de este campo, el aprendizaje automático automatizado (AutoML) permite desarrollar modelos predictivos sin requerir conocimientos avanzados en programación o ciencia de datos, lo que facilita su adopción en entornos clínicos. Landing Lens es una plataforma de visión artificial basada en AutoML que permite el entrenamiento y la validación de modelos de clasificación de imágenes mediante interfaces intuitivas.^{13,18}

Los objetivos de este estudio fueron desarrollar y evaluar un modelo de AutoML utilizando la plataforma Landing Lens para clasificar radiografías panorámicas de pelvis de niños normales o con DDC.

MATERIALES Y MÉTODOS

Este estudio se realizó de acuerdo con los principios éticos de la Declaración de Helsinki. Dado su diseño retrospectivo y el uso de imágenes radiográficas previamente adquiridas como parte de la práctica clínica habitual, se otorgó dispensa del consentimiento informado. Todas las imágenes fueron anonimizadas antes del análisis a fin de garantizar la confidencialidad y la protección de los datos personales de los pacientes.

Se realizó una búsqueda retrospectiva de todas las radiografías panorámicas de pelvis obtenidas en los pacientes pediátricos (de 1 mes a 14 años y 11 meses de edad) en la institución, entre el 1 de marzo de 2023 y el 1 marzo de 2025. Se excluyeron las radiografías de los niños con lesiones traumáticas (fracturas, luxaciones agudas), infecciosas (artritis séptica, osteomielitis), tumorales, neurológicas (encefalopatía crónica no evolutiva, mielomeningocele), con diagnóstico de artrogrifosis, deslizamiento de la epífisis capital femoral, enfermedad de Legg-Calvé-Perthes y antecedente de cirugía de cadera.

Se analizaron las variables demográficas edad en meses y sexo. Las imágenes radiográficas fueron convertidas del formato DICOM (*Digital Imaging and Communications in Medicine*) al formato JPG antes de cargarlas en la plataforma de IA. Durante este proceso, se mantuvieron parámetros estandarizados de visualización (ventana y nivel), con el objetivo de preservar la información anatómica relevante y homogeneizar el conjunto de datos. Si bien el formato DICOM permite un mayor control sobre los parámetros de imagen, la conversión a JPG fue necesaria para garantizar la compatibilidad con la plataforma AutoML utilizada.

El etiquetado de las radiografías (DDC o normal) estuvo a cargo de un equipo de cuatro especialistas en ortopedia infantil (un especialista con menos de 5 años de experiencia, dos especialistas con entre 10 y 15 años de experiencia y un especialista con más de 15 años de experiencia), en forma independiente.

Para la definición final del diagnóstico de cada radiografía (DDC o normal), se utilizó un criterio de mayoría simple entre los cuatro observadores independientes. Cuando había discrepancia entre las evaluaciones, el diagnóstico definitivo se estableció mediante la revisión conjunta y el consenso, con la participación del especialista con mayor experiencia clínica (más de 15 años en ortopedia infantil), quien actuó como árbitro final. Este criterio fue adoptado como estándar de referencia para el entrenamiento y la evaluación del modelo de IA.

Para implementar el modelo predictivo del proyecto, se utilizó la herramienta Landing Lens, plataforma de IA del sitio web Landing.ai. Las siguientes métricas empleadas para la evaluación del rendimiento del modelo de aprendizaje automático se detallan en la [Figura 1](#).

- Matriz de confusión: herramienta que permite visualizar el desempeño de un algoritmo obteniendo una visión más completa en la evaluación del rendimiento de un modelo.
- Precisión (*precision*): es la proporción de predicciones correctas de un conjunto de datos de prueba. Es la relación entre el número de predicciones correctas y el número total de ejemplos de entrada (casos positivos detectados).

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

Figura 1. Fórmula de las métricas para la evaluación del modelo de inteligencia artificial.
 TP = verdadero positivo, FN = falso negativo, FP = falso positivo, FN = falso negativo.

- Sensibilidad (*recall*): es la proporción de casos positivos que fueron correctamente identificados por el algoritmo. Representa los verdaderos positivos (radiografías con DDC).
- F1 *score*: combina precisión y sensibilidad en una sola métrica, creando una medida unificada que evalúa la eficacia del modelo para minimizar los falsos positivos y los falsos negativos. Un F1 *score* más alto indica que el modelo equilibra bien ambos factores.
- Exactitud (*accuracy*): determina la proporción de casos positivos y negativos correctamente identificados con respecto al número total de imágenes.
- Especificidad (*specificity*): mide cuántas radiografías normales se identifican correctamente como tales (verdaderos negativos).

Además, se calcularon el coeficiente kappa de Fleiss para evaluar la concordancia entre los especialistas y el coeficiente kappa de Cohen para valorar el acuerdo entre cada especialista y el modelo de IA.

RESULTADOS

Se obtuvieron 328 radiografías panorámicas de pelvis: 152 correspondían a niñas (46,34%) y 176, a niños (53,65%). El promedio de edad era de 66 meses (5.5 años; rango 2-179 meses), la mayor cantidad de radiografías eran de pacientes <24 meses (118 imágenes).

Al desarrollar el modelo de IA, 230 radiografías (70%) fueron muestras para entrenamiento; 65, para validación (20%) y 33, para prueba (10%). De las 328 radiografías, 320 fueron clasificadas correctamente y ocho, en forma errónea, por el modelo de IA (Figura 2).

La exactitud global del modelo fue del 97,6%; la precisión, del 81,2% y la sensibilidad, del 92,8%; identificó correctamente la mayoría de los casos de DDC (alta sensibilidad). El F1 *score* fue del 86,6%, que expresa un adecuado equilibrio entre precisión y sensibilidad; y la especificidad, del 98% lo que denota muy buena capacidad para distinguir radiografías normales (alta especificidad) (Figura 3).

Al evaluar la concordancia interobservador, se obtuvo un coeficiente kappa de Fleiss de 0,86 que corresponde a una fuerza de concordancia casi perfecta (Tabla). Cuando se calculó el coeficiente kappa de Cohen entre el especialista de menor experiencia (menos de 5 años) y el modelo de IA, se obtuvo 0,75. El coeficiente kappa de Cohen de ambos especialistas con 10-15 años de experiencia fue de 0,81 y el del especialista con más experiencia (>15 años), de 0,85.



Figura 2. Mapas de calor del modelo de inteligencia artificial, correspondiente al marcado de las áreas críticas de cada imagen para visualizar su proceso de toma de decisiones.

Figura 3. Matriz de confusión global del modelo de inteligencia artificial que incluye precisión y sensibilidad.

Ground truth			Ground truth	
Normal	294	6	0	98,0%
DDC	2	26	0	92,90%
No label	0	0	—	
Precisión	99,3%	81,30%		
	Normal	DDC	No prediction	Prediction

Tabla. Valoración del coeficiente kappa (Landis y Koch, 1977).

Coeficiente kappa	Fuerza de la concordancia
0,00	Pobre (<i>Poor</i>)
0,01-0,20	Leve (<i>Slight</i>)
0,21-0,40	Aceptable (<i>Fair</i>)
0,41-0,60	Moderada (<i>Moderate</i>)
0,61-0,80	Considerable (<i>Substantial</i>)
0,81-1,00	Casi perfecta (<i>Almost perfect</i>)

DISCUSIÓN

En diversos estudios, se ha evaluado el diagnóstico radiográfico de la DDC asistido por IA, en poblaciones pediátricas con amplios rangos etarios. Esta heterogeneidad, presente también en nuestra muestra, refleja las variaciones anatómicas propias del crecimiento, lo que representa un desafío adicional para los modelos de clasificación automática.^{2-5,7,9,10,12,15,16}

En concordancia con la literatura médica, la mayoría de las publicaciones reportan una mayor frecuencia de DDC en el sexo femenino. No obstante, en nuestra serie, predominó el sexo masculino, lo que podría estar relacionado con el perfil de derivación institucional y no necesariamente con la prevalencia real de la enfermedad.^{3,5,6,8,9,11,16}

El tamaño de la muestra de este estudio es comparable con la de múltiples investigaciones previas,^{2,4,16} aunque inferior a la de grandes series multicéntricas. Los estudios con mayor casuística son los de Zhang, con 10.219 radiografías y Liu, con 10.000 imágenes.^{6,8}

El etiquetado de las radiografías está a cargo de especialistas en ortopedia infantil con diferentes años de experiencia o de radiólogos con distintos años de formación en radiología pediátrica.⁷ Hay estudios en los que trabajan, en conjunto, ortopedistas infantiles con radiólogos pediátricos.^{3,6,12} En nuestro caso, el etiquetado lo realizaron cuatro ortopedistas infantiles: uno con menos de 5 años de experiencia (*junior*), dos con experiencia de entre 10 y 15 años (intermedio) y el último con más de 15 años de experiencia (*senior*), metodología que se observa en otras publicaciones.^{3,5,7,11}

El uso del coeficiente kappa permite valorar el grado de concordancia entre los observadores. En nuestro estudio, dicha concordancia fue casi perfecta (coeficiente kappa de Fleiss 0,86), semejante a la publicada por Xu y cols.³ Cuando se compara la concordancia entre el modelo de IA y el especialista, los valores más altos del coeficiente kappa de Cohen se dan en el especialista *senior*,⁵ coincidente, en nuestro caso, con un valor de 0,85.

El desempeño de nuestro modelo de IA mostró una exactitud global elevada, con alta sensibilidad y muy alta especificidad, valores acordes a los reportados en revisiones sistemáticas y estudios recientes. La diferencia observada entre exactitud y precisión puede explicarse por el desequilibrio relativo entre clases y el bajo número de falsos positivos, fenómeno ya descrito en modelos aplicados al diagnóstico de DDC.^{3-7,11,16,19,20}

Este estudio tiene limitaciones relacionadas con su diseño retrospectivo y unicéntrico, el tamaño de la muestra, la ausencia de validación externa y la conversión de imágenes al formato JPG. Estas limitaciones deberán ser abordadas en futuras investigaciones mediante estudios prospectivos y multicéntricos que permitan validar y generalizar los hallazgos.^{3,6,19-21}

CONCLUSIONES

El uso de plataformas accesibles de AutoML, como Landing Lens, representa una oportunidad para ampliar el acceso a herramientas de apoyo diagnóstico en entornos con recursos limitados, y así facilitar la estandarización de la lectura radiográfica y la reducción de la variabilidad interobservador. Si bien se requieren validaciones externas previas a su implementación clínica rutinaria, la IA se perfila como un complemento valioso del juicio clínico experto en el diagnóstico radiográfico de la DDC, sin necesidad de formación avanzada en ciencias de datos.

Uso de la inteligencia artificial

Se utilizó ChatGPT (OpenAI) para generar el 25% del contenido del resumen y la conclusión. Los autores verificaron la exactitud y originalidad del contenido generado por IA.

Conflicto de intereses: Los autores no declaran conflictos de intereses.

ORCID de R. Bertolotti: <https://orcid.org/0000-0003-1356-9741>
 ORCID de G. M. Ferrer: <https://orcid.org/0009-0007-4582-0255>

ORCID de M. C. González: <https://orcid.org/0009-0003-0230-2969>
 ORCID de A. E. Pedro Marade: <https://orcid.org/0000-0001-9016-6132>

BIBLIOGRAFÍA

1. Chen M, Cai R, Zhang A, Chi X, Qian J. The diagnostic value of artificial intelligence-assisted imaging for developmental dysplasia of the hip: a systematic review and meta-analysis. *J Orthop Surg Res* 2024;19(1):522. <https://doi.org/10.1186/s13018-024-05003-4>
2. Jan F, Rahman A, Busaleh R, Alwarthan H, Aljaser S, Al-Towailib S, et al. Assessing acetabular index angle in infants: A deep learning-based novel approach. *J Imaging* 2023;9(11):242. <https://doi.org/10.3390/jimaging9110242>
3. Xu W, Shu L, Gong P, Huang C, Xu J, Zhao J, et al. A deep-learning aided diagnostic system in assessing developmental dysplasia of the hip on pediatric pelvic radiographs. *Front Pediatr* 2021;9:785480. <https://doi.org/10.3389/fped.2021.785480>
4. Fraiwan M, Al-Kofahi N, Ibnian A, Hanatleh O. Detection of developmental dysplasia of the hip in X-ray images using deep transfer learning. *BMC Med Inform Decis Mak* 2022;22(1):216. <https://doi.org/10.1186/s12911-022-01957-9>
5. Chen J, Fan X, Chen Z, Peng Y, Liang L, Su C, et al. Enhancing YOLO5 for the assessment of irregular pelvic radiographs with multimodal information. *J Imaging Inform Med* 2024;37(2):744-55. <https://doi.org/10.1007/s10278-024-00986-2>
6. Zhang SC, Sun J, Liu CB, Fang JH, Xie HT, Ning B. Clinical application of artificial intelligence-assisted diagnosis using anteroposterior pelvic radiographs in children with developmental dysplasia of the hip. *Bone Joint J* 2020;102-B(11):1574-81. <https://doi.org/10.1302/0301-620X.102B11.BJJ-2020-0712.R2>
7. Park HS, Jeon K, Cho YJ, Kim SW, Lee SB, Choi G, et al. Diagnostic performance of a new convolutional neural network algorithm for detecting developmental dysplasia of the hip on anteroposterior radiographs. *Korean J Radiol* 2021;22(4):612-23. <https://doi.org/10.3348/kjr.2020.0051>
8. Liu C, Xie H, Zhang S, Mao Z, Sun J, Zhang Y. Misshapen pelvis landmark detection with local-global feature learning for diagnosing developmental dysplasia of the hip. *IEEE Trans Med Imaging* 2020;39(12):3944-54. <https://doi.org/10.1109/TMI.2020.3008382>
9. Qi G, Jiao X, Li J, Qin C, Li X, Sun Z, et al. A dataset for quality evaluation of pelvic X-ray and diagnosis of developmental dysplasia of the hip. *Sci Data* 2025;12(1):865. <https://doi.org/10.1038/s41597-025-05146-x>
10. Sha J, Huang L, Chen Y, Fan Z, Lin J, Yang Q, et al. Clinical thought-based software for diagnosing developmental dysplasia of the hip on pediatric pelvic radiographs. *Front Pediatr* 2023;11:1080194. <https://doi.org/10.3389/fped.2023.1080194>
11. Perry S, Folkman M, O'Brien T, Wilson L, Coyle E, Liu R, et al. Unaligned hip radiograph assessment utilizing convolutional neural networks for the assessment of developmental dysplasia of the hip. *J Eng Sci Med Diagn Ther* 2024;7(4):041003. <https://doi.org/10.1115/1.4064988>
12. Segura F, Segura P, Segura F. Determinación del grado de agresividad tumoral mediante técnicas de inteligencia artificial aplicadas a imágenes de resonancia magnética. *Rev Asoc Argent Ortop Traumatol* 2023;88(6):653-61. <https://doi.org/10.15417/issn.1852-7434.2023.88.6.1738>
13. Bayliss L, Jones LD. The role of artificial intelligence and machine learning in predicting orthopaedic outcomes. *Bone Joint J* 2019;101-B(12):1476-8. <https://doi.org/10.1302/0301-620X.101B12.BJJ-2019-0850.R1>
14. Segura FM, Segura FP, Zudaire MPL, Segura FV. Advances in Artificial Intelligence for automated knee osteoarthritis classification using the IKDC system. *Eur J Orthop Surg Traumatol* 2024;35(1):32. <https://doi.org/10.1007/s00590-024-04124-0>
15. Den H, Ito J, Kokaze A. Diagnostic accuracy of a deep learning model using YOLOv5 for detecting developmental dysplasia of the hip on radiography images. *Sci Rep* 2023;13(1):6693. <https://doi.org/10.1038/s41598-023-33860-2>

16. Lv J, Che J, Chen X. CBA-YOLOv5s: A hip dysplasia detection algorithm based on YOLOv5s using angle consistency and bi-level routing attention. *Biomed Signal Process Control* 2024;95:106482. <https://doi.org/10.1016/j.bspc.2024.106482>
17. Sha J, Huang L, Chen Y, Lin J, Fan Z, Li Y, et al. A novel approach for screening standard anteroposterior pelvic radiographs in children. *Eur J Pediatr* 2023;182(11):4983-91. <https://doi.org/10.1007/s00431-023-05164-0>
18. Gartland CN, Healy J, Lynham RS, Nowlan NC, Green C, Redmond SJ. Automated system of analysis to quantify pediatric hip morphology. *Comput Biol Med* 2025;197:110904. <https://doi.org/10.1016/j.combiomed.2025.110904>
19. van Kouswijk HW, Yazid H, Schoones JW, Witlox MA, Nelissen RGH, de Witte PB. Current and emerging applications of artificial intelligence in medical imaging for paediatric hip disorders: a scoping review. *Children* 2025;12(5):645. <https://doi.org/10.3390/children12050645>
20. Bhavsar S, Gowda BB, Bhavsar M, Patole S, Rao S, Rath C. Artificial intelligence to detect developmental dysplasia of hip: A systematic review. *J Paediatr Child Health* 2025;61(11):1712-27. <https://doi.org/10.1111/jpc.70172>
21. McArthur A, Wichuk S, Burnside S, Kirby A, Scammon A, Sol D, et al. Retuve: Automated multi-modality analysis of hip dysplasia with open source AI. *arXiv* 2025;:2504.06422. <https://doi.org/10.48550/ARXIV.2504.06422>