

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

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

Orthopedics and Traumatology Service, Hospital de Niños de la Santísima Trinidad de Córdoba, Córdoba, Argentina

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

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

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

Received on December 26th, 2025. Accepted after evaluation on July 7th, 2026 • Dr. MELINA I. LUNA • meli.itati4@gmail.com  <https://orcid.org/0000-0002-2466-9492>

How to cite this article: Luna MI, Bertolotti R, Ferrer GM, González MC, Pedro Maraude AE. Artificial Intelligence-Assisted Radiographic Diagnosis of Developmental Dysplasia of the Hip in Children: A Retrospective Study. *Rev Asoc Argent Ortop Traumatol* 2026;91(5):420-426. <https://doi.org/10.15417/issn.1852-7434.2026.91.5.2281>

INTRODUCTION

Developmental dysplasia of the hip (DDH) results from an abnormal relationship between the femoral head and the acetabulum in the immature hip joint. It encompasses a spectrum ranging from mild acetabular dysplasia to subluxation or complete dislocation. DDH is the most common pediatric musculoskeletal disorder, with an estimated incidence of 0.1–11%, and is more frequent in girls and in the left hip.¹⁻¹⁰ Pelvic radiography is the first-line imaging modality for diagnosing DDH in children older than 4-6 months.^{1,3-6,8,9,11,12}

Artificial intelligence (AI) is increasingly used in medicine, particularly for image analysis. Computer vision and machine learning techniques allow algorithms to be trained to recognize complex patterns and assist with diagnostic classification, helping to reduce interobserver variability.^{13,14}

In recent years, several studies have used deep learning models for the radiographic diagnosis of DDH, with encouraging results in terms of accuracy, recall, and specificity.^{3,5-7,9,11,12,15-17} Automated machine learning (AutoML) allows predictive models to be developed without advanced programming or data science expertise, making them more accessible in clinical settings. Landing Lens is an AutoML-based computer vision platform that provides intuitive interfaces for training and validating image classification models.^{13,18}

This study aimed to develop and evaluate an AutoML model using the Landing Lens platform to classify pediatric pelvic radiographs as normal or showing DDH.

MATERIALS AND METHODS

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Informed consent was waived because of the retrospective design and the use of radiographs previously obtained during routine clinical care. All images were anonymized before analysis to ensure confidentiality and protect patients' personal data.

A retrospective search was conducted for all pelvic radiographs obtained at the institution between March 1, 2023, and March 1, 2025, in pediatric patients aged 1 month to 14 years and 11 months. Radiographs were excluded if the children had traumatic conditions (fractures or acute dislocations), infections (septic arthritis or osteomyelitis), tumors, neurological conditions (chronic nonprogressive encephalopathy or myelomeningocele), a diagnosis of arthrogryposis, slipped capital femoral epiphysis, or Legg-Calvé-Perthes disease, or a history of hip surgery.

The demographic variables analyzed were age in months and sex. Radiographic images were converted from DICOM (*Digital Imaging and Communications in Medicine*) to JPG format before being uploaded to the AI platform. Standardized display settings (window width and level) were maintained during conversion to preserve relevant anatomical information and ensure consistency across the dataset. Although DICOM format offers greater control over image parameters, conversion to JPG was necessary for compatibility with the AutoML platform.

Four pediatric orthopedic specialists independently labeled the radiographs as DDH or normal: one had less than 5 years of experience, two had 10–15 years, and one had more than 15 years.

The final diagnosis for each radiograph (DDH or normal) was determined by a simple majority among the four independent observers. When their assessments differed, the final diagnosis was established through joint review and consensus, with the most experienced specialist (more than 15 years in pediatric orthopedics) acting as the final adjudicator. This approach served as the reference standard for training and evaluating the AI model.

The predictive model was implemented using Landing Lens, an AI platform available on the Landing.ai website. The following metrics were used to evaluate the machine learning model's performance and are detailed in Figure 1.

- Confusion matrix: A tool that displays an algorithm's performance, allowing a more comprehensive assessment of the model.
- Precision: The proportion of correct predictions in a test dataset. It is the ratio of the number of correct predictions to the total number of input examples (positive cases detected).

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

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

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

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

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

Figure 1. Formulas for the metrics used to evaluate the artificial intelligence model.

TP = true positive; TN = true negative; FP = false positive; FN = false negative.

- Recall: The proportion of positive cases correctly identified by the algorithm. It represents the true positives (radiographs showing DDH).
- F1 score: Combines precision and recall into a single metric to assess the model's ability to minimize false positives and false negatives. A higher F1 score indicates a good balance between these two factors.
- Accuracy: The proportion of correctly identified positive and negative cases relative to the total number of images.
- Specificity: Measures how many normal radiographs are correctly identified as such (true negatives).

Fleiss' kappa coefficient was also calculated to assess agreement among specialists, and Cohen's kappa coefficient was used to assess agreement between each specialist and the AI model.

RESULTS

A total of 328 pelvic radiographs were obtained: 152 from girls (46.34%) and 176 from boys (53.65%). The mean age was 66 months (5.5 years; range 2–179 months). The largest number of radiographs came from patients younger than 24 months (118 images).

For model development, 230 radiographs (70%) were used for training, 65 (20%) for validation, and 33 (10%) for testing. The AI model correctly classified 320 of the 328 radiographs and misclassified eight (Figure 2).

The model achieved an overall accuracy of 97.6%, a precision of 81.2%, and a recall of 92.8%, correctly identifying most cases of DDH (high recall). The F1 score was 86.6%, indicating an adequate balance between precision and recall. Specificity was 98%, indicating a very good ability to identify normal radiographs (high specificity) (Figure 3).

Interobserver agreement, assessed using Fleiss' kappa, was 0.86, corresponding to almost perfect agreement (Table). Cohen's kappa between the least experienced specialist (less than 5 years) and the AI model was 0.75. It was 0.81 for each of the two specialists with 10–15 years of experience and 0.85 for the most experienced specialist (>15 years).



Figure 2. Heatmaps generated by the artificial intelligence model, highlighting key areas in each image to visualize its decision-making process.

Figure 3. Overall confusion matrix for the artificial intelligence model, including precision and recall.

Ground truth			Ground truth	
Normal	294	6	0	98.0%
DDC	2	26	0	92.90%
No label	0	0	—	
Precision	99.3%	81.30%		
	Normal	DDC	No prediction	Prediction

Table. Interpretation of the kappa coefficient (Landis and Koch, 1977).

Kappa coefficient	Strength of Agreement
0.00	Poor
0.01-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

DISCUSSION

Several studies have evaluated AI-assisted radiographic diagnosis of DDH in pediatric populations with broad age ranges. This heterogeneity, also present in our sample, reflects the anatomical changes that occur during growth and poses an additional challenge for automated classification models.^{2-5,7,9,10,12,15,16}

Consistent with the medical literature, most publications report a higher frequency of DDH in girls. Boys predominated in our series, however, which may reflect institutional referral patterns rather than the actual prevalence of the disease.^{3,5,6,8,9,11,16}

Our sample size is comparable to those of several previous studies,^{2,4,16} although smaller than those of large multicenter series. The largest studies are those by Zhang, with 10,219 radiographs, and Liu, with 10,000 images.^{6,8}

Radiographs are labeled by pediatric orthopedic specialists with varying years of experience or by radiologists with different levels of training in pediatric radiology.⁷ In some studies, pediatric orthopedists and pediatric radiologists work together.^{3,6,12} In our study, four pediatric orthopedists performed the labeling: one with less than 5 years of experience (junior), two with 10-15 years (intermediate), and one with more than 15 years (senior). This approach has also been used in other publications.^{3,5,7,11}

The kappa coefficient measures the degree of agreement among observers. Agreement in our study was almost perfect (Fleiss' kappa 0.86), similar to that reported by Xu et al.³ When agreement between the AI model and individual specialists is assessed, the highest Cohen's kappa values are found for the senior specialist,⁵ consistent with the value of 0.85 in our study.

Our AI model achieved high overall accuracy, high recall, and very high specificity, in line with values reported in systematic reviews and recent studies. The difference between accuracy and precision may be explained by the relative class imbalance and the low number of false positives, a phenomenon previously described in models used to diagnose DDH.^{3-7,11,16,19,20}

This study is limited by its retrospective, single-center design, sample size, lack of external validation, and conversion of images to JPG format. Future prospective, multicenter studies should address these limitations to validate the findings and assess their generalizability.^{3,6,19-21}

CONCLUSIONS

Accessible AutoML platforms such as Landing Lens offer an opportunity to expand access to diagnostic support tools in resource-limited settings, helping to standardize radiographic interpretation and reduce interobserver variability. Although external validation is needed before routine clinical use, AI shows promise as a valuable complement to expert clinical judgment in the radiographic diagnosis of DDH, without requiring advanced training in data science.

Use of Artificial Intelligence

ChatGPT (OpenAI) was used to generate 25% of the content of the abstract and conclusion. The authors verified the accuracy and originality of the AI-generated content.

Conflicts of interest: The authors declare no conflicts of interest.

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

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

REFERENCES

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.compbimed.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>