Diagnostic accuracy of a deep learning model using YOLOv5 for detecting developmental dysplasia of the hip on radiography images

Hiroki Den (denh@med.showa-u.ac.jp)
National Rehabilitation Center for Children with Disabilities
Junichi Ito
National Rehabilitation Center for Children with Disabilities
Akatsuki Kokaze
Showa University School of Medicine

Article

Keywords:

Posted Date: September 23rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2067848/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on April 24th, 2023. See the published version at https://doi.org/10.1038/s41598-023-33860-2.

Abstract

Introduction

Developmental dysplasia of the hip (DDH) is a cluster of hip development disorders and one of the most common hip diseases in infants. Hip radiography is a convenient diagnostic tool for DDH, but its diagnostic accuracy is dependent on the interpreter's level of experience.

The aim of this study was to develop a deep learning model for detecting DDH using YOLOv5.

Methods

Patients younger than 12 months who underwent hip radiography between June 2009 and November 2021 were selected. Using their radiography images, transfer learning was performed to develop a deep learning model using YOLOv5.

Results

A total of 305 anteroposterior hip radiography images (205 normal hip images and 100 DDH hip images) were collected. Of these, 30 normal hip images and 17 DDH hip images were used as the test set. The sensitivity and the specificity of our best deep learning model (YOLOv5l) were 0.94(95%CI 0.73-1.00) and 0.96 (95%CI:0.89-0.99), respectively.

Conclusion

This is the first study to establish a model for detecting DDH using YOLOv5. Our deep learning models provided good diagnostic performance for DDH. We believe our model is a useful diagnostic assistant tool.

Introduction

Developmental dysplasia of the hip (DDH) is a cluster of hip developmental disorders, including dislocation, subluxation, and acetabular dysplasia. DDH is one of the most common hip diseases in infants. In our previous study, the incidence of DDH-related dislocations was 0.076% in Japan.[1] Early detection and treatment of DDH-related dislocations is highly effective, with >80% success rate.[2-4] However, treatment outcomes in patients diagnosed with DDH dislocation at the age of ≥1 year vary, suggesting that early detection and treatment is essential for good outcomes. In our previous study, we reported that the rate of late diagnosis (diagnosed ≥1 year) was 10-12% in Japan, which was concerning.

Hip radiography and ultrasonography are widely used screening tools for diagnosing DDH. Hip ultrasonography is an accurate modality, but it requires a certain level of experience to achieve acceptable performance.[5] Therefore, it may be difficult to popularize the hip ultrasonography technique for physicians who conduct hip screening for infants. Hip radiography is a convenient diagnostic tool since it is available for most hospitals and clinics. However, the diagnostic accuracy of hip radiography for DDH varies according to the interpreter's experience.[6] DDH is a relatively rare disease; in fact, a Japanese study revealed that 32% of orthopedic doctors have not encountered a patient with DDH in their career.[1] A tool that can help inexperienced doctors diagnose DDH may be necessary and can reduce the percentage of late diagnosis.

Deep learning technology has rapidly progressed and is widely used to detect or classify objects of images in many fields, such as face for recognition. Recent studies have shown that deep learning techniques can also be applied to radiographic images.[7-11]

To our knowledge, studies that have applied deep learning techniques to detect DDH using hip radiography images are limited.[12,13] In this study, the "You Only Look Once" v5 (YOLOv5) method, which is one of the most widely used object detection methods, was used.[14] This algorithm is popular because of its speed and accuracy. YOLOv5 has been used in various applications to detect traffic signals, people, parking meters, and animals.[15,16]

The aim of this study was to use YOLOv5 to develop a deep learning model that can distinguish between a normal hip and a hip with DDH using hip radiography images. We also sought to validate the diagnostic performance of the model.

Methods

This was a retrospective study. The study was approved by the Institutional Review Board of the National Rehabilitation Center for Children with Disabilities (2014-17), and all methods were carried out in accordance with their guidelines and regulations. Written informed consent was obtained from the parents or the legal guardians of all children.

Data collection

Patients younger than 12 months who were suspected to have DDH-related hip dislocation or hip subluxation and who had undergone anteroposterior(AP) view hip radiography between June 2009 and November 2021 were selected retrospectively. For patients who underwent several hip radiographies, the oldest image in the time series was selected. All patients were examined by the same experienced pediatric hip specialist for DDH screening. Patients were diagnosed with DDH based on physical examination findings, AP view hip radiography images, and hip ultrasonography images, if necessary. A positive DDH diagnosis using hip radiography images was based on the following criteria: a) lateralization of the epiphyseal ossification center, b) interruption of the Shenton line,[17] c) widened tear drop distance compared to that on the other side,[18] d) delayed femoral head ossification compared to that on the other side, e) high acetabular index (>30), and f) dulled edge of the acetabulum. The Graf method was used for hip ultrasonography.[19] The International Hip Dysplasia Institute (IHDI) classification was used to quantify DDH severity because the classification does not rely on the presence of the ossification center of the femoral head and it can be applied to patients of all ages.[20]

Basically, patients with DDH of IHDI grade 2 or worse and/or type 2c or worse Graf classification for hip ultrasonography images were considered as belonging to the DDH group since they required careful observation.[19]

Data preparation

The original images were 1430 × 1140 pixels in size. These images were changed into a square shape (1430 × 1430 pixels) by adding black regions to the top and bottom. Then, the images were resized to 864 × 864 pixels. Approximately 15 percent of normal images were randomly but equally distributed to validation and test datasets, and 15 percent of DDH images were also randomly but equally distributed to validation and test datasets, considering the equality of DDH severity based on IHDI classification.

For the training dataset, all images were augmented by flipping them horizontally. In addition, to avoid overfitting, the DDH images were augmented by 10° and -10°rotations.

Image annotation

Image annotation was performed with LablImg version 1.8.1.[21] Object bounding boxes were drawn with the following criteria: a) the inner boundary is drawn in anatomical regions deeper than the deepest region of the acetabulum, b) the outer boundary is drawn to include the greater trochanter, c) the upper boundary is drawn to include the acetabulum and the ossification center of the femoral head, and d) the lower boundary is drawn to include the lesser trochanter (Figure.1). Normal hips are labeled as “Normal” and DDH hips are labeled as “DDH”.

Deep learning algorithm

Transfer learning was performed using YOLOv5. Transfer learning is a technique in which a well-trained model from a large dataset is used for applications of interest with a small dataset.[22] Therefore, transfer learning can reduce the requirement of large datasets. YOLOv5 is the latest product in the YOLO series. YOLOv5 contains four different models: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The main difference between these models is the amount of feature extraction modules. YOLOv5s has the smallest size of modules and amount of module parameters, and YOLOv5x has the largest size of modules and amount of module parameters.[15] All four models were utilized for the present study and results were compared. For transfer learning, the first 10 layers of the YOLOv5 models were frozen in place, and the rest of the layers were retrained with our new datasets.

A learning rate of 0.01, mini-batch size of 32, and 100 epochs were used for the training.

The analyses were performed using Python 3.7.12 (Python Software Foundation, Wilmington, DA, U.S.). Consequently, the trained models could detect hips in AP view radiography images and label them as either “Normal” or “DDH”, with confidence scores (Figure.2). Confidence

The test set was evaluated using the trained models, with a 0.5 confidence score threshold.

If both “Normal” and “DDH” were labeled on the same hip, the hip evaluation was considered invalid (Figure.3). Sensitivity, specificity, positive predictive value, and negative predictive value were calculated for each trained model.

Results

Between June 2009 and November 2021, 305 hip AP X-ray images (205 normal hip images and 100 DDH hip images) were collected. Subsequently, 396 DDH and 290 normal images were used for training, and 30 normal and 17 DDH hip images were used for testing (Figure.4).

The basic characteristics of the datasets are shown in Table 1. In one case, the IHDI grade was grade 1 in the DDH group, but the patient’s Graf classification was type 2c. Therefore, this case was categorized as a DDH case.

Among the four models, four hips were labeled as both “Normal” and “DDH”; one hip in the YOLOv5x model and three hips in the YOLOv5s model. There was no double labeling in the other two models. Table 2 shows the diagnostic performance of each YOLOv5 model.

Table 1

Basic characteristics

		DDH Group N=100	Normal Group N=205
Girls, n (%)		88(88)	152(74.1)
Age(month) mean(min-max)		6.1 (1-12)	6.0 (0-12)
Affected side	Left	69	
	Right	26	
	Bilateral	5	
IHDI grade	1	1	
	2	46	
	3	37	
	4	21	

DDH: developmental dysplasia of the hip

IDHI: International Hip Dysplasia Institute.

Table 2

Diagnostic performance of YOLOv5 models.

	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
v5x	**0.83(0.59-0.96)**	**0.99(0.93-1.00)**	**0.94(0.70-1.00)**	**0.96(0.89-0.99)**
v5l	**0.94(0.73-1.00)**	**0.96(0.89-0.99)**	**0.85(0.62-0.97)**	**0.99(0.93-1.00)**
v5m	**0.89(0.65-0.99)**	**0.97(0.91-1.00)**	**0.89(0.65-0.99)**	**0.97(0.91-1.00)**
v5s	**0.94(0.73-1.00)**	**0.93(0.85-0.98)**	**0.77(0.55-0.92)**	**0.99(0.93-1.00)**

PPV: positive predictive value.

NPV: negative predictive value

In terms of false negatives, the YOLOv5s and YOLOv5l models mislabeled only one hip, and the same hip was found to have been mislabeled by the models (Figure.5).

All severe DDH cases (IHDI grade 3 and 4) were correctly labeled “DDH” in all four models. The labeling of mild DDH cases (IHDI grade 2) was variable, and no model labeled all IHDI grade 2 cases correctly (Table 3).

Table 3

DDH hips correctly labeled "DDH", according to IHDI grade.

IHDI grade	Total	N (%)			
		YOLOv5x	YOLOv5l	YOLOv5m	YOLOv5s
4	5 hips	5(100)	5(100)	5(100)	5(100)
3	7 hips	7(100)	7(100)	7(100)	7(100)
2	6 hips	3(50.0)	5(83.3)	4(66.7)	5(83.3)

DDH: developmental dysplasia of the hip

IDHI: International Hip Dysplasia Institute

Discussion

We developed deep learning models for detecting DDH using hip radiography images in the anteroposterior view. To the best of our knowledge, this is the first study to achieve this using YOLOv5. In addition, using the transfer learning technique, a good model could be constructed with a relatively small dataset. A benefit of using an object detection model rather than a classification model is the ability of the object detection model to evaluate both hips simultaneously without image processing. Studies that used YOLOv5 models have been reported in various medical fields.[23] One study used YOLOv5 models to detect brain abnormalities, and another study sought to detect lumbar spine deformities using YOLOv5 models.[24,25]

As a confirmed diagnostic tool, hip ultrasonography is probably the best modality in the hands of a well-trained operator. A study reported that the sensitivity and specificity of hip ultrasonography for detecting DDH were 93% and 97%, respectively.[5] The diagnostic performances of our trained models were comparable, but the diagnostic performance of hip ultrasonography would be better than that in the current study if the operator is well-trained.

The strength of a deep learning model compared to hip ultrasonography screening is that it can be applied in clinics who do not have a pediatric hip specialist close by. By using a deep learning model, a general practitioner with insufficient expertise in orthopedics might be able to diagnose DDH. Combining telemedicine with deep learning technology may decrease the rate of late DDH diagnosis. An autonomous artificial intelligence-based diagnosis system has already been applied in the field of medicine.[26]

A study showed the sensitivity and specificity of AP view hip radiography images reviewed by radiologists for the detection of DDH. Sensitivities and specificities of 96.0 and 89.0 were noted for a radiologist who had a 5-year experience in radiology, including pediatric radiology, and 84.0 and 85.8 were noted for a radiologist who had a 3-year experience in radiology, without any experience in pediatric radiology.[13] These results indicate that our deep learning model can be a useful screening tool for physicians who do not have sufficient experience regarding DDH.

High sensitivity is desirable for screening tools. Among our four trained models, the YOLOv5l and YOLOv5s models had the highest sensitivity (0.94). In contrast, YOLOv5x had the highest specificity (0.99) and YOLOv5s had the lowest specificity (0.93). Overall, of the four models, YOLOv5l was considered the best screening tool because of its high sensitivity and specificity. There was only one false-negative diagnosis for the YOLOv5l model. Furthermore, all four models mislabeled the same case (Figure.5). Another DDH case was mislabeled by the YOLOv5x and YOLOv5m models (Figure.6). Both patients were less than 2 months old, and hip epiphyseal ossification centers could not be detected in the hip radiography images.

Compared to our trained models, hip ultrasonography may be a superior tool for evaluating neonatal (before the appearance of the epiphyseal ossification center) and mild cases of DDH. However, our four trained models could correctly detect DDH in all severe DDH (IHDI grades 3 and 4) cases. Further, in two patients aged less than 4 months in whom the epiphyseal ossification centers could not be detected radiographically, DDH was correctly detected (Figure.7).

All four models did not display satisfactory diagnostic accuracy for IHDI grade 2 DDH. More images may be necessary to train the models to correctly detect young and mild DDH cases. Confidence scores were almost always high, and hips were either correctly labeled or not. For example, in the YOLOv5l model, all confidence scores were greater than 0.7, and 97% of the confidence scores were greater than 0.8. In addition, all three hips with confidence scores less than 0.8 were labeled correctly. Therefore, in our model, the confidence scores did not seem to play an important role in DDH diagnosis. Because the difference between hips with or without DDH was subtle, confidence scores were probably high in most cases. We found a few similar studies. One study reported the diagnostic performance of a convolutional neural network deep learning algorithm for detecting DDH. The sensitivity and specificity of the algorithm were 0.94 and 0.99, respectively.[13] In another study, a deep learning algorithm was applied to measure the acetabular index and center-edge angle and provide IHDI classification, and IHDI classification accuracies ranged from 0.86–0.95.[12] These study results are comparable to our results.

There were several limitations in this study. First, this was a single-center study and DDH is relatively rare; therefore, only 100 DDH hip images were collected. In general, the use of more data would lead to a better deep learning model. More DDH hip images may be required to improve our models. Second, the cutoff point for distinguishing normal and DDH hips may be different from those of other studies. In this study, we included patients who required careful observation. These cases may be considered as overdiagnosis. We believe that for a screening tool, overdiagnosis is better than underdiagnosis.

In conclusion, our deep learning models using YOLOv5 provided accurate diagnostic performances for DDH. We believe our model is a useful diagnostic assistant tool.

Declarations

Author contributions

H.D. designed and performed the research, reviewed the literature, performed data collection, analyzed the data, and wrote the manuscript. J.I. collected the data, reviewed the literature, and reviewed the manuscript. A.K. reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

The datasets analyzed in this study are available from the corresponding author upon a reasonable request.

References

1. Den, H., Ito, J. & Kokaze, A. Epidemiology of developmental dysplasia of the hip: analysis of Japanese national database. *J Epidemiol* (2021).
2. Cashman, J.P., Round, J., Taylor, G. & Clarke, N.M. The natural history of developmental dysplasia of the hip after early supervised treatment in the Pavlik harness. A prospective, longitudinal follow-up. *J Bone Joint Surg Br* **84**, 418-425 (2002).
3. Kotlarsky, P., Haber, R., Bialik, V. & Eidelman, M. Developmental dysplasia of the hip: What has changed in the last 20 years? *World J Orthop* **6**, 886-901 (2015).
4. Wada, I., *et al.* The Pavlik harness in the treatment of developmentally dislocated hips: results of Japanese multicenter studies in 1994 and 2008. *J Orthop Sci* **18**, 749-753 (2013).
5. Chavoshi, M., Mirshahvalad, S.A., Mahdizadeh, M. & Zamani, F. Diagnostic Accuracy of Ultrasonography Method of Graf in the detection of Developmental Dysplasia of the Hip: A Meta-Analysis and Systematic Review. *Arch Bone Jt Surg* **9**, 297-305 (2021).
6. Sahin, S., Akata, E., Sahin, O., Tuncay, C. & Ozkan, H. A novel computer-based method for measuring the acetabular angle on hip radiographs. *Acta Orthop Traumatol Turc* **51**, 155-159 (2017).
7. Xue, Y., Zhang, R., Deng, Y., Chen, K. & Jiang, T. A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. *PLoS One* **12**, e0178992 (2017).

8. Cicero, M., *et al.* Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs. *Invest Radiol* **52**, 281-287 (2017).
9. Mannil, M., von Spiczak, J., Manka, R. & Alkadhi, H. Texture Analysis and Machine Learning for Detecting Myocardial Infarction in Noncontrast Low-Dose Computed Tomography: Unveiling the Invisible. *Invest Radiol* **53**, 338-343 (2018).
10. Lotter, W., *et al.* Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat Med* **27**, 244-249 (2021).
11. Mori, M., *et al.* A deep transfer learning approach for the detection and diagnosis of maxillary sinusitis on panoramic radiographs. *Odontology* **109**, 941-948 (2021).
12. Xu, W., *et al.* A Deep-Learning Aided Diagnostic System in Assessing Developmental Dysplasia of the Hip on Pediatric Pelvic Radiographs. *Front Pediatr* **9**, 785480 (2021).
13. Park, H.S., *et al.* Diagnostic Performance of a New Convolutional Neural Network Algorithm for Detecting Developmental Dysplasia of the Hip on Anteroposterior Radiographs. *Korean J Radiol* **22**, 612-623 (2021).
14. Ultralytics. yolov5. Accessed 6/1, 2022. https://github.com/ultralytics/yolov5
15. Yan, B., Fan, P., Lei, X., Liu, Z. & Yang, F. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. *Remote Sensing* **13**, 1619 (2021).
16. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. in *Proceedings of the IEEE conference on computer vision and pattern recognition* 779-788 (2016).
17. Starr, V. & Ha, B.Y. Imaging update on developmental dysplasia of the hip with the role of MRI. *AJR Am J Roentgenol* **203**, 1324-1335 (2014).
18. Sweeney, J.P., Helms, C.A., Minagi, H. & Louie, K.W. The widened teardrop distance: a plain film indicator of hip joint effusion in adults. *AJR Am J Roentgenol* **149**, 117-119 (1987).
19. Graf, R. The diagnosis of congenital hip-joint dislocation by the ultrasonic Combound treatment. *Arch Orthop Trauma Surg (1978)* **97**, 117-133 (1980).
20. Narayanan, U., *et al.* Reliability of a New Radiographic Classification for Developmental Dysplasia of the Hip. *J Pediatr Orthop* **35**, 478-484 (2015).
21. tzitalin. labelImg. Accessed 6/1, 2022. https://github.com/tzutalin/labelImg.
22. Rahman, T., *et al.* Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray. *Applied Sciences* **10**, 3233 (2020).
23. Mohiyuddin, A., *et al.* Breast Tumor Detection and Classification in Mammogram Images Using Modified YOLOv5 Network. *Comput Math Methods Med* **2022**, 1359019 (2022).
24. Hossain, A., Islam, M.T. & Almutairi, A.F. A deep learning model to classify and detect brain abnormalities in portable microwave based imaging system. *Sci Rep* **12**, 6319 (2022).

25. Mushtaq, M., Akram, M.U., Alghamdi, N.S., Fatima, J. & Masood, R.F. Localization and Edge-Based Segmentation of Lumbar Spine Vertebrae to Identify the Deformities Using Deep Learning Models. *Sensors (Basel)* **22**(2022).
26. Abràmoff, M.D., Lavin, P.T., Birch, M., Shah, N. & Folk, J.C. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* **1**, 39 (2018).

Figures

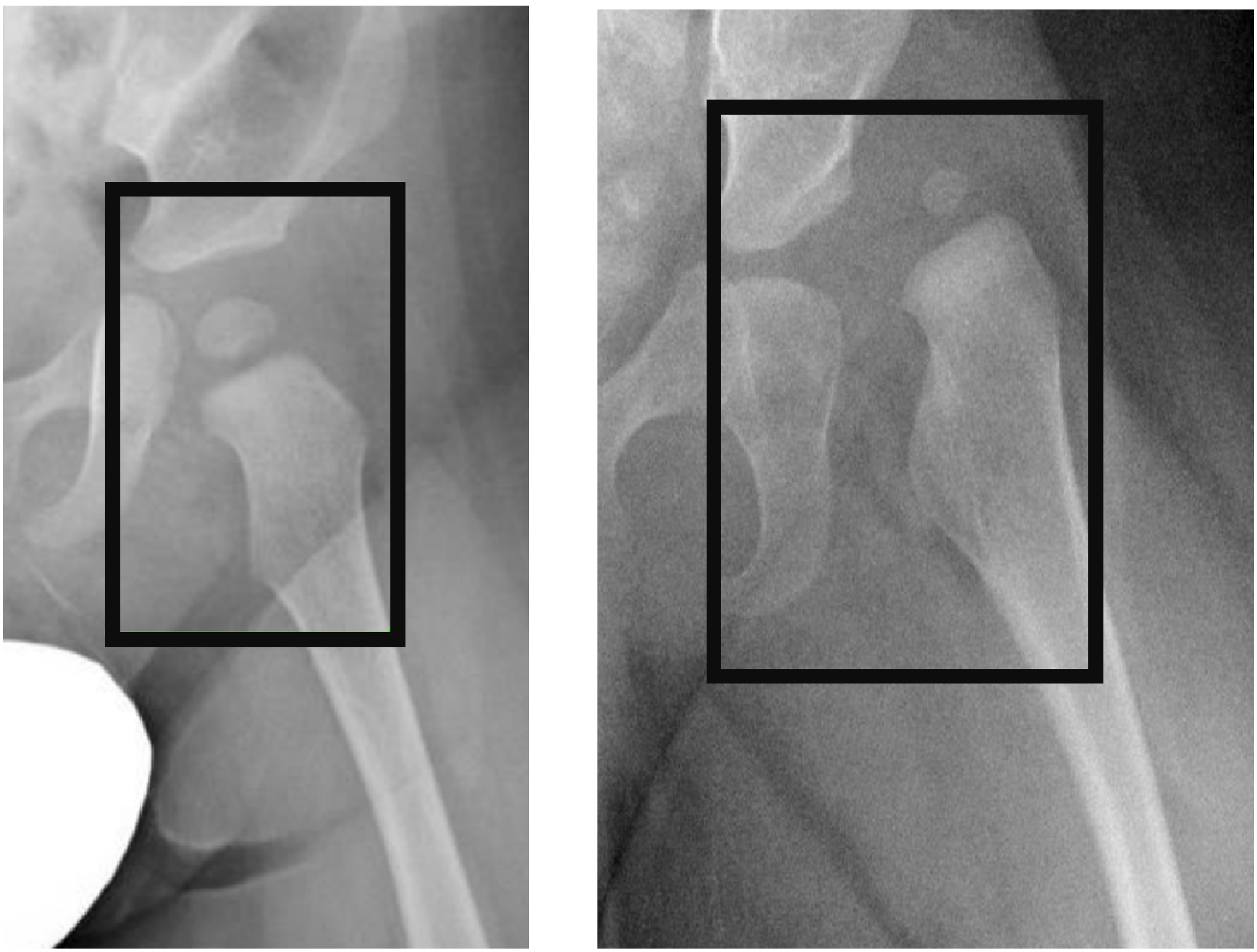

Figure 1

Annotation examples

Examples of bounding boxes for normal (left image) and DDH (right image) hips.

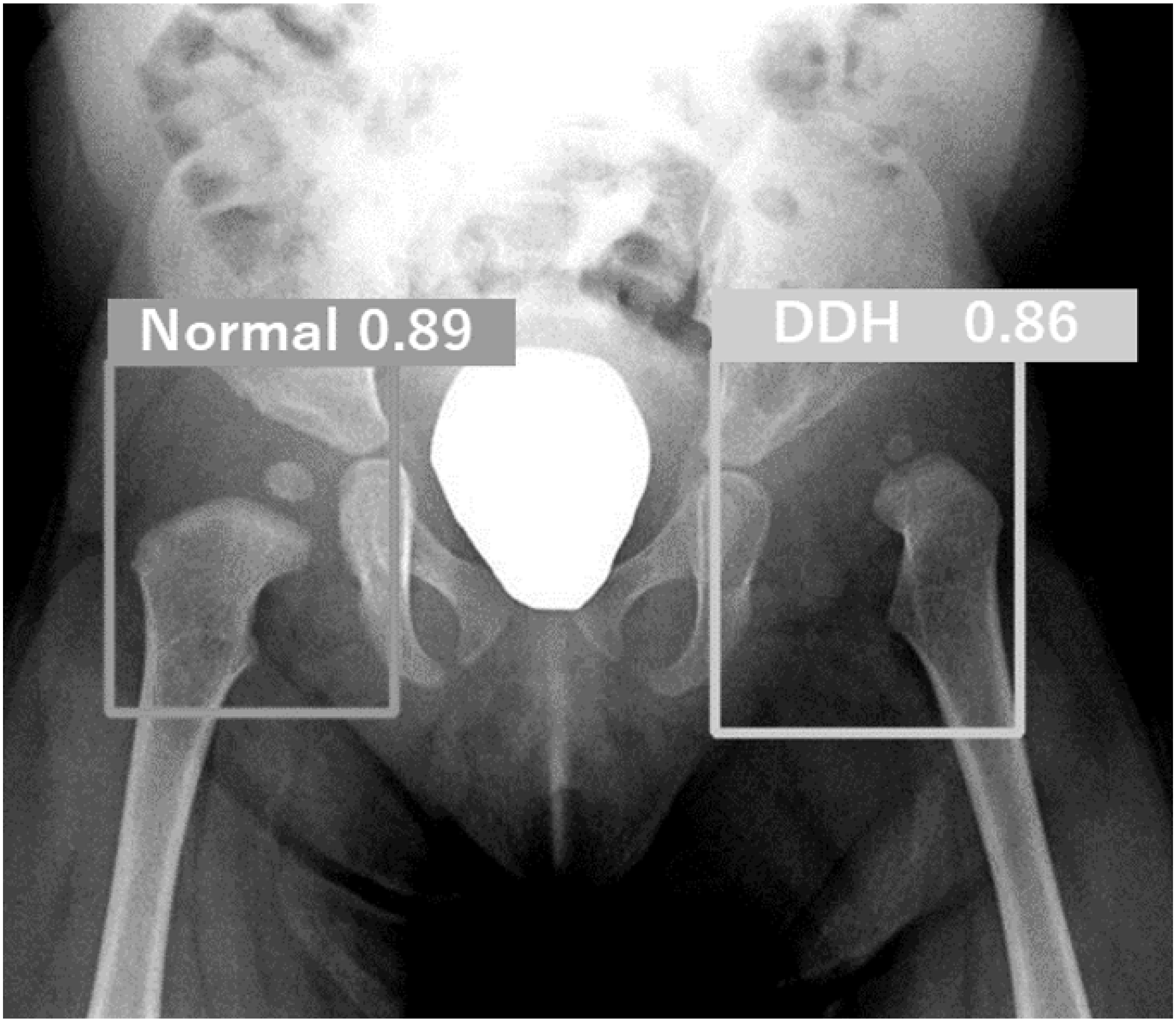

Figure 2

Example of a test image outcome. The right hip was labeled “Normal” and the left hip was labeled “DDH”, with confidence scores. Both hips were correctly labeled.

scores indicate the probability of correct class labeling and how well the predicted box fits

the labeled class.16

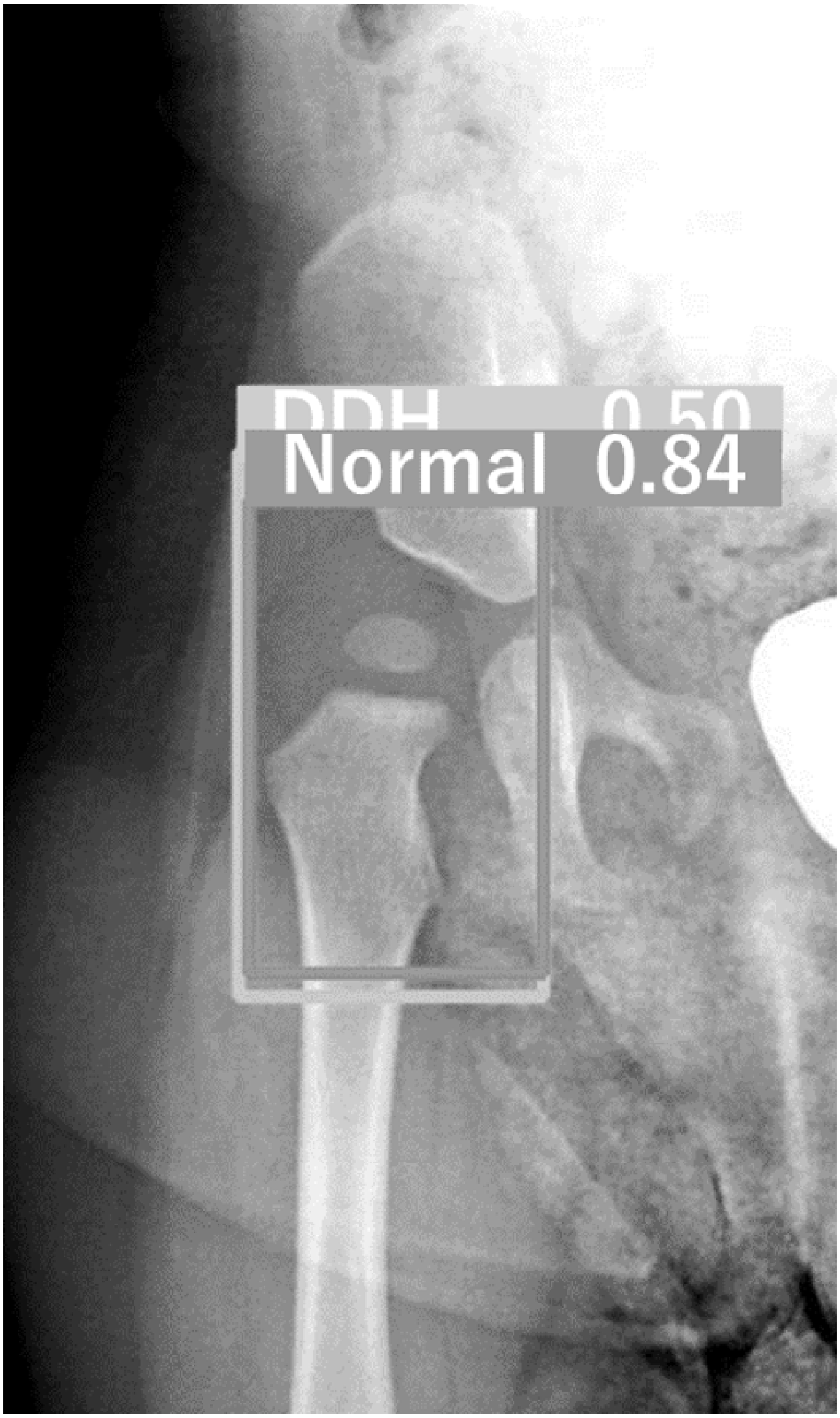

Figure 3

Example of double labeling.

The right hip was labeled as both “Normal” and “DDH”.

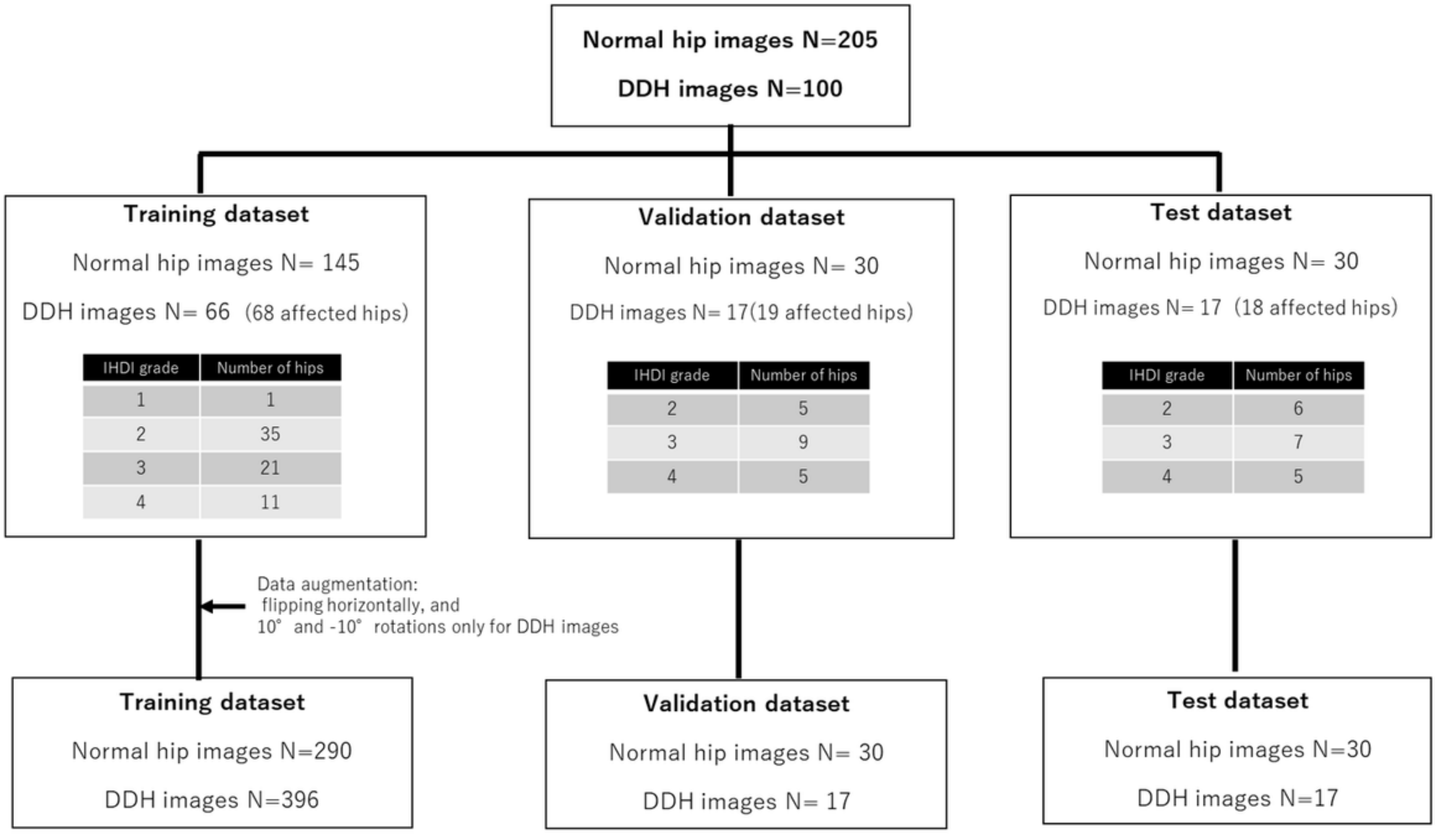

Figure 4

Flowchart of data preparation.

DDH: developmental dysplasia of the hip

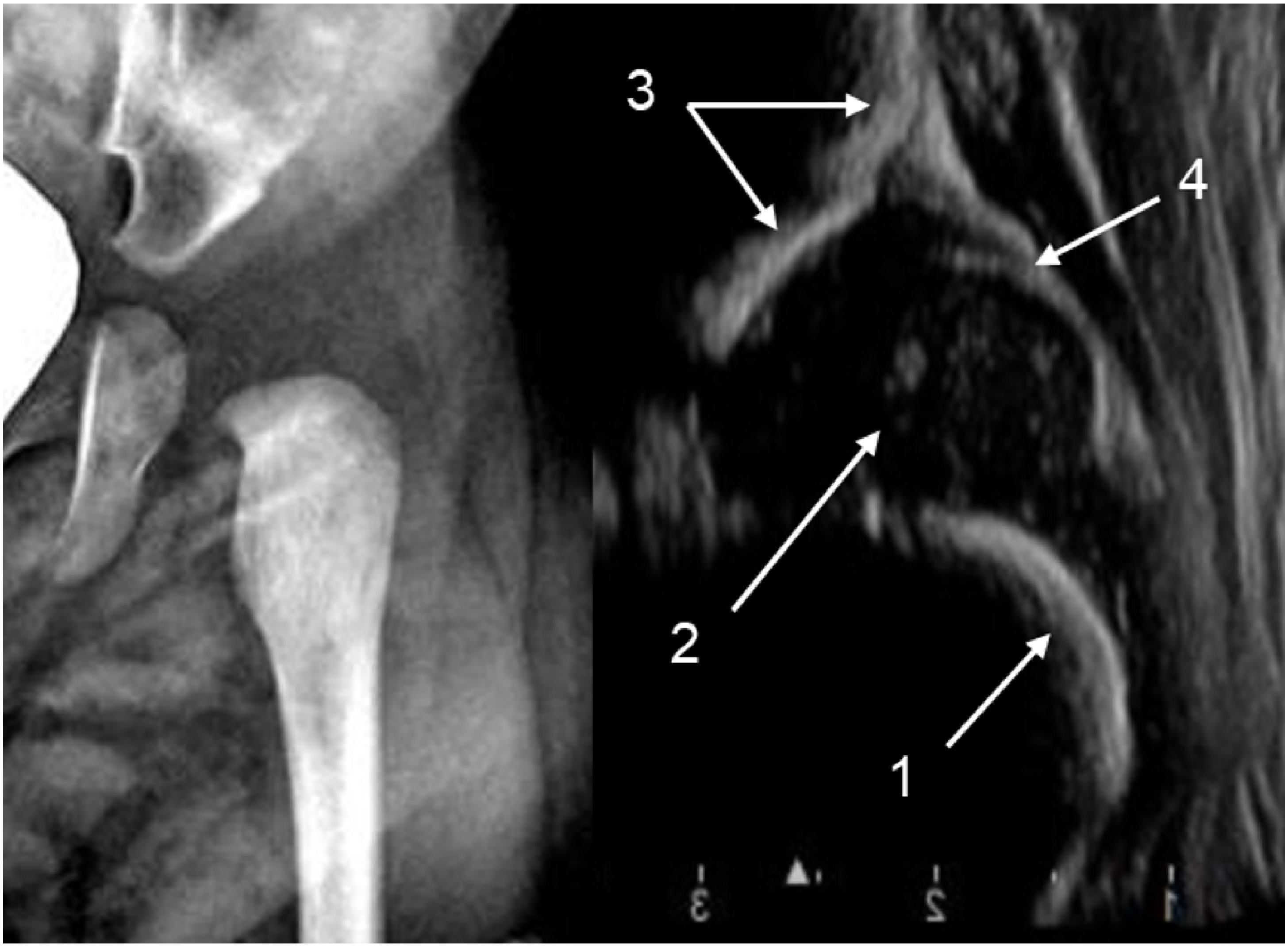

Figure 5

Hip anteroposterior radiography image (left) and hip ultrasonography image of the same hip (right). The patient was a 1-month-old boy. IHDI grade 2, Graf type D.

Anatomical interpretation of the ultrasonography image: 1, bony part of the femoral neck. 2, cartilaginous femoral head. 3, bony part of acetabular roof. 4, acetabular labrum.

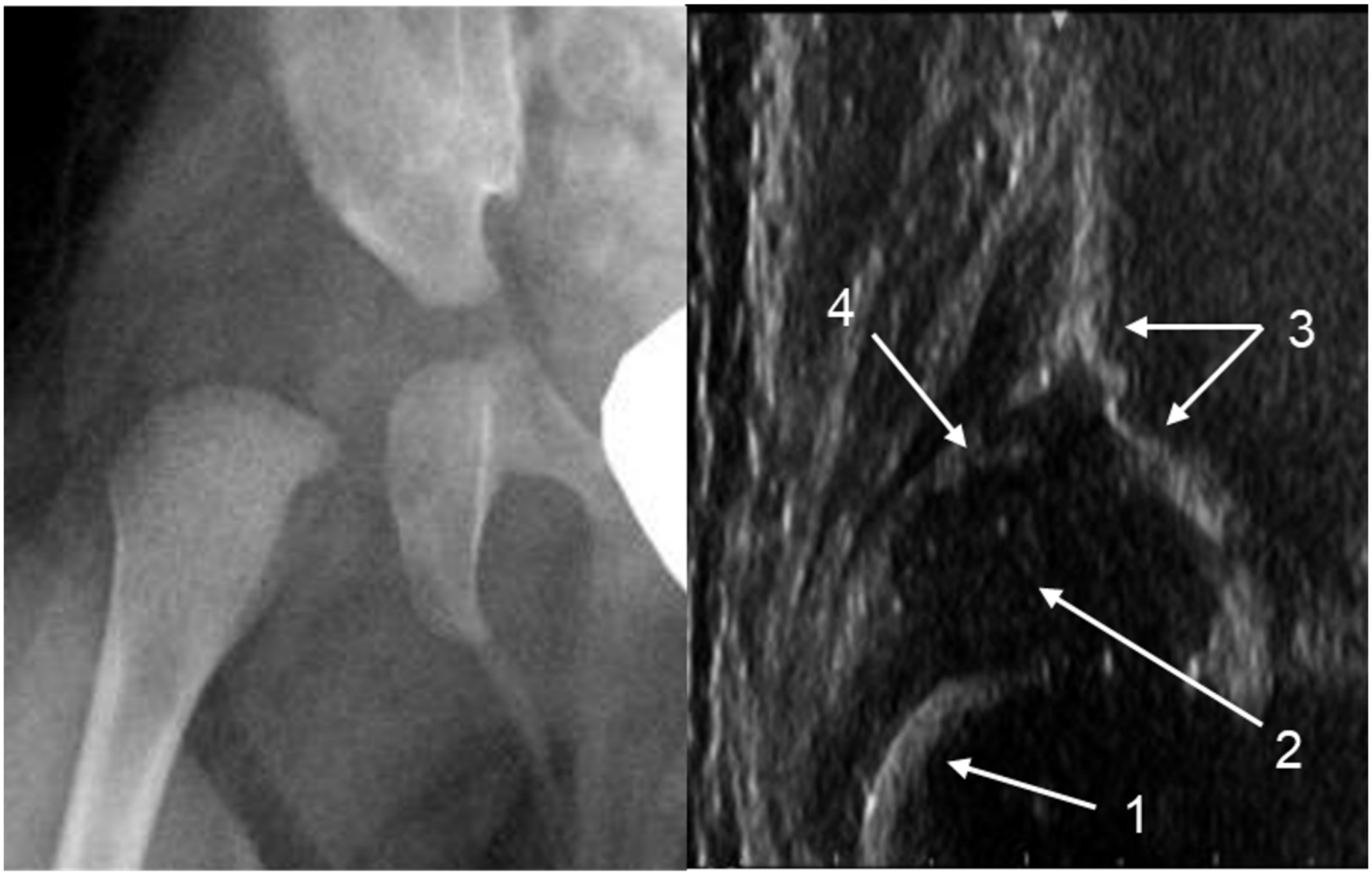

Figure 6

Hip anteroposterior radiography image (left) and hip ultrasonography image of the same hip (right).

The patient was a 2-month-old girl. IHDI grade 2, Graf type D.

Anatomical interpretation of the ultrasonography image: 1, bony part of the femoral neck. 2, cartilaginous femoral head. 3, bony part of acetabular roof. 4, acetabular labrum.

Figure 7

A 3-month-old girl with IHDI grade 3 DDH (left) and a 1-month-old girl with IHIDI grade 3 DDH (right).Both hips were correctly labeled as DDH, although epiphyseal ossification centers did not appear.