



OPEN

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

Hiroki Den^{1,2✉}, Junichi Ito¹ & Akatsuki Kokaze²

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. 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 the "You Only Look Once" v5 (YOLOv5) and single shot multi-box detector (SSD). A total of 305 anteroposterior hip radiography images (205 normal and 100 DDH hip images) were collected. Of these, 30 normal and 17 DDH hip images were used as the test dataset. The sensitivity and the specificity of our best YOLOv5 model (YOLOv5l) were 0.94 (95% confidence interval [CI] 0.73–1.00) and 0.96 (95% CI 0.89–0.99), respectively. This model also outperformed the SSD model. This is the first study to establish a model for detecting DDH using YOLOv5. Our deep learning model provides good diagnostic performance for DDH. We believe our model is a useful diagnostic assistant tool.

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 Japan¹. Early detection and treatment of DDH-related dislocations is highly effective, with a > 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 are 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 is concerning¹.

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 performance⁵. Therefore, it may be difficult to popularize the hip ultrasonography technique for physicians who conduct hip screening in infants. Hip radiography is a convenient diagnostic tool since it is available in most hospitals and clinics. However, the diagnostic accuracy of hip radiography for DDH varies according to the interpreter's experience⁶. DDH is a relatively rare disease; in fact, a Japanese study revealed that 32% of orthopedic doctors were yet to encounter a patient with DDH in their career as at the time of the study¹. A tool that can help inexperienced doctors diagnose DDH may be necessary and can reduce the percentage of late diagnoses.

Deep learning technology has rapidly progressed and is widely used to detect or classify objects on images in many fields, such as face 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 used¹⁴. Object detection methods include one- and two-stage

¹Department of Orthopaedic Surgery, National Rehabilitation Center for Children with Disabilities, 1-1-10 Komone, Itabashi-ku, Tokyo 173-0037, Japan. ²Department of Hygiene, Public Health, and Preventative Medicine, Showa University School of Medicine, 1-5-8 Hatanodai, Shinagawa-ku, Tokyo 142-8555, Japan. ✉email: denh@med.showa-u.ac.jp

approaches. The one-stage object detection methods mainly used are the YOLO series and single shot multi-box detector (SSD), while the faster R-CNN is a widely adopted two-stage object detection method^{15,16}. The one-stage methods can complete both classification and position detection tasks simultaneously, and are therefore popular because of their speed and accuracy^{17,18}. The YOLO series is particularly renowned for its flexible structure, and many researchers have conducted object detection using improved algorithms based on the YOLO algorithm. YOLOv5 was developed in 2020, and it has been used in various applications to detect traffic signals, people, parking meters, and animals^{19–21}.

The aim of this study was to use the YOLOv5 and SSD to develop deep learning models that can be used to distinguish between a normal and DDH hip, using hip radiography images. We also sought to validate the diagnostic performances of the developed models.

Methods

This retrospective study was approved by the Institutional Review Board of the National Rehabilitation Center for Children with Disabilities (approval no. 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. Radiography images were evaluated for appropriate standard AP view of the hip. The rotation quotients were obtained by dividing the transverse diameter of the obturator foramen on the right side by that of the left side. Radiography images with rotation quotients ranging from 0.5 to 2.0 were considered appropriate, while those taken in an inappropriate position were excluded^{22,23}. For patients who underwent several hip radiographs, the oldest image in the time series was selected, because the AP view hip radiographs were taken just once by the age of 12 months, in most patients. All patients were screened by the same experienced pediatric hip specialist for DDH. 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 line²⁴, (c) widened teardrop distance compared to that of the other side²⁵, (d) delayed femoral head ossification compared to that of the other side, (e) high acetabular index (> 30), and (f) dulled edge of the acetabulum. The Graf method was used for hip ultrasonography²⁶. 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 ages²⁷.

Patients with DDH of IHDI grade 2 or worse and/or type 2c or worse Graf classification based on the hip ultrasonography images were generally considered to have DDH, as they require close monitoring²⁶.

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% of each normal and DDH images were randomly but equally distributed between the 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 conducted using LabImg version 1.8.1 with reference to the diagnosis made by the experienced pediatric hip specialist (J.I.)²⁸. Object bounding boxes were drawn to encompass the entire hip joint, based on 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 (Fig. 1). Image annotations for the normal and DDH hips were performed in similar manner, based on the criteria above. Normal hips are labeled as “Normal” and DDH hips as “DDH”. Following the image annotation process, all images were reviewed by the senior orthopedic doctor (J.I.).

YOLOv5 algorithm. Transfer learning, a technique in which a well-trained model from a large dataset is used for applications of interest with a small dataset, was performed using YOLOv5²⁹. Therefore, transfer learning can reduce the requirement of large datasets. YOLOv5 contains four different models: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The main difference between these models is the amount of their 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 parameters¹⁹. 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 using our new datasets. YOLOv5 is composed of three types of loss functions: box loss, object loss, and class loss.

The box loss represents the localization loss of the positive predictions, and used complete IOU (CIOU). The object loss represents the CIOU loss of the detection and true boxes, defined as binary cross entropy loss. The class loss represents the classification loss of the positive predictions, defined as a binary cross-entropy loss^{30–32}. Our models used the stochastic gradient descent (SGD) as an optimizer, with a learning rate of 0.01, mini-batch size of 32, and 100 epochs, for the training.

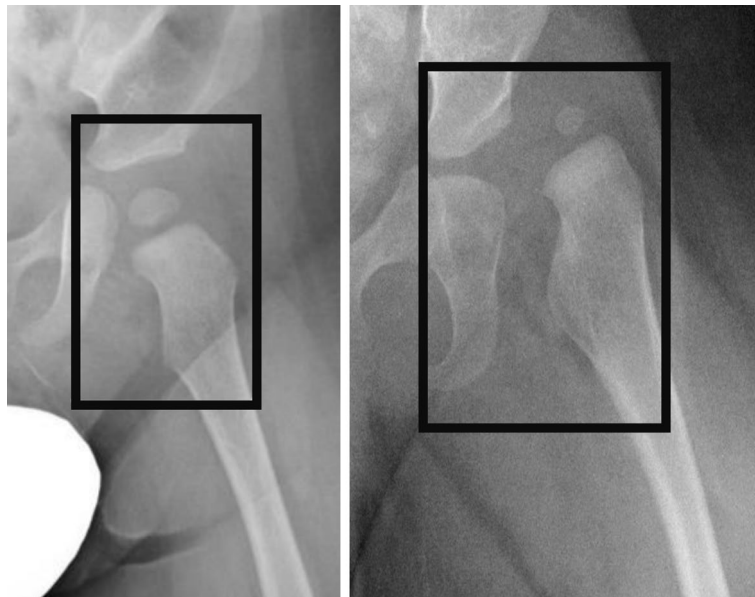


Figure 1. Annotation examples. Examples of bounding boxes for normal (left image) and DDH (right image) hips. *DDH* Developmental dysplasia of the hip.

SSD algorithm. The backbone network of SSD adopts the Visual Geometry Group (VGG) 16 feature extraction network, which is widely used in feature extraction, and the extra feature layers added after the backbone network is used to achieve the multi-scale detections^{15,33}. We also applied transfer learning by loading the pre-trained VGG16 model³⁴. The loss function of the SSD model consists of the localization and confidence losses. The overall loss function is the weighted sum of these two¹⁵. The localization loss is a smooth L1 loss between the predicted box and the ground truth box parameters, while the confidence loss is the softmax loss over multiple classes confidences¹⁵. Images were transformed to 300×300 pixels in size and the model used SGD as the optimizer with a learning rate of 0.01, mini-batch size of 32, and 200 epochs for the training. The same training and validation datasets used for YOLOv5 models were used.

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 class-specific confidence scores (Fig. 2).

The confidence scores indicate how the level of confidence that the box contains an object and how well the predicted box fitted the object. Class-specific confidence scores were calculated by multiplying the class probabilities and individual box confidence scores. These scores indicate the probability of a correct class labeling and how well the predicted box fitted the labeled class^{20,30}.

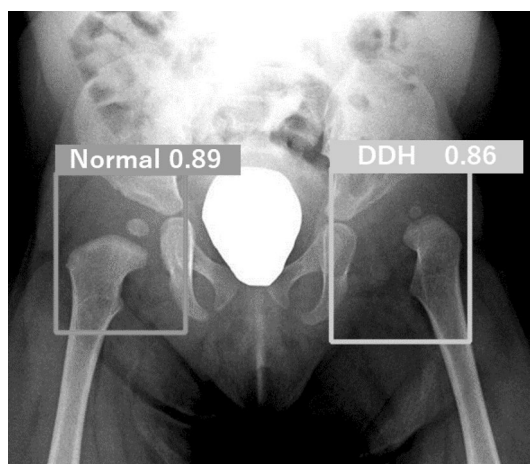


Figure 2. Example of a test image outcome. The right hip was labeled “Normal” and the left hip was labeled “DDH”, with class-specific confidence scores. Both hips were correctly labeled. *DDH* Developmental dysplasia of the hip.

The test dataset was evaluated using the trained models, at a 0.5 confidence score threshold. If both “Normal” and “DDH” were labeled on the same hip, the hip evaluation was considered invalid (Fig. 3). Accuracy, 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. Of the normal hip group, 171 patients (83.4%) underwent hip ultrasonography, while in the DDH group, 87 patients (87%) received hip ultrasonography to confirm the diagnosis. In particular, all IHDI grade 2 patients, who were under 6 months of age, underwent hip ultrasonography. As a result, 396 DDH and 290 normal images were utilized for training, while 30 normal and 17 DDH hip images were used for testing (Fig. 4). No age bias was observed between learning, validation, and test datasets.

The basic characteristics of the datasets are shown in Table 1. One case had an IHDI grade of 1 in the DDH group, but the patient’s Graf classification was type 2c. Therefore, this case was categorized as DDH. The training and validation loss curves are shown in Fig. 5.

Among the four YOLO 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, as well as in the SSD model. Table 2 shows the diagnostic performance of each model for the test dataset.

In terms of accuracy, all the models except YOLOv5s showed similar results.

For false negatives, the YOLOv5s and YOLOv5l models mislabeled only one hip, and the same hip was found to have been mislabeled by the models.

All severe DDH cases (IHDI grades 3 and 4) were correctly labeled as “DDH” in all YOLO models, but one IHDI grade 3 case was mislabeled in the SSD model. The labeling of mild DDH cases (IHDI grade 2) varied, and no model labeled all IHDI grade 2 cases correctly (Table 3).

Discussion

We developed deep learning models for detecting DDH using hip radiography images in the AP view. To the best of our knowledge, this is the first study to achieve this using YOLOv5 and SSD. In addition, using the transfer learning technique, a good model could be constructed with a relatively small dataset. The 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, and that the image outcomes are easy to understand.

The disadvantage of YOLOv5, in general, is that it is not optimized to detect small objects³⁵. Because the hips are large enough on hip radiography images, this drawback did not seem to affect our results. Studies that used YOLOv5 models have been reported in various medical fields³⁶. One study used YOLOv5 models to detect brain abnormalities, and another study sought to detect lumbar spine deformities using YOLOv5 models^{37,38}. As a confirmatory 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%, respectively⁵. 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.



Figure 3. Example of double labeling. The right hip was labeled as both “Normal” and “DDH”. DDH Developmental dysplasia of the hip.

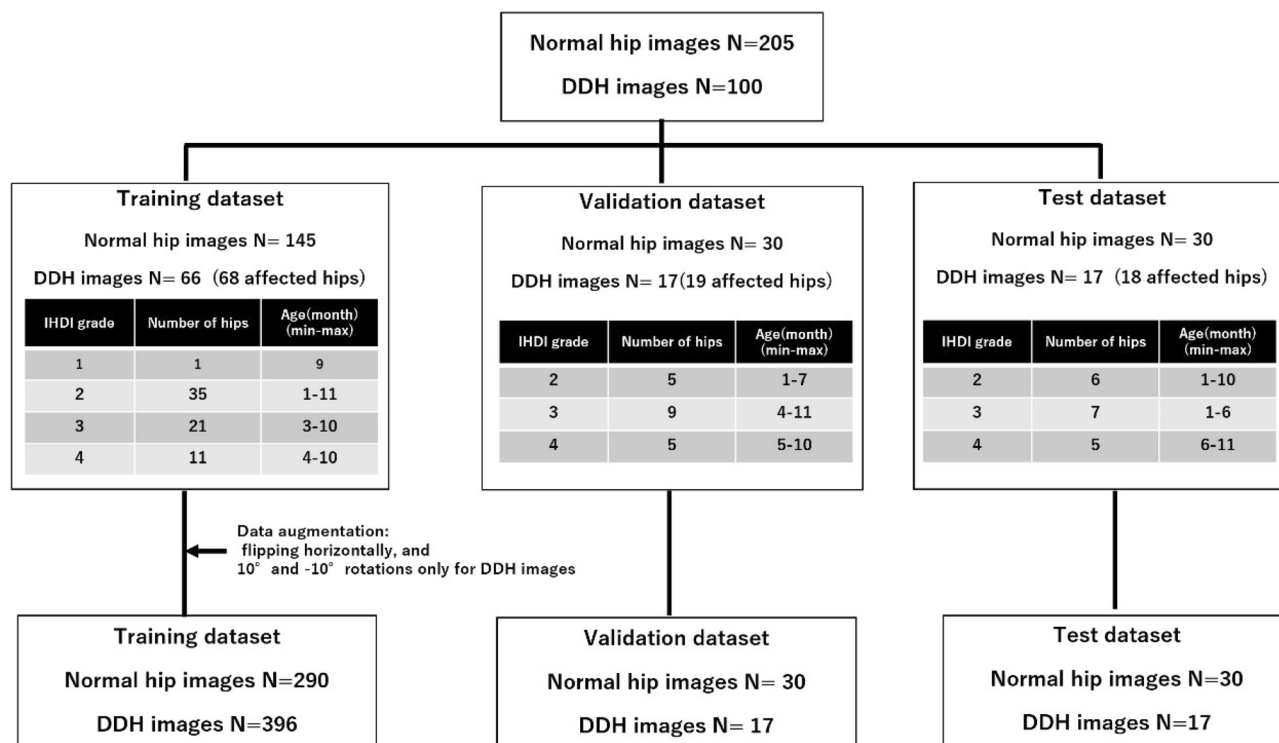


Figure 4. Flowchart of data preparation. DDH: developmental dysplasia of the hip.

		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	

Table 1. Basic characteristics. DDH Developmental dysplasia of the hip, IHDI International Hip Dysplasia Institute.

The strength of a deep learning model compared to hip ultrasonography screening is that it can be applied in clinics that do not have access to a pediatric hip specialist. 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 medicine³⁹.

A study reported the sensitivity and specificity of the AP view hip radiography images, reviewed by radiologists for the detection of DDH. The Sensitivity and specificity of 96.0% and 89.0%, respectively, were indicated for a radiologist with a 5-year radiology experience, including pediatric radiology, while 84.0% and 85.8% were reported for one with a 3-year radiology experience, but without any pediatric radiology experience¹³. 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 YOLO models, the YOLOv5l and YOLOv5s models had the highest sensitivity (0.94). In contrast, YOLOv5x had the highest specificity (0.99) while YOLOv5s had the lowest specificity (0.93). Overall, of the four YOLO models, YOLOv5l was considered the best screening tool because of its high sensitivity and specificity. The diagnostic performance of the SSD model is equivalent to that of the YOLOv5m model and considering that the one IHDI grade 3 DDH case was mislabeled, the YOLOv5l model also seems to be superior to the SSD model. One study reported the comparison of traffic

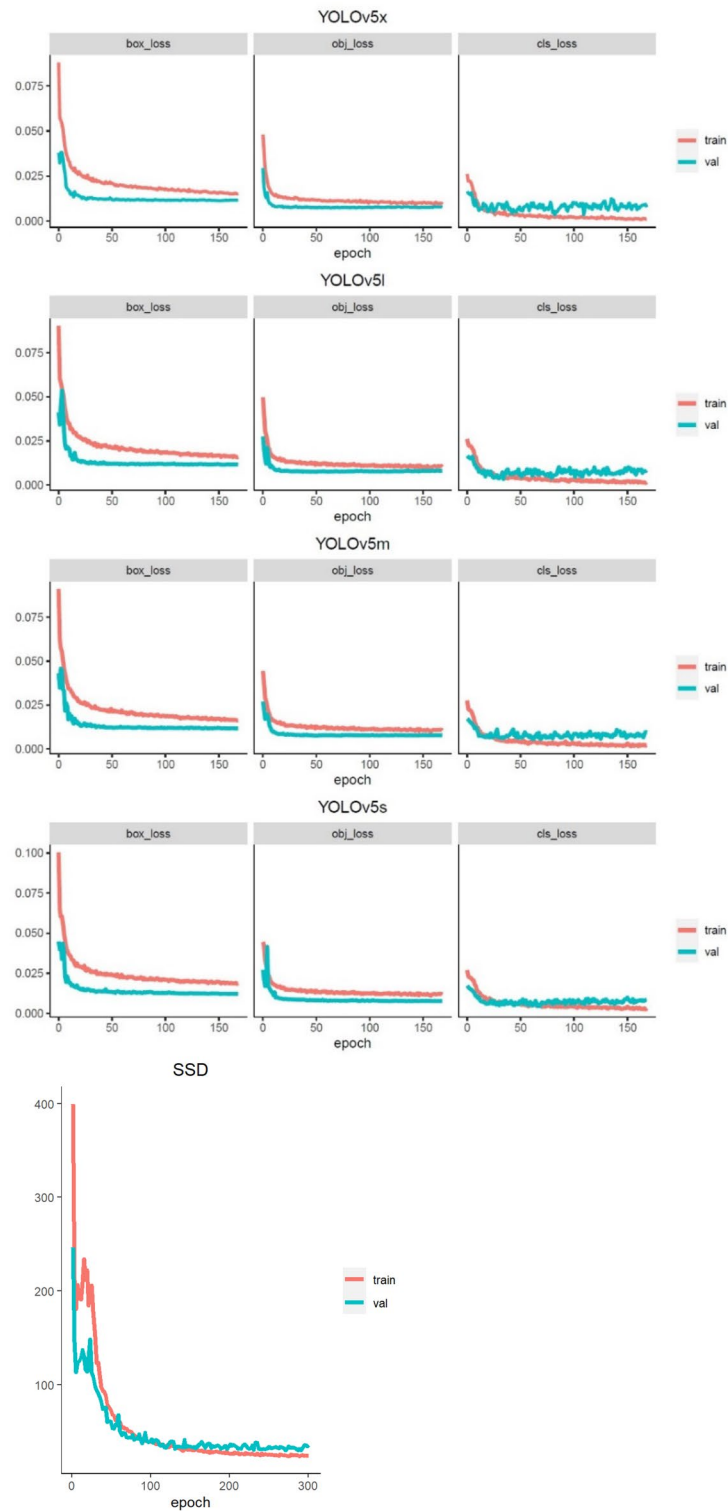


Figure 5. Training and validation curves for the YOLO and SSD models. Box_loss: box loss. Obj_loss: object loss. Cls_loss: class loss. SSD Single shot multi-box detector, YOLO “You Only Look Once”.

sign recognition performance between YOLOv5 and SSD using their own dataset, and YOLOv5 outperformed SSD in terms of the mean average precision (mAP)⁴⁰. Another study reported the comparison of pavement crack detection performance between YOLOv4 and SSD, and the YOLOv4 also outperformed SSD in terms of mAP¹⁷. These results are similar to ours.

	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
v5x	0.96 (0.89–0.99)	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	0.96 (0.89–0.99)	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	0.96 (0.89–0.99)	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 (0.87–0.98)	0.94 (0.73–1.00)	0.93 (0.85–0.98)	0.77 (0.55–0.92)	0.99 (0.93–1.00)
SSD	0.96 (0.89–0.99)	0.89 (0.65–0.99)	0.97 (0.91–1.00)	0.89 (0.65–0.99)	0.97 (0.91–1.00)

Table 2. Diagnostic performances of each model using the test dataset. *CI* Confidence interval, *PPV* Positive predictive value, *NPV* Negative predictive value, *SSD* Single shot multi-box detector.

IHDI grade	Total	N (%)				
		YOLOv5x	YOLOv5l	YOLOv5m	YOLOv5s	SSD
4	5 hips	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)
3	7 hips	7 (100)	7 (100)	7 (100)	7 (100)	6 (85.7)
2	6 hips	3 (50.0)	5 (83.3)	4 (66.7)	5 (83.3)	5 (83.3)

Table 3. DDH hips correctly labeled “DDH”, according to IHDI grade. *DDH* Developmental dysplasia of the hip, *IDHI* International Hip Dysplasia Institute, *SSD* Single shot multi-box detector, *YOLO* “You Only Look Once”.

There were four hips labeled as both “Normal” and “DDH” in the YOLOv5x and the YOLOv5s models. As a screening tool, high sensitivity is crucial. Therefore, in practice, these double-labeled cases should be treated as “positive”, and further examination should be performed.

There was only one false-negative diagnosis for the YOLOv5l model. Furthermore, all four YOLO models mislabeled the same case, which only the SSD model correctly labeled (Fig. 6). Another DDH case mislabeled by the YOLOv5x, YOLOv5m, and SSD models is shown in Fig. 7. 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 YOLO-trained models could correctly detect DDH in all severe DDH (IHDI grades 3 and 4) cases. Furthermore, in two patients aged less than 4 months in whom the epiphyseal ossification centers could not be detected radiographically, DDH was correctly detected (Fig. 8).

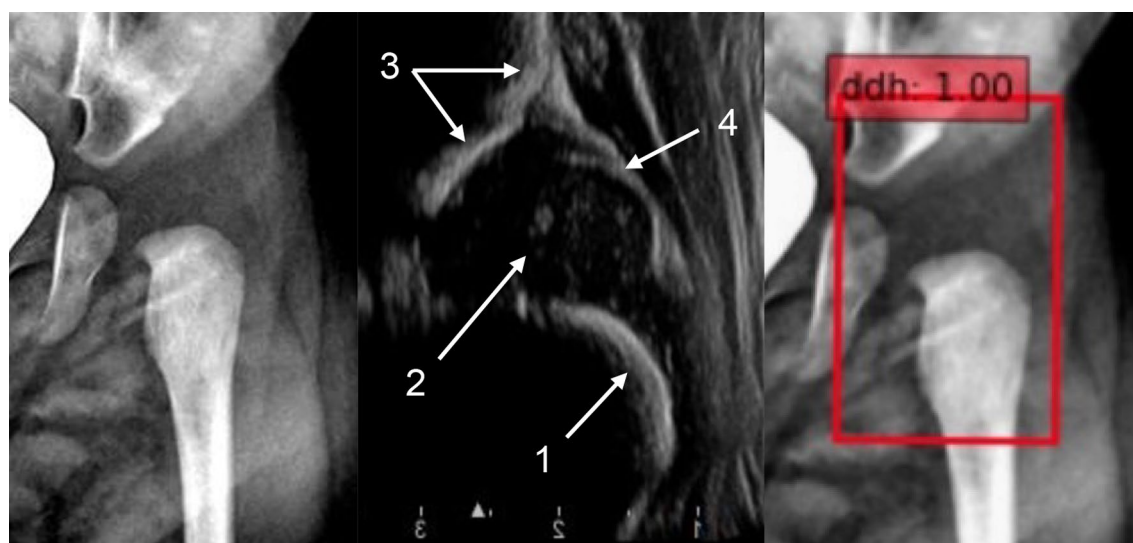


Figure 6. Hip anteroposterior radiography image (left), hip ultrasonography image of the same hip (middle), and the SSD model image output (right). The patient was a 1-month-old boy. IHDI grade 2, Graf type D. Anatomical interpretation of the ultrasonography image: 1, bony part of the femoral neck. 2, cartilaginous femoral head. 3, bony part of acetabular roof. 4, acetabular labrum. *IDHI* International Hip Dysplasia Institute, *SSD* Single shot multi-box detector.

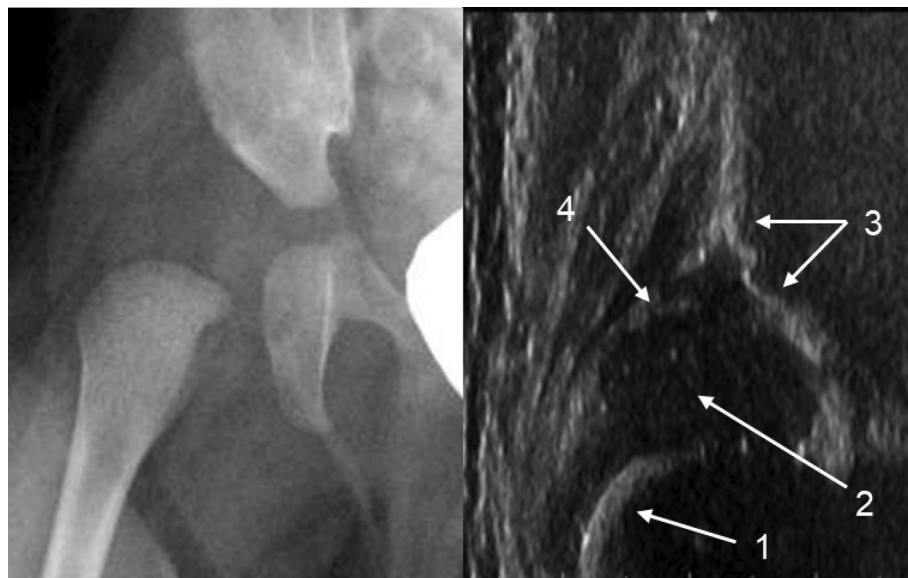


Figure 7. Hip anteroposterior radiography image (left) and hip ultrasonography image of the same hip (right). The patient was a 2-month-old girl. IHDI grade 2, Graf type D. Anatomical interpretation of the ultrasonography image: 1, bony part of the femoral neck. 2, cartilaginous femoral head. 3, bony part of acetabular roof. 4, acetabular labrum. *IDHI* International Hip Dysplasia Institute.



Figure 8. A 3-month-old girl with IHDI grade 3 DDH (left) and a 1-month-old girl with IHDI grade 3 DDH (right). Both hips were correctly labeled as DDH, although epiphyseal ossification centers did not appear. *DDH* Developmental dysplasia of the hip, *IDHI* International Hip Dysplasia Institute.

All trained models did not display satisfactory diagnostic accuracy for IHDI grade 2 DDH. However, as mentioned above, there was only one false-negative diagnosis for the YOLOv5l model, whereas, the SSD model could correctly label the same case. If hips labeled as “DDH” in either YOLOv5l or SSD model had been treated as “DDH”, the sensitivity of the test dataset would have achieved 1.00 while the specificity was still acceptable (0.93 [95% CI 0.85–0.98]). As such, our trained models could be a useful diagnostic assistant tool for any IHDI grade of DDH.

Even though YOLOv5x had the highest number of parameters, the YOLOv5l model outperformed the YOLOv5x model in terms of sensitivity. Similar results were found in another study that developed models to detect personal protective equipment using YOLOv5²¹. In that study, the YOLOv5m model outperformed the YOLOv5l and YOLOv5x models in terms of precision and recall. Class-specific confidence scores were almost always high whether the hips were 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 models, the class-specific 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, respectively¹³. 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 to 0.95¹². These study results are comparable to our results.

This study has several limitations. First, it was a single-center study with a relatively small sample size of only 100 DDH hip images were collected. Using more data would improve the deep learning model, and further studies with larger datasets are needed to evaluate different ages and grades of DDH more accurately. Second, in this study, we only developed the models using the one-stage object detectors, and studies using the two-stage object detectors such as Faster R-CNN should also be considered.

Third, the same physician made all diagnoses, which may have introduced bias. However, the diagnoses were based not only on radiographic findings but also on physical examination and hip ultrasonography findings, suggesting sufficient diagnostic accuracy. Finally, patients who required careful observation were included in the DDH group in this study, which may be considered an overdiagnosis. We believe that with 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.

Data availability

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

Received: 15 September 2022; Accepted: 20 April 2023

Published online: 24 April 2023

References

- Den, H., Ito, J. & Kokaze, A. Epidemiology of developmental dysplasia of the hip: Analysis of Japanese national database. *J. Epidemiol.* **33**, 186–192 (2021).
- 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 Jt. Surg. Br.* **84**, 418–425 (2002).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- ultralytics. *yolov5*. Accessed 6/1, 2022. <https://github.com/tzutalin/labelimg>.
- Liu, W. *et al.* SSD: Single shot multibox detector. In *Computer Vision—ECCV 2016* (eds Leibe, B. *et al.*) 21–37 (Springer International Publishing, 2016).
- Ren, S., He, K., Girshick, R. & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **28** (2015).
- Yan, K. & Zhang, Z. Automated asphalt highway pavement crack detection based on deformable single shot multi-box detector under a complex environment. *IEEE Access.* **9**, 150925–150938 (2021).

18. Nepal, U. & Eslamiat, H. Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs. *Sensors* **22**, 464 (2022).
19. 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 Sens.* **13**, 1619 (2021).
20. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* 779–788 (2016).
21. Ali, L. *et al.* Development of YOLOv5-based real-time smart monitoring system for increasing lab safety awareness in educational institutions. *Sensors*. **22**, 8820 (2022).
22. Lim, S. J. & Park, Y. S. Plain radiography of the hip: A review of radiographic techniques and image features. *Hip Pelvis*. **27**, 125–134 (2015).
23. Tonnis, D. Normal values of the hip joint for the evaluation of X-rays in children and adults. *Clin. Orthop. Relat. Res.* 39–47 (1976).
24. 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).
25. 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).
26. Graf, R. The diagnosis of congenital hip-joint dislocation by the ultrasonic compound treatment. *Arch. Orthop. Trauma Surg.* **1978**(97), 117–133 (1980).
27. Narayanan, U. *et al.* Reliability of a new radiographic classification for developmental dysplasia of the hip. *J. Pediatr. Orthop.* **35**, 478–484 (2015).
28. tzatalin. *labelimg*. Accessed 6/1, 2022. <https://github.com/tzatalin/labelImg>.
29. Rahman, T. *et al.* Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. *Appl. Sci.* **10**, 3233 (2020).
30. Yu, Z., Zhang, L., Gao, X., Huang, Y. & Liu, X. Research on non-pooling YOLOv5 based algorithm for the recognition of randomly distributed multiple types of parts. *Sensors* **22**, 9335 (2022).
31. Zhang, Y. *et al.* Real-time vehicle detection based on improved YOLOv5. *Sustainability* **14**, 12274 (2022).
32. Wang, Z., Wu, L., Li, T. & Shi, P. A smoke detection model based on improved YOLOv5. *Mathematics* **10**, 1190 (2022).
33. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. <http://arxiv.org/abs/1409.1556> (2014).
34. VGG16_features. Accessed 1/31,2023. https://s3.amazonaws.com/amdegroot-models/vgg16_reducedfc.pth.
35. Benjumea, A., Teeti, I., Cuzzolin, F. & Bradley, A. YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles. ArXiv. <http://arxiv.org/abs/2112.11798> (2021).
36. Mohiyuddin, A. *et al.* Breast tumor detection and classification in mammogram images using modified YOLOv5 network. *Comput. Math. Methods Med.* **2022**, 1359019 (2022).
37. 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).
38. 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**, 1547 (2022).
39. 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).
40. Zhu, Y. & Yan, W. Q. Traffic sign recognition based on deep learning. *Multimed. Tools Appl.* **81**, 17779–17791 (2022).

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.

Additional information

Correspondence and requests for materials should be addressed to H.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023