

Prior Knowledge and Adaptive Hyperparameter Optimization with Lightweight HRNet for Clinical Decision-Making of DDH IHDI Classification

Shuai Xue

Shanghai Ocean University

Xiaowei Yu

yuxw@sjtu.edu.cn

Shanghai Sixth People's Hospital

Research Article

Keywords: PCL-HRNet, Prior loss, Lightweight pipeline, QPSO-AWA, IHDI classification

Posted Date: August 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10532345/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Prior Knowledge and Adaptive Hyperparameter Optimization with Lightweight HRNet for Clinical Decision-Making of DDH IHDI Classification

Shuai Xue · Xiaowei Yu (Corresponding Author)

Abstract

Developmental Dysplasia of the Hip (DDH) is a prevalent congenital pediatric musculoskeletal deformity, and accurate early International Hip Dysplasia Institute (IHDI) grading is essential for timely clinical intervention and favorable prognosis. A novel framework for reliable IHDI classification of DDH from pelvic radiographs using a prior-constrained lightweight HRNet (PCL-HRNet) is presented to support clinical decision-making in primary care. This method builds multi-dimensional geometric prior loss to constrain keypoint estimation, and then implements lightweight optimization through knowledge distillation, structured channel pruning and INT8 static quantization sequentially. The compressed network is globally optimized by the Quantum-behaved Particle Swarm Optimization with Adaptive Weight Adjustment (QPSO-AWA) algorithm to search for optimal hyperparameters for stable multi-class IHDI grading. Keypoint localization error and anatomical logic error rate are adopted to evaluate the prior constraint performance, while classification performance is measured using Macro-F1, Kappa coefficient, mean distance error (MDE) of keypoints, and mean absolute error (MAE) of acetabular angle. The proposed method reduces the anatomical error rate to 0.24%, and achieves a Macro-F1 of 0.825 with merely 0.33% accuracy loss after lightweight compression. Finally, the proposed PCL-HRNet outperforms mainstream single-baseline deep networks in both diagnostic accuracy and CPU inference efficiency, providing a deployable intelligent decision support tool for DDH screening.

Keywords PCL-HRNet · Prior loss · Lightweight pipeline · QPSO-AWA · IHDI classification

1. Introduction

Developmental Dysplasia of the Hip (DDH) is one of the most common congenital musculoskeletal deformities in children. Studies have shown that its incidence in China is approximately 1‰ to 3‰, which is generally consistent with the epidemiological characteristics of Asian populations [1]. Without timely diagnosis and standardized intervention during infancy and early childhood, DDH will progressively develop into hip pain, claudication, secondary osteoarthritis and even lifelong disability. It not only severely impairs the quality of life of affected children, but also imposes heavy disease burden and public health pressure on families and society. Early screening and accurate classification are core measures to improve the prognosis of DDH and reduce the long-term disability rate. The development of efficient and reliable intelligent auxiliary diagnostic decision-making tools is of important clinical value and social significance for improving the management of children's skeletal health [2].

Anteroposterior (AP) pelvic radiographs serve as the gold-standard imaging modality for DDH diagnosis and classification decision-making in children over 4 months of age. The classification system of the International Hip Dysplasia Institute (IHDI) is currently the most widely used quantitative assessment system in clinical practice. The diagnostic logic of this classification system is entirely based on the spatial relationships of anatomical key points: by

identifying 8 core anatomical landmarks including the superolateral acetabular edge, the center of the Y-shaped cartilage, and the center of the femoral head, Hilgenreiner's line and Perkins' line are drawn to divide quadrants, and classification and grading are finally completed in combination with quantitative indicators such as the acetabular angle, as shown in Fig. 1. However, the accurate early diagnosis of DDH is highly dependent on clinicians' clinical experience and anatomical recognition ability. Inter-observer measurements of the acetabular angle are significantly affected by physician seniority [3]. Therefore, constructing an intelligent auxiliary decision-making system for DDH with high accuracy, interpretability and deployability is a critical pathway to address the insufficient diagnostic capacity in primary care and improve screening efficiency [4].

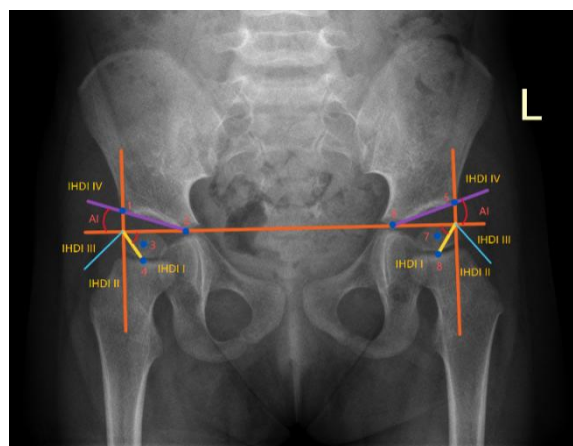


Fig. 1 IHD classification results

In recent years, deep learning technology has made breakthrough progress in the field of medical image analysis, providing a new technical approach for intelligent diagnostic decision-making of DDH [5]. Among them, the technical route based on keypoint detection has become the mainstream research direction of DDH intelligent diagnosis, as its diagnostic workflow is highly consistent with clinicians' decision-making logic and its results are traceable. Xu et al. [6] constructed a three-stage deep learning diagnostic system. They first performed instance segmentation of pelvic bony structures via Mask R-CNN, then accurately extracted key anatomical landmarks of the hip joint using HRNet, and finally achieved automated diagnosis of Tönnis classification and IHD classification. The diagnostic accuracy of IHD classification reached 86% to 95%, and the diagnostic consistency was close to the level of senior orthopedic surgeons. Jan et al. [7] proposed a technical framework integrating instance segmentation and keypoint localization. After segmenting the pelvic region based on Mask R-CNN, they combined keypoint detection to realize automated and rapid measurement of the infant acetabular index angle, with a mean angle measurement error of only 2.402° , which greatly improved the efficiency of quantitative DDH assessment. Chen et al. [8] incorporated multimodal clinical information such as age and gender into image input channels, and enhanced and optimized the YOLOv5 detection framework. They achieved a DDH diagnostic accuracy of 86.7% on a large-sample dataset containing complex scenarios such as skeletal deformities and postoperative internal fixation, significantly improving the generalization performance of the model in real clinical scenarios. The above studies have fully verified the application potential of deep learning technology in DDH imaging diagnosis, laying a research foundation for the realization of automated screening and early intervention.

1.1 Research Limitations and Motivation

Despite the remarkable progress achieved by existing methods, from the perspective of knowledge engineering and clinical decision-making implementation, current intelligent diagnostic models for DDH still have several limitations that restrict their translation from laboratory research to routine clinical application in primary care settings.

Most traditional DDH keypoint detection models are trained using the mean squared error (MSE) loss function. This loss independently optimizes the pixel-level localization error of each keypoint while ignoring the inherent geometric

constraints and physiological laws in the anatomical structure of the hip joint [9][10]. Such a learning paradigm tends to produce prediction results that violate basic anatomical logic, which not only directly reduces the accuracy of classification decisions but also undermines the credibility of the decision-making system due to its black-box nature [11][12]. Therefore, embedding clinical prior knowledge into the model training process to constrain the model's output space and improve decision interpretability and clinical credibility is a core problem to be urgently addressed in the field of DDH intelligent decision-making [13].

High-performance models generally suffer from high computational complexity and high deployment thresholds, making them difficult to adapt to decision-making scenarios in primary hospitals. Currently, the best-performing models in the DDH field usually have a large number of parameters and high computational overhead. Their inference latency for a single image is high under the hardware configurations commonly available in primary hospitals, which fails to meet the real-time requirements of large-scale population screening [14][15] and restricts the application of intelligent auxiliary decision-making tools in primary medical institutions. Model lightweighting technologies, through structured compression, parameter sharing and numerical precision optimization, can significantly reduce model parameters and computational complexity while minimizing accuracy loss, and improve edge-side inference speed. They serve as core technical means to break through hardware limitations in primary care settings [16].

Therefore, how to efficiently lightweight high-performance models while ensuring the accuracy of DDH lesion recognition and diagnosis, so as to adapt to low-configuration hardware environments in primary care, has become a key issue for the application of DDH intelligent auxiliary diagnosis technology at the grassroots level and the realization of large-scale clinical implementation.

The setting of model hyperparameters relies on manual experience-based tuning, with insufficient global optimization capability, which affects the performance stability and deployment time of decision-making models. In existing studies, the key hyperparameters of models are mostly tuned repeatedly based on researchers' experience, lacking systematic global optimization methods. This not only leads to low tuning efficiency and high labor costs, but also easily falls into local optimal solutions, making it difficult to fully exert the optimal performance of the model. Meanwhile, it results in poor reproducibility of experimental results, which is not conducive to adaptation and promotion in different clinical scenarios [17][18].

To address the above research limitations, targeting the practical demands of DDH clinical auxiliary decision-making, this paper proposes a Prior-Constrained Lightweight HRNet (PCL-HRNet) for intelligent diagnosis of DDH IHDI classification on pelvic radiographs, constructing a complete solution from three dimensions: prior embedding, model lightweighting, and performance optimization.

First, to meet the demand for decision reliability, a multi-dimensional geometric constraint loss function is proposed. Clinical prior knowledge such as pelvic symmetry, physiological range of acetabular angle, and relative position of the femoral head specified in the IHDI classification criteria is transformed into differentiable mathematical constraint terms, which form a joint loss function with the basic MSE loss for model training. By constraining the model's output space with domain knowledge, it not only improves keypoint localization accuracy but also ensures that the model's decision-making logic conforms to clinical anatomical laws. Meanwhile, interpretability analysis is conducted using the Gradient-weighted Class Activation Mapping (Grad-CAM) method to verify the consistency between the model's decision focus and clinicians' diagnostic basis, further enhancing the clinical credibility of the system.

Second, to meet the demand for primary care deployment, a lightweighting strategy is designed. Through three sequential steps—transferring clinical prior knowledge via knowledge distillation, pruning redundant channels, and accelerating inference with INT8 static quantization—the model parameter quantity and computational complexity are significantly reduced while minimizing diagnostic accuracy loss. This greatly improves inference speed under ordinary CPU environments and lowers the hardware threshold for deployment in primary hospitals.

Finally, to meet the demand for performance optimization, the QPSO-AWA global hyperparameter optimization algorithm is proposed. The key hyperparameters of the entire process covering model training and lightweighting are

encoded as particle position vectors, and global optimization is performed via the quantum particle swarm optimization algorithm with adaptive weight adjustment. The optimal hyperparameter combination is automatically searched, avoiding the subjectivity and limitations of manual parameter tuning, and effectively improving the comprehensive performance of the model and the reproducibility of experiments.

Experiments on the MTDDH public dataset and deployment tests in real-world environments demonstrate that the model proposed in this paper maintains high-precision diagnostic capability while significantly reducing deployment thresholds and application costs. It can provide reliable and efficient artificial intelligence auxiliary decision-making support for early screening and accurate diagnosis of DDH.

1.2 Paper Contributions

The main contributions of this paper are summarized as follows:

(1) To address the limitations of data-driven models, including the lack of domain knowledge constraints, the tendency to generate predictions violating medical logic, and insufficient decision interpretability, this study converts the core clinical priors from the IHDI classification criteria into differentiable constraint terms. These terms are combined with the baseline MSE loss to jointly optimize the model output space, effectively improving keypoint localization accuracy and the clinical rationality of diagnostic decisions.

(2) To tackle the drawbacks of existing high-performance models, such as large parameter scale, high inference latency, and poor compatibility with the hardware infrastructure of primary hospitals, a collaborative lightweighting strategy is designed. While preserving diagnostic accuracy, this strategy significantly reduces model parameter count and computational complexity, effectively accelerates CPU-side inference speed, and lowers the deployment threshold for primary medical institutions.

(3) To resolve the issue that hyperparameters of deep learning models rely on manual tuning with strong subjectivity and are susceptible to local optima, the QPSO-AWA global hyperparameter optimization algorithm is proposed. It enables automatic global search of hyperparameters across the full workflow, effectively enhancing the comprehensive performance of the model and reducing model deployment time simultaneously.

2. Theoretical Foundation

2.1 High-Resolution Network

High-Resolution Network (HRNet) serves as a mainstream backbone network for keypoint detection in medical imaging. Designed with parallel multi-resolution feature extraction and cross-scale information fusion, HRNet avoids the loss of subtle structural information caused by successive downsampling in conventional convolutional neural networks and enables accurate capture of structural details. Its architecture is illustrated in Fig. 2.

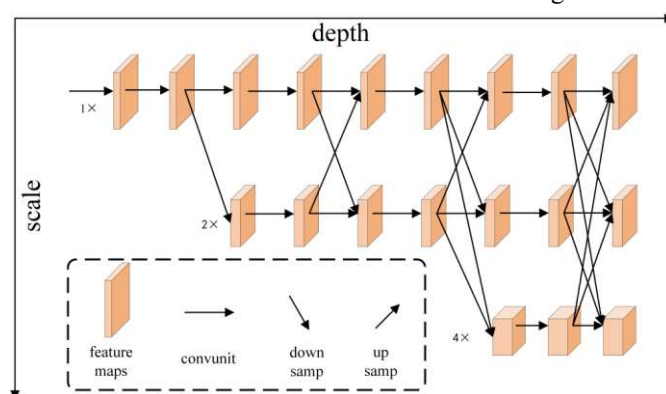


Fig. 2 Schematic diagram of HRNet parallel multi-resolution backbone network architecture

HRNet is composed of four sequentially connected stages, each of which comprises multiple resolution branches, and

feature fusion across branches is achieved through exchange units. Given an input image X , the first stage performs s -fold downsampling via P consecutive $k \times k$ convolutional layers. The forward propagation process of the convolution operation can be expressed as:

$$x_{l+1}^{(i)} = \sigma \left(\sum_{j=1}^{C_{in}} \omega_l^{(j,i)} * x_l^{(j)} + b_l^{(i)} \right) \quad (1)$$

Where $*$ denotes the two-dimensional convolution operation. $\omega_l^{(j,i)}$ and $b_l^{(i)}$ represent the convolution kernel weight and bias term from the j -th input channel to the i -th output channel of the l -th layer, respectively. $\sigma(\cdot)$ is the ReLU nonlinear activation function; $x_l^{(j)}$ denotes the output feature of the previous layer; and C_{in} is the number of input channels of the l -th layer.

At the end of each stage, bidirectional information interaction among all subnetworks is realized via cross-resolution feature exchange units. The mathematical formulation of feature exchange is expressed as:

$$F_{i,j}^s = \sum_{k=1}^s W_{i,j,k} \cdot U(F_{i,k}^{s-1}, r_{j,k}) \quad (2)$$

Where $F_{i,j}^s$ denotes the feature map of the j -th resolution in the i -th module of the s -th stage; $W_{i,j,k}$ is a learnable fusion weight matrix, $U(\cdot, r)$ represents the scaling transformation operation; and s denotes the number of resolutions in the current stage.

The high-resolution feature map F_{final} output by the fourth stage is processed by a heatmap generation convolutional layer g to generate heatmaps corresponding to K keypoints:

$$H = g(F_{final}) \quad (3)$$

Where H is the set of output heatmaps. The coordinates of keypoints in the feature map space are then obtained via heatmap decoding, and the position of the maximum value in each heatmap is taken as the predicted coordinate:

$$(x_k, y_k) = \arg \max_{x,y} H_k(x, y) \quad (4)$$

Finally, the coordinates in the feature map space are mapped back to the original image space, with the mapping ratio equal to the total downsampling factor S_{total} of the first stage, to obtain the final pixel coordinates of the keypoints:

$$(x'_k, y'_k) = (x_k \times S_{total}, y_k \times S_{total}) \quad (5)$$

2.2 Model Lightweighting Strategy

Model lightweighting reduces computational cost, memory consumption and inference latency via structural optimization, parameter compression and precision conversion while limiting performance degradation. In this work, knowledge distillation is adopted to reserve precision redundancy; network structures are subsequently streamlined by pruning, followed by quantization for numerical precision compression. The synergistic combination of multiple strategies achieves network simplification and improved inference efficiency with slight accuracy loss.

2.2.1 Knowledge Distillation

Knowledge distillation transfers feature distributions and diagnostic knowledge learned by the teacher model to a lightweight student model, so that high diagnostic accuracy can be maintained under a smaller model capacity. The corresponding pipeline is shown in Fig. 3

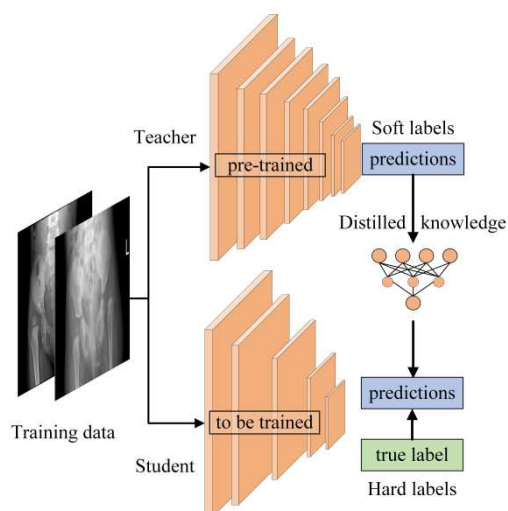


Fig. 3 Flowchart of knowledge distillation

Let Q_t denote the soft prediction probability distribution output by the original high-resolution network, and Q_s denote the distribution output by the lightweight network. Knowledge transfer in the distillation process is achieved by minimizing the distribution discrepancy between the two, and its loss function can be expressed as:

$$L_{kd} = (1 - \alpha_{kd})L_{hard} + \alpha_{kd} \cdot T^2 \cdot KL(Q_s(T), Q_t(T)) + \alpha_{feat}L_{feat} \quad (6)$$

Where L_{hard} denotes the cross-entropy loss between outputs of the student model and ground-truth labels, $KL(\cdot)$ represents the KL divergence loss between soft label distributions of the student and teacher models, T is the temperature coefficient, L_{feat} indicates the mean squared error loss of intermediate-layer features. α_{kd} and α_{feat} stand for the balance coefficients of each loss term.

2.2.2 Structured Channel Pruning

Structured pruning achieves model compression by eliminating redundant network units. The network maintains a regular structure after pruning and can be directly deployed on common inference frameworks. Its principle is illustrated in Fig. 4

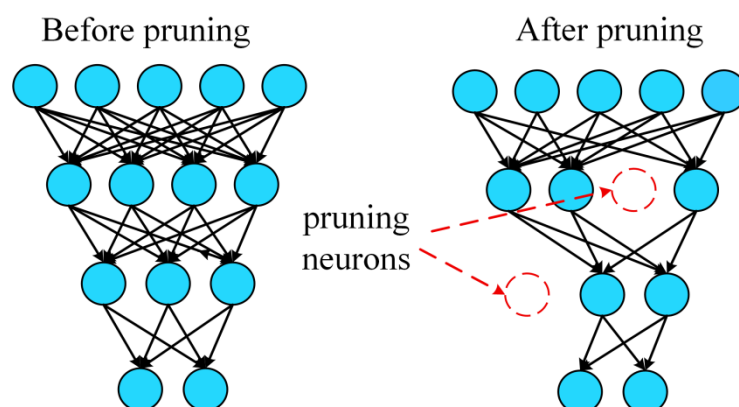


Fig. 4 Schematic diagram of the principle of structured pruning

Channel pruning streamlines network architectures by evaluating the contribution of each convolutional channel to model outputs and removing channels with low contribution. The channel importance A is quantitatively assessed using the L1 norm of convolution kernel weights, and the calculation formula is defined as:

$$S_c = \sum_{i=1}^{C_{in}} \sum_{k_h=1}^{K_h} \sum_{k_w=1}^{K_w} |W_{c,i,k_h,k_w}| \quad (7)$$

Where W_c denotes the weight tensor of the c -th convolutional channel, C_{in} represents the number of input channels. K_h and K_w are the height and width of the convolution kernel, respectively.

2.2.3 INT8 Static Quantization

Numerical quantization realizes model compression by reducing the precision of weights and activations. In this study, the KL divergence minimization approach is adopted to determine the optimal clipping threshold of activations, which minimizes the difference in feature distributions before and after quantization. INT8 quantization converts data from FP32 to INT8 format, compressing the model volume to one quarter of the original size and effectively accelerating inference. The quantization mapping process is depicted in Fig. 5.

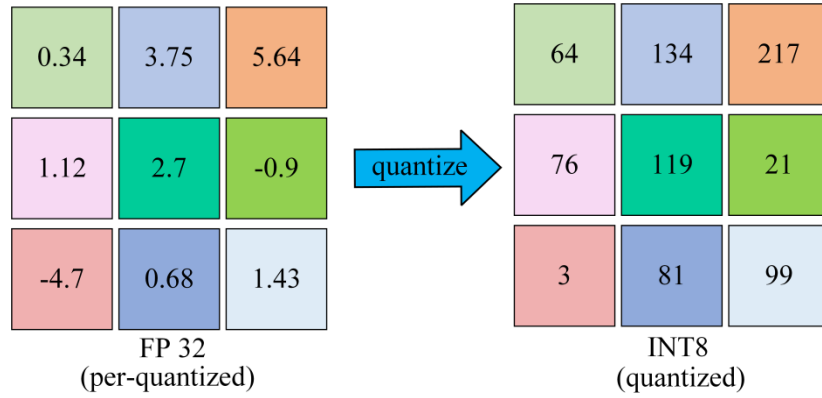


Fig. 5 Schematic diagram of numerical mapping for INT8 linear quantization

$$x_{int8} = \text{Round} \left(\frac{x_{fp32}}{scale} + zero_{point} \right) \quad (8)$$

Where $scale$ denotes the quantization scaling factor, which characterizes the proportional relationship between floating-point values and integer values. $zero_{point}$ is the zero offset representing the integer value corresponding to the floating-point zero.

2.3 Quantum-behaved Particle Swarm Optimization Algorithm

Quantum-behaved Particle Swarm Optimization (QPSO) is an improved algorithm that introduces a quantum mechanics potential well model on the basis of conventional PSO. The states of particles are described by wave functions, and positions are updated via Monte Carlo random sampling. Without the velocity term, QPSO exhibits remarkably stronger global optimization capability compared with standard PSO. In QPSO, the particle position update formula is expressed as:

$$x_i^{t+1} = p_i^t + (-1)^{rand(0,1)} \cdot \delta \cdot |m_{best}^t - x_i^t| \cdot \ln \left(\frac{1}{u} \right) \quad (9)$$

Where x_i^{t+1} is the position vector of the i -th particle in the t -th generation. δ is the contraction-expansion coefficient.

u is a random number within the interval $[0,1]$. p_i^t is the local attractor. m_{best}^t is the mean value of the personal best positions of all particles. The calculation formulas for these two variables are given as follows:

$$p_i^t = \varphi \cdot p_{best,i}^t + (1 - \varphi) \cdot g_{best}^t \quad (10)$$

$$m_{best}^t = \frac{1}{M} \sum_{i=1}^M p_{best,i}^t \quad (11)$$

Where $p_{best,i}^t$ denotes the historical best position of the i -th particle, g_{best}^t represents the historical best position of the entire population, and M is the population size.

3. Proposed Method

3.1 PCL-HRNet

To accurately extract hip anatomical features from pelvic radiographs and tackle the insufficient clinical constraints and high deployment barriers existing in conventional models, a Prior-Constrained Lightweight HRNet (PCL-HRNet) is proposed in this work. The overall architecture of PCL-HRNet is illustrated in Fig. 6.

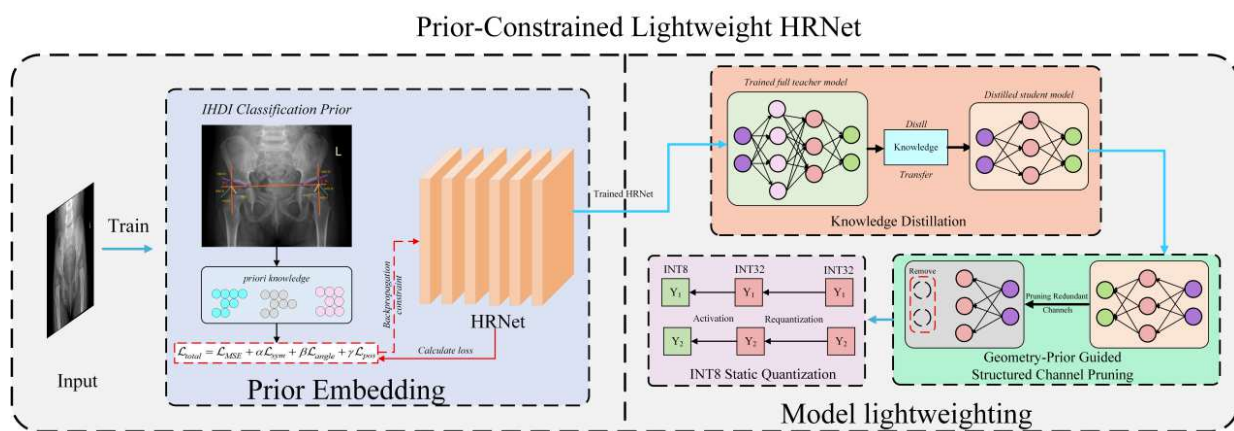


Fig. 6 Architecture of the PCL-HRNet model

HRNet is adopted as the backbone to extract multi-scale anatomical features and generate heatmaps of critical keypoints. The anatomical priors defined by the IHDI classification criteria, including pelvic symmetry, physiological range of acetabular angles and relative position of the femoral head, are transformed into differentiable geometric constraint losses. These losses are embedded into the training procedure to guide the network to learn anatomically consistent feature representations. A lightweighting pipeline is implemented for the model. First, knowledge distillation is employed to transfer diagnostic knowledge from the large teacher model to a student model with half channel dimensions. Channel pruning is subsequently conducted to preferentially retain channels closely associated with core diagnosis. Finally, INT8 static quantization is utilized to further compress the model size. The QPSO-AWA algorithm is adopted to automatically search the optimal hyperparameter combination for the entire pipeline, which eliminates the subjectivity caused by manual parameter tuning.

3.2 Prior-Constrained Loss Function

Mean squared error loss is adopted to train conventional keypoint detection models. Such loss fails to exploit inherent correlational constraints between keypoints, which may generate predictions violating anatomical principles and impair the computational accuracy of subsequent diagnostic indicators as well as clinical credibility. The prior-constrained loss converts medical geometric priors into differentiable mathematical constraint terms. When optimized jointly with the base loss, it constrains the model output space without modifying network architectures, ensuring that predictions satisfy both pixel-level accuracy and geometric rationality.

The symmetry constraint applies to bilaterally symmetric anatomical structures of human body. For pairs of symmetric anatomical landmarks, their spatial coordinates under standard imaging posture should comply with rigorous symmetric relations. Severe asymmetry among predicted landmarks implies the existence of localization errors. Its mathematical expression is given as:

$$L_{sym} = \alpha \sum_{i=1}^N (|x_i - x_{i+N/2}| + |y_i - y_{i+N/2}|) \quad (12)$$

where N denotes the total number of symmetric keypoints, (x_i, y_i) represents the predicted coordinates of the i -th keypoint, and α is the weight coefficient of the symmetry constraint.

The angular range constraint is applied to constrain the angles calculated from keypoints within the clinically normal range. Anatomical angles of the human body follow fixed physiological ranges; a predicted angle falling outside this range indicates erroneous keypoint localization. This constraint loss is formulated using hinge loss, with its mathematical expression as follows:

$$L_{angle} = \beta \cdot \max(0, |\theta - \mu_\theta| - k \cdot \sigma_\theta) \quad (13)$$

Where θ is the anatomical angle calculated based on the predicted keypoints, μ_θ is the population mean of this angle, σ_θ is the corresponding standard deviation, k is the confidence interval coefficient, and β is the weight coefficient of the angle constraint.

The relative position constraint serves to regularize the spatial ordering relationships among anatomical structures. Fixed relative positional relationships exist among human anatomical structures, and predictions violating such relationships will inevitably lead to diagnostic errors. This constraint loss imposes penalties on predictions that breach the prior positional relationships, with its expression as follows:

$$L_{pos} = \gamma \sum_{(i,j) \in R} \max(0, -g_{pos}(p_i, p_j)) \quad (14)$$

Where R is the set of keypoint pairs satisfying the prior positional relationships, g_{pos} is the function describing the relative positional relationship between two points, p_i and p_j are the coordinate vectors of keypoint i and keypoint j , respectively, and γ is the weight coefficient of the relative position constraint.

By integrating the aforementioned geometric constraint loss terms with the baseline mean squared error (MSE) loss function, a prior loss function for joint training is derived as follows:

$$L_{total} = L_{MSE} + \alpha L_{sym} + \beta L_{angle} + \gamma L_{pos} \quad (15)$$

Where L_{MSE} is the base MSE loss for keypoint detection. α , β , and γ are the weight coefficients of each constraint term.

3.3 Parameter Adaptive Optimization of QPSO-AWA

The model lightweighting strategy, geometric constraint loss, and training process involve multiple critical hyperparameters, the selection of which directly affects the final performance of the model. Traditional manual hyperparameter tuning relies on prior experience, which is not only inefficient but also prone to converging to local optima. To address this problem, this paper adopts the QPSO-AWA global hyperparameter optimization algorithm to automatically search for the optimal hyperparameter combination of each module. The overall optimization workflow of the algorithm is shown in Fig. 7.

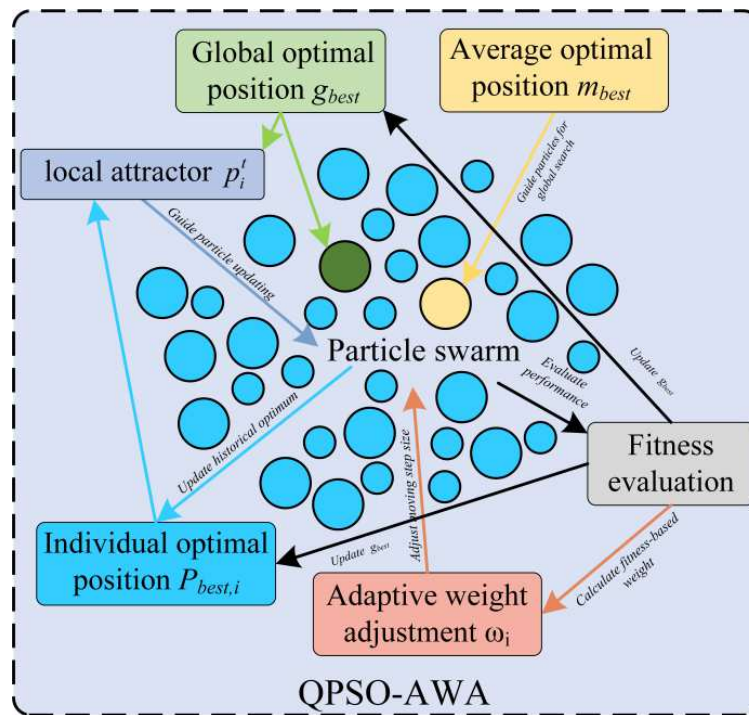


Fig. 7 Parameter adaptive optimization of QPSO-AWA

3.3.1 AWA Adaptive Weight Strategy

Traditional quantum-behaved particle swarm optimization (QPSO) mostly adopts fixed weights or uniformly linearly decreasing weights. Such a setting makes it difficult to balance global exploration and local exploitation, tends to ignore individual search differences among particles, and fails to exert the optimization characteristics of different particles in a differentiated manner. To this end, this paper proposes a fitness-based adaptive weight (AWA) strategy, which dynamically allocates inertia weights according to particle fitness, so as to accelerate the convergence of high-quality particles and strengthen global exploration for poor-performing particles. The update formula of the adaptive inertia weight is as follows:

$$\omega_i^t = \omega_{\max} - \frac{(\omega_{\max} - \omega_{\min}) \cdot (f_i^t - f_{\min}^t)}{f_{\max}^t - f_{\min}^t} \quad (16)$$

Where $\omega_{\max}$ and $\omega_{\min}$ denote the maximum and minimum values of the inertia weight, respectively. f_i^t represents the fitness value of the i -th particle in the t -th generation. $f_{\max}^t$ and $f_{\min}^t$ are the maximum and minimum fitness values of all particles in the t -th generation population, respectively.

3.3.2 Hyperparameter Encoding and Multi-objective Fitness Function

In this study, the QPSO-AWA algorithm optimizes all critical hyperparameters during model training, rather than network weights or post-deployment parameters. The hyperparameters to be optimized are encoded as the position vectors of particles, whose expression is as follows:

$$\Theta = [\alpha, \beta, \gamma, T, \lambda_{kd}, lr, batch_size, pruned_hreshold] \quad (17)$$

Where α denotes the weight of the symmetry constraint loss. β denotes the weight of the angle constraint loss. γ denotes the weight of the relative position constraint loss. T is the temperature coefficient for knowledge distillation. λ_{kd} denotes the weight of the distillation loss. lr represents the initial learning rate. $batch_size$ is the batch size, and $pruned_threshold$

refers to the channel pruning threshold.

To guarantee the clinical accuracy and rationality of DDH diagnosis, a multi-objective fitness function is designed as follows:

$$F_{search}(\Theta) = \lambda_1 \cdot (1 - Accuracy(\Theta)) + \lambda_2 \cdot MAE_{acetabular}(\Theta) + \lambda_3 \cdot Error_{logic}(\Theta) \quad (18)$$

Where $Accuracy(\Theta)$ denotes the IHDI classification accuracy of the candidate hyperparameter configuration Θ on the validation set. $MAE_{acetabular}(\Theta)$ represents the mean absolute error of the acetabular angle calculated from the predicted key points, and $Error_{logic}(\Theta)$ is the proportion of predicted samples that violate clinical geometric logic.

3.3.3 Adaptive Optimization Workflow of QPSO-AWA

Compared with manual trial-and-error and grid search, the QPSO-AWA algorithm integrated with the adaptive weight strategy delivers stronger search capability in hybrid hyperparameter spaces. During the iterative process, the algorithm dynamically allocates inertia weights according to particle fitness, achieves population evolution in conjunction with the position update rule of standard QPSO, and continuously updates the individual best and global best hyperparameter combinations. The complete execution workflow is presented as follows:

- (1) Set the population size $N=30$ and the maximum number of iterations $T_{max}=20$. The position vectors of all particles are initialized to random values near empirical default values..
- (2) For each particle, the PCL-HRNet model is fully trained with its corresponding hyperparameter configuration Θ .
- (3) Calculate the fitness value $F_{search}(\Theta)$ of this configuration on the validation set.
- (4) Compare the current fitness of each particle with its historical best fitness to update the personal best position $p_{best,i}$. Then compare the personal best fitness of all particles with the global best fitness to update the global best position g_{best} .
- (5) Calculate the adaptive inertia weight of each particle via Equation (16).
- (6) Update the position vector of each particle via Equations (9)-(11).
- (7) Terminate the iteration if the maximum number of iterations is reached or the global best fitness shows no improvement for 5 consecutive generations, and output the global best position g_{best} as the final hyperparameter configuration.
- (8) Retrain the final model using the optimal hyperparameter configuration for subsequent evaluation and deployment.

4. Experiments and Result Analysis

4.1 Dataset and Preprocessing

In this study, model performance evaluation is conducted using the publicly available Multitasking Developmental Dysplasia of the Hip (MTDDH) dataset, which is currently the first authoritative public dataset worldwide integrating both pelvic X-ray quality assessment and keypoint annotations for DDH diagnosis [19]. All images undergo standardized preprocessing and are uniformly resized to an input resolution of 1333×800 pixels. Typical anteroposterior pelvic X-ray images and keypoint annotation examples from the MTDDH dataset are presented in Fig. 8.

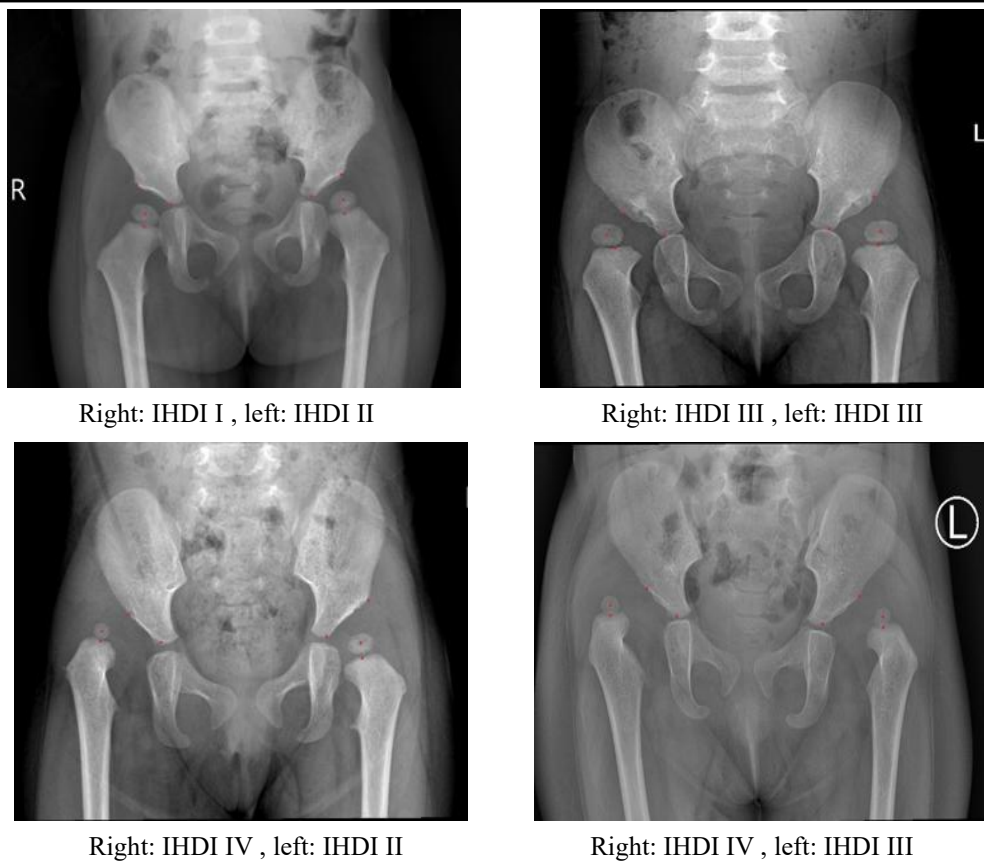


Fig. 8 Examples of MTDDH dataset with anatomical keypoint annotations.

The MTDDH dataset (available at: <https://doi.org/10.57760/sciencedb.24372>) was constructed by the Children's Hospital of Zhejiang University School of Medicine and approved by the hospital's Medical Ethics Committee (approval No.: 2023-IRB-0287-P-01). All data have been anonymized, and the requirement for informed patient consent was waived. The dataset comprises two subsets. In this study, Dataset 2, which is dedicated to computer-aided DDH diagnosis, is selected. This subset contains a total of 906 clinically eligible pediatric anteroposterior pelvic X-ray images. The age of the patients ranges from 5 to 40 months, with a mean age of 14.32 months, including 558 female and 348 male cases.

Eight core anatomical keypoints are annotated on each image by three senior pediatric orthopedic specialists, namely the center of the right triradiate cartilage, center of the left triradiate cartilage, superolateral edge of the right acetabulum, superolateral edge of the left acetabulum, center of the right femoral head, center of the left femoral head, midpoint of the superior margin of the right proximal femoral metaphysis, and midpoint of the superior margin of the left proximal femoral metaphysis. These keypoints fully cover all anatomical landmarks required for IHDI classification and acetabular angle measurement.

The annotation procedure follows a standardized two-stage workflow consisting of independent annotation by two junior physicians and final quality control by a senior physician:

(1) Keypoint annotation of all samples is independently performed by two junior physicians with over 3 years of clinical experience in pediatric orthopedics.

(2) Consistency testing is conducted on the annotation results from the two physicians. Samples with an intraclass correlation coefficient (ICC) greater than 0.85 are directly accepted.

(3) For samples with inconsistent annotations, a chief physician with more than 15 years of clinical experience makes the final adjudication.

(4) The average error of all annotation results is controlled within 2 pixels, ensuring that the precision of the ground truth meets the requirements of clinical diagnosis.

In strict accordance with the gold-standard splitting principle for medical imaging research, a patient-level

independent splitting strategy is adopted in this study. All images from the same patient are assigned to only one of the training set, validation set, or test set, which thoroughly eliminates the risk of data leakage caused by the presence of data from the same patient in both the training and test sets.

Dataset 2 is ultimately divided into a training set (634 images), a validation set (181 images), and a test set (91 images) at a ratio of 7:2:1. A stratified random splitting method based on IHDI grades is employed to ensure that the disease distribution of the three subsets is fully consistent with that of the overall dataset, avoiding excessive concentration or absence of minority class samples due to random splitting. The detailed splitting statistics are presented in Table 1.

Table 1 MTDDH Dataset 2 description.

IHDI Classification	Training Set	Validation Set	Test Set
I 度	352	101	50
II 度	176	50	25
III 度	70	20	10
IV 度	36	10	6

For grade IV DDH with an extremely small sample size, leave-one-out cross-validation (LOOCV) is additionally adopted for separate evaluation to obtain more reliable performance estimation. All splitting procedures are performed with 10 fixed sets of distinct random seeds. The final experimental results are averaged over 10 independent trials, and the corresponding 95% confidence intervals (CIs) are calculated to ensure the stability and reproducibility of the results.

All images are processed via a standardized preprocessing pipeline. All preprocessing parameters are derived exclusively from the training set, and only the parameters obtained from the training set are applied to transform the validation and test sets, so as to strictly avoid data leakage. The specific steps are as follows:

- (1) Convert original DICOM images into 8-bit grayscale PNG format, while retaining the original window width and window level information.
- (2) Maintain the original aspect ratio, and uniformly resize all images to 1333×800 pixels via zero-padding on the shorter edge, consistent with the input size of the original dataset.
- (3) Enhance bone edge contrast using Contrast Limited Adaptive Histogram Equalization.
- (4) Normalize pixel values to the interval $[0, 1]$ to facilitate model training.

To mitigate overfitting during model training and improve model robustness, lightweight unsupervised data augmentation strategies including random rotation within $\pm 15^\circ$, horizontal flipping, and random scaling by a factor of 0.8–1.2 are only applied to the training set. None of the augmentation operations alter the relative positional relationships of key points, thereby guaranteeing the accuracy of annotation information.

4.2 Experimental Environment and Hyperparameter Configuration

To ensure the stability of model training and the reproducibility of results, this study is implemented based on Python 3.9 and the PyTorch 2.0.1 framework. Training and inference are performed on an Intel Core i5-10400 CPU, which matches the hardware configuration of primary hospitals. The key hyperparameters of the model are set as follows: in the basic training phase, the batch size is set to 8, and the maximum number of training epochs is 200. An early stopping strategy is adopted to prevent overfitting. The Adam optimizer is selected with a weight decay coefficient of $1e-5$. The initial learning rate is set to 0.001, which is reduced to a minimum learning rate of $1e-6$ via a cosine annealing scheduling strategy. For knowledge distillation, the original HRNet is used as the teacher model with a distillation temperature $T = 10$. The threshold of structured channel pruning is set to 0.2. For INT8 static quantization, 10% of samples from the training set are used as the calibration set, and the output layer of the model is retained at FP32 precision to mitigate accuracy loss caused by quantization. This design significantly improves model deployment performance while maintaining diagnostic accuracy.

4.3 Diagnostic Performance Analysis of the PCL-HRNet Model

To verify the effectiveness of the proposed PCL-HRNet model on the MTDDH dataset, training loss curves are applied to evaluate the training stability of the model in the experiments. Combined with clinical metrics and IHDI classification-specific performance on the test set, the reliability of the model for DDH diagnosis is comprehensively validated. The mean Euclidean distance error (MDE) of keypoints, the mean absolute error (MAE) of the acetabular angle, classification accuracy, and recall are selected as the core evaluation metrics. MDE and MAE are used to quantify the accuracy of anatomical keypoint localization and clinical key parameter measurement, while classification accuracy and recall are adopted to measure the diagnostic effectiveness of the model for different IHDI classifications. The results of PCL-HRNet under each classification are presented as follows.

Table 2 Performance of the PCL-HRNet model across different IHDI grades

IHDI classification	Keypoint MDE	Acetabular angle MAE (°)	Classification accuracy (%)	Recall (%)
Type I	3.92±0.23	1.51±0.30	91.38	92.16
Type II	4.21±0.26	1.67±0.33	65.21	68.73
Type III	5.47±0.37	2.15±0.44	84.73	89.51
Type IV	6.89±0.45	2.92±0.54	78.79	79.73

As shown in Table 2, the PCL-HRNet model achieves effective identification and accurate classification of hip lesions across all IHDI grades. For IHDI grade I normal hips, the model attains optimal keypoint localization and acetabular angle measurement accuracy, with a classification accuracy of 91.38% and a recall of 92.16%. The performance slightly declines for IHDI grade II mild dysplasia cases, with a classification accuracy of 65.21% and a recall of 68.73%, which is attributed to the high overlap in imaging features between mild lesions and normal hip joints. For IHDI grade III moderate dislocation and grade IV severe complete dislocation, the classification accuracies reach 84.73% and 78.79%, with corresponding recalls of 89.51% and 79.73% respectively, indicating the model's stable recognition capability for severe lesions.

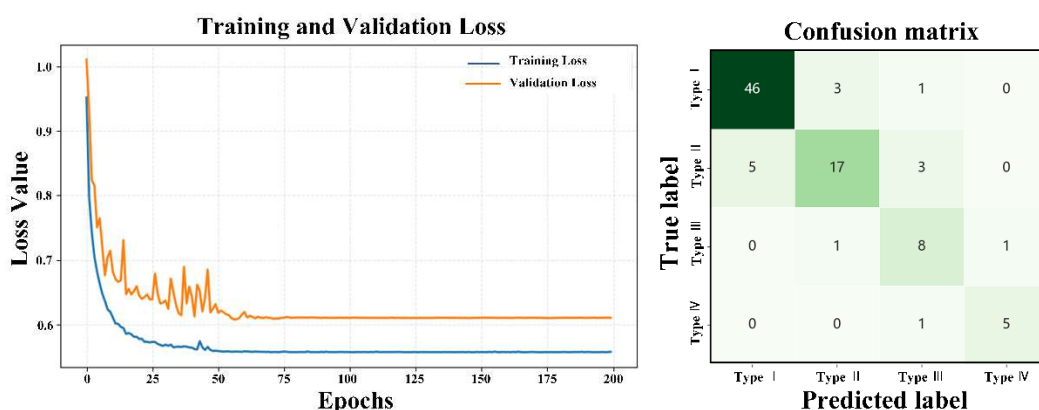


Fig. 9 Examples of MTDDH dataset with anatomical keypoint annotations.

The loss curves during model training are illustrated in Fig. 9 Both training loss and validation loss present a typical monotonically decreasing trend with the increase of training epochs: the training loss converges from an initial value of 0.96 to 0.56, and the validation loss converges from an initial value of 1.01 to 0.61. The validation loss shows minor fluctuations in the first 50 training epochs, reflecting the dynamic learning process of the model on boundary samples with low feature discriminability. After approximately 75 epochs, the loss curves stabilize without obvious fluctuation or upward trend. Meanwhile, the training loss remains consistently lower than the validation loss, and no overfitting or data leakage is observed, which verifies the stability of the model training process. The confusion matrix results further validate the

diagnostic rationality of the model. All mispredictions occur exclusively between adjacent grades, and no cross-grade errors such as misclassifying grade IV as grade I or II are detected.

4.4 Ablation Studies

4.4.1 Ablation Study on Prior-Constrained Loss Function

To systematically verify the contributions and synergistic optimization effects of three geometric priors (symmetry constraint, relative position constraint and angle constraint), this study designs single-variable controlled ablation experiments. Four experimental groups are set up: Group 1 is the HRNet baseline model trained with only mean squared error (MSE) loss; Group 2 adds the symmetry constraint on the basis of the baseline; Group 3 further incorporates the relative position constraint; Group 4 finally supplements the angle constraint to form the complete prior constraint combination. Training hyperparameters, hardware environment and dataset partitioning are kept completely consistent across all groups. The improvement effects of different priors on model performance are comprehensively validated through quantitative metric evaluation and qualitative error pattern analysis.

The logical error rate is defined as the proportion of samples whose predictions violate the anatomical geometric rules of the hip joint. The judgment criteria include severe asymmetry of bilateral keypoints, acetabular angle exceeding the physiological range of children, and relative position between the femoral head and acetabulum violating the spatial logic of IHDI classification. The keypoint F1 score is calculated with an OKS threshold of 0.5, representing the average F1 value of 8 anatomical keypoints.

Table 3 Performance Comparison of Different Loss Function Combinations

Group	Keypoint F1	Logical Error Rate (%)	Acetabular Angle MAE (°)
Group1	0.8762	3.4	2.87
Group2	0.8825	1.7	2.51
Group3	0.8892	0.3	2.32
Group4	0.8915	0.0	1.96

The performance comparison of different loss function combinations is shown in Table 3. The three geometric constraints progressively correct prediction deviations from different dimensions. Group 1 uses only MSE loss; lacking global structural constraints, it achieves an F1 score of only 0.8762, a logical error rate of 3.41%, and an acetabular angle MAE of 2.87°. Group 2 introduces the symmetry constraint, which suppresses unilateral abnormal offset by leveraging bilateral anatomical homology. Its F1 score rises to 0.8825, the error rate drops to 1.62%, and MAE decreases to 2.51°. Group 3 adds the relative position constraint, which focuses on the femoral head–acetabulum spatial relationship and reduces cross-grade classification errors, reaching an F1 score of 0.8892, an error rate of 0.73%, and an MAE of 2.32°. Group 4 integrates the angle constraint, which regularizes the physiological range of the acetabular angle and compensates for the insufficient angle sensitivity of the former two constraints, achieving the optimal performance. The three priors synergistically complement each other to jointly improve detection accuracy and clinical quantitative consistency.

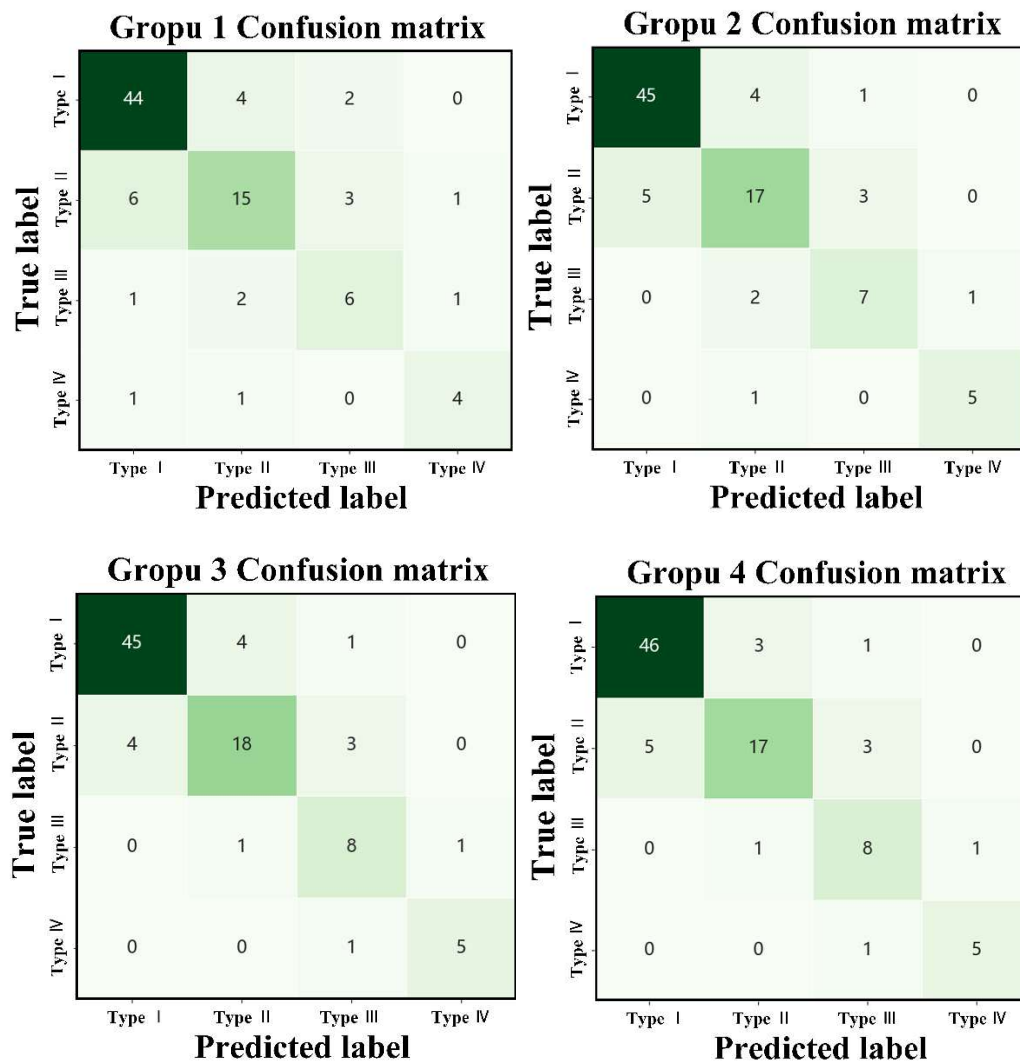


Fig. 10 Comparison of confusion matrices for IHDI classification under different loss function combinations.

(a) Group 1 (b) Group 2 (c) Group 3; (d) Group

The comparison of IHDI classification confusion matrices under different loss combinations is presented in Fig. 10, which qualitatively demonstrates the progressive improvement of the model's error patterns. Three severe cross-grade logical errors exist in Group 1, and such errors would cause fundamental deviations in treatment regimens. In Group 2, cross-grade errors are reduced to 1 case, and asymmetry errors of bilateral hip joints are basically eliminated. No cross-grade logical errors are observed in Group 3, and all 12 mispredictions are strictly confined to adjacent grades, verifying the mandatory regulatory effect of the relative position constraint on the femoral head–acetabulum spatial relationship. Group 4 further improves the discriminability between adjacent grades: compared with Group 1, the confusion between Type I and Type II decreases from 10 cases to 8 cases, and the number of correctly classified Type III cases increases from 6 to 8. The error distribution conforms to the objective pattern of blurred boundaries between adjacent grades in clinical diagnosis.

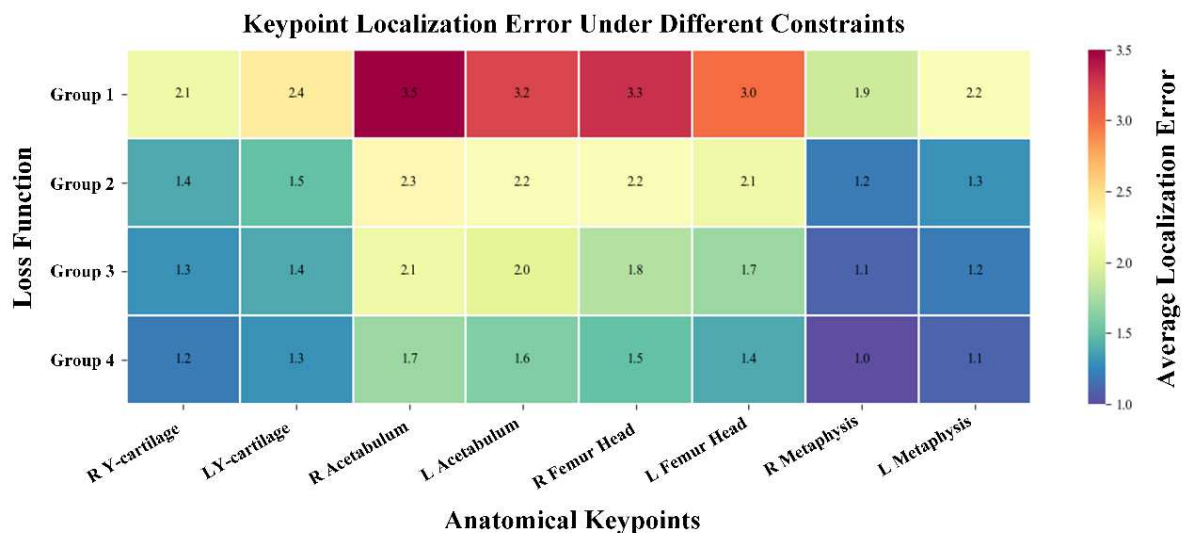


Fig. 11 Heatmaps of average localization errors for each anatomical keypoint under different geometric constraints

The heatmap of mean localization errors for each anatomical keypoint under different constraints is shown in Fig. 11, illustrating the specific effects of each constraint on different anatomical structures. In Group 1, the localization errors of the superolateral acetabular edge and the femoral head center are the largest. These regions are characterized by blurred bone margins and severe soft tissue overlap, which is consistent with the limitation that purely data-driven methods struggle to accurately capture anatomical boundaries. In Group 2, the mean localization error of the 8 bilaterally symmetric keypoints is reduced by approximately 35% compared with the baseline, and the error difference between corresponding left and right keypoints narrows from 0.3 pixels to within 0.1 pixel, effectively resolving the unilateral keypoint localization deviation. In Group 3, the localization error of the femoral head center is further reduced by approximately 18%, verifying the precise regulatory effect of this constraint on the spatial position of the femoral head. In Group 4, the error of the superolateral acetabular edge is reduced by another 20%, improving the measurement accuracy of core clinical indicators.

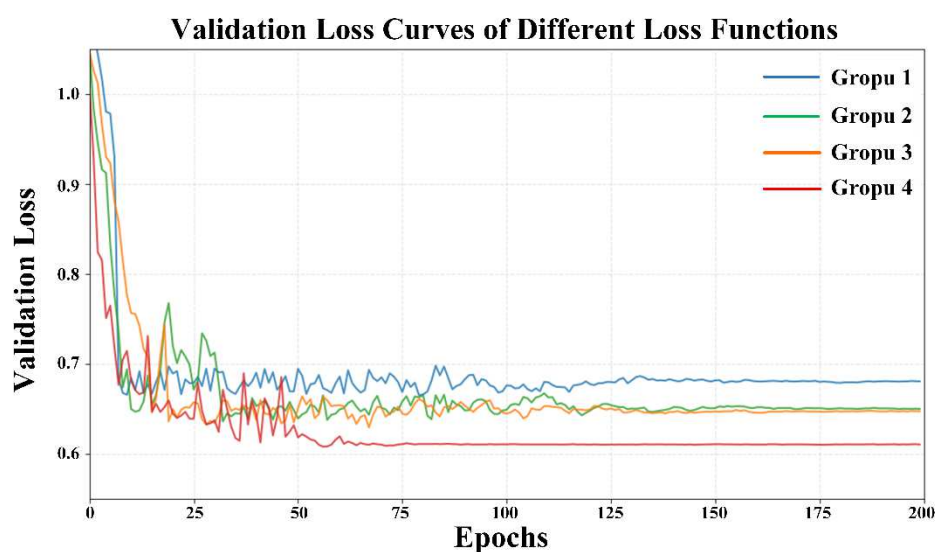


Fig. 12 Convergence curves of validation loss under different loss function combinations

The convergence curves of validation loss under different combinations of loss functions are illustrated in Fig. 12

Group 1 lacks regularization mechanisms, resulting in the slowest convergence and conspicuous fluctuations in the later training stage, with a final loss of approximately 0.68. After introducing symmetry constraints, Group 2 achieves faster convergence and weaker oscillation, and its loss declines to around 0.65. Group 3 further reduces the loss to roughly 0.63 with a much more stable training process. Group 4 reaches a steady state at about 70 epochs and obtains the minimum final loss of nearly 0.61, without obvious fluctuations throughout training. By supplying abundant supervisory signals and inherent regularization effects, multi-dimensional geometric priors effectively suppress overfitting and enable the model to converge to the global optimum more rapidly and steadily.

4.4.2 Ablation Study on Lightweighting Strategy

To verify the effectiveness of the proposed model lightweighting strategy, single-variable controlled ablation experiments are conducted in this study. A total of 5 experimental groups are established: Group 5 serves as the HRNet baseline model embedded with geometric prior constraints; Group 6 is the accuracy-enhanced model with knowledge distillation introduced on the baseline, designed to verify the accuracy compensation effect of distillation. Group 7 adopts pruning to achieve structural-level compression based on Group 6, and Group 8 is the full lightweight model with INT8 static quantization superimposed on Group 7. For all experimental groups, training hyperparameters, hardware environment and dataset partitioning are kept completely consistent with those in the aforementioned experiments. The keypoint F1 score is adopted as the core accuracy metric, and deployment efficiency is comprehensively evaluated from four dimensions: parameter count, computational complexity (FLOPs), inference speed and model size. All experimental results are presented as the average of 10 independent repeated experiments.

Table 4 Performance Comparison of Different Lightweight Strategies

Group	Keypoint F1	Params (M)	FLOPs(G)	Single-image Inference Latency on CPU (ms)	Model Size (MB)	Accuracy Drop (%)
Group5	0.8915±0.0018	14.3	7.8	96±3	57	0.0
Group6	0.8921±0.0015	14.3	7.8	95±2	57	-0.1
Group7	0.8863±0.0032	8.3	3.7	51±2	33	0.6
Group8	0.8908±0.0021	7.8	3.4	51±2	31	0.1
Group9	0.8912±0.0019	7.8	3.4	45±2	8	0.0

The comprehensive performance comparison of different lightweighting strategies is shown in Table 4. The results demonstrate that the proposed lightweighting strategy can achieve deep model compression and inference acceleration with negligible diagnostic accuracy loss. The accuracy loss is calculated relative to the Group 5 baseline model, and negative values indicate higher accuracy of the experimental group than the baseline.

For Group 6 with knowledge distillation, the parameter count, FLOPs, inference speed and model size remain completely identical to the baseline, while the keypoint F1 score rises from 0.8915 (baseline) to 0.8921, representing a 0.1% relative accuracy improvement over the baseline. This verifies the effective accuracy compensation effect of knowledge distillation. On this basis, Group 7 implements structural-level compression via pruning: the parameter count is reduced to 7.8 M, FLOPs to 3.4 G, single-image CPU inference latency to 51 ms, and model size to 31 MB, with an accuracy loss of only 0.1%, striking a balance between structural compression and accuracy preservation. Finally, Group 8 incorporates INT8 static quantization on the basis of pruning. Its parameter count and FLOPs stay consistent with Group 7, while the model size is further compressed from 31 MB to 8 MB, and CPU inference latency is further reduced to 45 ms.

Overall, the proposed lightweighting strategy achieves remarkable deep compression effects: a 45.5% reduction in parameter count, a 56.4% reduction in FLOPs, a 53.1% improvement in CPU inference speed, and an 86.0% compression

in model size, with only 0.33% accuracy loss introduced. It achieves an excellent balance between diagnostic accuracy and deployment efficiency, verifying the effectiveness and clinical implementation value of the proposed lightweighting scheme.

4.5 Model Interpretability Verification and Analysis

In this paper, the Gradient-weighted Class Activation Mapping (Grad-CAM) [20] method is adopted to verify the interpretability of the proposed model. As shown in Figure 14, this study visualizes the attention activation regions of four typical cases (IHDI grades I, II, III and IV) to verify whether the multi-dimensional geometric prior constraints achieve deep embedding of clinical anatomical knowledge, and whether the model's decision-making logic conforms to clinicians' diagnostic criteria for DDH classification.

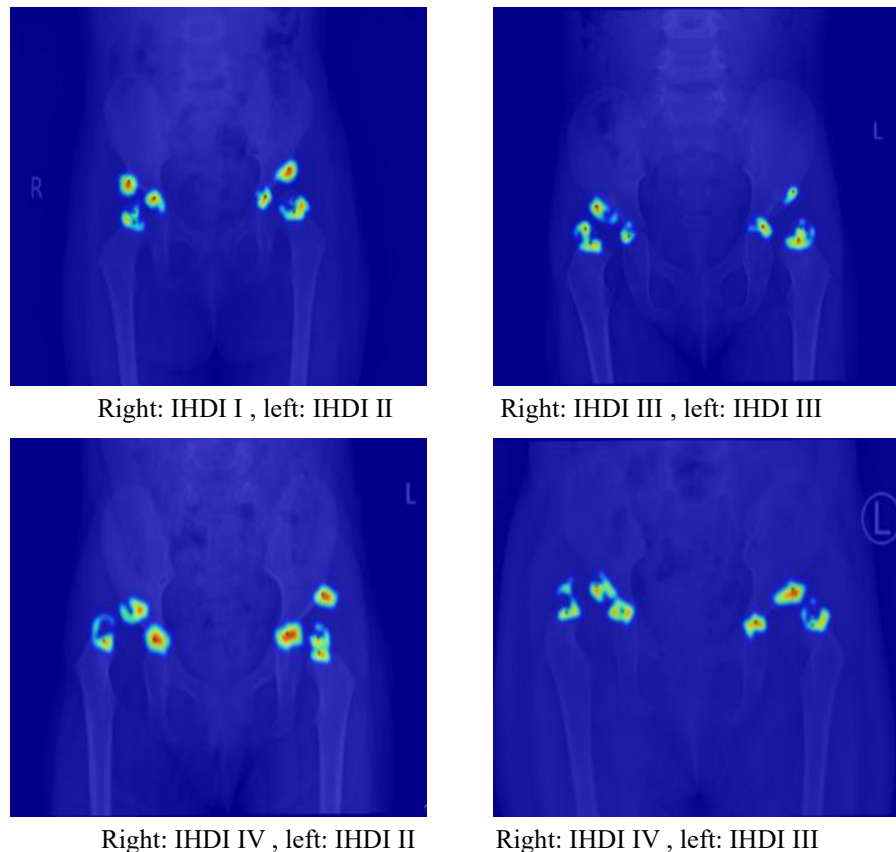


Fig. 13 Grad-CAM visualization of anatomical keypoints for DDH IHDI classification guided by prior constraints

As shown in Fig. 13, the high-attention regions of the model during keypoint localization and classification decision-making exhibit substantial spatial consistency with the eight core anatomical keypoints manually annotated by clinicians. The model's activation regions are primarily concentrated in the core structures of bilateral hip joints that are critical for IHDI classification diagnosis, and accurately cover key anatomical landmarks including the superolateral acetabular rim, the center of the femoral head, and the center of the Y-shaped cartilage. Even for severe IHDI grade IV cases characterized by high complete dislocation of the femoral head and marked anatomical deformation, the model's attention activation regions can still shift synchronously with the dislocated osseous structures and align precisely with core anatomical landmarks, with no evident attention dispersion or deviation.

The loss function incorporating multi-dimensional geometric prior constraints effectively embeds clinical anatomical prior knowledge into the entire pipeline of model feature learning and decision-making, and reasonably calibrates the model's learning objectives and decision-making logic. Instead of relying on pseudo-features such as background grayscale and bone texture for localization and classification, the model performs spatial relationship inference and classification output primarily based on clinically accepted core anatomical landmarks. This effectively mitigates the black-box decision-

making issue prevalent in deep learning models, and provides crucial support for the safety and credibility of the model's clinical implementation in primary healthcare institutions.

4.6 Comparative Experiments

4.6.1 Comparative Experiment on QPSO-AWA Adaptive Hyperparameter Optimization Algorithm

To verify the effectiveness and efficiency advantages of the proposed QPSO-AWA adaptive hyperparameter optimization algorithm, this study designs single-variable controlled comparative experiments. Four optimization schemes are established, namely manual tuning, grid search optimization, standard QPSO optimization[21], and the QPSO-AWA optimization scheme. All algorithms adopt exactly the same hyperparameter search space, and the hardware environment, dataset partitioning and model architecture are kept consistent across all groups. The keypoint F1 score and acetabular angle MAE are selected as core performance metrics, while tuning time consumption and the degree of manual intervention are also recorded to ensure a fair and rigorous comparison.

Table 5 Performance Comparison of Different Hyperparameter Optimization Methods

Optimization Method	Keypoint F1	Acetabular Angle MAE (°)	Tuning Time (h)	Manual Intervention Required
Manual parameter tuning	0.8915±0.0018	1.96±0.08	8	Yes
Grid search optimization	0.8931±0.0016	1.91±0.07	12	Slight
QPSO	0.8962±0.0014	1.81±0.05	3	No
QPSO-AWA	0.8975±0.0012	1.77±0.05	0.5	No

The experimental results are presented in Table 5. The results demonstrate that QPSO-AWA outperforms all other schemes in both optimization accuracy and operational efficiency. Manual tuning relies on researchers' empirical experience, resulting in the longest total time consumption. Furthermore, its tuning outcomes are constrained by subjective experience and prone to falling into local optima. Grid search achieves slightly higher accuracy than manual tuning, but its computational load grows combinatorially with the increase of hyperparameter dimensions, making it unable to efficiently handle high-dimensional continuous hybrid hyperparameter spaces. The standard QPSO introduces the quantum potential well model, endowing it with stronger global search capability and faster convergence speed. However, it adopts a globally unified linearly decreasing contraction–expansion coefficient, which fails to account for differences in search states among individual particles, and thus cannot fully unleash the convergence potential of high-quality particles and the exploration capability of low-quality particles.

Building upon the standard QPSO, QPSO-AWA introduces an adaptive contraction–expansion coefficient adjustment strategy based on particle fitness. For particles with high fitness that are close to the optimal solution, the contraction coefficient is reduced to strengthen local exploitation and accelerate convergence. For particles with low fitness and poor search performance, the contraction coefficient is increased to expand the search scope and enhance global exploration capability, ultimately achieving a dynamic balance between global exploration and local exploitation. This approach effectively mitigates the drawbacks of manual tuning, such as strong subjectivity, low efficiency and poor reproducibility, and provides an automated and standardized solution for model performance optimization.

4.7 Comparative Experiment on Performance and Deployment Efficiency of Different Models

To verify the comprehensive diagnostic performance and clinical deployment practicability of the proposed PCL-

HRNet model, comparative experiments are designed in this section from two dimensions: diagnostic accuracy and deployment efficiency. Traditional medical image diagnosis models including SimpleBaseline (ResNet50)[22][23], the lightweight model YOLOv8-Pose-n[24][25], and medical image keypoint detection baseline models HRNet [26]and HR-SCNet [27][28]are selected for comparison. All models are validated on the same MTDDH test set with a unified input pixel size of 1333×800, and core hyperparameters such as training batch size, optimizer type and learning rate scheduling strategy are kept completely consistent. All experiments are conducted with 10 independent repeated tests. The experimental results are presented as follows.

Table 6 Comparison of IHDI classification performance among different models

Model	Type I Recall (%)	Type II Recall (%)	Type III Recall (%)	Type IV Recall (%)	Macro-F1	Kappa Coefficient
SimpleBaseline(ResNet50)	88.12	63.15	78.22	72.41	0.762	0.738
YOLOv8-Pose-n	89.35	65.47	80.16	73.58	0.781	0.755
HRNet	88.66	61.90	83.33	76.56	0.803	0.779
HR-SCNet	91.86	66.67	83.33	77.78	0.812	0.806
PCL-HRNet	91.38	65.21	84.73	78.79	0.825	0.822

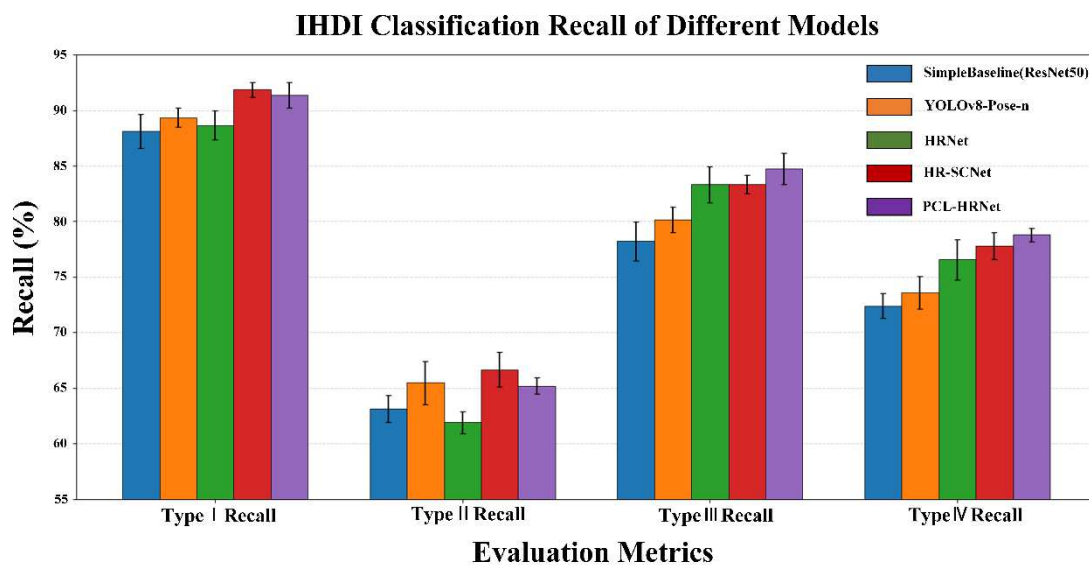


Fig. 14 Comparison of recall rates of different models in the IHDI classification task

The comparison of diagnostic performance among different models is shown in Table 6 and Fig. 14. SimpleBaseline (ResNet50) adopts a single-path downsampling encoding structure, which causes the loss of fine anatomical features during deep feature extraction. It is prone to keypoint localization deviations and performs weakly in identifying grade III and IV severe dislocation cases with abnormal anatomical structures, yielding the lowest overall classification accuracy. As a lightweight keypoint detection model, YOLOv8-Pose-n adopts a shallow network and channel compression design, resulting in insufficient receptive field and high-order feature representation capacity. It fails to fully capture the complex anatomical correlation features required for DDH classification and can hardly meet the demand for high-precision clinical classification. HRNet effectively preserves spatial detail features through parallel multi-resolution branches, addressing the feature loss problem of single-path downsampling, and its classification accuracy is significantly superior to the

aforementioned two models. However, it lacks anatomical prior constraints and tends to produce localization results inconsistent with clinical logic, so there remains a notable bottleneck in its comprehensive classification performance.

HR-SCNet optimizes multi-scale feature extraction and fine-grained representation capability via the structural innovation of spatial-channel reconstruction convolution. It achieves a recall rate of 91.86% for grade I normal hip joints, which account for the largest proportion of samples, higher than that of the proposed PCL-HRNet. The PCL-HRNet proposed in this paper converts prior knowledge of IHDI classification into differentiable constraint loss and embeds it into the model training pipeline. For grade III and IV severe DDH classification, which bears higher clinical intervention value, the recall rates reach 84.73% and 78.79% respectively, both surpassing HR-SCNet. In terms of comprehensive evaluation metrics, PCL-HRNet attains a Macro-F1 value of 0.825 and a Kappa coefficient of 0.822, both higher than all other comparison models. This fully verifies that PCL-HRNet achieves superior comprehensive classification accuracy and diagnostic reliability through the deep integration of clinical priors and model architecture.

Table 7 Comparison of Deployment Performance Among Different Models

Model	Params (M) (M)	FLOPs(G)	Single-image Inference Latency on CPU (ms)	Model Size (MB)
SimpleBaseline(ResNet50)	24.5	12.8	136	98
YOLOv8-Pose-n	3.2	2.1	35	13
HRNet	14.3	7.8	96	57
HR-SCNet	16.0	8.8	110	64
PCL-HRNet	7.5	3.2	45	8

As indicated by the deployment performance comparison in Table 7, there are marked differences in computational complexity and inference efficiency across different models. HR-SCNet embeds a spatial-channel reconstruction convolution module on the basis of HRNet to enhance feature expression capability, with its parameter count and computational complexity increased to 16.0 M and 8.8 G respectively, and single-image CPU inference latency reaching 110 ms. Although it preserves the diagnostic performance of the high-resolution network, its large parameter scale and high computational complexity make it difficult to satisfy the deployment requirements of low-computing hardware in primary medical institutions. SimpleBaseline (ResNet50) has even higher parameter count and computational complexity with longer inference latency, which cannot meet the efficiency demand of rapid screening for large-scale populations. While YOLOv8-Pose-n boasts extreme lightweight advantages, its limited model capacity and insufficient feature extraction capability result in diagnostic accuracy that fails to reach the clinical classification standard for DDH[29].

The proposed PCL-HRNet adopts a progressive lightweight strategy guided by clinical geometric priors, and effectively optimizes model complexity while retaining the key anatomical features and core diagnostic performance of DDH classification. Compared with HR-SCNet, PCL-HRNet reduces parameter count by 51.3%, cuts FLOPs by 61.4%, improves CPU inference throughput by approximately 2.4 times, and compresses model storage size by 87.5%. The experimental results demonstrate that PCL-HRNet achieves a more favorable balance between diagnostic accuracy and deployment efficiency, and can effectively adapt to the low-computing hardware conditions and real-time diagnosis demands of primary clinical scenarios[30][31].

5. Conclusion

Developmental Dysplasia of the Hip is a high-incidence congenital pediatric skeletal disorder, and reliable radiological classification is critical for early intervention and reducing long-term disability rates. This work presents a prior-constrained lightweight high-resolution network (PCL-HRNet) for intelligent IHDI grading of DDH based on pelvic X-ray images.

The proposed approach introduces multi-dimensional geometric constraint loss, which converts clinical priors including pelvic symmetry, physiological range of acetabular angle, and relative position of the femoral head into differentiable constraint terms optimized jointly with MSE loss; meanwhile, a prior-guided three-stage progressive lightweight strategy and the QPSO-AWA global hyperparameter optimization algorithm are integrated to balance diagnostic accuracy and deployment efficiency. The prior constraints reduce the anatomical logic error rate to 0.24% and the acetabular angle MAE to 1.96°, while the lightweight strategy achieves a 45.5% reduction in parameter count and a 53.1% improvement in CPU inference speed with only 0.33% accuracy loss. On the MTDDH dataset, the model achieves a Macro-F1 of 0.825 and a Kappa coefficient of 0.822, outperforms existing baseline models in both diagnostic reliability and deployment feasibility, and can provide interpretable and efficient intelligent decision support for DDH screening in primary care settings.

Acknowledgements The authors would like to sincerely thank all co-authors for their valuable and constructive suggestions during the planning and development of this research work. This work was supported by the National Natural Science Foundation of China (Grant Nos. 82372387, 82502878) and the Fundamental Research Funds for the Central Universities (Grant No. YG2024ZD19).

Author Contributions Shuai Xue: Conceptualization, methodology, software, Data curation, writing—original draft preparation, formal analysis, Validation, writing and editing. Xiaowei Yu: Funding Acquisition, Resources, Investigation, Project administration.

Funding This work was supported by the National Natural Science Foundation of China (Grant Nos. 82372387, 82502878) and the Fundamental Research Funds for the Central Universities (Grant No. YG2024ZD19).

Data Availability In this study, the performance of the proposed model was evaluated on the publicly available Multitasking Developmental Dysplasia of the Hip (MTDDH) dataset. Developed by the Children's Hospital, Zhejiang University School of Medicine, the MTDDH dataset is accessible at <https://doi.org/10.57760/sciencedb.24372>. The data acquisition and utilization protocol was reviewed and approved by the hospital's Medical Ethics Committee (approval No. 2023-IRB-0287-P-01). All data were fully anonymized prior to release, and the requirement for individual patient informed consent was formally waived by the ethics committee.

Declarations

Competing Interests The authors declare no competing interests.

Ethical Approval Not applicable.

Informed Consent Not applicable.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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 <https://creativecommons.org/licenses/by-nc-nd/4.0/>

References

- [1] Xu, Na, et al. "Epidemiological investigation and ultrasonic diagnosis of developmental dysplasia of the hip in Chinese infants: A large multi-center cohort study." *Medicine* 101.2 (2022): e28320.
- [2] Franco H, Saxby N, Corlew D S, et al. An assessment of the impact of developmental dysplasia of the hip on patients' wellbeing: calculation of disability weight and country-specific utility weights[J]. *Bone & joint open*, 2023, 4(3): 120-128.
- [3] Roposch, Andreas, and James G. Wright. "Increased diagnostic information and understanding disease: uncertainty in the diagnosis of developmental hip dysplasia." *Radiology* 242.2 (2007): 355-359.
- [4] Bhavsar, Suketu, et al. "Artificial intelligence to detect developmental dysplasia of hip: A systematic review." *Journal of Paediatrics and Child Health* 61.11 (2025): 1712-1727.
- [5] Litjens, Geert, et al. "A survey on deep learning in medical image analysis." *Medical image analysis* 42 (2017): 60-88.
- [6] Xu W, Shu L, Gong P, et al. A Deep-Learning Aided Diagnostic System in Assessing Developmental Dysplasia of the Hip on Pediatric Pelvic Radiographs[J]. *Frontiers in Pediatrics*, 2022, 9: 785480. DOI: 10.3389/fped.2021.785480.
- [7] an F, Rahman A, Busaleh R, et al. Assessing Acetabular Index Angle in Infants: A Deep Learning-Based Novel Approach[J]. *Journal of Imaging*, 2023, 9(11): 242.
- [8] Chen J, Fan X, Chen Z, et al. Enhancing YOLO5 for the Assessment of Irregular Pelvic Radiographs with Multimodal Information[J]. *Journal of Imaging Informatics in Medicine*, 2024, 37(2): 744-755.
- [9] Liu, C. et al. Misshapen Pelvis Landmark Detection With Local-Global Feature Learning for Diagnosing Developmental Dysplasia of the Hip. *IEEE Transactions on Medical Imaging* 39(12), 3944–54, <https://doi.org/10.1109/TMI.2020.3008382> (2020).
- [10] Yang, Y. et al. How to judge pelvic malposition when assessing acetabular index in children? Three simple parameters can determine acceptability. *Journal of Orthopaedic Surgery and Research* 15(1), 12, <https://doi.org/10.1186/s13018-020-1543-9> (2020).
- [11] Qi G. Q. et al. The MTDDH dataset for quality evaluation of pelvic X-ray and diagnosis of developmental dysplasia of the hip[Internet]. *Science Data Bank*. <https://doi.org/10.57760/sciencedb.24372> (2025).
- [12] Salahuddin, Zohaib, et al. "Transparency of deep neural networks for medical image analysis: A review of interpretability methods." *Computers in biology and medicine* 140 (2022): 105111.
- [13] St George, J. et al. Importance of early diagnosis for developmental dysplasia of the hip: A 5-year radiological outcome study comparing the effect of early and late diagnosis. *Journal of Paediatrics and Child Health* 57(1), 41–5, <https://doi.org/10.1111/jpc.15111> (2021).
- [14] 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. *Scientific Reports* 13(1), 6693, <https://doi.org/10.1038/s41598-023-33860-2> (2024).
- [15] Zhang, S.-C. et al. Clinical application of artificial intelligence-assisted diagnosis using anteroposterior pelvic radiographs in children with developmental dysplasia of the hip. *Bone & Joint Journal* 102B(11), 1574–81, <https://doi.org/10.1302/0301-620X.102B11.BJJ-2020-0712.R2> (2020).
- [16] Cheng, Yu, et al. "Model compression and acceleration for deep neural networks: The principles, progress, and challenges." *IEEE Signal Processing Magazine* 35.1 (2018): 126-136.
- [17] Yang, L. & Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415, 295-316.
- [18] Dhanka, Sanjay, et al. "Advancements in hybrid machine learning models for biomedical disease classification using integration of hyperparameter-tuning and feature selection methodologies: A comprehensive review." *Archives of Computational Methods in Engineering* 33.1 (2026): 289-324.
- [19] Qi, Guoqiang, et al. "A dataset for quality evaluation of pelvic X-ray and diagnosis of developmental dysplasia of the

hip." *Scientific Data* 12.1 (2025): 865.

- [20] Bimodal machine learning model for unstable hips in infants: integration of radiographic images with automatically-generated clinical measurements[J]. *Scientific Reports*, 2025, 15(1): 16227.DOI: 10.1038/s41598-025-16227-2
- [21] Sha S, Li X, Wang Z, et al. Quantum intelligence in medicine: Empowering thyroid disease prediction through advanced machine learning[J]. *IET Quantum Communication*, 2024, 5(3): 12078.DOI: 10.1049/qtc2.12078
- [22] Xiao B, Wu H, Wei Y. Simple Baselines for Human Pose Estimation and Tracking[C]//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich, Germany: Springer, 2018: 466-481. DOI: 10.1007/978-3-030-01231-1_29
- [23] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [24] Liu B, Song J. Improved YOLOv8n-Pose Lightweight Joint Point Detection Algorithm[C]//*2025 IEEE International Conference on Computer Vision, Image and Deep Learning (CVIDL)*. Xiamen, China: IEEE, 2025.
- [25] Lai Q, Chen W, Ding X, et al. Quality control of elbow joint radiography using a YOLOv8-based artificial intelligence technology[J]. *European Radiology Experimental*, 2024, 8: 107.DOI: 10.1186/s41747-024-00504-724.
- [26] Wang J, Sun K, Cheng T, et al. Deep High-Resolution Representation Learning for Visual Recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 43(10): 3349-3364.DOI: 10.1109/TPAMI.2020.3017468
- [27] Liu J, Hou Q, Cheng M, et al. Improving Convolutional Networks with Self-Calibrated Convolutions[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, 2020: 10096-10105.DOI: 10.1109/CVPR42600.2020.01011
- [28] Zhang L, Wang H, Li Y, et al. HR-SCNet-based computer-aided diagnosis model for developmental dysplasia of the hip in pediatric pelvic radiographs[J]. *Chinese Journal of Digital Medicine*, 2024, 19(10): 1001-1008.
- [29] Van Campenhout C, De Maeseneer M, Van Glabbeek F. Tönnis and the Novel IHDI Radiographic Classification Systems for the Developmental Dysplasia of The Hip (DDH): Evaluation of 406 hips with DDH[J]. *Acta Orthopaedica Belgica*, 2023, 89(1): 21-28.DOI: 10.53156/aob.2023.89.1.21
- [30] Li Y, Wang H, Liu X, et al. Deep Learning-Based Automatic Diagnosis System for Developmental Dysplasia of the Hip with Explainable Multi-Parameter Scoring[J]. *Journal of Clinical Medicine*, 2025, 14(11): 3267.
- [31] Zhang Q, Xu L, Shen H, et al. Clinical thought-based software for diagnosing developmental dysplasia of the hip on pediatric pelvic radiographs[J]. *Frontiers in Pediatrics*, 2023, 11: 1080194.

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

Authors and Affiliations

Shuai Xue¹² · Xiaowei Yu²¹ (Corresponding Author)

✉ Shuai Xue

xueshuai0925@163.com

✉ Xiaowei Yu

yuxw@sjtu.edu.cn

¹ College of Engineering, Shanghai Ocean University, Shanghai 201306, P. R. China

² Department of Orthopedics, Shanghai Sixth People's Hospital, Shanghai 201306, P. R. China