



QualityDDH: visualized standardization of neonatal hip ultrasound via a structural prior regression framework

Ruhan Liu¹ · Yuan Zhang³ · Xiaoxiao Luo³ · Yiwen Zheng³ · Qirong Liu⁵ · Mengyao Liu^{2,3} · Lixin Jiang^{3,4}

Accepted: 14 July 2025

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2025

Abstract

Early ultrasound screening of developmental dysplasia of the hip (DDH) is crucial for timely intervention and preventing hip replacement. However, the lack of standardization in image acquisition during DDH ultrasound screening often hinders the accuracy and consistency of the screening process, posing a significant challenge. Therefore, there is an urgent need to develop automated, visual, and high-precision methods to assist in image standardization. Conventional quality classification methods, which rely on the recognition and judgment of anatomical structures, struggle to achieve accurate results. To address this, we propose the QualityDDH framework, a visual quality assessment tool. It uses structural priors to obtain key structural segmentation maps and assesses anatomical availability based on standardized guidelines. We applied the QualityDDH framework to clinical prospective validation. It assisted ultrasound physicians of different levels in making standardized judgments using an independent external validation dataset of 600 infants. The QualityDDH framework improved performance for ultrasound physicians at all levels: Expert (Area Under the Receiver Operating Characteristic Curve, AUC, increased by 4.70%), Attending (AUC increased by 12.95%), and Resident (AUC increased by 20.85%). This lays the foundation for the clinical application of intelligent auxiliary screening for DDH. Code available at: <https://github.com/Liuruhan/QualityDDH>.

Keywords Ultrasound image (US) · Developmental dysplasia of the hip (DDH) · Dual attention · Detection network

1 Introduction

Developmental dysplasia of the hip (DDH) is a prevalent pediatric musculoskeletal disorder, where abnormal hip joint development can cause long-term functional issues without early diagnosis and treatment [1, 2]. Ultrasound (US) imaging, particularly the Graf method, is a reliable diagnostic approach for DDH [3–5]. Created by Graf in 1980, Graf method assesses hip joint development and stability via US, aiding DDH management [6].

As technology progresses [7, 8], standards for DDH US examinations, championed by Professor Graf, have evolved. The current Graf method stresses obtaining standard images for DDH diagnosis [8], which involves two checklists: one for anatomical identification (Checklist I) and another for usability (Checklist II), as detailed in Fig. 1. However, accurately identifying these fine details during image acquisition is time-consuming and depends on the operator's expertise [9]. Consequently, less experienced sonographers may produce substandard scans, risking misdiagnosis. Literature [10] shows that nearly half of studies fail to meet Graf quality criteria, often due to inadequate anatomical marker description

✉ Mengyao Liu
mengyao_liu08@163.com

✉ Lixin Jiang
jinger_28@sina.com

Ruhan Liu
223101@csu.edu.cn

Yuan Zhang
columbianzhang@163.com

Qirong Liu
liuqirong5833@link.tyut.edu.cn

¹ Furong Laboratory, Central South University, Hunan, China

² Department of Ultrasound, Shanxi Provincial People's Hospital, Affiliate of Shanxi Medical University, Taiyuan 030012, Shanxi, China

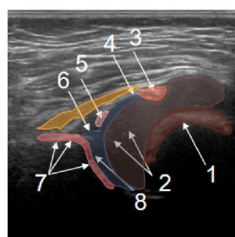
³ Department of Ultrasound, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

⁴ Jiading District Central Hospital Affiliated Shanghai University of Medicine and Health Sciences, Shanghai, China

⁵ College of Software, Taiyuan University of Technology, Taiyuan, China

Checklist 1:**Anatomical identification:**

- 1 – chondroosseous border (CB),
- 2 – femoral head (FH),
- 3 – synovial fold (SF),
- 4 – joint capsule (JCP),
- 5 – labrum (La),
- 6 – cartilaginous roof (CR),
- 7 – bony roof (BR),
- 8 – bony rim (concavity–convexity)

Meet the **Checklist 1****Checklist 2:****Usability Check:**

- 1 – the lower limb of the os ilium in the fossa acetabuli must be visible
- 2 – the os ilium contour must be straight or parallel to the probe
- 3 – the labrum must be visible

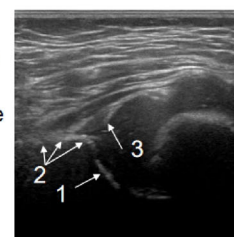
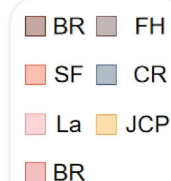
Meet the **Checklist 2****Key structures in hip joint**

Fig. 1 Standards for DDH US examinations, including Checklist I for anatomical identification and Checklist II for usability assessment. Checklist I requires the visibility of eight key hip joint structures, including the chondro-osseous border (CB), femoral head (FH), synovial fold (SF), joint capsule (JCP), labrum (La), cartilaginous roof (CR), bony

roof (BR), and bony rim (concavity–convexity). Checklist II focuses on usability assessment, requiring that the lower limb of the os ilium in the fossa acetabuli must be visible, the os ilium contour must be straight or parallel to the probe, and the labrum must be visible

or non-standard images. Moreover, inconsistent image acquisition can lead to misdiagnosis, highlighting the need for standardized imaging techniques. Jaremkó et al. [11] found that variations in US probe direction could cause misdiagnoses in many cases, further emphasizing the necessity of standardized imaging techniques for accurate DDH diagnosis.

While many researchers are considering using deep learning technology for medical imaging [12–14], existing methods based on binary classification models and key point detection have limitations in clinical practice [15–18]. These methods concentrate more on diagnosing and classifying DDH in standardized images, which limits their usability in actual clinical settings with non-standardized US interference. Also, our previous work [17, 18] put forward the segmentation of seven key hip joint structures for visualizing standardization and DDH severity classification for clinicians. However, subsequent clinical use showed that merely offering key structure segmentation results, while somewhat reducing non-standard imaging issues, is still user-dependent and has limited benefits for inexperienced junior doctors. To address these challenges, we propose the QualityDDH framework, which, on the basis of key hip joint structure segmentation, adds visual quality assistance evaluation on external dataset, and compare the framework with six sonographers of different levels in actual clinical practice. The contributions of the framework are as follows:

- We proposed an interactive learning-based visual quality prompt annotation strategy in line with the quality assessment guidelines. This helps experts reach a consensus on the standardized assessment of DDH US.
- We designed a relative position-sensitive iliac structure assessment module. It can quantify the standardization criteria using the segmentation map of the iliac structure.

Also, it provides a visual standardization judgment for Checklists I and II.

- We carried out methodological validation of the proposed method on an external dataset named SPD dataset of 600 standard planes from 600 participants. Our method achieved an accuracy of 85.94% (internal test set) and 97.33% (external set) in standardization judgment, which is superior to traditional classification networks.

The remainder of this paper is organized as follows: Sect. 2 presents the literature review; Sect. 3 elaborates on the proposed methodology; and Sect. 4 details the experimental setup, results, as well as the comparison and analysis.

2 Related work

This section systematically examines the role of computer-aided algorithms in enhancing the standardization of medical images, with a particular focus on DDH US images. The analysis is structured around three key dimensions: first, the implementation of quality assessment and control techniques in medical imaging; second, the advancement of ultrasound-based recognition and segmentation methodologies for neonatal hip joint and cartilage structures; and third, the development of automated DDH screening and diagnostic grading systems grounded in ultrasonic imaging.

2.1 Medical image quality assessment and control methods

Medical image quality control is crucial for accurate diagnosis and effective treatment planning. Many studies have categorized quality assessment methods into visual quality-based and task-based approaches [19], highlighting the significance of traditional image quality metrics and empha-

sizing the importance of specialized metrics for medical tasks. For example, Chow et al. [20] evaluated the impact of various full-reference metrics on MRI images and found that noise quality measurement (NQM) had the best correlation with subjective ratings. This underscores the importance of quality assessment in medical contexts.

Ultrasound imaging, favored for its non-invasive and real-time capabilities, faces quality challenges like speckle noise and motion artifacts. Razaak et al. [21] developed the Cardiac Ultrasound Video Quality Index (CUQI), which effectively assesses cardiac ultrasound video quality by integrating cardiac motion and structural information, showing strong correlation with subjective scores. Panayides et al. [22] found the Weighted Signal-to-Noise Ratio (WSNR) to be the most effective metric for atherosclerotic plaque ultrasound videos. However, more technologies focus on the noise interference of image acquisition, and less consideration is given to the measurement of the standardization of anatomical structures in ultrasonic images. The research on related quality standards and methods is lacking.

2.2 Hip joint structure segmentation methods

Accurate segmentation of anatomical structures in non-standardized images is crucial for developing standardized quality assessment indicators. Early research focused on ilium segmentation using manually extracted features [23–28]. However, these methods had poor generalization and were complex, mainly focusing on high-echo structures with limited applicability. Subsequently, deep learning techniques were applied to anatomical structure segmentation [17, 29], but most were used on manually standardized ultrasound images, making it hard to evaluate their effectiveness in segmenting structures in unstandardized images. Given the limitations of existing methods in handling non-standardized images, there is an urgent need to develop more robust and adaptive segmentation approaches. The automatic DDH diagnosis approaches should be capable of accurately identifying anatomical structures across a wider range of image conditions. This would significantly enhance the reliability and effectiveness of automated diagnostic systems in real-world clinical settings. Future research should focus on improving the generalization and applicability of segmentation algorithms, particularly for non-standardized images. This could involve exploring advanced deep learning architectures or incorporating additional contextual information to better handle image variations and improve segmentation accuracy.

2.3 Deep-learning methods in DDH severity classification

Currently, there is a growing focus among researchers on developing automatic DDH diagnosis methods based on US images, with deep learning-based approaches gaining particular attention. These automated diagnostic methods can be classified into three main types. The first type is end-to-end classification methods, such as the two-stage meta-learning-based deep exclusive regularization machine presented by Gong et al. [15]. The second type relies on anatomical structure segmentation and key point detection [16, 30]. The third type is based on multi-task networks, like the AutoDDH method [18] we proposed in our previous study, which combines anatomical structure features for accurate DDH classification. However, similar to the segmentation methods for hip joint structures, these approaches generally do not address the standardization of ultrasound images and often assume that the input images are standardized. This assumption poses challenges to the stability of the models. Consequently, developing a DDH quality assessment method based on anatomical structures and standardized guidelines [6] is of great significance and requires urgent research. Such a method could enhance the reliability and effectiveness of automated DDH diagnosis in clinical practice.

3 Method

The QualityDDH framework is a visual AI tool designed to assist sonographers in ensuring the collection of ultrasonic standard planes for DDH screening. As illustrated in Fig. 2, the framework comprises three modules. First, NHBS-Net [17] performs anatomical structure segmentation and automatically segments seven structures related to standardization [7, 8]. Second, Checklist I module calculates the number of pixels in each structure and provides non-standard prompts based on the threshold values of each structure. Third, Checklist II module extracts the iliac bone edge line and classifies it into the iliac bone plane and the lower limb of the iliac bone based on its position coordinates. The iliac bone plane angle is calculated based on the position coordinates of the iliac bone plane, and parallel non-standard prompts are provided. Additionally, visual reminders are given based on the length of the lower limb of the iliac bone.

3.1 Key structure segmentation in hip joint

The key structure segmentation model is based on the NHBS-Net network framework. In addition to the original NHU

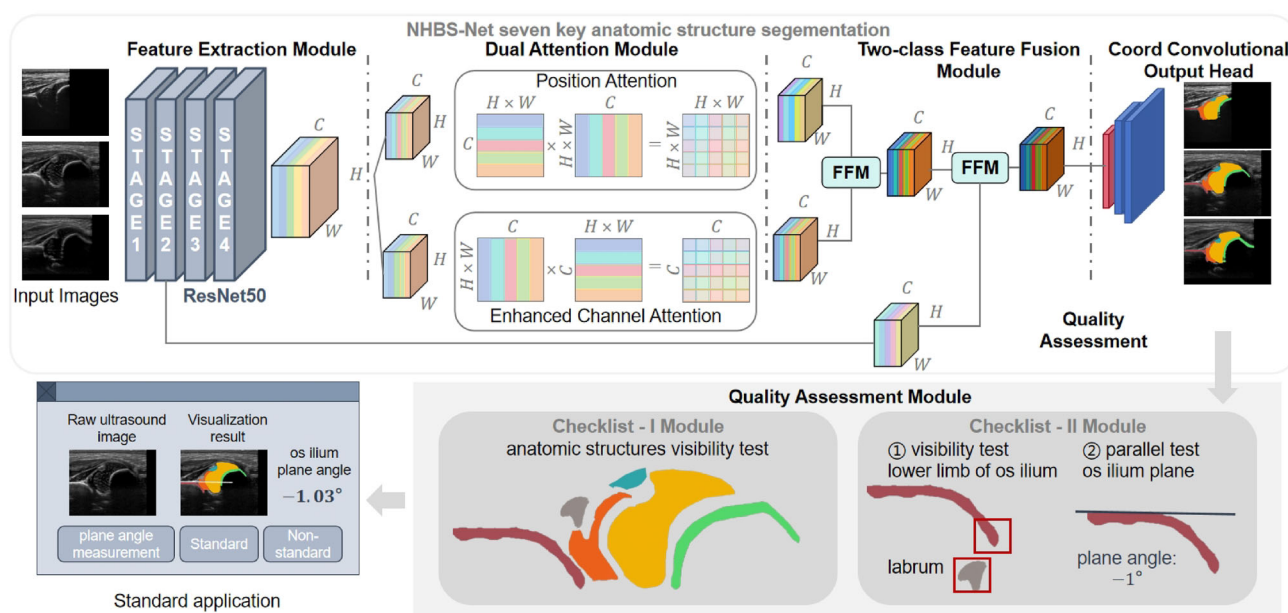


Fig. 2 Overview of QualityDDH. QualityDDH system contains three parts: NHBS-Net for key structures segmentation, Quality Assessment Module (Checklist I module and Checklist II Module) for evaluating the quality of input US images, and Standard application to provide the visual quality assessment results

data set, some non-standard images are added for training to improve the model's performance in non-standard images. See [17] for the specific network architecture.

3.2 Interactive quality assessment based on checklists

Previous studies on quality assessment have shown significant individual differences among experts [14, 31]. In response, we propose an AI-assisted visual quality assessment method for human–computer interaction with ultrasound physicians. This method not only enhances the objectivity and consistency of quality assessment but also provides real-time visual feedback to assist physicians in making more accurate judgments. Furthermore, we conduct a detailed comparison between this AI-assisted method and the traditional expert back-to-back consensus approach. The results demonstrate that our proposed method achieves higher consistency and reliability in quality assessment. It effectively reduces the impact of individual differences among experts. This makes it a promising alternative to traditional expert consensus methods.

In line with DDH ultrasound standards, we've developed the Checklist-I module (Fig. 3a). It calculates pixel counts for each structure using the seven key structural diagrams from our key structure segmentation. We also determined the pixel counts for all structures in the NHU dataset. Structures with pixel counts below 60% of the standard image's average are highlighted. Beyond pixel counts, we supply the average

pixel value per structure and flag those below 60% of the standard. These dual highlights aid ultrasound physicians in visually assessing Checklist-I standardization.

Our proposed Checklist-II module (Fig. 3b) focuses on anatomical structure availability identification, based on structural visibility judgment. Since the visibility of structure La has been confirmed in the Checklist-I module, the Checklist-II module concentrates on the morphology of RF. The coordinates of RF's upper edge line are extracted from its segmented image. A coordinate classification model is designed based on these coordinates. Regression model processes the edge line coordinates to predict the parallel line and the lower edge of the ilium. Additionally, a linear regression model predicts the angle of the parallel line using the coordinates of the ilium's parallel line. If the angle error exceeds 40% of the standard image, it is highlighted. For the lower edge of the ilium, if its length is below the empirical threshold, it is also highlighted. These dual prompts offer ultrasonic physicians visual standards for Checklist-II.

3.3 Structural prior position regression (SPPR) module

The Checklist-II module incorporates the SPPR module to analyze the ilium structure, and the upper edge line of the ilium distinguishes the plane of the os ilium and the lower limb of the os ilium via SPPR. The parallel angle is calculated through the plane of the os ilium, and the lower edge angle is determined through the lower limb of the os ilium.

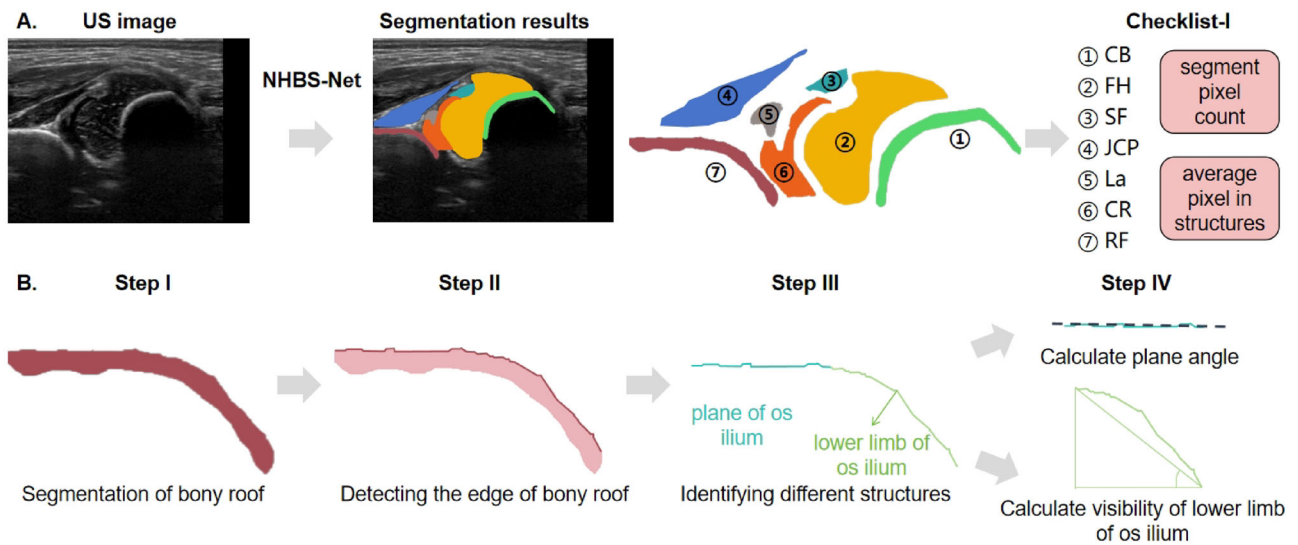


Fig. 3 Procedures in Checklist-I and Checklist-II module. **a** The Checklist-I module involves several procedures. First, key structure segmentation is performed using NHBS-Net. Then, the module conducts segment pixel counting and average pixel value calculation for each structure. These steps help in assessing the standardization of the

structures based on pixel-related metrics. **b** The Checklist-II module includes four steps: segmenting bony roof, detecting the upper edge of bony roof, identifying the parallel line and the lower edge of the ilium, calculating the plane angle, and calculating visibility of lower limb of os ilium

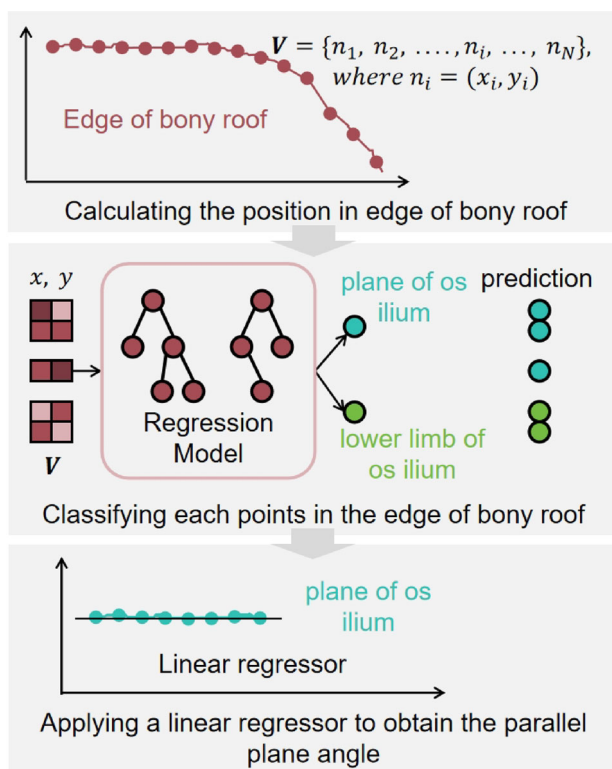


Fig. 4 Schematic diagram of calculating parallel angle in SPPR module. Calculating parallel angle via SPPR module includes three steps: calculating the position in edge of bony roof, classifying each point into plane of os ilium or lower limb of os ilium, and applying a linear regressor to obtain the parallel plane angle

Firstly, the absolute position coordinates of the upper edge line $V = \{n_1, n_2, \dots, n_i, \dots, n_N\}$ are obtained, where $n_i = (x_i, y_i)$ represents the coordinates of the i -th point on the upper edge line. Secondly, the relative position coordinates $V' = \{n'_1, n'_2, \dots, n'_i, \dots, n'_N\}$ are calculated, where $n'_i = (x_i - x_1, y_i - y_1)$. Thirdly, the relative position coordinate point set V' is sampled by sliding window to generate the sampled point set matrix $M \in \mathbb{R}^{L_{sw} \times (N - L_{sw})}$. The point set matrix M is used for outputting the regression model to classify the point coordinates. To calculate the parallel angle, absolute coordinates of certain points on the iliac plane are input into a linear regressor to fit a parallel line. The slope of this line determines the parallel angle of the iliac plane. Additionally, the angle of the lower iliac edge is found by calculating the angle of the line connecting the start and end points of the lower edge (Figs. 4 and 5).

3.4 Training and testing strategy of QualityDDH

During the training process, the regression model classifier is fed with the coordinates of the iliac edge line extracted from structural segmentation annotations. The input comprises fixed-length point coordinate pairs, with the maximum length set to the number of points on the longest iliac edge line. If the input length is shorter than this maximum, it is padded with zeros to meet the required length. In the testing phase, NHBS-Net predicts the structural segmentation. Based on this predicted segmentation, the iliac edge line is extracted. The coordinates of this edge line are then extended

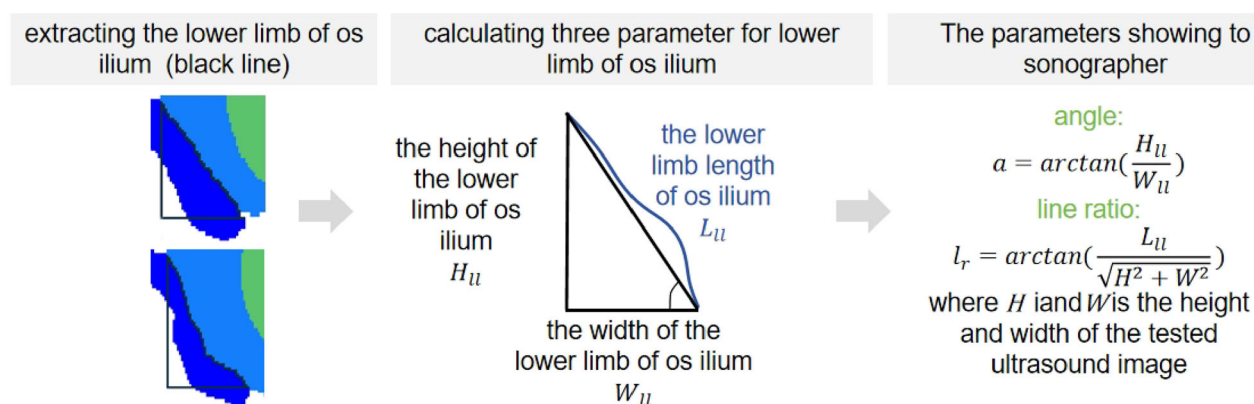


Fig. 5 Schematic diagram of calculating the angle of the lower edge of the iliac bone in SPPR module. Calculating the angle of the lower edge of the iliac bone via SPPR module includes three steps: extracting the

lower limb of os ilium, calculating three parameters for lower limb of os ilium, and calculating the line ratio

to form a specified-length point coordinate pair. This pair is input into the trained regression model classifier to determine whether the points belong to the iliac plane or the lower limb of the ilium. The parallel angle of the iliac bone is calculated using the points identified as part of the iliac plane. Meanwhile, the length and angle of the lower edge of the iliac bone are computed based on the points classified as part of the lower limb of the ilium.

In clinical practice, ultrasound image data is first input into NHBS-Net to obtain segmentation results for seven anatomical structures. Then, based on these results, the pixel counts of the seven structures are calculated, and visibility alerts are provided according to a threshold (structures with values below the given threshold receive an 'unstandard' alert), which involves the Checklist I judgment. Next, the point coordinates that form the iliac bone edge line are extracted from the bony roof segmentation results. These coordinates are extended to a specified length and fed into a trained regression model classifier to obtain two categories: plane of os ilium and lower limb of os ilium. The parallel angle of the iliac bone is calculated using the points in the plane of os ilium category, while the length and angle of the lower edge of the iliac bone are determined using the points in the lower limb of os ilium category. Parallelism and lower edge visibility alerts are provided, which involves the Checklist II judgment. Finally, the visual standards from the Checklist I and Checklist II judgments are used to assist ultrasound doctors in image acquisition.

4 Experiments

4.1 Dataset

This study was approved by the Ethics Committee of the medical institution (KY2021-286-B). The study involved three datasets retrospective collected from two hospitals, namely the Neonatal Hip US (NHU) dataset, the standard plane judgment (SPJ) dataset (https://github.com/hidden-ops/NHBS-Net_SPJ_dataset) (Fig. 6), and the Standard Plane Determination (SPD) dataset. All datasets were NHU dataset and SPJ dataset were used for model development and internal validation. SPD dataset was used for external validation. As the study utilized fully anonymized reports and US images, the need for informed consent was waived.

The NHU and SPJ datasets from Shanghai Sixth People's Hospital were employed for constructing and internal evaluating the preliminary application value of the QualityDDH system. NHU dataset including 563 images with seven key structures annotated were marked and proofread by two sonographers with years of experience. NHU dataset was used for QualityDDH construction and SPJ dataset includes a total of 50 cases, of which 25 cases are standard and 25 cases are non-standard (also seen in the NHBS-Net paper [17]). A total of 613 US images, each of which has a seven-structure segmentation tag, and in the predicted bony roof structure, the edge of bony roof is classified as the classification of plane of os ilium and the lower limb of os ilium. The combination of the NHU and SPJ datasets was randomly divided into training and internal testing sets in a ratio of 8:2.

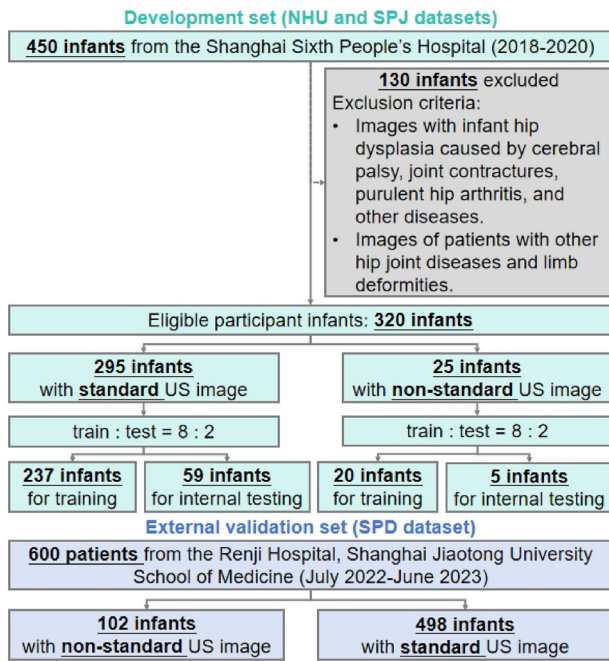


Fig. 6 Flowchart showing the datasets involved in this study and the exclusion criteria in development set

The SPD dataset was obtained from Renji Hospital, Shanghai Jiaotong University School of Medicine and served as an external validation set for StandardDDH system and for six sonographers' reader study. The SPD dataset was retrospectively collected from Renji Hospital, Shanghai Jiao Tong University School of Medicine during July 2022 to June 2023, and included a total of 600 two-dimensional US images from children aged 0 to 6 months. A total of 498 images were randomly selected according to the incidence rates of different types of DDH, among which 291 were type I, 190 were type IIa/IIb, 17 were type IIc, and 2 were type III/IV. In addition, 20% of non-standard images (102 images) were specially selected. All images were selected by independent sonographers with more than three years' experience in DDH diagnosis who were not involved in this study. During hip ultrasonography, infants were positioned in a lateral decubitus position with a slight hip flexion, and a 5–7.5 MHz linear probe was placed at the greater trochanter of the femur on the lateral side of the hip, maintaining parallel alignment with the body's long axis.

4.2 Implementation details

All experiments in our study were conducted on an Intel (R) Core (TM) CPU i9-14900K @ 3.20GHz and two NVIDIA GeForce RTX 4090D GPUs, operating on the Ubuntu Linux platform. Both the proposed QualityDDH framework and other state-of-the-art methods utilized the same loss function design and grid search strategy for hyperparameter tuning.

The segmentation labels were provided by two senior sonographer experts. These experts determined the boundary between the iliac plane and the lower edge based on the iliac crest line, dividing it into the plane of os ilium and the lower limb of os ilium. If the boundary points marked by the two experts differed by more than 10 pixels, a third senior expert was involved to mark the boundary. The midpoint of the boundary points marked by the two closest experts was used as the gold-standard boundary. Conversely, if the difference was 10 pixels or less, the midpoint of the boundary points marked by the two experts served as the gold-standard boundary.

4.3 Evaluation metrics

In our study, statistical analyses were performed using MedCalc 20.0 software. Categorical data were reported as frequencies (percentages), while measurement data following a normal distribution were expressed as mean standard deviation. Fleiss kappa is used to represent the reliability of multiple observers' assessments of binary categorical variables. Cohen kappa is used to represent the reliability of two observers' assessment of binary categorical variables. AUC (area under the curve of receiver operating characteristic), sensitivity, and specificity were used to evaluate the diagnostic performance of sonographers in accurately identifying DDH standard images. A p value of < 0.05 was considered indicative of a statistically significant difference. A kappa value less than 0.20, 0.40, 0.60, 0.80, and 1.0 indicates Poor, Slight, Moderate, Substantial, and Excellent agreement, respectively.

4.4 Experimental results

In the experimental results, we first compared the QualityDDH framework with other traditional quality classification models in terms of DDH standardization. The compared classification models included: ResNet-50 [32], DenseNet-121 [33], Dilated ResNet-50 [34], visual transformer (ViT) [35], BiSe-Net [36], and NHBS-Net [17]. Next, we further evaluated the performance of the QualityDDH framework in anatomically related usability discrimination for Checklist-I through independent external validation. Then, we compared the accuracy of ultrasound physicians in standardization discrimination before and after AI assistance. Finally, we conducted ablation experiments to analyze the impact of module design and training strategies on the performance of the QualityDDH model.

4.4.1 Comparison performance of DDH standardization

We designed a comparative study to evaluate the classification accuracy of the QualityDDH framework against six

Table 1 Comparison performance of our QualityDDH framework and other state-of-the-art methods, including ResNet-50 [32], DenseNet-121 [33], Dilated ResNet-50 [34], visual transformer (ViT) [35], BiSe-Net [36], NHBS-Net [17], in internal and external test datasets

Methods	Acc (%)	Spe (%)	Sen (%)	Pre (%)	F1 (%)	κ
Internal-test dataset						
ResNet-50	76.56	79.66	40.00	14.29	21.05	0.1078
DenseNet-121	79.69	83.05	40.00	16.67	23.53	0.1405
Dilated ResNet-50	81.25	83.05	60.00	23.08	33.33	0.2485
ViT	82.81	84.75	60.00	25.00	35.29	0.2727
NHBS-Net	79.69	81.36	60.00	21.43	31.58	0.2268
BiSe-Net	79.69	81.36	60.00	21.43	31.58	0.2268
Ours (QualityDDH)	85.94	86.44	80.00	33.33	47.06	0.4050
External-test dataset						
ResNet-50	90.33	92.97	77.45	69.30	73.15	0.6728
DenseNet-121	92.93	95.53	80.39	78.85	79.61	0.7534
Dilated ResNet-50	93.27	95.73	81.37	79.81	80.58	0.7651
ViT	94.11	96.34	83.33	82.52	82.93	0.7937
NHBS-Net	93.60	95.93	82.35	80.77	81.55	0.7768
BiSe-Net	93.43	95.93	81.37	80.58	80.98	0.7701
Ours (QualityDDH)	97.33	97.79	95.10	89.81	92.38	0.9077

*Each network is the segmentation model added with a classification output header (one average pooling layer and one linear output layer)

Boldface values denote the highest (best) score for the given metric

commonly used quality assessment models. The framework showed superior performance in both internal and external validation datasets, achieving the highest accuracy, sensitivity, specificity, F1 score, and kappa value. In internal validation, limited by the low proportion of non-standard data, all models obtained a low kappa κ value. Compared to other models, the QualityDDH model demonstrated a 3.13% increase in accuracy, a 1.69% increase in specificity, a 20.00% increase in sensitivity, an 8.33% increase in precision rate, an 11.77% increase in F1 score, and a 0.1323 improvement in kappa κ value. In external validation, it showed a 3.22% increase in accuracy, a 1.45% increase in specificity, an 11.77% increase in sensitivity, a 7.29% increase in precision rate, a 9.45% increase in F1 score, and a 0.114 improvement in kappa κ value. Table 1 details each model's classification results on both validation datasets, clearly showing that the QualityDDH model significantly outperforms other classification algorithms in sensitivity, precision rate, F1 score, and kappa κ value.

As shown in Fig. 7, we compared the confusion matrices of these methods both internally and externally. Compared to other quality assessment classification networks, our QualityDDH performed best in both internal and external validation, achieving accuracy rates of 85.94% and 97.33%, respectively. The quality assessment framework implemented by QualityDDH significantly reduces false positives and false negatives, not only providing visual discrimination prompts but also better capturing the features of non-standard images compared to traditional classification networks.

4.4.2 Performance of identifying landmark in Checklist-I

Figure 8 shows examples of diagnostic process using the QualityDDH model for standard plane determination and compares them with golden standard, including examples of common standard and non-standard images. The QualityDDH system was used to evaluate the 600 images in the SPD dataset to obtain seven different structural landmarks in Checklist-I. The QualityDDH AI system achieved 0.967 accuracy rate, 0.987 precision rate, 0.970 sensitivity rate, 0.894 specificity rate, and 0.979 F1 score. The final results showed that the AUC of the QualityDDH model was 0.941, as shown in Fig. 9 and Table 2. The AUC values of the QualityDDH system in S1-S8 were 0.694, 0.671, 0.906, 0.832, 0.939, 0.825, 0.927 and 0.841, respectively, the sensitivity of its structure were higher, ranging from 0.987 to 1.000, and the specificity were lower (0.389–0.893). In addition, by comparing with the golden standard, the final judgment results of the QualityDDH system have excellent agreement (0.900), also with excellent agreement in SF, FH, and JCP, and substantial agreement in CR and CB, with moderate agreement in BR and La.

4.4.3 Performance of DDH standardization in AI-aided sonographers

In addition, we conducted a detailed diagnostic performance and diagnostic consistency study on eight important anatomical structures and landmarks in the standard plane determination process (Fig. 9). Without the assistance of

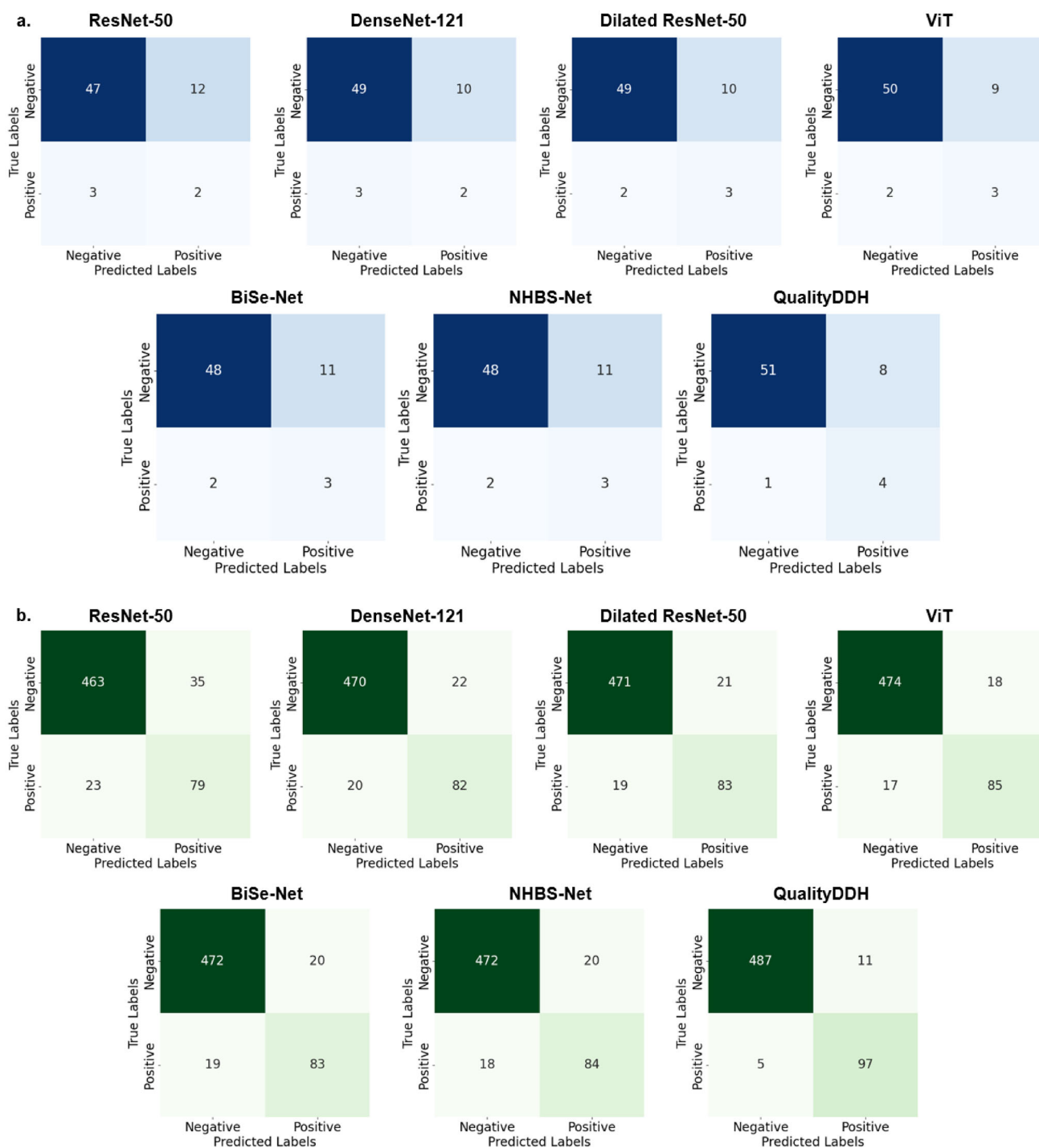


Fig. 7 Confusion matrixes of compared classification models, including QualityDDH, ResNet-50 [32], DenseNet-121 [33], Dilated ResNet-50 [34], visual transformer (ViT) [35], BiSe-Net [36], and NHBS-Net [17]

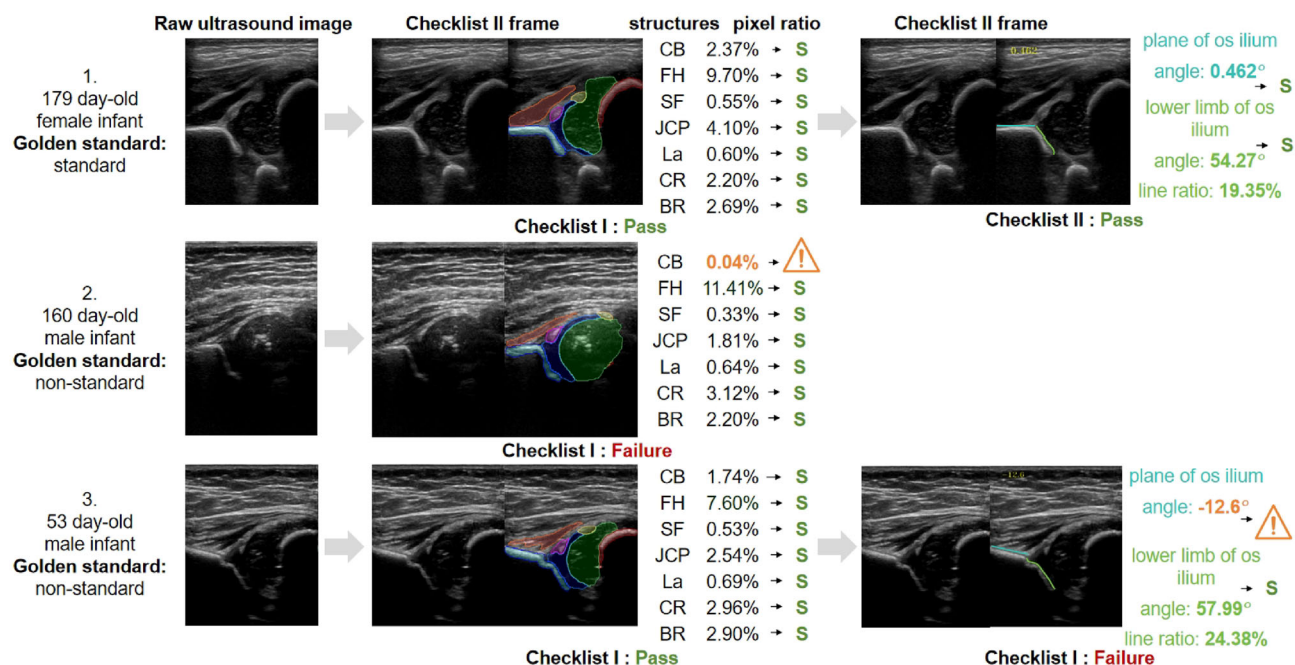


Fig. 8 Visualization examples of QualityDDH framework aid sonographers. Example 1 in row 1 shows the AI-aided process in a standard US images. Example 2 and 3 in row 2 and 3 show the AI-aided steps in non-standard US images, which are fail to pass checklist I and checklist II, respectively

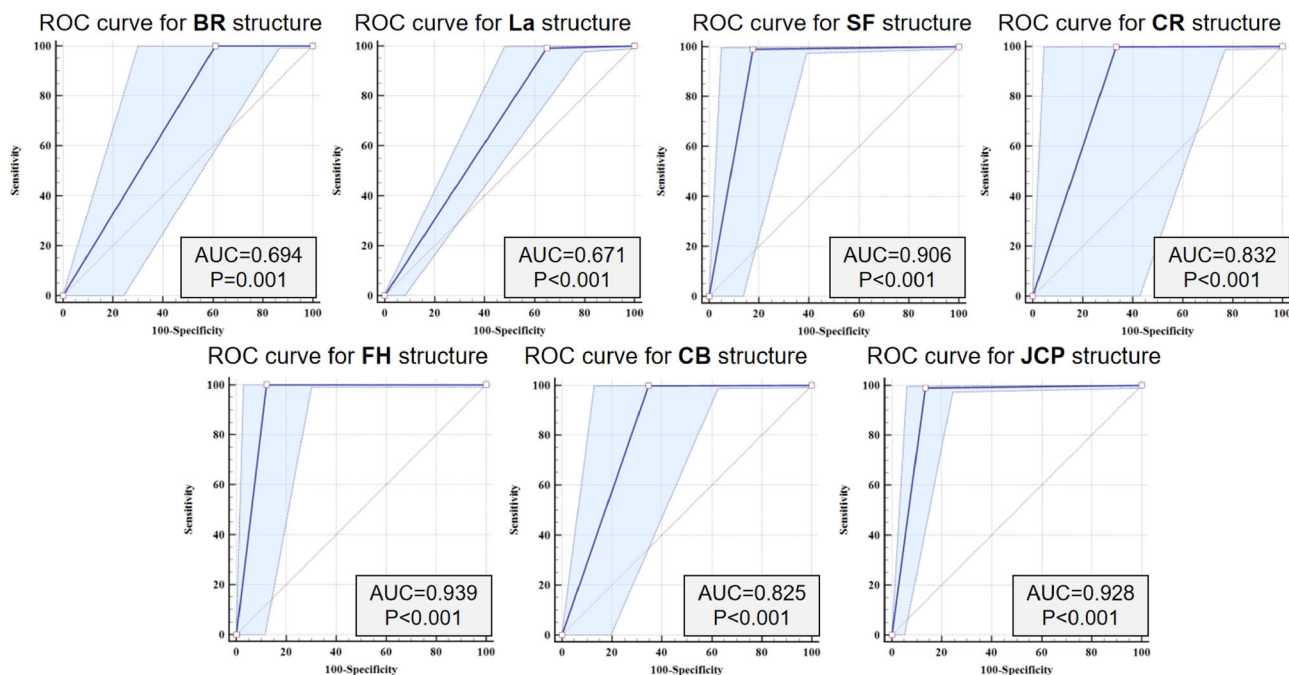


Fig. 9 ROC curves of identifying seven key structure in external validation dataset

Table 2 Identification performance of the QualityDDH framework compared with the golden standard standardization evaluations

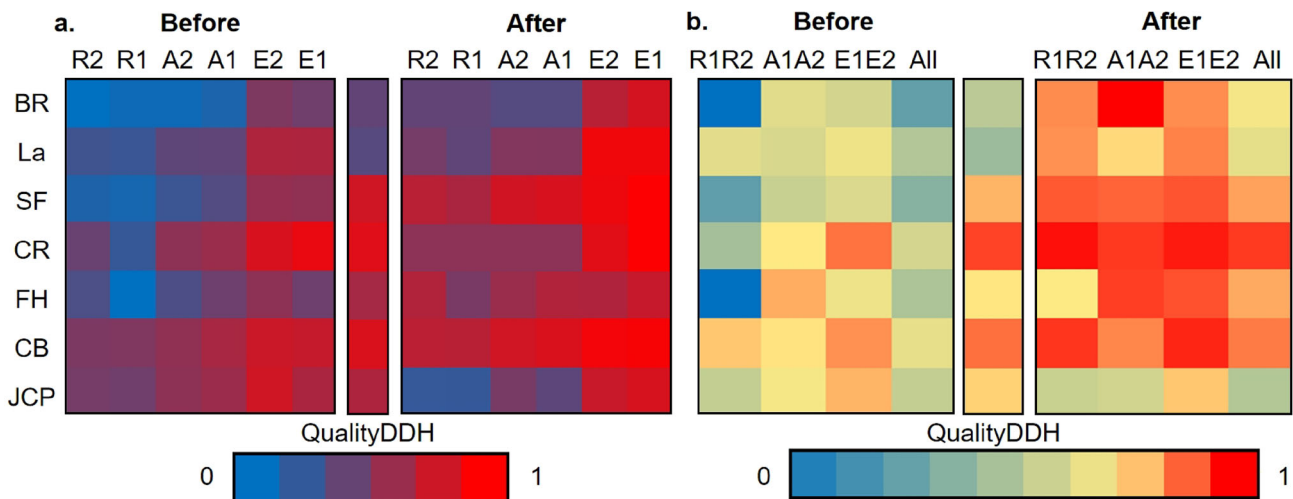
Structures	AUC	Sen	Spe	κ
BR	0.694	1.000	0.389	0.552
95% CI	(0.656 0.731)	(0.994 1.000)	(0.173 0.643)	(0.318 0.787)
La	0.671	0.992	0.35	0.462
95% CI	(0.632 0.709)	(0.981 0.998)	(0.231 0.484)	(0.330 0.595)
SF	0.906	0.989	0.823	0.813
95% CI	(0.880 0.929)	(0.977 0.996)	(0.655 0.932)	(0.710 0.916)
CR	0.832	0.998	0.667	0.747
95% CI	(0.800 0.862)	(0.991 1.000)	(0.299 0.925)	(0.507 0.986)
FH	0.939	1.000	0.878	0.931
95% CI	(0.917 0.957)	(0.993 1.000)	(0.738 0.959)	(0.870 0.991)
CB	0.825	0.998	0.652	0.762
95% CI	(0.792 0.855)	(0.990 1.000)	(0.427 0.836)	(0.611 0.912)
JCP	0.927	0.990	0.864	0.884
95% CI	(0.904 0.947)	(0.977 0.997)	(0.780 0.926)	(0.832 0.937)

*95% CI: 95% confidence interval

QualityDDH system, the experts had the best diagnostic ROC for structures, with a range of 0.500–0.960, especially in the identification of BR, La, SF, FH, and JCP, the experts

had a significantly improved ability to identify anatomical structures than the attending sonographers and residents ($p < 0.001$). In the identification of the structures of La, FH, and JCP, the diagnostic AUC values of experts could reach more than 0.800. However, the diagnostic performance of attending and resident sonographers in the identification of BR, SF, CR, and CB were poor, with AUC values ranging from 0.500 to 0.600. However, after the assistance of QualityDDH model, the recognition ability of all sonographers was improved, and the trend was still the same: Expert > Attending > Resident sonographer. The AUCs of SF, FH, and JCP were higher than 0.800, and the performance of BR, La, CR, and CB were higher than or close to 0.700 for all sonographers. With the assistance of the QualityDDH model, experts still have a significant advantage over other sonographers in the recognition performance in BR, La, and CR.

Without the assistance of the QualityDDH model, the final judgment results of all sonographers had moderate to substantial consistency (kappa values: 0.688, 0.771–0.840), and the consistency of the final judgment results within experts, attending and resident sonographers were 0.840, 0.771, and 0.775, respectively. There was slight to moderate agreement among the six sonographers in the identification of all key structures, ranging from 0.291 to 0.688. The agreement between two experts and two attending sonographers on the eight structures were high, generally high above 0.6. However, the difference in consistency of these structures varied considerably between the two residents, ranging from 0.000 to 0.793. After the assistance of the QualityDDH model, the consistency of the final judgment results were excellent among six sonographers, between two experts, and between two attending sonographers and two residents, ranging from

**Fig. 10** Performance of six sonographers before and after QualityDDH's assistance compared with golden standard. **a** Identification AUC values of the six radiologists before/after AI's assistance compared with

the golden standard. **b** Identification consistency of the six radiologists before/after AI's assistance compared with the golden standard

Table 3 Classification comparison shown in ablation study results based on different regressors, including support vector machine (SVM), AdaBoost, XGBoost, and multilayer perceptron (MLP) in external test dataset

Model	Acc (%)	Sen (%)	Spe (%)	κ
SVM	94.67	88.24	95.98	0.8167
AdaBoost	95.67	91.18	96.59	0.8511
XGBoost	95.00	89.22	96.18	0.8282
MLP	97.33	95.10	97.79	0.9077

0.932 to 0.995. In contrast, the agreement of all sonographers in all structures was generally improved, as shown in Fig. 10, especially in the identification of these structures in SF, CR, FH, CB, and JCP, which reached excellent agreement. The differences in internal consistency of the three different levels of sonographers were gradually reduced, and the differences in recognition consistency within a single structure were also reduced. In particular, the recognition consistency of residents was significantly improved, which was generally increased to above 0.85 (except CB).

4.4.4 Ablation study

QualityDDH is a comprehensive framework for visualizing and evaluating the quality of DDH ultrasound images. Ablation experiments were performed on the SPPR module to explore its role within the QualityDDH framework. Different point coordinate regression models, including support vector machine (SVM), AdaBoost, XGBoost, and Multilayer perceptron (MLP), were compared. The results are presented in Table 3, which shows how the quality classification evaluation metrics changed with the use of different point coordinate regression models. Among these models, the MLP model achieved the highest accuracy in quality assessment.

5 Conclusions

In this study, we developed QualityDDH, a visual quality assessment framework based on the segmentation of key structures in ultrasound images. It evaluates multi-structure segmentation maps from the segmentation model against standardized identification guidelines. QualityDDH achieves high-precision and consistent DDH ultrasound image quality assessment. It offers visual and interactive quality assessment prompts, addressing the low robustness of existing methods in clinical practice. By ensuring the standardization and consistency of DDH ultrasound image acquisition, QualityDDH helps doctors obtain standard planar images, thereby enhancing the accuracy of both doctors and AI-assisted screening.

Acknowledgements This work was supported by the National Natural Science Foundation of China (62402530, 82171936), the Shanghai Jiao Tong University “Jiao Tong University Star” plan key project of the Medical and Industrial Cross (YG2022ZD007), the Program of Shanghai Academic/Technology Research Leader (23XD1403100), the Science and Technology Commission of Shanghai Municipality (23DZ2202200), and the Key Discipline Construction Project of Jiading District Health System (XK202403). We also thank the Bioinformatics Center, Furong Laboratory for providing computational resources.

Author Contributions R. L. and M. L. were responsible for drafting the main manuscript text, while Y. Z. and L. J. oversaw the research activities. Data collection and method implementation were conducted by R. L., M. L., X. L., Y., Z., and Q. L., with all authors participating in the review of the manuscript.

Data availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

- Schaeffer, E.K., Mulpuri, K., IHDI Study Group: Developmental dysplasia of the hip: addressing evidence gaps with a multicentre prospective international study. *Med. J. Aust.* **208**(8), 359–364 (2018)
- O’Beirne, J.G., Chlapoutakis, K., Alshryda, S., Aydingoz, U., Bauermann, T., et al.: International interdisciplinary consensus meeting on the evaluation of developmental dysplasia of the hip. *Ultraschall Med.* **40**(4), 454–464 (2019)
- 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**(11), 886–901 (2015)
- Jackson, J.C., Runge, M.M., Nye, N.S.: Common questions about developmental dysplasia of the hip. *Am. Fam. Physician* **90**(12), 843–850 (2014)
- Roposch, A., Moreau, N.M., Uleryk, E., Doria, A.S.: Developmental dysplasia of the hip: quality of reporting of diagnostic accuracy for US. *Radiology* **241**(3), 854–860 (2006)
- Graf, R.: The diagnosis of congenital hip-joint dislocation by the ultrasonic compound treatment. *Arch. Orthop. Trauma Surg.* **97**(2), 117–133 (1980)
- Graf, R., Mohajer, M., Plattner, F.: Hip sonography update. Quality-management, catastrophes—tips and tricks. *Med. Ultrason.* **15**(4), 299–303 (2013)
- Graf, R.: Hip sonography: background; technique and common mistakes; results; debate and politics; challenges. *Hip Int.* **27**(3), 215–219 (2017)
- Ömeroğlu, H., Köse, N., Akceylan, A.: Success of Pavlik harness treatment decreases in patients ≥ 4 months and in ultrasonographically dislocated hips in developmental dysplasia of the hip. *Clin. Orthop. Relat. Res.* **474**(5), 1146–1152 (2016)
- Tao, C., Yuxiao, Z., Bo, W., Jian, W., Ligang, C., Jingnan, H., Cong, L.: Development of a fully automated Graf standard plane and angle evaluation method for infant hip ultrasound scans. *Diagnostics (Basel)* **12**(6), 1423 (2022)
- Jaremko, J.L., Mabee, M., Swami, V.G., Jamieson, L., Chow, K., Thompson, R.B.: Potential for change in US diagnosis of hip dysplasia solely caused by changes in probe orientation: patterns

- of alpha-angle variation revealed by using three-dimensional US. *Radiology* **273**(3), 870–878 (2014)
12. Huang, S., Liu, X., Tan, T., Menghan, H., Wei, X., et al.: Transmrsr: transformer-based self-distilled generative prior for brain MRI super-resolution. *Vis. Comput.* **39**(8), 3647–3659 (2023)
 13. Al-Jebrni, A.H., Ali, S.G., Li, H., Lin, X., Li, P., et al.: Sthy-net: a feature fusion-enhanced dense-branched modules network for small thyroid nodule classification from ultrasound images. *Vis. Comput.* **39**(8), 3675–3689 (2023)
 14. Dai, L., Liang, W., Li, H., Cai, C., Qiang, W., et al.: A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nat. Commun.* **12**(1), 3242 (2021)
 15. Gong, B., Shi, J., Han, X., Zhang, H., Huang, Y., et al.: Diagnosis of infantile hip dysplasia with b-mode ultrasound via two-stage meta-learning based deep exclusivity regularized machine. *IEEE J. Biomed. Health Inform.* **26**(1), 334–344 (2022)
 16. Chen, Y.-P., Fan, T.-Y., Chu, C.-C., Lin, J.-J., Ji, C.-Y., et al.: Automatic and human level Graf's type identification for detecting developmental dysplasia of the hip. *Biomed. J.* **47**(2), 100614 (2024)
 17. Liu, R., Liu, M., Sheng, B., Li, H., Li, P., et al.: NHBS-Net: a feature fusion attention network for ultrasound neonatal hip bone segmentation. *IEEE Trans. Med. Imaging* **40**(12), 3446–3458 (2021)
 18. Liu, M., Liu, R., Shu, J., Liu, Q., Zhang, Y., Jiang, L.: AutoDDH: a dual-attention multi-task network for grading developmental dysplasia of the hip in ultrasound images. *Vis. Comput.* (2025)
 19. Sze, C.L., Paramesran, R.: Review of medical image quality assessment. *Biomed. Signal Process. Control* **27**, 145–154 (2016)
 20. Sze, C.L., Rajagopal, H., Paramesran, R.: Alzheimer's disease neuroimaging initiative et al correlation between subjective and objective assessment of magnetic resonance (mr) images. *Magn. Reson. Imaging* **34**, 820–831 (2016)
 21. Razaak, M., Martini, M.G., Savino, K.: A study on quality assessment for medical ultrasound video compressed via HEVC. *IEEE J. Biomed. Health Inform.* **18**(5), 1552–1559 (2014)
 22. Panayides, A., Pattichis, M.S., Pattichis, C.S., Loizou, C.P., Pantziaris, M., Pitsillides, A.: Atherosclerotic plaque ultrasound video encoding, wireless transmission, and quality assessment using IEEE Trans h.264. *Inf. Technol. Biomed.* **15**(3), 387–397 (2011)
 23. Kolb, A., Chiari, C., Schreiner, M., Heisinger, S., Willegger, M., et al.: Development of an electronic navigation system for elimination of examiner-dependent factors in the ultrasound screening for developmental dysplasia of the hip in newborns. *Sci. Rep.* **10**(1), 16407 (2020)
 24. Komatsu, M., Sakai, A., Dozen, A., Shozu, K., Yasutomi, S., et al.: Towards clinical application of artificial intelligence in ultrasound imaging. *Biomedicine* **9**(7), 720 (2021)
 25. Mabee, M.G., Hareendranathan, A.R., Thompson, R.B., Dulai, S., Jaremko, J.L.: An index for diagnosing infant hip dysplasia using 3-D ultrasound: the acetabular contact angle. *Pediatr. Radiol.* **46**(7), 1023–1031 (2016)
 26. de Luis-García, R., Alberola-López, C.: Parametric 3D hip joint segmentation for the diagnosis of developmental dysplasia. In: *International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4807–4810 (2006)
 27. Quader, N., Hodgson, A.J., Mulpuri, K., Schaeffer, E., Abugharbieh, R.: Automatic evaluation of scan adequacy and dysplasia metrics in 2-D ultrasound images of the neonatal hip. *Ultrasound Med. Biol.* **43**(6), 1252–1262 (2017)
 28. Pandey, P.U., Quader, N., Guy, P., Garbi, R., Hodgson, A.J.: Ultrasound bone segmentation: a scoping review of techniques and validation practices. *Ultrasound Med. Biol.* **46**(4), 921–935 (2020)
 29. El-Hariri, H., Mulpuri, K., Hodgson, A.J., Garbi, R.: Comparative evaluation of hand-engineered and deep-learned features for neonatal hip bone segmentation in ultrasound. In: *Proc. MICCAI* vol. 11765, pp. 12–20 (2019)
 30. Sezer, A., Sezer, H.B.: Deep convolutional neural network-based automatic classification of neonatal hip ultrasound images: a novel data augmentation approach with speckle noise reduction. *Ultrasound Med. Biol.* **46**(3), 735–749 (2020)
 31. Jiang, S.W., Flynn, M.S., Kwock, J.T., Liu, B., Quow, K., Blanchard, S.K., Breglio, K.F., Fresco, A., Jamison, M.O., Lesesky, E., Bellet, J.S., Green, C.L., Shearer, S.M., Nicholas, M.W.: Quality and perceived usefulness of patient-submitted store-and-forward teledermatology images. *Nat. Commun.* **158**(10), 1183–1186 (2022)
 32. He, K., Zhang, X., Ren, S., Sun, J.: Jian: Deep residual learning for image recognition. In: *Proc. IEEE CVPR*, pp. 770–778 (2016)
 33. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proc. IEEE CVPR*, pp. 2261–2269 (2017)
 34. Yu, F., Koltun, V., Funkhouser, T.: Dilated residual networks. In: *Proc. IEEE CVPR*, pp. 636–644 (2017)
 35. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., et al.: An image is worth 16 × 16 words: Transformers for image recognition at scale. In: *Proc. ICLR* (2021)
 36. Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., et al.: Bisenet: Bilateral segmentation network for real-time semantic segmentation. In: *Proc. ECCV*, volume 11217 of *Lecture Notes in Computer Science*, pp. 334–349 (2018)

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Ruhan Liu received her Ph.D. degree in computer science and engineering from Shanghai Jiao Tong University, Shanghai. Currently, she is a lecturer at Central South University, and affiliated with Furong Laboratory. Her research interests include medical image analysis and medical signal processing.



Yuan Zhang received his M.D. degree in Pediatrics from Fudan University, Shanghai, China, in 2022. Currently, he is a doctor in Department of Ultrasound Medicine, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. His major interests include musculoskeletal diseases, ultrasound evaluation of inflammatory bowel disease and congenital malformation.



Qirong Liu is currently pursuing a B.S. degree in College of Software, Taiyuan University of Technology. His research interests include sport medicine, computer vision and medical ultrasound image processing.



Xiaoxiao Luo is currently pursuing a Master degree in the Department of Ultrasound, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai. Her current research interests include ultrasound shear wave imaging, intestinal ultrasound diagnosis and artificial intelligence ultrasound diagnosis.



Mengyao Liu is currently pursuing a Ph.D. degree in the Department of Ultrasound, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai. Her current research interests include musculoskeletal ultrasound diagnosis and treatment, medical image analysis, and artificial intelligence ultrasound diagnosis.



Yiwen Zheng is currently a resident physician undergoing standardized residency training in the Department of Ultrasound, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai. Her research focuses on ultrasound image analysis, ultrasound diagnosis of tumors, and mechanisms of ultrasound therapy.



Lixin Jiang received a Ph.D. degree from Shanghai Jiao Tong University School of Medicine, Shanghai, China, in 2006. He is currently the director of the Department of Ultrasound Medicine, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. He has been awarded the honorary title of Shanghai Academic/Technology Research Leader and has contributed more than 60 articles and coauthored 12 books. His research interests include sports

medicine ultrasound, high-intensity focused ultrasound (HIFU) treatment of tumors, ultrasound diagnosis of endocrine and metabolic diseases, ultrasound evaluation of rheumatic immune diseases and artificial intelligence in ultrasound medicine.