BioMedical Engineering OnLine

RESEARCH

Open Access

Using 2D U-Net convolutional neural networks for automatic acetabular and proximal femur segmentation of hip MRI images and morphological quantification: a preliminary study in DDH

Dian Zhang[1], Hongyan Zhou[2,3,5], Tianli Zhou[2,3], Yan Chang[3], Lei Wang[4], Mao Sheng[1], Huihui Jia[1*] and Xiaodong Yang[2,3*]

*Correspondence:
jiahuihui0405@163.com;
xiaodong.yang@sibet.ac.cn

[1] Department of Radiology, Children's Hospital of Soochow University, Suzhou 215025, China
[2] School of Medical Imaging, Xuzhou Medical University, Xuzhou 221004, China
[3] Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163, China
[4] School of Ophthalmology and Optometry, Eye Hospital of Wenzhou Medical University, Wenzhou 325027, China
[5] Patent Examination Cooperation (Jiangsu) Center of The Patent Office, China National Intellectual Property Administration, Suzhou, 215163, China

Abstract

Background: Developmental dysplasia of the hip (DDH) is a common pediatric orthopedic condition characterized by varying degrees of acetabular dysplasia and hip dislocation. Current 2D imaging methods often fail to provide sufficient anatomical detail for effective treatment planning, leading to higher rates of misdiagnosis and missed diagnoses. MRI, with its advantages of being radiation-free, multi-planar, and containing more anatomical information, can provide the crucial morphological and volumetric data needed to evaluate DDH. However, manual techniques for measuring parameters like the center–edge angle (CEA) and acetabular index (AI) are time-consuming. Automating these processes is essential for accurate clinical assessments and personalized treatment strategies.

Methods: This study employed a U-Net-based CNN model to automate the segmentation of hip MRI images in children. The segmentation process was validated using a leave-one-out method during training. Subsequently, the segmented hip joint images were utilized in clinical settings to perform automated measurements of key angles: AI, femoral neck angle (FNA), and CEA. This automated approach aimed to replace manual measurements and provide an objective reference for clinical assessments.

Results: The U-Net-based network demonstrates high effectiveness in hip segmentation compared to manual radiologist segmentations. In test data, it achieves average DSC values of 0.9109 (acetabulum) and 0.9244 (proximal femur), with a 91.76% segmentation success rate. The average ASD values are 0.3160 mm (acetabulum) and 0.6395 mm (proximal femur) in test data, with Ground Truth (GT) edge points and predicted segmentation maps having a mean distance of less than 1 mm. Using automated segmentation models for clinical hip angle measurements (CEA, AI, FNA) shows no statistical difference compared to manual measurements ($p > 0.05$).

Conclusion: Utilizing U-Net-based image segmentation and automated measurement of morphological parameters significantly enhances the accuracy and efficiency of DDH assessment. These methods improve precision in automatic measurements and provide an objective basis for clinical diagnosis and treatment of DDH.

Keywords: Developmental dysplasia of the hip, Magnetic resonance imaging, Image segmentation, Convolutional neural network, Morphological quantification

Introduction

Developmental Dysplasia of the Hip (DDH) is a common pediatric orthopedic condition characterized primarily by inadequate coverage of the femoral head due to acetabular dysplasia, potentially leading to hip joint instability and possible deformities [1]. DDH is a major cause of hip pain and functional impairment in children, imposing significant economic burdens on families. The unreliable treatment process may contribute to high complication rates, such as recurrent dislocation, avascular necrosis of the femoral head, and hip joint stiffness [2, 3]. Currently, conventional X-ray imaging remains the main diagnostic tool; however, its reliance on two-dimensional images and parameters makes it challenging to accurately depict the spatial anatomical details of the hip joint [4]. Therefore, achieving quantification of DDH's three-dimensional anatomical information is crucial.

While evidence suggests that computed tomography (CT) with its three-dimensional reconstruction technology is valuable for personalized clinical assessment and tracking of DDH [5, 6], its use in pediatric imaging is limited due to concerns about radiation exposure. Magnetic resonance imaging (MRI), on the other hand, offers significant advantages as it does not involve ionizing radiation and provides detailed visualization of soft tissues, making it suitable for assessing hip dysplasia and cartilage conditions [2, 7, 8]. This has led many researchers to favor MRI for studying hip morphology [9, 10]. Central edge angle (CEA), acetabular index (AI), and femoral neck anteversion (FNA) are widely recognized morphological parameters for evaluating DDH [11, 12]. However, CEA and AI are typically based on manual measurements from two-dimensional radiographs, while FNA relies on measurements from three-dimensional images such as MRI or CT. Conventional manual measurement of these parameters is time-consuming and highly dependent on experienced clinicians. Therefore, it is essential to develop fully automated segmentation methods for the acetabulum and proximal femur, alongside reliable measurement of hip joint angles.

In recent years, several methods using MRI images for automatic segmentation and angle measurement of the hip have been proposed. Researchers has proposed [13] the use of multi-atlas and active shape model (ASM) approaches for automatic hip segmentation. However, this method relies on generative models based on prior anatomical features, necessitating prior anatomical knowledge or templates to generate images, which may be limited by the accuracy and completeness of the chosen templates or prior knowledge. In contrast, other researchers [14] have suggested random forest classifiers for femur segmentation, which belong to a discriminative model. While these classifiers do not require prior anatomical knowledge, they typically involve manual feature extraction from images before training. Convolutional neural network (CNN) models,

on the other hand, are a promising machine learning-based approach that automatically extracts features from images through multiple layers of convolution and pooling operations [15, 16]. They do not require manual feature design and offer stronger capabilities for automatic feature learning and adaptation to complex data distributions compared to generative and discriminative models. Some researchers [17] have applied CNNs to achieve automatic measurement of critical hip angles, primarily in general healthy populations. However, their ability to measure angles on MRI images for specific hip joint diseases such as DDH has not been extensively studied. Therefore, this study would utilize deep CNN for automatic hip MRI segmentation and further attempt to automatically quantify MRI morphological parameters in children with DDH.

The purposes of our study were to: (1) advance the automation of clinical practices for DDH, including MRI segmentation of the acetabulum and proximal femur, as well as measurement of bone morphological parameters; (2) compare and assess the reliability of automated segmentation against manual segmentation; (3) quantitatively evaluate the hip joint using MRI three-dimensional models, comparing automatically measured angles between normal children and those with DDH. The workflow for MRI segmentation and morphological quantification of the hip is shown in Fig. 1.

Results

As shown in Fig. 2, it was obvious that the predicted segmentation maps generated by the CNN-based network were successfully segmented, as shown by comparison with the GT segmentation maps manually segmented by two experienced radiologists. In this preliminary study, both the validation data and the testing data were fed into 11 segmentation models trained by the way of leave-one-child-out evaluation to analyze the performance of the U-Net-based segmentation. And the results are shown in Table 1 with the metrics of DSC and ASD values for the comparison of the predicted maps with their one-to-one corresponding GT maps. It showed that the average DSC values from 11 trained models were 0.9178 for the acetabulum and 0.9368 for the proximal femur of the validation data and 0.9109 for the acetabulum and 0.9244 for the proximal femur of the testing data, respectively, meaning that a success rate of 92.73% for the validation data and 91.76% for the testing data was achieved for hip segmentation from a holistic perspective. Moreover, the average values of ASD were 0.3595 mm for the acetabulum and 0.2902 mm for the proximal femur of the validation data and 0.3160 mm for the

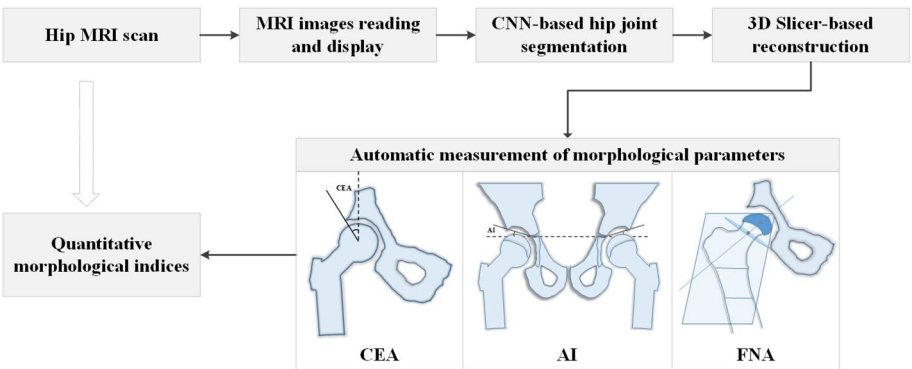

Fig. 1 Quantitative processes of MRI-based morphological evaluation for children with DDH in this research

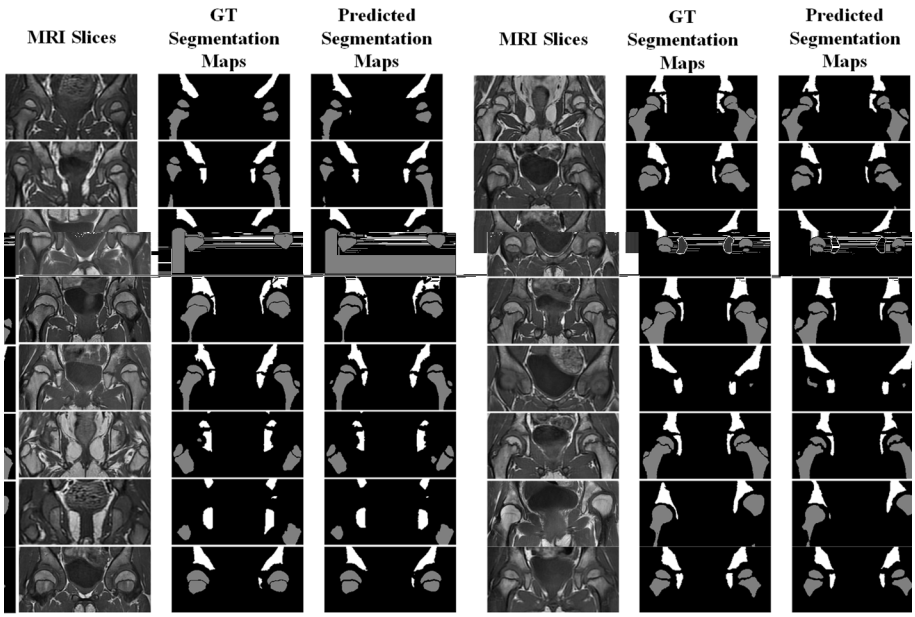

Fig. 2 The comparison of manual labels and CNN-based segmentation maps in hip MRI images

acetabulum and 0.6395 mm for the proximal femur of the testing data, respectively, and the average distances between the edge points of the GT maps and the predicted segmentation maps were less than 1 mm. Furthermore, a non-parametric statistical analysis of the Mann–Whitney *U* test was adopted to comparing the DSC and ASD values of two samples. The results showed that the DSC values of the acetabulum were significantly lower than those of the proximal femur both for the validation or testing data ($P_{\text{Validation}} = 0.003 < 0.05$ and $P_{\text{Testing}} = 0.000 < 0.05$). However, there was no significant difference of the average ASD values between of the acetabulum and of the proximal femur for the validation data ($P_{\text{Validation}} = 0.250 > 0.05$), but the segmented performance of the acetabular edge was better than that of the proximal femur for the testing data ($P_{\text{Testing}} = 0.000 < 0.05$).

The testing data consisting of the hip MRI images of 14 children were available to quantify the various changes in hip morphology comparing hip joints with DDH to normal hip joints. Considering that the segmentation performance of hip image as well as the acetabular edge important to carry out further morphological quantification, the predicted maps from the sixth model were taken advantage of in the following research study. After the predicted maps, a point cloud model of hip was firstly constructed by 3D Slicer. Then, the measurements of three parameters in 3D space are shown in Fig. 3, comparing seven children with healthy bilateral hips (Fig. 3a) to seven children with DDH or post-operation (Fig. 3b). The transverse coordinates were in natural numbers corresponding to the number of hip cases while arranged according to the increasing age. The specific ages corresponding to the x-axis in Fig. 3 could be observed in Additional file 1: Table S1 (where the green and blue shaded areas labeled 1–7 in Fig. 3, respectively, correspond to 'number 34–40' and 'number 41–47' in Additional file 1: Table S1). From the curve changes in the results, it was found that the CEA values of normal children with bilateral hip joints gradually increased with age, while the AI and FNA values decrease

Table 1 The average values of DSC and ASD with respect to the segmentation performance of 11 models on the validation data and testing data

			M-1	M-2	M-3	M-4	M-5	M-6	M-7	M-8	M-9	M-10	M-11
Validation data	DSC	Acetabulum	0.9135	0.916	0.9201	0.923	0.9121	0.9281	0.9225	0.9291	0.912	***0.9357***	0.8837
		Proximal femur	0.9161	0.9345	0.9432	0.9196	0.9303	0.942	0.9507	***0.948***	0.927	0.9474	0.9461
		Average	0.9148	0.9253	0.9316	0.9213	0.9212	0.9351	0.9366	0.9386	0.9195	***0.9415***	0.9149
	ASD/mm	Acetabulum	0.701	0.2348	0.2445	***0.1176***	0.3204	0.4448	0.2268	0.3851	0.4632	0.2418	0.5743
		Proximal femur	0.4328	0.2509	0.1919	0.2785	0.2172	0.2692	***0.1674***	0.1797	0.7393	0.2624	0.2031
		Average	0.5669	0.2429	0.2182	0.198	0.2688	0.357	***0.1971***	0.2824	0.6012	0.2521	0.3887
Testing data	DSC	Acetabulum	0.9096	0.9122	0.9017	0.9087	0.9108	0.9136	0.9117	0.9132	0.9099	***0.9148***	0.9133
		Proximal femur	0.9267	0.923	0.9265	***0.9279***	0.9226	0.9278	0.923	0.9235	0.9225	0.9238	0.9211
		Average	0.9182	0.9176	0.9141	0.9183	0.9167	***0.9207***	0.9173	0.9184	0.9162	0.9192	0.9172
	ASD/mm	Acetabulum	0.3433	0.2776	0.34	0.4047	0.3899	***0.2112***	0.3955	0.2482	0.1877	0.2226	0.4553
		Proximal femur	0.7098	***0.4579***	0.7855	0.6167	0.6029	0.7436	0.6737	0.5176	0.638	0.543	0.7458
		Average	0.5266	***0.3677***	0.5628	0.5107	0.4964	0.4774	0.5346	0.3829	0.4129	0.3828	0.6006

The values with bold italics indicated that the optimal performance was carried out by one model, as well as the larger DSC and the smaller ASD values indicated the better segmentation

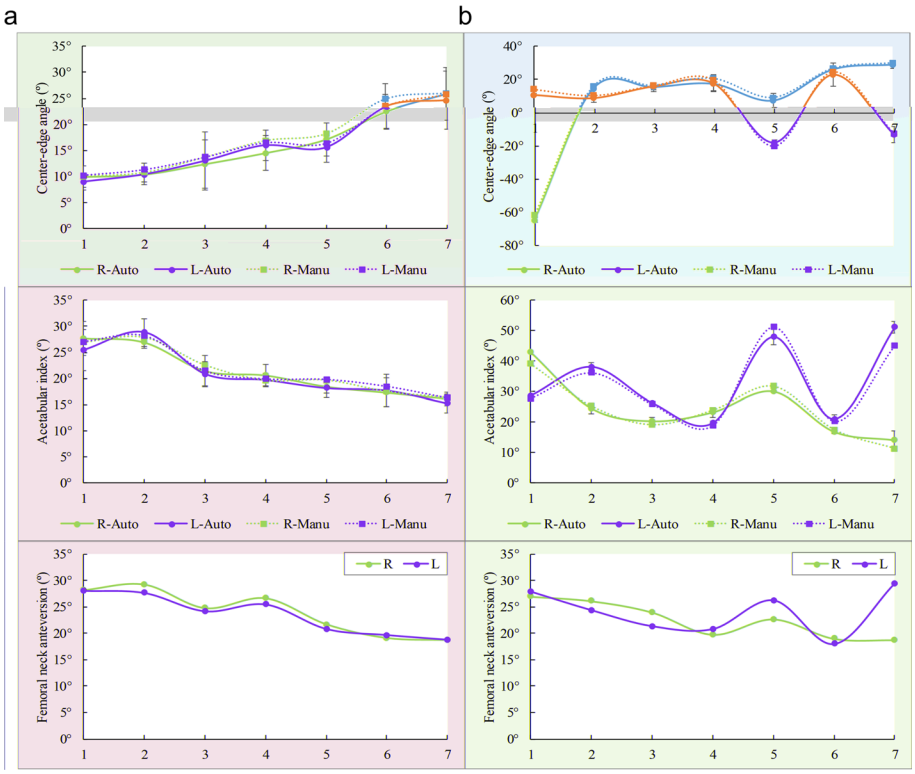

Fig. 3 The measurement of three parameters including CEA, AI and FNA for seven children with bilateral healthy hip (left column with green shading) and seven children with DDH or post-operation (right column with blue shading). The solid blue (donated as right hip) and orange (donated as left hip) lines represented the results with the approach of automatic measurement while the dotted lines represented the results with the approach of manual measurement. R and L represented as "Right" and "Left", respectively. Auto and Manu represented as "Automatic" and "Manual", respectively

with age. Furthermore, compared to normal hips, the magnitude of three parameters in DDH hips showed obvious differences. The CEA values in the affected hips with DDH were lower than the normal range and even negative. And it was found that the AI and FNA values in DDH hips were higher than those of normal hips. At the same time, the measurement could be used to observe the postoperative recovery of DDH hips. The CEA values of the operated hips were largely close to those of the healthy hips. However, there were higher values of AI and lower values of FNA in the operated hips compared to the healthy hips. Finally, to compare the relationship between automatic and manual measurements of CEA and AI in 14 children with 28 hips in total, independent sample t-test was used to reveal the results of no significant differences between two methods (P_{CEA}=0.808 > 0.05, P_{AI}=0.950 > 0.05).

Discussion

We have developed a model for automatic segmentation of the hip joint on MRI and automated measurement of critical angles for DDH. Manual segmentation methods typically require significant time and effort. Moreover, due to the complexity of tissue structures and minimal differences in grayscale values between adjacent tissues, traditional segmentation methods may struggle to accurately separate targets from raw MRI

images. Therefore, in this preliminary study, we employed a 2D CNN-based segmentation approach to automatically segment the acetabulum and proximal femur from hip MRI images. Segmentation was performed on 14 new hip joint datasets under two structural environments, achieving a satisfactory success rate of 91.76%. However, segmentation of the proximal femur was objectively superior to the acetabulum, mainly due to the irregular contour of the acetabulum bone and minimal grayscale differences compared to surrounding tissues. Given the relatively small sample size in this study, we utilized leave-one-out cross-validation to assess the model's performance. This approach maximizes the use of limited sample data and provides an unbiased estimate of model performance, ensuring reliability and accuracy in evaluation results. Throughout this process, we generated 11 different types of predictive segmentation maps. Following training and validation, we selected the best predictive map for automatic measurement of hip joint morphological parameters and subsequent evaluation.

The CEA, AI, and FNA are three critical parameters widely recognized for evaluating DDH. CEA and AI are conventionally measured using 2D X-ray images, while FNA involves assessing the degree of anterior tilt of the femoral neck by measuring the angle between the femoral neck axis and the femoral shaft axis. Given that these measurements exist in different planes, 3D modeling is essential to mitigate measurement errors arising from variations in angles or projection methods in 2D images [18, 19]. Our developed 3D measurement approach for DDH provides comprehensive spatial anatomical information, allowing for personalized assessment of the hip joint across multiple angles and planes. This method circumvents the limitations of 2D imaging, facilitating dynamic observation of the hip joint and enabling tailored treatment planning based on individual anatomical structure and developmental conditions.

In our study, the CEA values for 4-year-old children were $21.62^{\circ} \pm 6.23^{\circ}$, while for children aged 10–14 years, CEA values ranged from $24.54^{\circ} \pm 5.47^{\circ}$ to $31.11^{\circ} \pm 6.20^{\circ}$. There was an approximate 6° difference in CEA values with increasing age [20]. In a previous research [21], the AI values were $26.75^{\circ} \pm 2.57^{\circ}$ for children aged 1, $18.22^{\circ} \pm 2.2.26^{\circ}$ for children aged 4–8, a decrease of 2°–3° for children aged 8 to adolescence and then in a dynamic balance. The AI values in our study roughly correspond to the ranges mentioned above. In our study, the average FNA values were 27° for newborns and 22° for 5-year-old children, consistent with previous research findings [22]. Our study also found that children with DDH under 4 years old did not show significant differences in FNA values compared to healthy hip joints, suggesting that DDH in younger children may not always lead to increased FNA [23, 24].

Additionally, in our study, we compared three morphological parameters among children with DDH, normal children, and those who underwent hip surgery. Among normal children, bilateral hip CEA values increase gradually with age, indicating improved coverage of the femoral head by the acetabulum. AI reflects acetabular morphology, with decreasing values indicating deeper and larger acetabula as the hip matures. FNA measures the angle between the femoral head and neck, which decreases with hip joint maturation and growth, indicating a more mature orientation of the femoral head towards the acetabulum [25]. These changes reflect morphological adjustments of the hip joint during childhood growth and development. Compared to normal hips, DDH hips exhibit significant differences in these parameters. The CEA values of DDH hips are notably

below the normal range, often negative, primarily due to complete femoral head dislocation towards the lateral side of the acetabulum. Additionally, AI and FNA values in DDH hips are higher than those in normal hips. In DDH, the acetabulum is typically shallow and irregularly shaped, resulting in inadequate coverage of the femoral head, hence the higher AI values. The increased FNA values reflect a more pronounced anterior tilt of the femoral head relative to the femoral neck, possibly due to the directional change caused by acetabular deformities [26]. Post-surgical measurements of DDH hips can also be used to monitor postoperative recovery. Post-surgical CEA values in the hips generally approach those of healthy hips, indicating improved acetabular coverage of the femoral head post-surgery. However, post-surgical AI values remain higher, and FNA values are lower. This may result from surgical interventions impacting the morphological and biomechanical properties of the hip joint. Surgery typically alters acetabular morphology to some extent, maintaining higher AI values, while the decreased FNA values may reflect partial improvement in hip joint stability post-surgery, though not fully restored to normal levels [27]. These measurement outcomes are crucial for assessing postoperative hip joint function and predicting long-term outcomes.

In our study, we also compared the performance of automatic measurements with manual measurements. A good correlation was found between manual and automatic segmentation based on MRI for the measurement of CEA, AI, and FNA. The automatically measured CEA and AI values in healthy, affected, and post-surgical hips of children were roughly equivalent to those manually measured by radiologists. To reduce errors, automated measurements in this study involve averaging a series of angles taken along the midportion of the acetabular edge [28]. Theoretically, the automatic results should be larger than the manual angles on MRI slices, including the maximum hip slice. The average value of the automatic measurements compensates to some extent for the errors caused by CNN-based segmentation, which explains the high correlation between automatic and manual results. Therefore, the algorithm for measuring these three angles has a certain degree of validity.

There are also some limitations in this primary study. This preliminary study has several limitations. Firstly, given the lack of newer algorithms, we employed the classic and robust U-Net segmentation algorithm due to its previously demonstrated significant advantages and success in medical imaging. In the future, more advanced algorithms could be designed to segment pediatric hip MRI images, aiming to improve the accuracy of hip joint tissue segmentation and the calculation of morphological parameters. Secondly, future studies should include a larger sample of children with DDH to further investigate the characteristics of the disease. Thirdly, this study only examined the morphology of the hip joint. In subsequent research, the advantages of high-resolution MRI in displaying hip cartilage can be utilized to automatically reveal early pathophysiological changes in hip diseases such as DDH.

Conclusion

The automatic segmentation of the hip joint and automated measurement of morphological parameters based on our developed U-Net neural network have significantly enhanced the accuracy and efficiency of evaluating DDH. This advancement

allows clinicians to make more informed decisions regarding diagnosis and treatment strategies.

Materials and methods

Subjects

From August 2023 to May 2024, 47 cases who underwent hip MRI scanner (GE Discovery MR750, USA) at the Children's Hospital of Soochow University were enrolled into this study. And then 33 cases were used for hip segmentation and 14 cases were used for morphological quantification (Table 2). The hip was imaged bilaterally using a coronal 3D CUBE T1 sequence and the subject was kept in a supine position with the legs in a symmetrical neutral position during a scan. The parameters were set as follows: field of view (FOV) 380 × 380 mm, slice thickness 0.8 mm, pixel spacing 0.74 mm, repetition time 400–650 ms, echo time 10–15 ms, and flip angle 90°. Half an hour before MRI examination, all subjects were given 10% chloral hydrate sedative orally at a dose of 0.5 ml/kg to avoid the influence of some factors such as motion artifacts on image quality. All parents of children signed informed consent to participate in this study with the approval of the hospital institutional review board.

Image annotation and preprocessing

In this paper, serial MRI slices that could reconstruct a geometric hip model were selected from each child for spatial morphological quantification. The CNN-based segmentation of the acetabulum and proximal femur is essentially a pixel-based image classification task, also known as a supervised learning problem. In this environment, the pixel labels of hip bones manually annotated by using ITK-SNAP software with the guidance of two experienced radiologists were indispensable for model training. However, the area occupied by bone structures in hip MRI slices might be various due to the age, and the other tissue information introduced by the large FOV would make segmentation more challenging. Therefore, the extraction of region of interest (ROI) was implemented in each MRI slice in the pre-processing process. As in Fig. 4, the boundaries of the ROI (the yellow rectangular area, width = 2 × height) determined by the largest cross-sectional hip MRI image ranged as follows: the upper boundary no lower than the outer contour of the acetabular fossa, the lower boundary no higher than the lowermost

Table 2 Clinical statistics of the subjects

Age	MRI data used for hip segmentation model ($n=33$)			MRI data used for morphological quantification ($n=14$)		
	0–5 ($n=13$)	6–10 ($n=15$)	11–15 ($n=5$)	0–5 ($n=5$)	6–10 ($n=5$)	11–15 ($n=4$)
Sex						
Male	$n=7$	$n=11$	$n=3$	$n=4$	$n=4$	$n=2$
Female	$n=6$	$n=4$	$n=2$	$n=1$	$n=1$	$n=2$
Health status						
Health	$n=13$	$n=15$	$n=5$	$n=3$	$n=2$	$n=2$
DDH	$n=0$	$n=0$	$n=0$	$n=1$	$n=1$	$n=1$
Surgery	$n=0$	$n=0$	$n=0$	$n=1$	$n=2$	$n=1$

In total, 47 cases were included in this study, with 33 used for training and validation of the automatic hip joint segmentation model, and 14 for automated measurement of critical angles post-segmentation

Fig. 4 The ROI applied to one hip MRI slice and its corresponding GT segmentation map

edge of the femoral neck, and the left and right boundaries no smaller than the lateral edge of the greater trochanter. The number of available MRI slices in each dataset and the corresponding ROI sizes are shown in Additional file 1: Table S1. Furthermore, the ROIs were uniformly adjusted to 128×256 (height $\times$ width) and the signal intensity of MRI images after cropped were normalized to the range of [0, 1] to meet the input of the neural network.

In the segmentation process, the acetabular bone (white, value = 255) and the proximal femur (grey, value = 127) were target regions, while other tissues were background regions (black, value = 0). It meant that hip segmentation was actually a tri-classification problem, while the feature attributes were discrete and unordered. Therefore, in order to map the values of the discrete features into Euclidean space to calculate feature similarity, a one-hot coding operation was used for labels with multi-task targets during pre-processing, a process where the tri-classified GT map was converted into a matrix consisting of only zeros and ones in three layers, which would then be seen as multiple binary classification problems.

Model architecture and segmentation implementation

One problem focused on this article was 2D bone segmentation of the acetabulum and proximal femur in hip MRI images, which was solved by means of a CNN model based on the U-Net architecture [29]. U-Net is one of the deep learning architectures based on encoder-decoder for the segmentation of the anatomical tissues in medical images, and it has also been proved to achieve a promising results even though the size of the sample is small [30]. In the training process of the constructed U-Net-based CNN structure, MRI images of size 128×256 with the corresponded pixel labels were input to CNN. In this way, the visualization process for training the CNN model was shown in Fig. 5. In the training block, the optimal model parameters obtained from the training were saved, and then the saved model was loaded to segment new MRI images to generate prediction maps in the segmentation block.

In this paper, consecutive MRI slices were selected from each child's volumetric images and a child independent based evaluation approach was adopted to train the constructed network. For this purpose, 33 children data were divided into 11 groups (Table 3) and an approach called Leave-One-Out Cross-Validation [31, 32] was applied for this small sample, which is deterministic in the classification of samples and has the highest utilization of samples. In this way, 10 groups were taken advantage of for

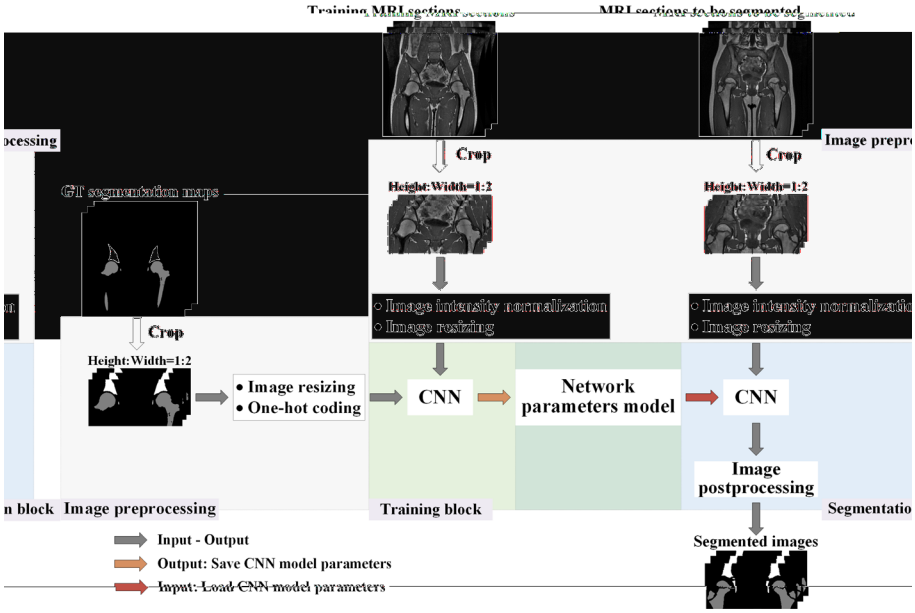

Fig. 5 The training process of U-Net architecture for hip MRI images segmentation

Table 3 The groups of hip MRI images and the information about age and slice numbers of each group

Groups	Serial number combination	Age corresponding to serial number	Number of slices in validation data	Number of slices in training data
Group-1	1 + 15 + 26	12, 1, 7	124	1260
Group-2	2 + 3 + 29	6, 3, 11	121	1263
Group-3	4 + 12 + 24	1, 8, 12	129	1255
Group-4	5 + 25 + 28	4, 10, 5	126	1258
Group-5	6 + 16 + 30	1, 11, 8	122	1262
Group-6	7 + 13 + 14	9, 2, 8	122	1262
Group-7	8 + 11 + 23	6, 10, 4	128	1256
Group-8	9 + 10 + 32	9, 3, 8	129	1255
Group-9	17 + 19 + 22	5, 9, 5	125	1259
Group-10	18 + 27 + 33	7, 8, 2	121	1263
Group-11	20 + 21 + 31	7, 13, 3	137	1247

MRI slices from 33 children were used with Leave-One-Out Cross-Validation (LOOCV). Data were divided into 11 groups, training 11 models, with 10 groups for training and 1 for validation to maximize sample utilization

network training and the remaining one was used for the validation at a training process, which means that 11 distinct segmentation models would be obtained through 11 cycles of training stages. For each training process, the specific parameters were set as follows: the epochs = 30, the batch size = 4, and the initial learning rate = 1e−4. Moreover, the optimizer was Adam, the loss function was DiceLoss, and the weight decay term was selected as the value of 1e−5 to reduce overfitting. The detailed grouping information and the leave-one-out model training method are provided in Additional file 1: Fig. S1. Each sample consisting of ten training datasets and one validation dataset.

Evaluation metrics

To evaluate the performance of these segmentation models, two metrics with regard to dice similarity coefficient (DSC) [33, 34] and average surface distance (ASD) [35, 36] were given in this paper. DSC is a commonly quantitative index for segmentation performance in medical image analysis, which characterizes the information about the overlapping area of two sample regions by measuring the ratio of the intersection to the concatenated area. ASD is a distance index, which mainly assess the segmentation performance of image edge by measuring the average distances of all edge points from both two sample regions. And both metrics could be respectively calculated by means of Eq. (1) and Eq. (2).

$$\mathrm{DSC}(S_{\mathrm{A}}, S_{\mathrm{P}}) = \frac{2 \cdot |S_{\mathrm{A}} \cap S_{\mathrm{P}}|}{|S_{\mathrm{A}}| + |S_{\mathrm{P}}|} \tag{1}$$

$$\mathrm{ASD}(S_{\mathrm{A}}, S_{\mathrm{P}}) = \frac{1}{|B_{S_{\mathrm{P}}}|} \sum_{x \epsilon b_{S_{\mathrm{P}}}} d\left(x, B_{S_{\mathrm{A}}}\right) \tag{2}$$

where S_{A} is the GT segmentation map, S_{P} is the predicted segmentation map, and $B_{S_{\mathrm{A}}}$ and $B_{S_{\mathrm{P}}}$ are all the edge points of the GT map and predicted map respectively. The value of DSC is in the range of [0, 1]. If the DSC value is close to 1, the predicted map is larger similar to the GT map indicating that the segmentation is more successful. But for ASD, the smaller the value, the better the segmented performance.

Quantitative measurement methods of morphological parameters

For the special measurement of morphological parameters, 3D Slicer 4.13.0 (https://www.slicer.org), a free and open-source medical image processing software, was first used to perform a 3D hip reconstruction after a set of consecutive predicted segmentation maps, belonging to a single child, was transformed into a NiFti (*.nii) file as the input. Then, a geometric model of the hip was created by assigning a value of 255 to the acetabulum and a value of 127 to the proximal femur in the "Segment Editor" module (Additional file 1: Fig. S2). Finally, hip model was exported into a point cloud model (*.ply) with relatively denser points to accurately measure the angular values of morphological parameters.

In a point cloud model of the hip, each point corresponds to a unique coordinate in a 3D coordinate system. To facilitate this, approaches for the automatic measurement of morphological parameters, including CEA, AI, and FNA, were developed in MATLAB. CEA is defined as the angle between a line from the center of the femoral head to the lateral edge of the acetabulum and a vertical line through the femoral head's center. It measures the lateral coverage of the femoral head by the acetabulum, crucial for assessing acetabular coverage depth and diagnosing conditions like developmental dysplasia of the hip (DDH). AI measures the angle between a line through the triradiate cartilage and another line through the lateral margin of the acetabulum. It indicates the orientation and depth of the acetabulum. AI helps in evaluating the development of the acetabulum, particularly in conditions where acetabular dysplasia may lead to hip instability. FNA measures the angle between the axis of the femoral neck and a line perpendicular to

the femoral head axis. It indicates the rotational orientation of the femoral neck relative to the femoral head, which is crucial for understanding hip function and stability. FNA helps doctors assess hip joint development and identify any abnormal anteversion. According to the schematic definitions of these three parameters provided in Fig. 1, the outer edge points and the central point of the acetabulum, as well as the central points of the femoral head and femoral neck, are crucial for measuring morphological angles.

In Fig. 6, we primarily employ a three-step approach to determine the positions of key anatomical points: First, Acetabular Outer Edge Points and Central Point Determination [39]: Based on the acetabular bone's growth pattern, which initially extends medially and then laterally along the ilium, specific parameters are constrained to pinpoint outer edge points (marked in green). The central point coincides with the intersection of the Y-shaped cartilage and acetabular bone, typically the lowest point of the V-shaped bone. Second, Femoral Head Central Point Determination: The femoral head, situated atop the femur and crucial for connecting to the acetabulum, is approximated using a spherical Hough transform (SHT) [37] to identify the center of the nearest 2/3 spherical shape. Third, Femoral Neck Central Point Determination:

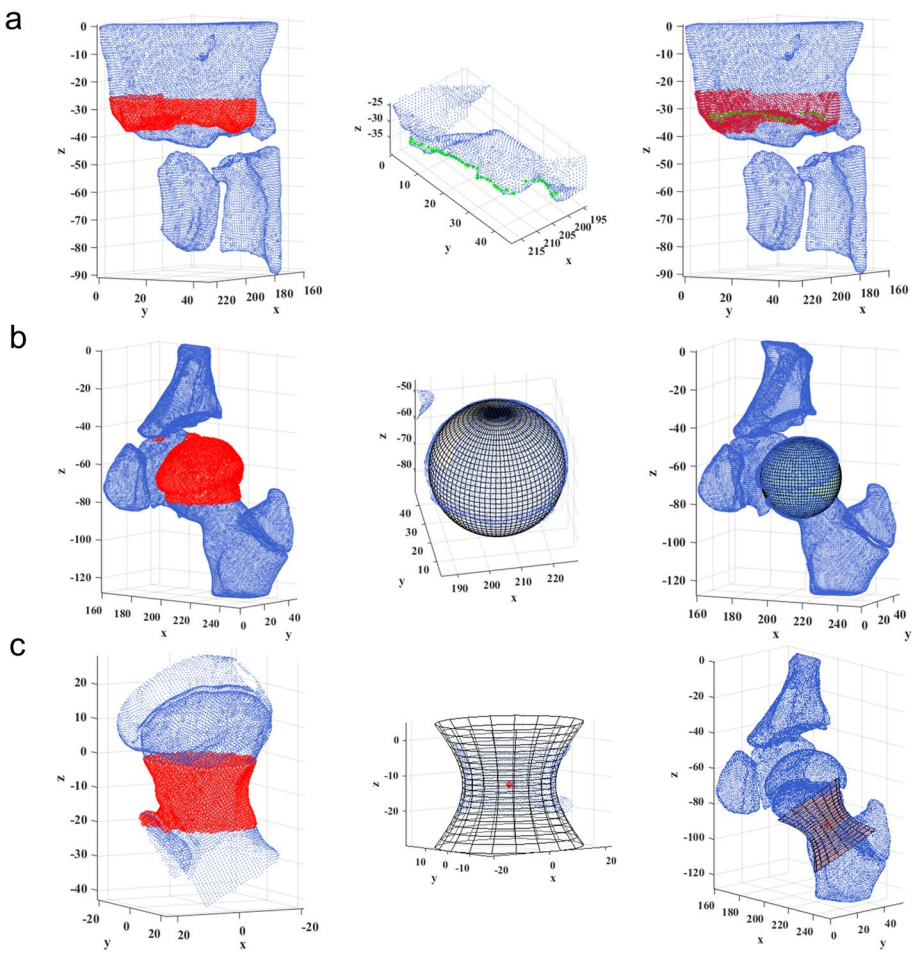

Fig. 6 Steps to determine key points: the determination of the outer edge points and the central point of the acetabular bone (**a**). The detection of the optimal sphere fitting to the femoral head (**b**). The fitting of the point cloud intercepted from the femoral neck (**c**)

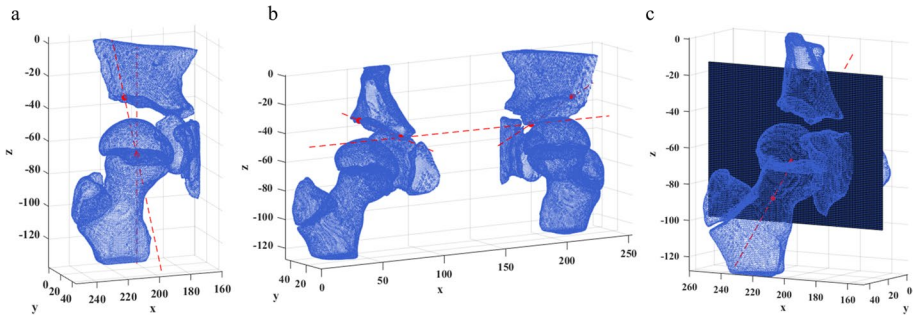

Fig. 7 The presentation of automatic measurement results for morphological parameters

The central point of the femoral neck is determined by fitting a hyperboloid model, tailored to the femoral neck's morphology, using a spatial quadratic surface least-squares fitting algorithm [38, 39]. This involves optimizing parameters through minimizing the sum of squared errors, derived from fitting the model to a set of points representing the femoral neck. These methods integrate anatomical knowledge with mathematical modeling techniques to accurately define crucial landmarks of the acetabulum and femur, providing essential data for medical research and diagnostic applications. The detailed determination process of key points and the mathematical modeling methods can be found in the Additional file 1.

The measurement of CEA, AI and FNA would be transformed to solve the angle between two lines or between one line and one plane on the spatial coordinate system. According to the above corresponding key points and the definition of the above three parameters, the results showed as in Fig. 7.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12938-024-01291-3.

Additional file 1. Table S1 Total number of the MRI sections, number of the MRI sections used, MRI image matrix sizes and the sizes of the region of interests (ROIs) of the proximal femur and femoral heads in MRI sequences. **Fig. S1** The approach of leave-one-child-out evaluation used in the training splitting of the evaluated MRI database into the training and validation data. **Fig. S2** The 3D model of hip joint (left) and 3D point cloud model (right).

Acknowledgements
Funding from Scientific Research Project of the Suzhou Health Commission (M2021029) is gratefully acknowledged.

Author contributions
DZ, HZhou, HJ and XY contributed to the study design, participated in the review process, and prepared the manuscript. DZ, TZ, YC, LW were responsible for data collection, statistical analysis, figure preparation, and critical interpretation. DZ, MS, HJ and XY conceptualized the paper and revised the manuscript. All authors read and approved the final manuscript.

Funding
This work was partially supported by Scientific Research Project of the Suzhou Health Commission (M2021029).

Availability of data and materials
The datasets generated and analyzed during the current study are not publicly available due to local restrictions of data protection but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
This study was conducted in accordance with the Declaration of Helsinki. The authors are accountable for all aspects of the work, ensuring that any questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Competing interests
The authors declare no competing interests.

Received: 11 June 2024 Accepted: 12 September 2024
Published online: 05 October 2024

References

1. Yang S, Zusman N, Lieberman E, Goldstein RY. Developmental dysplasia of the hip. Pediatrics. 2019;143(1):e20181147.
2. Madadi F, Shamsian BS, Alavi S, Madadi F, Eajazi A, Aslani A. Avascular necrosis of the femoral head in children with acute lymphoblastic leukemia: a 4-to 9-year follow-up study. Orthopedics. 2011;34(10):e593–7.
3. Naci E, Kadri Y. A study on the complications of surgical treatment for bilateral developmental dysplasia of the hip and a comparison of two osteotomy techniques. Eurasian J Med. 2011;43(3):162–8.
4. Starr V, Ha BY. Imaging update on developmental dysplasia of the hip with the role of MRI. AJR Am J Roentgenol. 2014;203(6):1324–35.
5. Meesters A, Kraeima J, Banierink H, Slump C, De Vries J, Ten Duis K, Witjes M, IJpma F. Introduction of a three-dimensional computed tomography measurement method for acetabular fractures. PLoS ONE. 2019;14(6):e0218612.
6. IJpma FFA, Meesters AML, Merema BBJ, Ten Duis K, de Vries JPM, Banierink H, Wendt KW, Kraeima J, Witjes MJH. Feasibility of imaging-based 3-dimensional models to design patient-specific osteosynthesis plates and drilling guides. JAMA Netw Open. 2021;4(2):e2037519.
7. Stelzeneder D, Hingsammer A, Bixby SD, Kim Y-J. Can radiographic morphometric parameters for the hip be assessed on MRI? Clin Orthop Relat Res. 2013;471(3):989–99.
8. Nishii T, Sugano N, Sato Y, Tanaka H, Miki H, Yoshikawa H. Three-dimensional distribution of acetabular cartilage thickness in patients with hip dysplasia: a fully automated computational analysis of MR imaging. Osteoarthr Cartil. 2004;12(8):650–7.
9. Lazik A, Theysohn JM, Geis C, Johst S, Ladd ME, Quick HH, Kraff O. 7 Tesla quantitative hip MRI: T1, T2 and T2* mapping of hip cartilage in healthy volunteers. Eur Radiol. 2016;26(5):1245–53.
10. Luitjens J, Gassert FG, Patwardhan V, Bhattacharjee R, Joseph GB, Zhang AL, Souza RB, Majumdar S, Link TM. Is hip capsule morphology associated with hip pain in patients without another structural correlate? Eur Radiol. 2024;34(7):4321–30.
11. Li Y, Guo Y, Li M, Zhou Q, Liu Y, Chen W, Li J, Canavese F, Xu H. Acetabular index is the best predictor of late residual acetabular dysplasia after closed reduction in developmental dysplasia of the hip. Int Orthop. 2018;42(3):631–40.
12. Li H, Wang Y, Oni J, Qu X, Li T, Zeng Y, Liu F, Zhu Z. The role of femoral neck anteversion in the development of osteoarthritis in dysplastic hips. Bone Joint J. 2014;96(12):1586–93.
13. Xia Y, Fripp J, Chandra SS, Schwarz R, Engstrom C, Crozier S. Automated bone segmentation from large field of view 3D MR images of the hip joint. Phys Med Biol. 2013;58(20):7375.
14. Damopoulos D, Lerch TD, Schmaranzer F, Tannast M, Chênes C, Zheng G, Schmid J. Segmentation of the proximal femur in radial MR scans using a random forest classifier and deformable model registration. Int J Comput Ass Rad. 2019;14(3):545–61.
15. Hephzibah R, Anandharaj HC, Kowsalya G, Jayanthi R, Chandy DA. Review on deep learning methodologies in medical image restoration and segmentation. Curr Med Imaging. 2023;19(8):844–54.
16. Deng Y, Wang L, Zhao C, Tang S, Cheng X, Deng HW, Zhou W. A deep learning-based approach to automatic proximal femur segmentation in quantitative CT images. Med Biol Eng Comput. 2022;60(5):1417–29.
17. Fischer M, Walter SS, Hepp T, Zimmer M, Notohamiprodjo M, Schick F, Yang B. Automated morphometric analysis of the hip joint on MRI from the German National Cohort Study. Radiol Artif Intell. 2021;3(5):e200213.
18. Kohli A, Xia S, Wells JE, Chhabra A. Three-dimensional CT and 3D MRI of hip-important aids to hip preservation surgery. Semin Ultrasound CT MR. 2023;44(4):252–70.
19. Nishimura Y, Ogawa M, Okamura K, Yamasaki T, Inagaki Y, Tanaka Y. Validation of compositional MRI from a histological standpoint: advantages of three-dimensional T1ρ mapping for quantitative evaluation of articular cartilage. Magn Reson Imaging. 2024;110:210–7.
20. Shi YY, Liu TJ, Zhao Q, Zhang LJ, Ji SJ. The measurement of center-edge angle of hips in a healthy Chinese population. Chia J Pediatr Surg. 2010;31(2):118–22.
21. Li LY, Zhang LJ, Li QW, Zhao QWJC, Jia J, Huang T. Development of the osseous and cartilaginous acetabular index in normal children and those with developmental dysplasia of the hip: a cross-sectional study using MRI. J Bone Joint Surg Br. 2012;94(12):1625–31.
22. Souza A, Ankolekar VV, Padmashali S, Das A, Souza AED, Hosapatna M. Femoral neck anteversion and neck shaft angles: determination and their clinical implications in fetuses of different gestational ages. Malays Orthop J. 2015;9(2):33–6.
23. Ruwe PA, Gage JR, Ozonoff MB, DeLuca PA. Clinical determination of femoral anteversion. A comparison with established techniques. J Bone Joint Surg Am. 1992;74(6):820–30.
24. Sarban S, Ozturk A, Tabur H, Isikan UE. Anteversion of the acetabulum and femoral neck in early walking age patients with developmental dysplasia of the hip. J Pediatr Orthop B. 2005;14(6):410–4.
25. Scorcelletti M, Reeves ND, Rittweger J, Ireland A. Femoral anteversion: significance and measurement. J Anat. 2020;237(5):811–26.
26. Boughton OR, Uemura K, Tamura K, Takao M, Hamada H, Cobb JP, Sugano N. Gender and disease severity determine proximal femoral morphology in developmental dysplasia of the hip. J Orthop Res. 2019;37(5):1123–32.
27. Õunpuu S, DeLuca P, Davis R, Romness M. Long-term effects of femoral ferotation osteotomies: an evaluation using three-dimensional gait analysis. J Pediatr Orthop. 2002;22(2):139–45.
28. Wylie JD, Kapron AL, Peters CL, Aoki SK, Maak TG. Relationship between the center-edge angle and 3-dimensional acetabular coverage. Orthop J Sports Med. 2017;5(4):2325967117700589.

29. Rong Y, Xiang D, Zhu W, et al. Deriving external forces via convolutional neural networks for biomedical image segmentation. Biomed Opt Express. 2019;10(8):3800–14.
30. Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell. 2017;39(12):2481–95.
31. Wong T-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recogn. 2015;48(9):2839–46.
32. Cawley GC, Talbot NLC. Fast exact leave-one-out cross-validation of sparse least-squares support vector machines. Neural Netw. 2004;17(10):1467–75.
33. Kiser KJ, Barman A, Stieb S, Fuller CD, Giancardo L. Novel autosegmentation spatial similarity metrics capture the time required to correct segmentations better than traditional metrics in a thoracic cavity segmentation workflow. J Digit Imaging. 2021;34(3):541–53.
34. Yoshida A, Kondo Y, Yoshimura N, Kuramoto T, Hasegawa A, Kanazawa T. U-Net-based image segmentation of the whole heart and four chambers on pediatric X-ray computed tomography. Radiol Phys Technol. 2022;15(2):156–69.
35. Wang L, Ye M, Lu Y, et al. A combined encoder-transformer-decoder network for volumetric segmentation of adrenal tumors. Biomed Eng Online. 2023;22(1):106.
36. Dodin P, Martel-Pelletier J, Pelletier JP, Abram F. A fully automated human knee 3D MRI bone segmentation using the ray casting technique. Med Biol Eng Comput. 2011;49(12):1413–24.
37. Memiş A, Albayrak S, Bilgili F. 3D detection of spheric and aspheric femoral heads in coronal MR images of patients with Legg-Calve-Perthes disease using the spherical Hough transform. In: Proceedings of the 2018 3rd international conference on biomedical imaging, signal processing, Bari, Italy. 2018. p. 46–52.
38. Masjedi M, Harris SJ, Davda K, Cobb JP. Mathematical representation of the normal proximal human femur: application in planning of cam hip surgery. Proc Inst Mech Eng H. 2013;227(4):421–7.
39. Bektas S. Fitting hyperboloid and hyperboloid structures. Art Des. 2017;1(1):14–22.

Publisher's Note