RESEARCH ARTICLE

Deep Learning-based Detection of Hip Anatomical Landmarks in X-ray Images

Mahdie Aghasizade, PhD; Morad Karimpour, PhD; Masoud Shariat Panahi, PhD; Mohammad Akhlaghi, MSc; Alireza Almasi Nokiani, PhD; Ramin Jafarzadeh Khaledi, MSc; Hadi G. Kashani, PhD

Research performed at The Musculoskeletal Laboratory, Faculty of Mechanical Engineering, University of Tehran, Tehran, Iran

Received: 16 September 2024 *Accepted: 4 January 2025*

Abstract

Objectives: The accurate quantification of hip deformities in medical images presented a significant challenge for radiologists. Inaccurate quantifications can lead to misclassification of diseases and improper treatment. This study introduced two-stage deep learning (DL) models designed for the detection of landmarks in hip X-ray images, aiming to improve the accuracy of deformity quantification.

Methods: This study employed two two-stage deep learning (DL) models based on VGG16 and ResNet50. The first stage aimed to detect the bounding box for each landmark, while the second stage focused on pinpointing the exact location of the landmark within the magnified box. The model automatically identified 16 hip landmarks and measures 12 specific dimensions on each side. Training, validation, and testing were conducted on a dataset comprising 854 3-joint lower limb X-ray images depicting various anomalies.

Results: The model's measurements were compared with those obtained manually to evaluate performance. The resulting average error ranged from 0.56 to 2.33 mm for different landmarks in the ResNet50-based model. The most accurate measurements were obtained for the femoral head radius and shaft width, with average errors of 0.49 mm and 0.64 mm, respectively. Conversely, the least accurate results were observed for the alpha angle, with an error of 7.37°.

Conclusion: This dataset can be valuable for diverse research endeavors related to the hip region. The 2-stage model demonstrated a brief learning time due to the small image size loaded in both steps, while also offering high precision through the use of a bounding box.

Level of evidence: V

Keywords: Anatomical landmark detection, Deep learning (DL), Hip abnormalities, Two-stage DL model

Introduction

Hip joint deformities are among the most complex musculoskeletal disorders, requiring comprehensive examinations, both radiological and clinical, for accurate diagnosis.[1] Among the various imaging modalities, plain radiography (X-ray) is considered the most fundamental and crucial technique for detecting structural diseases of the hip, as most deformities occur in the coronal plane.[2] The orthopedic community has identified a set of anatomic features that characterize hip abnormalities, which are categorized into angular, length, and proportional measurements.[2,3] Angular and longitudinal metrics are the most commonly used features. Angular measurements involve drawing lines between two specific anatomical landmarks. In comparison, longitudinal measurements determine the distance between two landmarks or between a landmark and a line drawn by two other landmarks.[4] The accurate identification of these anatomical landmarks is essential in medical imaging, as inaccuracies in hip metrics can lead to variations in surgical planning.[5] Changes in hip replacement surgery can result in leg length inequality, rotation, and misalignment of the lower limb, as well as increased pressure on the hip and

Corresponding Author: Mahdie Aghasizade, School of Mechanical Engineering, University of Tehran, Tehran, Iran

Email: m.aghasizade@ut.ac.ir

THE ONLINE VERSION OF THIS ARTICLE ABJS.MUMS.AC.IR

knee during the gait cycle.

Currently, orthopedic surgeons commonly manually draw and measure specific landmarks, as well as perform various radiographic measurements on X-rays.[6] This method is time-consuming, dependent on the surgeon's experience, and may result in variability between different observers.[7] As such, there is a need for an automated approach to accurately and efficiently identify anatomical landmarks to assess lower extremity alignment. Researchers have explored the use of automatic landmark detection, including a model developed by Bekkouch et al. (2022) for the detection and quantification of pelvic abnormalities. To overcome these challenges, experts advocate the use of Deep Learning (DL) algorithms to automatically detect anatomical landmarks, enabling precise and rapid evaluation of lower extremity alignment.[8-10]

DL methods can be classified into two categories: supervised and unsupervised learning. Unsupervised machine learning methods focus on detecting landmarks based on rules, such as the structure of bones and borders, to identify anatomical landmarks. These methods are particularly accurate when applied to healthy geometries. In contrast, supervised deep learning (DL) methods are commonly used for medical image recognition, classification, and landmark detection in abnormal geometries, particularly when trained on datasets that contain abnormal cases.[11-15]

Research on anatomical landmark detection in the hip is limited, as most studies focus on detecting dysplasia in children, typically using a limited number of landmarks and measurements.[16,17] Ha et al. identified hip abnormalities in adults using a restricted set of metrics, focusing on the femoral head center and trochanters.[18,19] Wu et al.'s studies revealed five landmarks on each leg for detecting two common abnormalities.[20]

Predicting lower limb alignment has been a prominent area of research for years, with similarities to the detection of hip landmarks. Tack et al., Schock et al., and Nguyen et al. developed a combined deep learning (DL) model to measure the hip-knee-ankle angle.[21-23] Research on the hip-knee-ankle angle typically focuses on a limited number of landmarks and metrics, meaning that the speed of automatic landmark detection is less critical in this area. Furthermore, alignment angles are determined by four key landmarks, resulting in a minimal impact of landmark errors on angular accuracy.

Two-step models have been proposed for locating hip joint landmarks using full-leg X-rays to assist in hip X-ray analysis, employing VGG-16 and ResNet 50 architectures. Aghasizade et al. demonstrated that two-step VGG-based models are more efficient and accurate in processing medical images compared to direct, one-step deep learning (DL) models. This is due to their ability to handle low-quality images in stages and refine a bounding box in the second stage.[22] Previous studies on hip anatomical landmark detection have primarily focused on detecting limited deformities and highlighted the inadequacy of landmarks on X-ray images.[18,23-26] Clinical reports for hip joint replacement surgery typically involve measuring various parameters, many of which are not addressed in similar artificial intelligence studies, as these studies primarily focus on hip dysplasia. Our approach differs from previous work by utilizing an optimal number of landmarks to calculate comprehensive anatomical measurements, enabling the generation of a holistic report on hip joint features. Furthermore, our dataset includes a diverse range of patients with various hip abnormalities, as well as standard cases. We enhance the generalizability of the DL model by diversifying and increasing the training data, compared to previous studies that have used limited datasets from healthy and dysplastic patients.[17,23] The model's proficiency is validated by comparing the landmark annotations and measurements to those performed by specialists. Consequently, the DL-based approach holds significant promise in improving the accuracy and efficiency of hip joint deformity diagnosis and treatment by automating the process of anatomical landmark detection.

Materials and Methods

Material

This dataset includes anonymized standing X-ray images of three joints, taken at a local imaging center during routine medical services between May 2020 and December 2021. All data were gathered retrospectively from patients with lower limb issues. Each patient case includes a radiograph of the entire leg, with a total of 430 patients. Three images of poor quality were removed, resulting in a dataset of 427 images [Figure 1]. Since the size of the final dataset was insufficient for training the DL model and most measurements were taken from one side, we mirrored the remaining data to create an 850-image dataset. The training and validation set consists of 85% of the study cases, with the remaining 15% randomly assigned to the test set. Demographic data for the final cohort can be found in [Table 1].

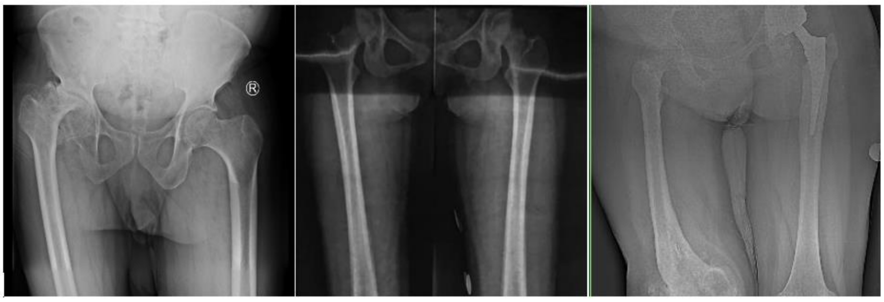

Figure 1. The excluded data of the X-ray dataset. Hip area landmarks were indistinguishable in these images. The femoral head is not viable in the right one, the femoral head is destroyed in the left one, and a bold image scratch is in the middle one

Table 1. Characteristics of the patient database used in this study

Characteristics	Developmental cohort
Number	854
Sex	
Female	604
Male	250
Prosthesis	
Total hip	4
Other hip prostheses	4
Spine	22
Neck shaft angle	
Normal (124-136)	250
Coxa vara (Low)	109
Coxa valga (High)	495
Leg Length Discrepancy	
Lower than 3 mm	96
Higher than 3 mm	758
Medial proximal femoral angle (°)	
Normal (80-89)	249
Low	144
High	457
Center edge angle (°)	
Normal (25-40)	444
Dysplasia (Low)	316
Excessive (High)	94

In the current study, we are examining the lower limbs, specifically the ilium, hip joint, and proximal femur. A skilled specialist manually annotates 28 anatomical landmarks on each image for hip pathology assessment using ImageJ software.[27] The pixel size of the images is 0.1 mm. The average width is 3092 pixels (minimum 2548, maximum 3359, standard deviation (SD) 44), and the average height is 8259 pixels (minimum 6993, maximum 10050, SD 558). To focus on the hip region, we cropped one-third of the images, from 1/30 to 11/30 of the height, while maintaining the full width. The average height of the resulting cropped images is 2753 pixels (minimum 2331, maximum 3350, SD 186).

Quantities of interest

A total of 16 anatomical landmarks are marked on each image: 14 on the left side and two on the right side [Figure 2a]. These landmarks are used to calculate 12 lower limb quantities of interest. The axes involved in this calculation include the head-neck axis, the anatomical axis of the shaft, the hip joint orientation axis, the horizontal axis, and the head-tangent axis. The horizontal axis, or intra-ischial (IIL) line, runs from the most distal part of the ischium on one side to the corresponding part on the other side; the vertical axis is defined as running perpendicular to the horizontal axis. The head-neck axis runs between the center of the femoral head and the center of the narrowest part of the femoral neck. The anatomical axis of the femur refers to the orientation of the mid-shaft. The hip joint orientation axis extends from the head of the femur to the most proximal point of the greater trochanter. Based on the frontal radiographs of the lower limbs, the typical parameters that characterize lower limb alignment are determined using the following quantities shown in [Figure 2]:

1- Femoral Head Radius (FHR) defines the radius of the best circular fit between points surrounding the femoral head.

2-Neck-Shaft Angle (NSA) is the angle between the head-neck axis and the anatomical axis of the shaft.

3-The Ischium Angle is the angle between the intra-ischial (IIL) line and the horizontal axis of the image.

4-Leg Length Discrepancy (LLD) is the vertical (perpendicular to IIL) distance between the most prominent parts of the lesser trochanters.

5- The Center-Edge (CE) Angle is the angle between a line containing the femoral head and the most lateral point of the acetabulum and the horizontal axis.

6- Medial Proximal Femoral Angle (MPFA) defines the angle between the hip joint orientation and the anatomical axes of the shaft.

7- Acetabular Coverage (AC) defines the horizontal distance between the most lateral point of the acetabulum and the femur, divided by the distance between the most lateral and medial points of the femoral head.

8- The Alpha Angle is the angle between the head-neck axis and the axis drawn between the femoral head and the first point on the lateral part of the femoral head, whose distance from the femoral head exceeds the diameter of the femoral head.

9- Horizontal Offset (HO) is the vertical distance from the femoral head to the anatomical axis of the shaft.

10- Intertrochanteric Distance (ITD) is the distance between the most prominent parts of the lesser and greater trochanters.

11- Neck Width (NW) defines the width of the neck at its narrowest part.

12- Shaft Width (SW) refers to the diameter of the shaft below the lesser trochanter.

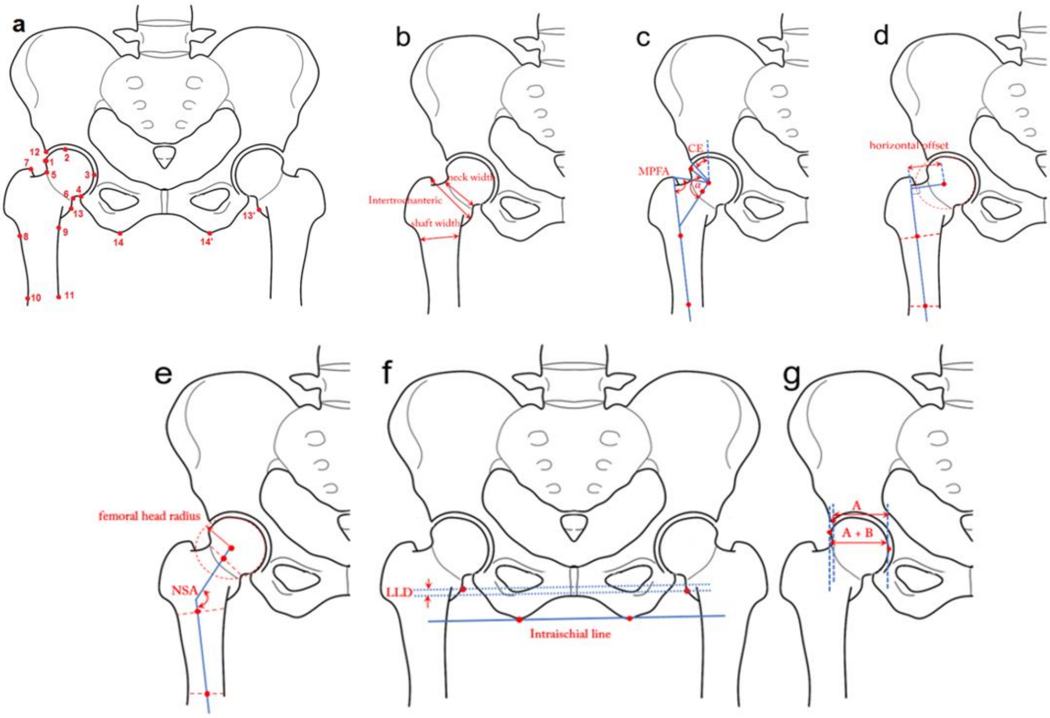

Figure 2. Anatomical landmarks and hip coronal measurements. a) Anatomical landmarks, 1, 2, 3, and 4. The most lateral, proximal, medial, and distal points of the circular part of the femoral head. 5 and 6. Lateral and medial points of the femoral neck. 7. Greater trochanter. 8 and 9. The most proximal points of the femoral shaft are lateral and medial. 10 and 11. The most distal points of the femoral shaft are lateral and medial. 12. The most lateral point of the acetabular curve. 13. Lesser trochanter. 14. The most distal point of the ischium. a, c, d, e, f, and g) All 12 measurements including neck width, intertrochanteric distance, shaft width, medial proximal femoral angle (MPFA), center edge angle (CE), alpha angle (α), horizontal offset, femoral head radius, neck shaft angle (NSA), leg length discrepancy (LLD), instar-ischial distance, and acetabular coverage (A/(A+B))

However, some of the landmarks used for measuring the quantities of interest are not explicitly defined. For instance, the position and diameter of the femoral head are determined by averaging the centers of four circles, each formed by three points corresponding to the four femoral head landmarks. In other words, four circles are drawn using four sets of three points from the femoral head landmarks, and their centers and diameters are then averaged to define the final femoral head position. Similarly, the anatomical axis of the femoral shaft is represented by four landmarks. Two of these landmarks indicate the lateral and medial aspects of the femoral shaft at the entrance of the image. In contrast, the other two indicate the medial and lateral points at the most proximal portion of the femoral shaft. Lastly, the center of the narrowest part of the femoral neck is defined by averaging the positions of two landmarks located on the lateral and medial sides of the narrowest portion. Other landmarks, including the most lateral point of the acetabulum, the greater and lesser trochanter, and the most distal part of the ischium, are directly indicated without the need for auxiliary points.

Landmark detection algorithm for hip X-ray images

The purpose of the developed two-stage deep learning (DL) models is to accurately locate anatomical landmarks on X-ray images for determining hip measurements. These landmarks are located at the edges of the bones, with some situated at prominent points (such as the lateral acetabulum, trochanters, and two neck landmarks). In contrast, others are positioned at the medial, lateral, proximal, and distal edges. To achieve accurate landmark localization, the implemented two-stage models consist of two deep learning (DL) networks connected in series. The models described below are based on VGG-16 and ResNet-50, respectively, named TSVM (two-stage VGG-based model) and TSRM (two-stage ResNet-based model). These methods are implemented separately to evaluate their performance on lower limb X-ray images.

TSVM

In the TSVM, two deep learning (DL) models are connected in series. The first and second networks are identical, except for the input image. The first-stage network in TSVM receives a whole, low-quality (512×512) image as input and provides an initial estimate for the landmark's X and Y coordinates. This preliminary prediction results in a sub-region containing the landmark. The sub-region is then used as input for the second network. By using a more detailed image and avoiding the larger field of view of the entire image, the second-stage network produces the final prediction of the landmark's position. A detailed description of the model is provided in the appendix.[28]

TSRM

In this study, a two-stage deep learning (DL) model similar to TSVM is proposed for precise landmark area estimation. The initial stage involves providing an accurate estimate of the landmark area by identifying the landmark and assuming the surrounding region as the landmark area. The subsequent stage focuses on determining the exact position of the landmark within the estimated area. The Python code for this model has been implemented in Google Colab and is publicly available. A detailed description of the model is provided in the appendix [Table 2].

Table 2. Specification of layer type, shape, and number of trainable parameters in two-stage VGG-based model (TSVM) and two-stage ResNet-based model (TSRM). This table applies to both DL model stages.

TSVM			TSRM		
Layer	**Output shape**	**Param #**	**Layer**	**Output shape**	**Param #**
Input	(224,224,3)	0	Input	(224,224,3)	0
Vgg16	(7,7,512)	14714688	ResNet50	(7,7,2048)	23000000
conv2D	(7,7,1024)	4719616	Average Pooling	2048	0
conv2D	(7,7,512)	4719104	Dense	256	524544
conv2D	(7,7,9)	4617	Dropout	256	0
Flatten	441	0	Dense	64	16448
Dropout	441	0	Dropout	64	0
Dense	128	56576	Dense	8	520
Dropout	128	0	Dense	2	18
Dense	16	2064	----	----	----
Dropout	16	0	----	----	----
Dense	9	153	----	----	----
Dense	2	20	----	----	----

Training

For both TSVM and TSRM, 15% of the data is reserved for testing, while the remaining 85% is divided into training and validation sets. The training/validation set is randomly split into five equal groups, with each group serving as the validation set in one of the five training loops. The Adam optimizer, with a learning rate of 1e-4, is used to optimize these two-stage models. Both TSVM and TSRM employ a custom loss function in the first stage to calculate the proximity to the correct position, as defined in Equation 1, thereby achieving the highest similarity in learning parameters. In this equation, x sub p and y sub p are predicted coordinates, and x_a and y_a are actual ones. In the first stage of learning, where the focus is on identifying the landmark region, implementing this loss function helps by disregarding nearly found landmarks (assigning them a weight of zero) and making a better estimation of the landmark area by weighing more distant landmarks. A threshold of 6 pixels is defined to find the surrounding box of all landmarks by testing the error in landmark annotation by the DL models. The loss function in the second stage of learning is the mean squared error, used to determine the precise location of the landmarks.

$$\text{loss} = \begin{cases} 0 & \text{if distance} < (\frac{6}{512})^2 \\ (x_p - x_a)^2 + (y_p - y_a)^2 & \text{if distance} > (\frac{6}{512})^2 \end{cases} \qquad \text{Equation 1}$$

The Python code and the complete dataset, including two different qualities, are accessible at https://github.com/MahdieAghasi/HipLandmarkDetection.git

Results

The effectiveness of the landmark localization approach proposed in TSVM, as well as a comparable single-stage VGG-based model (SSVM), was assessed through experimental evaluation. This evaluation involved comparing the predicted landmarks with those identified by specialists. The entire dataset was annotated by a radiologist and subsequently by the authors under the supervision of a radiology professor. The specialists manually marked landmarks for all images using ImageJ software. They placed 16 manual landmarks and performed 12 measurements based on these landmarks for each image.

The comparison between the results of TSVM, SSVM, and the manually marked landmarks is presented in [Figure 3]. TSVM refers to the proposed model that comprises two sequential predictions. In the second stage, the focus is on the landmark bounding box, while SSVM represents the first stage of the proposed model. Figure 3 illustrates a notable improvement in landmark detection accuracy using TSVM compared to SSVM.

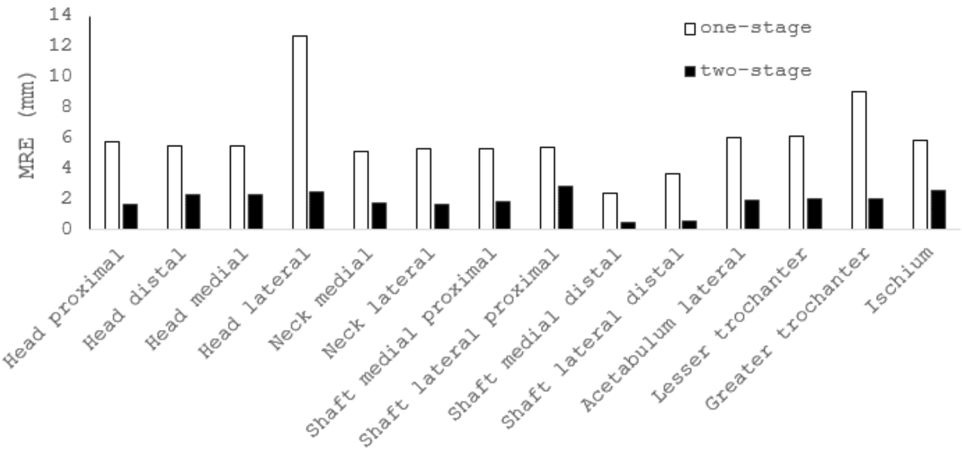

Figure 3. The mean radial error of each landmark applying single-stage and two-stage similar models based on VGG16. The two-stage model achieves significantly greater accuracy by focusing on the bounding box around the landmark detected in the single-stage model

Both TSVM and TSRM were tested using the same criteria, and the predictive results were used for performance evaluation. The error in landmark localization was calculated by determining the Euclidean distance between the model's prediction and the data provided by the specialist. For the left leg, the mean relative error (MRE), standard deviation (SD), median error, maximum error, and success detection rate (SDR) for 4 mm errors were computed for a total of 14 direct and four indirect landmarks [Table 3]. The indirect landmarks (denoted by asterisks) include the center of the femoral head, the center of the femoral neck, and the proximal and distal center points of the femoral shaft. These are determined by four femoral head landmarks, two femoral neck landmarks, and four associated landmarks on the femoral shaft. The average SDR for the landmarks is 85.4% for TSVM and 92.7% for TSRM, with MRE values of 2.30 mm and 1.63 mm, respectively. A statistical comparison of the metrics is presented in [Table 3 and Figure 2]. Additionally, Table 4 provides the mean absolute error (MAE), standard deviation (SD), median error, and maximum error for all metrics on unseen test data. The measurements are categorized into angles (intra-ischial angle, NSA, MPFA, alpha angle, CE angle) and lengths (FHR, LLD, ITD, HO, NW, SW, A, B), as well as length proportions (AC) [Table 4].

Table 3. The comparison of predicted quantities of interest with those provided by specialists in 127 unseen test images

Landmark	MRE±SD(mm)		Median (mm)		Maximum (mm)		4 mm (SDR %)	
	TSVM / TSRM		*TSVM / TSRM*		*TSVM / TSRM*		*TSVM / TSRM*	
Head center*	2.10 (2.19)	1.06 (0.81)	1.44	0.84	17.73	4.81	90	100
Proximal	1.99 (1.82)	1.10 (0.93)	1.47	0.87	12.45	5.05	90	98
Distal	2.58 (1.81)	1.91 (1.25)	2.15	1.67	9.05	7.04	83	94
Medial	2.60 (1.80)	2.05 (1.51)	2.14	1.50	8.08	6.48	79	87
Lateral	3.15 (2.72)	2.30 (1.59)	2.33	1.83	13.75	7.60	78	85
Neck center*	1.68 (1.54)	1.17 (1.02)	1.30	0.93	10.66	7.68	94	97
Medial	2.01 (1.66)	1.53 (1.07)	1.62	1.24	9.15	5.92	90	97
Lateral	1.95 (1.85)	1.42 (1.39)	1.41	1.06	12.74	8.41	90	96
Shaft proximal*	1.24 (1.26)	1.12 (0.76)	0.9	0.93	8.67	4.48	100	98
Medial	2.24 (2.31)	1.65 (1.95)	1.62	1.04	13.17	13.51	88	92

Table 3. Continued								
Lateral	3.35 (3.12)	2.33 (2.04)	2.38	1.83	15.97	14.09	71	85
Shaft distal*	0.66 (0.75)	0.53 (0.50)	0.42	0.43	4.48	3.72	99	100
Medial	0.93 (2.21)	0.56 (0.58)	0.43	0.41	15.44	5.07	96	99
Lateral	0.97 (2.21)	0.59 (0.55)	0.40	0.50	19.36	5.19	96	99
Acetabulum lateral	2.35 (1.95)	1.67 (1.48)	1.78	1.25	13.28	9.57	88	94
Lesser trochanter	2.71 (2.63)	2.26 (1.65)	2.06	1.71	17.74	7.96	83	87
Greater trochanter	2.35 (1.93)	1.83 (1.25)	1.67	1.50	11.30	6.61	87	91
Ischium	3.05 (2.71)	1.67 (1.39)	2.32	1.41	14.18	9.28	77	94

*These landmarks are found using four or two other landmarks named below them. Their position is not the model's

Table 4. Statistical comparison between measurements by the model-predicted landmarks and the specialist-annotated

Angles	MAE±SD(°)		Median (deg)		Maximum (deg)	
	TSVM / TSRM		*TSVM / TSRM*		*TSVM / TSRM*	
Intra Ischial	1.07 (1.35)	1.10 (6.00)	0.56	0.39	7.61	67.61
NSA	4.61 (5.40)	2.86 (3.65)	3.07	1.87	37.26	28.07
MPFA	2.29 (6.87)	1.76 (1.44)	1.42	1.31	28.26	16.70
Alpha	9.12 (9.85)	7.37 (6.07)	6.55	6.39	81.28	36.18
CE	1.38 (1.61)	1.31 (5.97)	0.84	0.51	9.25	67.04
Lengths	**MAE±SD(mm**		**Median (mm)**		**Maximum (mm)**	
	TSVM / TSRM		*TSVM / TSRM*		*TSVM / TSRM*	
FHR	0.74 (1.73)	0.49 (0.41)	0.43	0.40	18.95	2.20
LLD	4.03 (2.77)	5.52 (5.07)	2.92	3.94	26.25	37.17
ITD	2.02 (2.06)	1.98 (1.57)	1.43	1.48	11.33	7.73
HO	1.90 (2.43)	1.18 (1.40)	1.14	0.80	16.87	9.64
NW	1.01 (1.07)	0.75 (0.91)	0.71	0.52	7.86	6.10
SW	0.93 (0.84)	0.64 (0.65)	0.67	0.43	3.88	3.73
A	1.70 (1.61)	1.47 (1.26)	1.36	1.14	9.32	6.49
B	1.52 (1.56)	1.47 (1.53)	1.07	1.02	7.35	8.76
proportions	**MAE±SD (%)**		**Median (%)**		**Maximum (%)**	
	TSVM / TSRM		*TSVM / TSRM*		*TSVM / TSRM*	
AC	4.33 (4.57)	3.97 (4.32)	3.09	2.63	25.06	23.40

Figure 4 shows separate plots of specialist-reported ground truth and model prediction results for both methods across all 12 measurements [Figure 4]. The horizontal axis represents the ground truth, while the vertical axis indicates the predicted values. The closer the points are to the 45-degree hypothetical line, the lower the mean relative error (MRE), and the closer they are to the trendlines, the lower the standard deviation (SD).

Discussion

The current study highlights the effectiveness of TSVM and TSRM in accurately locating anatomical landmarks using whole-leg radiographs. The proposed models achieved low average errors of 2.30 mm for TSVM and 1.63 mm for TSRM, demonstrating their potential for precise landmark localization. Furthermore, the inclusion of a comprehensive dataset containing various anatomical variations, such as prostheses, coxa vara and valga, leg length discrepancies, and abnormal MPFA and CE angles, enhances the reliability and applicability of the models in radiology practice.

However, several limitations must be addressed. First, while the presence of foreign objects, such as orthopedic fixators, was briefly mentioned, their impact on the model's performance warrants further investigation. These objects can introduce artifacts into radiographs, potentially affecting the accuracy of landmark detection due to the creation of white areas and the obstruction of significant portions of the X-ray images. Loss of image data means that the model must estimate the location of the landmark, which can result in decreased accuracy. Future studies should focus on quantifying the extent of this effect and exploring methods to minimize its impact, such as developing techniques to detect and compensate for these artifacts in the dataset.

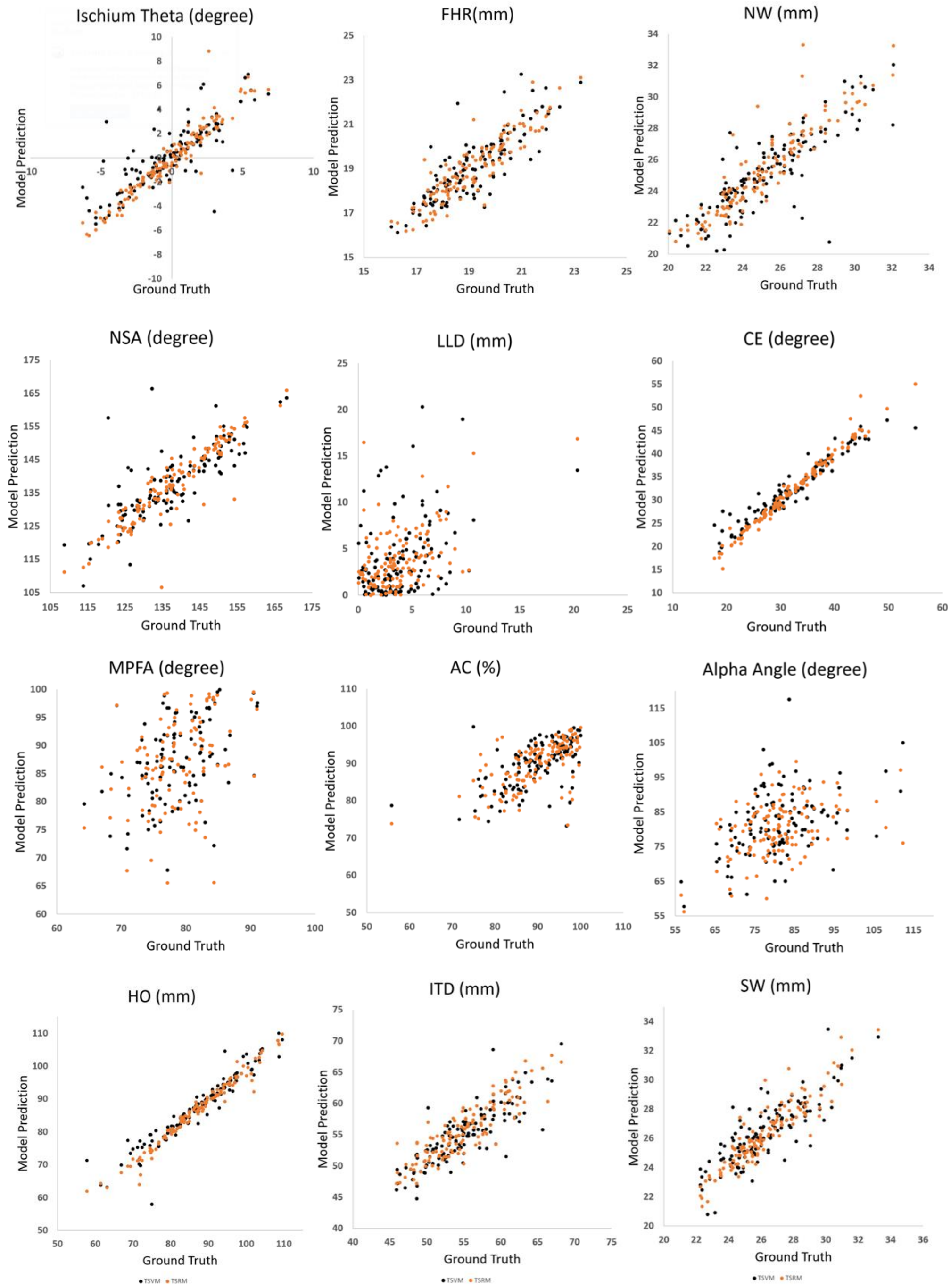

Figure 4. Scatterplots displaying the amount of each measurement calculated with specialist landmarks and automatically detected landmarks by a two-stage VGG-based model (TSVM) and a two-stage ResNet-based model (TSRM), colored in black and orange. Measurements are femoral head radius (FHR), neck width (NW), ischium theta, center edge angle (CE), leg length discrepancy (LLD), neck shaft angle (NSA), inter-trochanteric distance (ITD), acetabular coverage (AC), medial proximal femoral angle (MPFA), horizontal offset (HO), shaft width (SW), and alpha angle

Second, the study noted that AC measurements were susceptible to issues arising from the proximity of landmarks in low-quality images. This limitation should be further investigated, as it could impact the clinical utility of the model in cases involving lower-quality radiographs. Future work could address this by incorporating advanced image enhancement or preprocessing techniques to improve the quality of input images, potentially reducing AC sensitivity and improving overall accuracy.

Third, the retrospective nature of this study may introduce biases in dataset selection, such as an overrepresentation of certain conditions or imaging types. Retrospective studies often face challenges in ensuring a balanced dataset, which could affect the model's generalizability. Future research should aim to include a more diverse and balanced dataset, comprising a larger proportion of normal cases and a broader range of abnormalities, to more effectively evaluate the model's performance across a wider variety of clinical scenarios.

A notable strength of this study is the use of auxiliary markers on the femoral head, neck, and shaft peripheries to infer the center and orientation of these anatomical structures. This approach yielded promising results in predicting measurements, demonstrating its potential for clinical applications. The evaluation of hip measurement parameters revealed high accuracy, with minimal divergence from measurements taken by orthopedic specialists. Despite slightly higher errors associated with proximal landmarks on the femoral shaft, the overall trajectory of the femoral shaft was accurately identified, leading to reliable predictions.

The comparison of the two methodologies for landmark detection highlighted the superiority of TSRM, while also demonstrating the acceptable performance of TSVM [Figure 5]. The inclusion of multiple auxiliary landmarks contributed to improved accuracy, as evidenced by the lower errors associated with the additional landmarks.

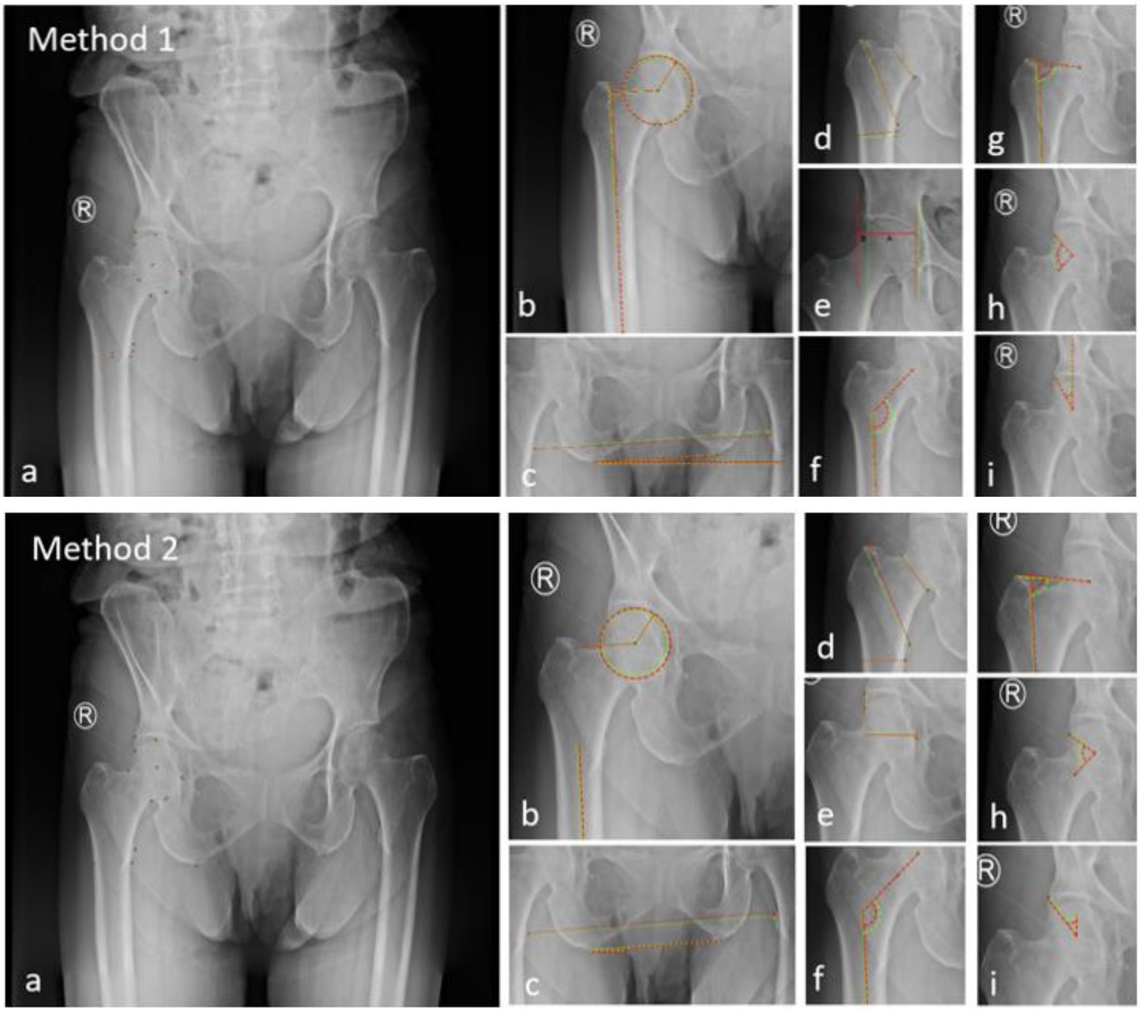

Figure 5. The result of hip landmark locating and related measurements by each method for one randomly selected hip X-ray image. The results presented for the model predicted (red lines and smaller radius arc) and specialist ground truth (green lines and larger radius arc) (a) All 14 direct and three indirect landmarks, (b) Femoral head circle, femoral head radius, shaft anatomical axis, and horizontal offset, (c) intra-ischial length and leg length discrepancy, (d) neck width, shaft width and inter-trochanteric distance, (e) acetabular coverage (A/(A+B)) and the related parameters (A and B), (f) neck shaft angle, (g) medial proximal femoral angle, (h) alpha angle, and (i) center edge

In conclusion, the study highlights the potential of the proposed model for accurate landmark localization and measurement prediction using whole-leg radiographs, with significant implications for enhancing diagnostic accuracy in clinical practice. However, the limitations discussed should be addressed in future work. Exploring methods to mitigate the impact of artifacts, improving sensitivity to AC, and enhancing dataset diversity will contribute to refining and advancing this model for broader clinical applicability. Furthermore, additional investigation into the integration of auxiliary landmarks could further improve the accuracy of landmark detection and measurement prediction in medical imaging.

We conducted a comparative analysis of the mean error in identifying hip landmarks across our models, referencing methods outlined in nine other publications.[18,21,23-26,29-31] While there have been limited efforts to detect hip landmarks, some studies have focused on knee or alignment detection.[32-36] Additionally, some publications that concentrate on the hip do not cover as many landmarks as our study.[18,19,30,32] Nevertheless, similar studies have presented comparable landmarks and measurements.[6,23,37] In Table 5, we present the results from other research studies alongside those from our proposed method, based on 127 unseen data points. Our two-stage ResNet50-based model demonstrates a lower average error in pinpointing the femoral head center, lateral point of the head, femoral neck center, FHR, ITD, and HO when compared to other models. However, the comparative accuracy for AC is notably lower than the best result [Table 5].

Table 5. Comparison between our proposed methods and existing methods for AC (Acetabular Coverage), the 2D position of the center of the femoral head, the most lateral point of the femoral head, the middle of the narrowest part of the femoral neck, the most lateral point of the acetabulum, FHR (Femoral Head Radius), ITD (Intertrochanteric Distance), HO (Horizontal Offset), SW (Shaft Width), NW (Neck Width), NSA (Neck Shaft Angle), and CE (Center Edge). The mentioned parameters regarding MRE for landmarks and metrics. Errors indicate the discrepancy between the predicted landmark position/metrics and the manual ground truth.

Parameter	Unit	Proposed method		21	32	34	26	27	18	25	33	28
		TSRM	*TSVM*									
AC	%	3.97	4.33				L:2.0/R:2.0			5.1		
Femoral head center	mm	1.06	2.10	1.72	L:3.0/R:5.0	3.09			1.41			
Head lateral		2.30	3.15						3.6			
Femoral neck		1.17	1.68					2.00	1.68			
Acetabulum lateral		1.67	2.35					3.43	1.17			
FHR		0.49	0.74					1.3				0.8
ITD		1.98	2.02					2.5				
HO		1.18	1.90					3.1				
SW		0.64	0.93					0.6				
NW		0.75	1.01					0.5				
NSA	degree	2.86	4.61					3.55		2.6		5.7
CE		1.31	1.38				L:1.55/R:1.79	1.63	2.6	1.2	4.1	1.1

Conclusion

In this study, we introduced two two-stage deep learning (DL) methods designed to identify 14 anatomical landmarks in the hip region, which are crucial for calculating various hip measurements, including angles (NSA, MPFA, alpha, and CE), distances (FHR, ITD, HO, and LLD), and proportions (AC). The first method utilized the pretrained VGG16 model, while the second method was based on the pretrained ResNet50 model. Both methods followed a two-stage approach: in the first stage, a landmark is detected, and in the second stage, the area around the detected landmark is used for more accurate localization. Both models outperformed manual annotations.

The TSRM achieved more accurate landmark detection than the TSVM. A comparison of our results with those from other studies demonstrated superior accuracy in measurements for the TSRM in most cases. Furthermore, the study emphasizes that two-stage deep learning (DL) models significantly enhance landmark detection accuracy compared to single-stage models.

The novelty of this work lies in improving hip landmark detection accuracy through a hybrid deep learning architecture that is capable of simultaneously detecting all critical and auxiliary hip landmarks. While some measurements exhibited acceptable errors, AC proved to be more sensitive to issues arising from the proximity of landmarks in low-quality images. Our findings also suggest that considering four landmarks for head center prediction, rather than just one, significantly improves prediction accuracy, highlighting the importance of incorporating additional landmarks for more accurate diagnosis and treatment planning.

Ultimately, our study highlights the importance of expanding research beyond specific conditions, such as hip dysplasia, to ensure clinical applicability across a broader range of cases. The identified landmarks are versatile in evaluating a range of features, such as angles, lengths, and proportions, making the proposed method well-suited for diagnosing various hip deformities when applied to diverse datasets.

Acknowledgement

All people make contributions are named as author.

Authors Contribution: Authors who conceived and designed the analysis: M. Aghasizade, M. Karimpour, M. Shariat Panahi/ Authors who collected the data: M. Karimpour, A. Almasi Nokiani, R. Jafarzadeh/ Authors who contributed data or analysis tools: M. Aghasizade, M. Karimpour, A. Almasi Nokiani, R. Jafarzadeh/ Authors who performed the analysis: M. Aghasizade, M. Akhlaghi/ Authors who wrote or revised the paper: All Authors

Declaration of Conflict of Interest: The author(s) do NOT have any potential conflicts of interest for this manuscript.

Declaration of Funding: The author(s) received NO financial support for the preparation, research, authorship, and publication of this manuscript.

Declaration of Ethical Approval for Study: is study was conducted in accordance with the ethical principles of the Declaration and was approved by the Ethics Committee of University of Tehran (Approval ID: IR.UT.SPORT.REC.1402.131). Since the data used in this research were retrospectively extracted from routine clinical records collected during standard patient care, and no additional data collection or intervention was performed specifically for this study, the requirement for informed consent was waived by the Ethics Committee. All patient data were anonymized and handled in strict compliance with confidentiality guidelines.

Declaration of Informed Consent: Declaration of Informed Consent:

The submitted manuscript contains no information that could identify any patient, including names, initials, hospital identification numbers, dates of birth, or photographs. All data have been fully anonymized, and informed consent was not required for this retrospective analysis of de-identified clinical data.

Mahdie Aghasizade PhD [1]
Morad Karimpour PhD [1]
Masoud Shariat Panahi PhD [1]
Mohammad Akhlaghi MSc [1]
Alireza Almasi Nokiani PhD [2]
Ramin Jafarzadeh Khaledi MSc [3]
Hadi G. Kashani PhD [1]

1 School of Mechanical Engineering, University of Tehran, Tehran, Iran

2 Firoozabadi Clinical Research Development Unit (FACRDU), Iran. University of Medical Sciences (IUMS), Tehran, Iran

3 Rad Radiology and Sonography Clinic, Tehran, Iran

References

1. Wilson JJ, Furukawa M. Evaluation of the patient with hip pain. Am Fam Physician. 2014;89(1):27-34.
2. Lim SJ, Park YS. Plain radiography of the hip: a review of radiographic techniques and image features. Hip Pelvis. 2015;27(3):125-34. doi: 10.5371/hp.2015.27.3.125.
3. Kehr P. The hip joint in adults: advances and developments edited by K. Mohan Iyer. European Journal of Orthopaedic Surgery & Traumatology. 2018;28(8):1645-6.
4. Liu W, Wang Y, Jiang T, Chi Y, Zhang L, Hua XS. Landmarks detection with anatomical constraints for total hip arthroplasty preoperative measurements. InInternational Conference on Medical Image Computing and Computer-Assisted Intervention 2020 (pp. 670-679). Cham: Springer International Publishing.
5. Torosdagli N, Liberton DK, Verma P, Sincan M, Lee JS, Bagci U. Deep geodesic learning for segmentation and anatomical landmarking. IEEE Trans Med Imaging. 2019;38(4):919-931. doi: 10.1109/TMI.2018.2875814.
6. Bekkouch IEI, Maksudov B, Kiselev S, Mustafaev T, Vrtovec T, Ibragimov B. Multi-landmark environment analysis with reinforcement learning for pelvic abnormality detection and quantification. Med Image Anal. 2022:78:102417. doi: 10.1016/j.media.2022.102417.
7. Jo C, Hwang D, Ko S, et al. Deep learning-based landmark recognition and angle measurement of full-leg plain radiographs can be adopted to assess lower extremity alignment. Knee Surg Sports Traumatol Arthrosc.2023;31(4):1388-1397. doi: 10.1007/s00167-022-07124-x.
8. Juneja M, Garg P, Kaur R, et al. A review on cephalometric landmark detection techniques. Biomedical Signal Processing and Control. 2021;66:102486.
9. Schwendicke F, Chaurasia A, Arsiwala L, et al. Deep learning for cephalometric landmark detection: systematic review and meta-analysis. Clin Oral Investig. 2021;25(7):4299-4309. doi: 10.1007/s00784-021-03990-w.
10. Khalid H, Hussain M, Al Ghamdi MA, et al. A comparative systematic literature review on knee bone reports from MRI, X-rays, and CT scans using deep learning and machine learning methodologies. Diagnostics (Basel). 2020;10(8):518. doi: 10.3390/diagnostics10080518.
11. Hernigou P, Barbier O, Chenaie P. Hip arthroplasty dislocation risk calculator: evaluation of one million primary implants and twenty-five thousand dislocations with deep learning artificial intelligence in a systematic review of reviews. Int Orthop. 2023;47(2):557-571. doi: 10.1007/s00264-022-05644-2.

12. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health. 2019;1(6):e271-e297. doi: 10.1016/S2589-7500(19)30123-2.
13. Kalmet PH, Sanduleanu S, Primakov S, et al. Deep learning in fracture detection: a narrative review. Acta Orthop. 2020;91(2):215-220. doi: 10.1080/17453674.2019.1711323.
14. Chea P, Mandell JC. Current applications and future directions of deep learning in musculoskeletal radiology. Skeletal Radiol. 2020;49(2):183-197. doi: 10.1007/s00256-019-03284-z.
15. Fischer MC, Krooß F, Habor J, Radermacher K. A robust method for automatic identification of landmarks on surface models of the pelvis. Sci Rep. 2019;9(1):13322. doi: 10.1038/s41598-019-49573-4.
16. Thompson P, Medical Annotation Collaborative, Perry DC, Cootes TF, Lindner C. Automation of clinical measurements on radiographs of children's hips. InInternational Conference on Medical Image Computing and Computer-Assisted Intervention 2022 (pp. 419-428). Cham: Springer Nature Switzerland.
17. Liu C, Xie H, Zhang S, Mao Z, Sun J, Zhang Y. Misshapen pelvis landmark detection with local-global feature learning for diagnosing developmental dysplasia of the hip. IEEE Trans Med Imaging. 2020;39(12):3944-3954. doi: 10.1109/TMI.2020.3008382.
18. McCouat J, Voiculescu I, Glyn-Jones S. Automatically diagnosing hip conditions from x-rays using landmark detection. In2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) 2021 (pp. 179-182). IEEE.
19. Ha A, Vorhies J, Campion A, et al. Automatic Extraction of Skeletal Maturity from Whole Body Pediatric Scoliosis X-rays Using Regional Proposal and Compound Scaling Convolutional Neural Networks. In2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2020 (pp. 996-1000). IEEE.
20. Wu H, Xie H, Lin F, Zhang S, Sun J, Zhang Y. WaveCSN: cascade segmentation network for hip landmark detection. InProceedings of the 1st ACM International Conference on Multimedia in Asia 2019 (pp. 1-6).
21. Tack A, Preim B, Zachow S. Fully automated assessment of knee alignment from full-leg X-rays employing a "YOLOv4 And Resnet Landmark regression Algorithm"(YARLA): data from the Osteoarthritis Initiative. Comput Methods Programs Biomed. 2021:205:106080. doi: 10.1016/j.cmpb.2021.106080.
22. Aghasizade M, Kiyoumarsioskouei A, Hashemi S, et al. A coordinate-regression-based deep learning model for catheter detection during structural heart interventions. Applied Sciences. 2023;13(13):7778.
23. Archer H, Reine S, Alshaikhsalama A, et al. Artificial intelligence-generated hip radiological measurements are fast and adequate for reliable assessment of hip dysplasia: an external validation study. Bone Jt Open. 2022;3(11):877-884. doi: 10.1302/2633-1462.311.BJO-2022-0125.R1.
24. Yang W, Ye Q, Ming S, et al. Feasibility of automatic measurements of hip joints based on pelvic radiography and a deep learning algorithm. Eur J Radiol. 2020:132:109303. doi: 10.1016/j.ejrad.2020.109303.
25. Hussain D, Han S-M, Kim T-S. Automatic hip geometric feature extraction in DXA imaging using regional random forest. J Xray Sci Technol. 2019;27(2):207-236. doi: 10.3233/XST-180434.
26. Mast NH, Impellizzeri F, Keller S, Leunig M. Reliability and agreement of measures used in radiographic evaluation of the adult hip. Clin Orthop Relat Res. 2011;469(1):188-99. doi: 10.1007/s11999-010-1447-9.
27. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671-5. doi: 10.1038/nmeth.2089.
28. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014 Sep 4.
29. Grupp RB, Unberath M, Gao C, et al. Automatic annotation of hip anatomy in fluoroscopy for robust and efficient 2D/3D registration. Int J Comput Assist Radiol Surg. 2020;15(5):759-769. doi: 10.1007/s11548-020-02162-7.
30. Jensen J, Graumann O, Overgaard S, et al. A Deep Learning Algorithm for Radiographic Measurements of the Hip in Adults-A Reliability and Agreement Study. Diagnostics (Basel). 2022;12(11):2597. doi: 10.3390/diagnostics12112597.
31. Jang SJ, Kunze KN, Vigdorchik JM, Jerabek SA, Mayman DJ, Sculco PK. John Charnley Award: deep learning prediction of hip joint center on standard pelvis radiographs. J Arthroplasty. 2022;37(7S):S400-S407.e1. doi: 10.1016/j.arth.2022.03.033.
32. Jo C, Hwang D, Ko S, et al. Deep learning-based landmark recognition and angle measurement of full-leg plain radiographs can be adopted to assess lower extremity alignment. Knee Surg Sports Traumatol Arthrosc. 2023;31(4):1388-1397. doi: 10.1007/s00167-022-07124-x.
33. Nguyen TP, Chae D-S, Park S-J, Kang K-Y, Lee W-S, Yoon J. Intelligent analysis of coronal alignment in lower limbs based on radiographic image with convolutional neural network. Comput Biol Med. 2020:120:103732. doi: 10.1016/j.compbiomed.2020.103732.
34. Wojciechowski W, Molka A, Tabor Z. Automated measurement of parameters related to the deformities of lower limbs based on x-rays images. Comput Biol Med. 2016:70:1-11. doi: 10.1016/j.compbiomed.2015.12.027.
35. Cakmak GR, Hamamci IE, Yilmaz MK, Alhajj R, Azboy I, Ozdemir MK. AutoCOR: Autonomous Condylar Offset Ratio Calculator on TKA-Postoperative Lateral Knee X-ray. arXiv preprint arXiv:220403120. 2022;doi:10.2139/ssrn.4856904.
36. Wu H, Xie H, Liu C, Zha ZJ, Sun J, Zhang Y. Circlenet for hip landmark detection. InProceedings of the AAAI Conference on Artificial Intelligence 2020 (Vol. 34, No. 07, pp. 12370-12377).
37. Stotter C, Klestil T, Röder C, et al. Deep Learning for Fully Automated Radiographic Measurements of the Pelvis and Hip. Diagnostics (Basel). 2023;13(3):497. doi: 10.3390/diagnostics13030497.

Appendix

TSVM details: As previously mentioned, the two networks share similarities. We implemented a Python code in Google Colab to promote a two-stage VGG-based DL method. Each stage includes a VGG-16. ^28 backbone (frozen with weights pre-trained on ImageNet),28 followed by three convolutional layers and four fully connected (FC) layers. The kernel sizes of the first and second convolutional layers are 3×3, while the third is 1×1. The convolution layers have 1024, 512, and 9 channels, respectively. The activation function for the first two convolutions is Rectified Linear Unit (ReLU), and the third is Sigmoid. Before the first three of the four FC layers, a dropout of 0.2 is applied. These FC layers have 128, 16, 9, and 2 neurons, respectively, all of which utilize the ReLU activation function. The input images are reshaped to 224×224 pixels since the input of the VGG network is a 224×224×3 image. The output consists of two numbers, which show the X and Y components of the landmark in the input image. Table 2 lists the type, size, and number of trainable parameters for each layer.
TSRM details: Both the first and second stages of the DL model share the same architecture, with a detailed explanation provided for one of the stages. The input image, sized 224×224, is fed into a ResNet 50 model pre-trained on ImageNet weights. The output undergoes global average pooling before being processed by a fully connected (FC) layer with 256 neurons. Subsequently, a dropout rate of 0.2 is employed to reduce the computational load and prevent overfitting, followed by another fully connected (FC) layer with 64 neurons. After a dropout of 0.1, the last FC layer with eight neurons is utilized, and the final FC layer with two neurons indicates the X and Y coordinates of the landmark. The activation function for the first three fully connected (FC) layers is ReLU, while the last layer employs a linear function. Detailed information on these stages is provided in Table 2.