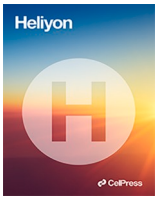
Heliyon 10 (2024) e29677

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Automated assessment of pelvic radiographs using deep learning: A reliable diagnostic tool for pelvic malalignment

Ki-Ryum Moon [a], Shi Sub Byon [b], Sung Hyun Kim [c], Byoung-Dai Lee [a,*]

[a] Department of Computer Science, Kyonggi University, Suwon-si, Gyeonggi-do, 16227, Republic of Korea
[b] AI Lab., HealthHub, Co. Ltd., Seoul, Republic of Korea
[c] Human Medical Imaging and Intervention Center, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Standing anteroposterior
Pelvic radiographs
Pelvic displacement
Landmark detection
Deep learning

ABSTRACT

Pelvic malalignment leads to general imbalance and adversely affects leg length. Timely and accurate diagnosis of pelvic alignment in patients is crucial to prevent additional complications arising from delayed treatment. Currently, doctors typically assess pelvic alignment either manually or through radiography. This study aimed to develop and assess the validity of a deep learning-based system for automatically measuring 10 radiographic parameters necessary for diagnosing pelvic displacement using standing anteroposterior pelvic X-rays. Between March 2016 and June 2021, pelvic radiographs from 1215 patients were collected. After applying specific selection criteria, 550 pelvic radiographs were chosen for analysis. These data were utilized to develop a deep learning-based system capable of automatically measuring radiographic parameters relevant to pelvic displacement diagnosis. The system's diagnostic accuracy was evaluated by comparing automatically measured values with those assessed by a clinician using 200 radiographs selected from the initial 550. The results indicated that the system exhibited high reliability, accuracy, and reproducibility, with a Pearson correlation coefficient of $\geq$0.9, an intra-class correlation coefficient of $\geq$0.9, a mean absolute error of $\leq$1 cm, mean square error of $\leq$1 cm, and root mean square error of $\leq$1 cm. Moreover, the system's measurement time for a single radiograph was found to be 18 to 20 times faster than that required by a clinician for manual inspection. In conclusion, our proposed deep learning-based system effectively utilizes standing anteroposterior pelvic radiographs to precisely and consistently measure radiographic parameters essential for diagnosing pelvic displacement.

1. Introduction

The pelvis, a wide bowl-shaped bone structure, comprises the hip bones on either side, the sacrum in the rear, and the coccyx, serving as a vital link between the body's torso and lower extremities at the lower abdomen while safeguarding abdominal organs [1]. It plays a crucial role in dispersing and absorbing forces from the head and upper body, as well as those exerted from the lower extremities, while also contributing to body movement through coordinated muscle activity alongside the lumbar region and femurs. Additionally, it maintains the body in an upright position [2].

However, instances of pelvic torsion and displacement are on the rise among populations, attributed to declining physical activity

* Corresponding author.
E-mail address: blee@kgu.ac.kr (B.-D. Lee).

https://doi.org/10.1016/j.heliyon.2024.e29677
Received 11 December 2023; Received in revised form 12 April 2024; Accepted 12 April 2024
Available online 16 April 2024

due to economic development and societal structural changes. Consequently, individuals develop poor posture habits and encounter environmental factors in their work and daily lives that contribute to bodily imbalances, potentially leading to stress, abdominal fat accumulation, and weight issues [3]. Pelvic displacement can induce lumbar region rotation and alter spine curvature, disrupting the spine's natural support structure [4]. As pelvic and spinal deformation progresses, it places strain on nearby soft tissues like muscles and ligaments [5], affecting lower body balance and causing fatigue and discomfort in areas such as knees and ankles during routine movements [6]. In severe cases, musculoskeletal disorders like back pain, sciatica, neck pain, and shoulder stiffness can arise, adversely impacting nearby organs and resulting in various disorders [7]. Hence, accurate and timely diagnoses of pelvic displacement progression are imperative for prevention.

Various tests can proactively assess pelvic displacement and its progression, with common examples including leg length checking [8] and plain radiograph analysis [9]. Leg length checking involves measuring the length of the legs, typically conducted as an initial assessment to identify general structural defects in the body [10]. However, there is currently no universally accepted diagnostic standard or specific clinical practice method proposed by researchers [11]. Plain radiograph analysis is a method used in clinical settings to diagnose overall pelvic displacement, encompassing the ilium, sacrum, and ischium, by observing and measuring pelvic alignment and quantitative parameters through radiographic imaging of the pelvic structure. This method is favored for its cost-effectiveness [12].

Nevertheless, physicians encounter challenges when personally examining pelvic displacement using radiographs. Firstly, distinguishing between different types and stages of pelvic displacement proves difficult due to variations among patients. Secondly, a wide array of parameters are utilized to assess pelvic displacement. Additionally, discrepancies in skeletal structures and cartilage conditions among patients make it challenging to discern subtle changes in images, leading to potential difficulties in accurate interpretation of structural alterations. Furthermore, the proficiency and condition of the physician during measurements may influence assessment outcomes.

In recent years, medical imaging has emerged as a pivotal area undergoing extensive research and development to integrate artificial intelligence, particularly deep learning (DL) technology. Notably, methods utilizing convolutional neural networks (CNN) have shown promising results in image recognition and are being explored as diagnostic aids in medical imaging [13]. Consequently, this study sought to create a landmark-based automatic measurement system using standing anteroposterior (AP) pelvic radiographs and a DL algorithm. Subsequently, we evaluated the system's reliability and its concordance with a radiologist's assessment regarding radiographic parameters pertinent to pelvic displacement.

2. Related work

Accurately assessing skeletal displacement and malalignment is paramount for the precise diagnosis and treatment of musculoskeletal disorders. These assessments primarily rely on anatomical parameters measurable from medical images. However, traditional manual measurement methods are time-consuming and exhibit somewhat inconsistent results. Consequently, ongoing research has focused on utilizing image processing techniques to automatically measure these parameters [14–16].

Historically, techniques such as canny edge detection [17] and Hough transform [18] were employed to comprehend the anatomical structure of the pelvic region and diagnose disorders like fractures and arthritis [19]. However, these methods have limitations; they may struggle to accurately identify targeted straight lines or landmarks due to inherent algorithmic constraints. Additionally, the configuration parameters for each method often require fine-tuning based on specific image characteristics or patterns. As current DL technologies advance, researchers are increasingly exploring neural networks to identify anatomical landmarks and analyze X-ray image patterns.

Jodeiri et al. [20] proposed a method utilizing the Mask R-CNN to separate iliac regions in pelvic X-ray images and measure pelvic sagittal inclination (PSI) using a CNN on the segmented iliac regions. While this method demonstrated high accuracy for normal pelvic X-ray images, it yielded relatively inaccurate segmentation results for iliac regions in radiographs containing non-pelvic elements due to patient movement.

Yang et al. [21] proposed a system utilizing a network comprising several stacked convolutional layers to detect five femur-related anatomical landmarks and utilize them to measure three angles crucial for diagnosing and monitoring hip joint disorders, including osteoarthritis [22], developmental dysplasia of the hip [23], and femoroacetabular impingement syndrome [24]. However, this study had limitations as it utilized a relatively small dataset of only 100 pelvic radiographs. Furthermore, it did not verify the accuracy and consistency of the system by comparing the results of the DL model with angles personally measured by radiographic specialists.

Bier et al. [25] introduced a method for detecting anatomical landmarks in X-ray images to measure parameters used in determining the necessity for pelvic trauma surgery. Their study generated 2D X-ray images projected from various directions based on 3D whole-body CT images and employed a DL network model, consisting of six stages, to extract 23 landmarks from these X-ray images. This technique enabled the detection of landmarks in X-ray images captured from different angles and directions. However, the accuracy of the technique was affected when images not included in the training, such as those containing medical implants, were provided as input. Additionally, the prediction performance of the technique was somewhat lower for pelvic images captured from the sides compared to those captured from the front or rear.

Stotter et al. [26] introduced a system utilizing a DL network to automatically detect major landmarks associated with the pelvis and hip joints and utilize these landmarks to compute anatomic distances and angles. Their approach employs an AI-based software trained on over 4000 pelvis and hip joint images. The software utilizes a CNN-based classification model on pelvic X-ray images to ascertain the presence of medical implants. If none are detected, it employs a U-Net-based model on relevant pelvic region segments to segment the images and predict the landmarks. These acquired hip joint and pelvis anatomical landmarks are then utilized to measure

various parameters, including the caput-collum-diaphyseal angle, lateral-center-edge angle, acetabular index, and femoral head extrusion index.

Pei et al. [27] proposed a method for measuring center-edge angles, sharp angles, and Tönnis angles in pelvic X-rays, crucial for determining appropriate treatment times for patients with developmental dysplasia of the hip. Landmarks were detected beforehand to measure these angles, employing an attentional algorithm that considers spatial information and correlations between channels to establish long-distance pixel correlations, resulting in high landmark detection performance. However, these results were obtained from experiments conducted on high-quality X-rays, suggesting potential performance degradation if X-ray image contrast is low or the intensity deviates significantly from that of the training data.

3. Material and methods

3.1. Study participants and datasets

In this study, pelvic X-ray images were gathered from March 2016 to June 2021, and a rigorous data selection criterion was established to ensure data accuracy. Images were scrutinized for the presence of orthopedic devices and hip joint-related disorders, such as developmental dysplasia of the hip, osteoarthritis, and femoroacetabular impingement syndrome, and those from patients with these disorders were excluded. Consequently, 550 pelvic X-ray images obtained from 550 patients after this data selection process constituted the dataset for system development. The entire dataset was sourced from HPACS (HealthHub, Co., Ltd., Seoul, Korea), with all identifiable information removed from the images to safeguard personal privacy. Fig. 1 provides an overview of the data organization process.

The pelvic X-rays in the dataset had an average resolution of approximately 3000 × 3000 and were 24-bit black-and-white images in JPEG format. The dataset was partitioned into training, validation, and test sets, comprising 245, 105, and 200 images, respectively. Table 1 presents basic patient information included in the dataset.

3.2. Radiographic parameters for assessing pelvic displacement

A set of commonly used radiographic parameters for assessing pelvic displacement was selected to evaluate the model [28–30]. Table 2 summarizes these radiographic parameters, while Fig. 2 illustrates the 16 landmarks necessary for calculating them. A total of 8320 landmarks from 550 pelvic radiographs, labeled by a board-certified radiologist, were utilized for training and validating the DL algorithm for landmark detection.

3.3. Model architecture, data preprocessing, and augmentation

Our proposed dual-stage pelvic landmark detection network conducts a sequential learning process comprising two stages (see Fig. 3). Initially, the image preprocessing stage generates five cropped images, encompassing the input image's four corners and the middle part of the image. These five cropped images, along with the original images, are utilized for training purposes. Incorporating

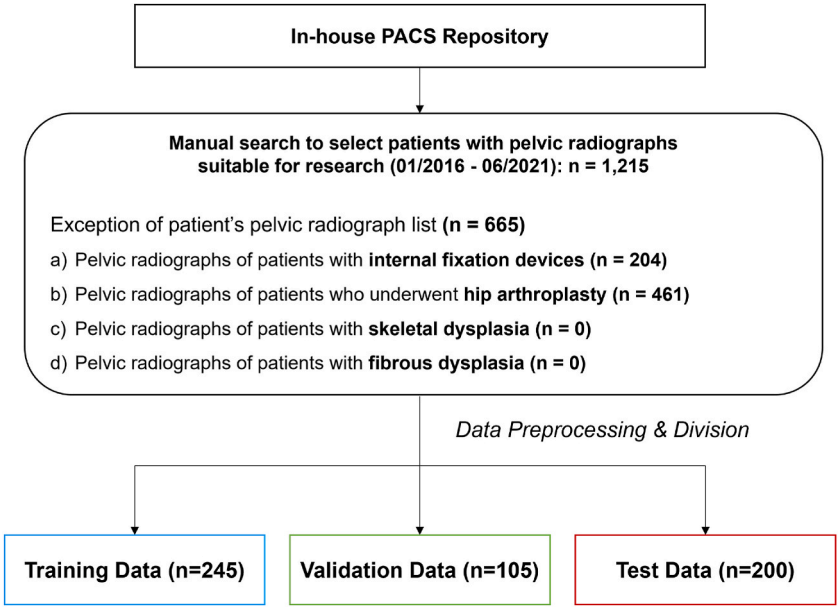


Fig. 1. Flowchart illustrating the number of patients, reflecting the pre-defined exclusion criteria for this study.

Table 1
Characteristics of study participants.

		Training set	Validation set	Test set
Demographics	Total no. of patients	245	105	200
	No. of male patients	103	36	93
	No. of female patients	142	69	107
	Mean age (year)	52.62 ± 15.56	51.94 ± 17.71	52.44 ± 14.96
	Age range (year)	14–89	12–87	13–89
Parameter	FHHD (mm)	4.03 ± 3.32	4.67 ± 3.71	3.96 ± 3.36
	ALFHRF (°)	1.15 ± 0.96	1.33 ± 1.04	1.15 ± 0.98
	Left IM (mm)	225.08 ± 14.03	224.92 ± 14.16	223.81 ± 13.90
	Right IM (mm)	225.34 ± 14.16	226.14 ± 14.79	224.30 ± 14.46
	ICHD (mm)	4.72 ± 3.81	5.65 ± 4.55	4.62 ± 3.95
	DOCS (mm)	4.55 ± 3.36	4.36 ± 3.35	4.25 ± 3.36
	Left SAM (mm)	64.04 ± 4.94	63.35 ± 5.15	63.46 ± 4.92
	Right SAM (mm)	62.74 ± 4.66	63.43 ± 4.92	62.81 ± 4.42
	Left ISM (mm)	122.42 ± 10.23	122.43 ± 10.01	121.25 ± 11.59
	Right ISM (mm)	118.96 ± 10.68	119.50 ± 11.00	118.49 ± 10.51

FHHD: Femoral head height difference; ALFHRF: Angle between the left femoral head horizontal line and the right femoral head horizontal line; Left IM: Left innominate measurement; Right IM: Right innominate measurement; ICHD: Iliac crest height difference; DOCS: Distance between off-centering measurement and the symphysis pubis; Left SAM: Left sacral ala measurement; Right SAM: Right sacral ala measurement; Left ISM: Left ilium shadow measurement; Right ISM: Right ilium shadow measurement.

Table 2
Radiographic parameters for determining the presence of pelvic displacement [28–31].

Parameter	Definition	Normal Range
FHHD (mm)	Femoral top horizontal lines are drawn perpendicular to the pelvic plumb line to intersect the tops of the femoral heads. FHHD is the distance between these two horizontal lines.	<10 mm
ALFHRF (°)	It is the angle between the straight line following the upper femoral of both sides and the horizontal line of the standing anteroposterior (AP) pelvic radiograph.	<1°
Left IM (mm)	At the apex of the iliac crest and the inferior margin of the ischial tuberosity, draw lines parallel to the femur base line (the line connecting the two femoral bases). The left IM is the distance between these two lines.	The absolute difference between Left IM and Right IM: <5 mm
Right IM (mm)	From the apex of the left iliac crest and the inferior aspect of the ischial tuberosity, draw lines parallel to the femoral base line (the line connecting the two femoral bases). The right IM is the distance between these two lines.	
ICHD (mm)	From the apex of the right iliac crest and the inferior aspect of the ischial tuberosity, draw a line horizontal to the femoral base line (the line connecting the two femoral bases). ICGD is the distance between these two lines.	<5 mm
DOCS (mm)	It is the distance between the symphysis pubis and the line drawn from the second sacral nodule at right angles to the femoral head line.	<3 mm
Left SAM (mm)	It is the distance between the lines drawn perpendicular to the femur base line at the outermost part of the left sacral ala in the second sacral nodule.	The absolute difference between Left SAM and Right SAM: <5 mm
Right SAM (mm)	It is the distance between the lines drawn perpendicular to the femur base line at the outermost part of the right sacral ala in the second sacral nodule.	
Left ISM (mm)	It is the distance between the lines drawn perpendicular to the femoral base line at the left outermost part of the iliac bone in the left Posterior Superior Iliac Spine.	The absolute difference between Left ISM and Right ISM: <5 mm
Right ISM (mm)	It is the distance between the lines drawn perpendicular to the femoral base line at the right outermost part of the iliac bone in the right Posterior Superior Iliac Spine.	

additional images as training data enables a data augmentation effect, enhancing the model's generalization performance by providing varied feature information regarding landmark location [32].

Contrast limited adaptive histogram equalization (CLAHE) [33] is employed to eliminate elements that impede learning by preventing excessive contrast enhancement of feature regions within the input images for each of the stages containing the generated images. CLAHE employs two parameters: the block size for processing block units and the clipping value for avoiding instances where the intensity of certain pixels within a block is exceedingly high or low after segmentation into blocks. Empirically determined as the optimal values, a block size of (8,8) and a clipping value of 2.0 are utilized for these parameters. Subsequently, the resolutions of all images are resized to 224 × 224, and learning and inference processes are executed.

In the first step of landmark detection, the coarse-grained network predicts the approximate or tentative locations of 16 landmarks in the input images. Subsequently, 16 patch images sized at 448 × 448 are generated, with their center points set as the location coordinates of each predicted landmark from the original pelvic X-ray images. CLAHE is applied to each of the generated patch images using the same block sizes and clipping values employed for the original pelvic X-ray images. Finally, the resolution of all patch images is resized to 224 × 224 again, serving as input to the fine-grained networks constituting stage two. Independent fine-grained networks are designated for each landmark, aiming to refine landmark locations using the patch images. Stage two operates to enhance landmark detection accuracy and efficiency concurrently. Specifically, the coarse-grained network's approximate estimations facilitate

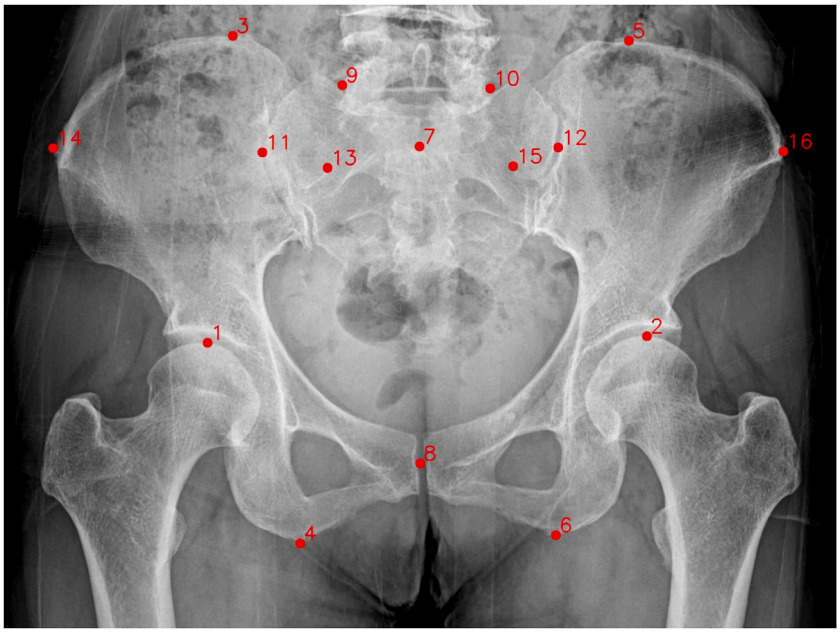


Fig. 2. Landmarks used to measure radiographic parameters for assessing pelvic displacement.

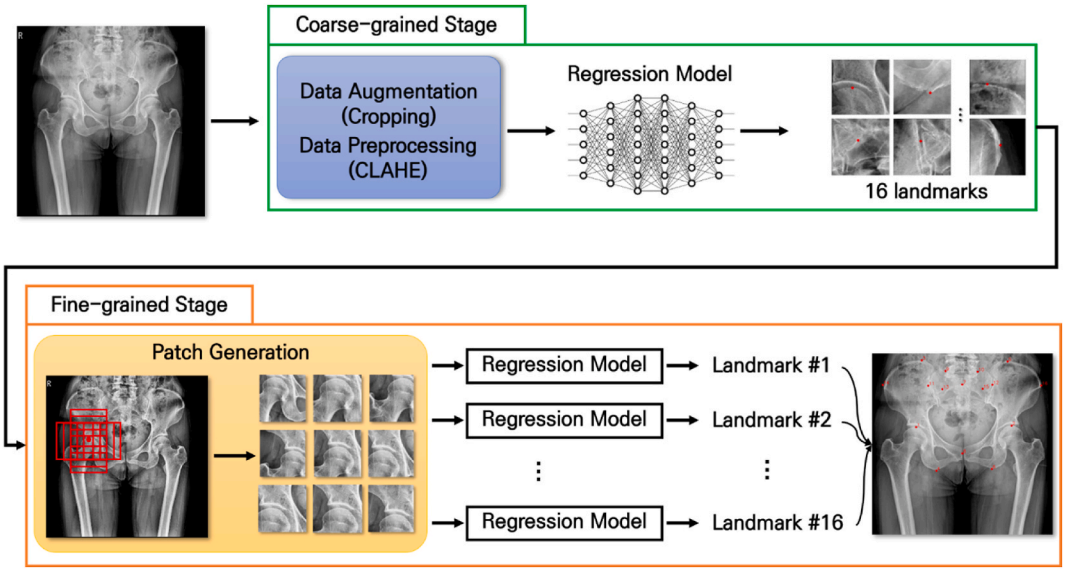


Fig. 3. Architecture of the dual-stage pelvic landmark detection network.

swift identification of general landmark locations, while the fine-grained networks correct previous results to ascertain final landmark locations, thus achieving the aforementioned objective.

3.4. Training strategy

The coarse-grained network employs a ResNet50 [34] network for regression analysis to detect 16 rough landmarks in the original images. Learning is conducted over 100 epochs with a learning rate set at 0.001. If performance does not improve within three epochs, the *ReduceLROnPlateau* learning rate scheduler reduces the next epoch's learning rate by 0.1 times the existing one. The Adam function serves as the gradient-based optimizer function, with a batch size of eight. Validation using a separate dataset occurs at the end of each epoch to prevent overfitting.

Similarly, the fine-grained network in stage one employs a ResNet50 network for each landmark independently to extract precise

landmarks. Learning spans 100 epochs per landmark, with a learning rate of 0.001. If performance stagnates for three consecutive epochs, the learning rate decreases by 0.1 times the existing rate. The Adam function is utilized as the optimization function, and the batch size remains eight.

The computational resources utilized for the landmark detection model include an Intel® Core™ i9-10980XE CPU @ 3.00 GHz, 24-GB NVIDIA RTX, and 64 GB of RAM. These resources are employed for model learning, performance evaluation, and inference time measurements.

3.5. Statistical analysis

Various evaluation metrics were utilized to assess the performance of the proposed DL model, and the results of each calculated parameter were analyzed. Firstly, the mean radial error (MRE) and successful detection rate (SDR) [35–37], commonly employed to evaluate landmark detection models, were selected to gauge the prediction accuracy of the landmark detection model. The MRE denotes the product of the Euclidean distance between the predicted and actual landmarks and the pixel spacing, representing the physical interval between pixels. The SDR denotes the ratio of successfully predicted landmarks within a specified error range from the actual landmark (see Eqs. (1) and (2)). x_i^{gt} and x_i^{pred} are predicted value and the corresponding ground truth, respectively.

$$\widehat{x}_i = \left|x_i^{gt} - x_i^{pred}\right| \cdot x_i^{ps}$$

$$\widehat{y}_i = \left|y_i^{gt} - y_i^{pred}\right| \cdot y_i^{ps}$$

$$dist_i = \sqrt{\widehat{x}_i^2 + \widehat{y}_i^2}$$

$$MRE = \frac{1}{N}\sum_{i=1}^{N} dist_i \tag{1}$$

$$SDR = \frac{dist_i}{N} \leq Z \ (Z \in \{1mm, 2mm, 2.5mm, 3mm\ 4mm\}) \tag{2}$$

To evaluate the reliability and accuracy of the quantitative measurements automatically determined by the proposed system, the Pearson correlation coefficient (PCC) and intra-class correlation coefficient (ICC) were calculated. Furthermore, to assess biases and similarities between the results measured by a clinician and those predicted by the system, Bland-Altman charts and paired T-tests were employed. These methodologies quantitatively and intuitively confirm similarities and biases between values measured by a clinician and those predicted by the system.

Several error calculation methods were employed to evaluate the system's accuracy. The mean absolute error (MAE) signifies the absolute difference between actual and predicted values, indicating overall divergence of model predictions from actual values. Additionally, the mean square error (MSE) and root mean square error (RMSE) were calculated; RMSE, the square root of MSE, returns to the original units, facilitating easier analysis. MSE emphasizes larger errors due to the squared prediction error, while RMSE aids in interpreting results in the original units. In summary, utilizing these diverse metrics deepens understanding of the system's performance and establishes a comprehensive comparison standard, facilitating verification of the proposed system's robustness and reliability compared to clinician observations.

Table 3
Mean radial error for 16 landmarks.

Landmark	Mean radial error (mm)
Top of the left femoral head (1)	1.07
Top of the right femur head (2)	1.20
Top of the left iliac crest (3)	4.50
Bottom of the left ischial tuberosity (4)	1.69
Top of the right iliac crest (5)	3.92
Bottom of the right ischial tuberosity (6)	1.88
Second sacral tubercle (7)	3.44
Center of the symphysis pubis (8)	1.31
Left sacral groove (9)	2.31
Right sacral groove (10)	1.99
Lateral aspect of the left sacrum (11)	4.55
Lateral aspect of the right sacrum (12)	4.65
Medial aspect of the left sacrum (13)	4.23
Lateral aspect of the ilium (14)	3.06
Medial aspect of the right sacrum (15)	4.14
Lateral aspect of the right ilium (16)	2.45
Average	2.89

Numbers in the parentheses indicate the landmarks' numbers in Fig. 2.

4. Results

4.1. Study participants

In this study, standing anteroposterior (AP) pelvic radiographs from 550 individuals were utilized (average age ± standard deviation (SD): 52.43 ± 15.78, age range: 12–89), comprising 232 men and 318 women. The female-to-male ratio in the collected radiographs was 1.37. All radiographs definitively included the region from the iliac crest to the femoral head, with exclusions made for images containing medical implants, skeletal abnormalities, or osteodysplasia.

4.2. Performance of the DL models for landmark detection

A test dataset was employed to assess the landmark detection model's performance. Mean radial error was measured for each landmark, and the successful detection rate (SDR) was assessed at distances of 1, 2, 2.5, 3, and 4 mm (refer to Tables 3 and 4). Notably, there was a discernible trend indicating larger errors for landmarks corresponding to the sacrum region compared to others. This discrepancy can be attributed to the sacrum's position and characteristics within the pelvis. Situated centrally, the sacrum is surrounded by numerous bones, organs, and tissues, leading to challenges in discerning detailed sacral information in radiographs. Other organs and tissues projected alongside the sacrum in radiographs further complicate accurate landmark prediction, underscoring the relatively inaccurate results observed for sacral locations.

4.3. Accuracy and reliability of pelvic tilt index measurement

Our proposed system can measure 10 parameters closely linked to pelvic displacement diagnoses. The automatically measured parameters were compared with those personally measured by a clinician, demonstrating consistency and a strong correlation between the two sets of results. Firstly, high correlations were evident between the clinician's measurements and the system's predictions, as indicated by the ICC and PCC results (refer to Table 5).

Next, paired T-tests were conducted to assess the consistency of measurement values obtained by the clinician and the system for the same objects. This comparison aimed to identify significant differences between the two sets of measurements. For parameters excluding those related to the sacrum (e.g., left SAM, right SAM), p-values were 0.05 or greater, confirming high similarity between the two measurement values (see Table 6).

Additionally, Bland-Altman plots were utilized to visually analyze the consistency between result values measured by the clinician and those generated by the system. These plots facilitated an intuitive understanding of differences between the two measurement methods for all objects. By calculating the average of the differences, systematic trends in values produced by the system compared to those measured by the clinician were determined. Consequently, Bland-Altman plots affirmed high consistency between the system's predictions and the clinician's measurements, while also confirming the absence of bias between the two measurement values (refer to Fig. 4).

Additionally, the model's performance was assessed by calculating the mean differences (MD), standard deviation (SD), MAE, (MSE, and RMSE. Analysis of MD and SD values revealed only slight differences between the model predictions and the clinician's measurements, indicating the proposed model's clinically effective predictive capability. MAE, MSE, and RMSE highlighted the magnitude of differences between the two measurements, with these metrics displaying relatively low values. These findings, summarized in Table 5, suggest that the proposed system consistently provides highly accurate predictions. Fig. 5 showcases the execution results when all landmarks are accurately detected, while Fig. 6 demonstrates instances of mis-detection, particularly in the landmarks

Table 4
Successful detection rates within different distances at the 16 landmarks.

Landmark	≤ 1 mm	≤ 2 mm	≤ 2.5 mm	≤ 3 mm	≤ 4 mm
Top of the left femoral head	0.58	0.88	0.95	0.98	1.00
Top of the right femur head	0.54	0.86	0.93	0.97	0.99
Top of the left iliac crest	0.16	0.34	0.44	0.54	0.65
Bottom of the left ischial tuberosity	0.35	0.77	0.82	0.90	0.94
Top of the right iliac crest	0.18	0.41	0.49	0.56	0.67
Bottom of the right ischial tuberosity	0.45	0.74	0.80	0.84	0.90
Second sacral tubercle	0.17	0.44	0.54	0.60	0.71
Center of the symphysis pubis	0.54	0.82	0.87	0.95	0.98
Left sacral groove	0.38	0.76	0.82	0.88	0.94
Right sacral groove	0.37	0.73	0.84	0.90	0.96
Lateral aspect of the left sacrum	0.09	0.21	0.30	0.41	0.56
Lateral aspect of the right sacrum	0.06	0.21	0.30	0.41	0.52
Medial aspect of the left sacrum	0.01	0.32	0.37	0.48	0.60
Lateral aspect of the ilium	0.30	0.57	0.66	0.73	0.82
Medial aspect of the right sacrum	0.13	0.32	0.395	0.53	0.64
Lateral aspect of the right ilium	0.26	0.57	0.70	0.80	0.87
Average	0.29	0.56	0.64	0.72	0.80

Table 5
Correlation coefficients (ICC, PCC), error metrics (MAE, MSE, RMSE), mean, and SD of values measured by a clinician and those predicted by the proposed system.

Parameter	ICC	PCC	MAE (mm)	MSE (mm)	RMSE (mm)	Mean (SD)		Proposed vs. radiologist
						Ours	Radiologist	Mean difference ± SD (95 % CI)
FHHD	0.984	0.984	0.47	0.36	0.60	3.96 (± 3.36)	4.04 (± 3.32)	0.07 ± 0.60 (0.00, 0.16)
ALFHRF	0.984	0.984	0.14[a]	0.03[a]	0.17[a]	1.15 (± 0.98)	1.17 (± 0.98)	0.02 ± 0.17 (0.00, 0.04)
Left IM	0.997	0.996	0.87	2.11	1.46	223.81 (± 13.90)	223.82 (± 13.95)	0.01 ± 1.45 (−0.19, 0.21)
Right IM	0.998	0.998	0.79	2.22	1.49	224.30 (± 14.46)	224.33 (± 14.60)	0.03 ± 1.49 (−0.17, 0.23)
ICHD	0.972	0.971	0.72	1.16	1.08	4.62 (± 3.95)	4.61 (± 3.89)	−0.02 ± 1.08 (−0.16, 0.13)
DOCS	0.920	0.921	1.03	1.80	1.35	4.25 (± 3.36)	4.41 (± 3.26)	1.33 ± 0.34 (−0.02, 0.34)
Left SAM	0.889	0.894	1.81	6.53	2.56	63.46 (± 4.92)	64.12 (± 5.07)	0.66 ± 2.47 (0.32, 1.00)
Right SAM	0.832	0.850	2.01	7.44	2.73	62.81 (± 4.42)	63.69 (± 4.87)	0.88 ± 2.58 (0.52, 1.24)
Left ISM	0.977	0.976	1.87	8.56	2.91	121.25 (± 11.59)	120.99 (± 11.87)	−0.25 ± 2.92 (−0.65, 0.15)
Right ISM	0.977	0.976	1.67	5.65	2.36	118.49 (± 10.51)	118.51 (± 10.70)	0.01 ± 2.38 (−0.32, 0.34)

[a] MAE, MSE, and RMSE in ALFHRF are in °, not mm.

Table 6
Results of paired T-tests on parameters for determining the presence of pelvic displacement.

Parameter	p-value
FHHD	0.08
ALFHRF	0.07
Left IM	0.91
Right IM	0.79
ICHD	0.81
DOCS	0.09
Left SAM	0.0001
Right SAM	0.000006
Left ISM	0.23
Right ISM	0.94

located in the sacrum region.

Furthermore, we evaluated the time required for each stage performed by the system and the time until the final output was computed. These measurements were conducted three times and categorized based on whether only CPU operations or both GPU and CPU operations were utilized. On average, the total detection time was approximately 4.93 s when solely CPU operations were employed, whereas including GPU operations reduced the time to approximately 4.04 s.

In conclusion, our proposed system demonstrates sufficient performance and stability across various metrics. These results underscore the efficacy and robustness of the system in diverse real-life scenarios.

5. Discussion

We proposed a system that utilizes DL technology to automatically measure radiographic parameters for assessing pelvic displacement in X-ray images. The developed system autonomously detects 16 landmarks on the pelvis and computes 10 radiographic parameters (i.e., FFHD, ALFHRF, Left IM, Right IM, ICHD, DOCS, Left SAM, Right SAM, Left ISM, Right ISM) based on these landmarks.

To validate the proposed DL-based model, 200 standing anteroposterior (AP) pelvic radiographs not used during training were employed to verify the reliability and reproducibility of the landmark detection model. Additionally, ICC and PCC were utilized to assess the correlation between each automatically calculated radiographic parameter and those personally measured by a clinician using the same test data. The results confirmed striking similarities between the two measurement values.

Through Bland-Altman plot analysis, we observed minimal variability and mean difference between the two measurements. While a few observations deviated slightly from the central line for certain parameters, these deviations were negligible. Moreover, paired T-test results corroborated the absence of statistically significant differences between the two sets of values. Finally, MAD, MAE, MSE, and RMSE were computed, reaffirming close agreement between the values predicted by the system and those determined by the clinician. These parameters are crucial as they allow for intuitive analysis of the error magnitude in predicted values.

However, it was observed that the parameters related to the sacrum (i.e., left SAM, right SAM) were comparatively low. This is attributed to inherent limitations in the radiographic images, where bones in the same vicinity as the sacrum are simultaneously projected, making it challenging to obtain accurate imaging information for the sacrum. This could potentially lead to a deterioration in the landmark detection performance in this region, ultimately compromising the robustness and consistency of the automatic measurement system. To address these limitations, the advancement of sophisticated DL models coupled with the development of extensive datasets that encompass a wide range of these challenging cases is crucial. In particular, incorporating anatomical information from the pelvic region, as demonstrated in numerous prior studies [38–40], is anticipated to enhance the consistency and

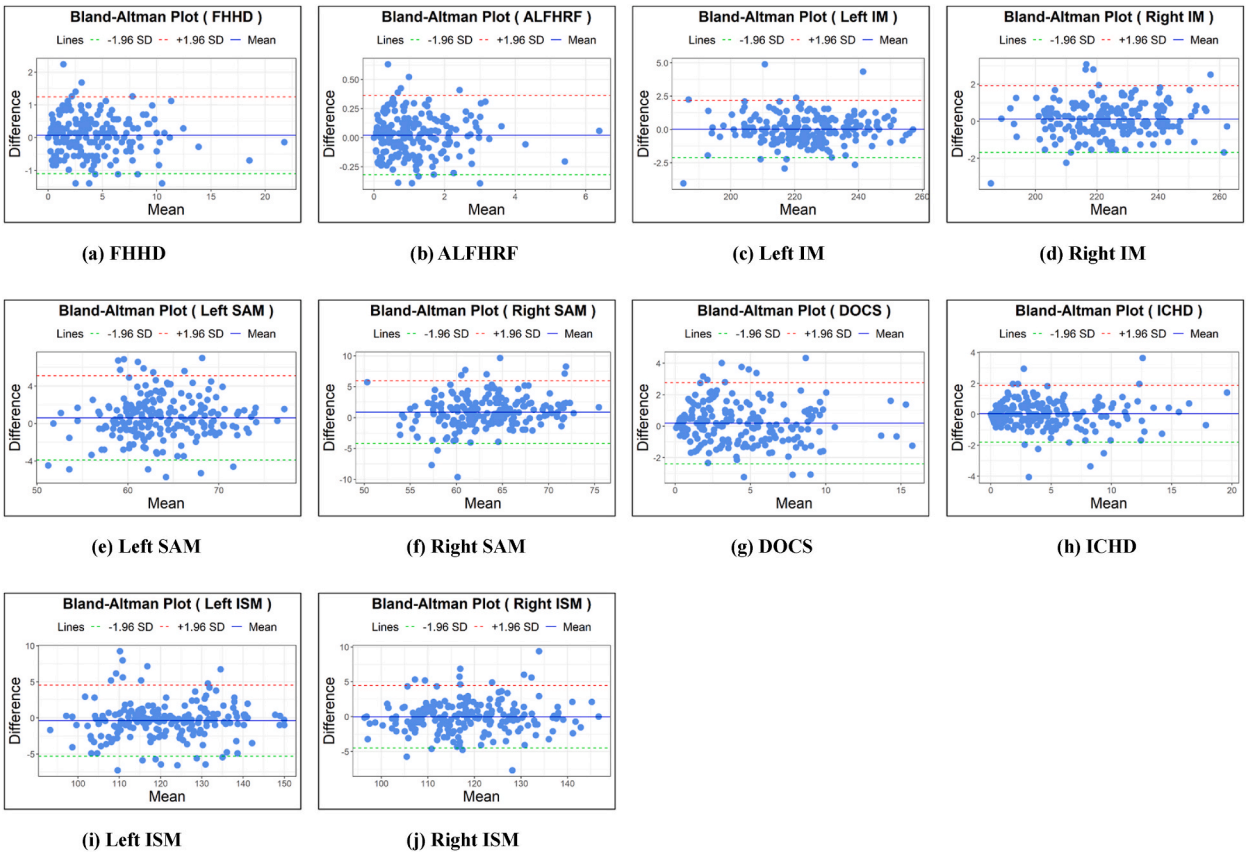


Fig. 4. Bland – Altman plots for 10 pelvic displacement assessment parameters (a) FHHD (b) ALFHRF (c) Left IM (d) Right IM. (e) Left SAM (f) Right SAM (g) DOCS (h) ICHD. (i) Left ISM (j) Right ISM.

robustness of landmark detection.

The time required for the system to perform the entire calculation was confirmed to be approximately 4 s. This is significantly faster than the time taken by a clinician to manually assess a radiograph (approximately 90 s per image). Thus, this system can accurately perform measurements at a faster pace and more consistently than a clinician, particularly when large numbers of images are provided as input.

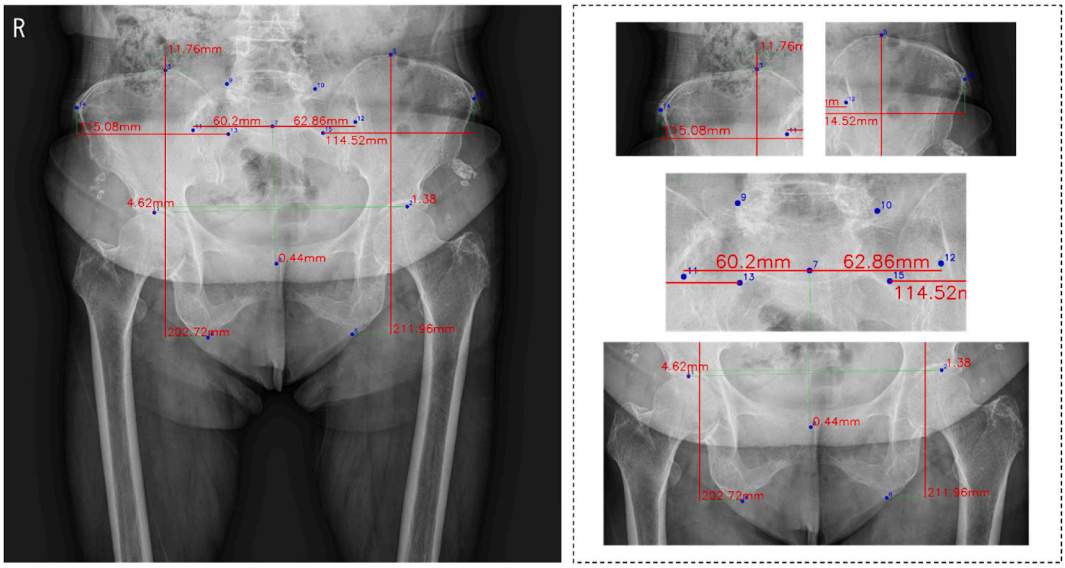
This study employed pelvic radiographs of patients exhibiting radiographic parameters within the abnormal range, excluding those with severe abnormalities such as skeletal dysplasia and fibrous dysplasia. While an overall accuracy analysis of automatic measurement algorithm was conducted, there is lack of more detailed analysis. For instance, performance verification related to the consistency and accuracy of the automatic measurement system based on the degree of abnormality in patients was not included. Moreover, a comprehensive analysis according to various demographic characteristics, such as age, sex, and race, is warranted. These limitations may restrict the proposed system's scope of application.

The data used in this study was measured by a single clinician. However, for additional verification, other methods could be considered, such as having the same clinician measure an identical verification dataset at different times and comparing the results with those predicted by the proposed system to comprehensively confirm the system's consistency. Furthermore, the system's generality could be definitively confirmed by having more than one clinician analyze the same dataset.

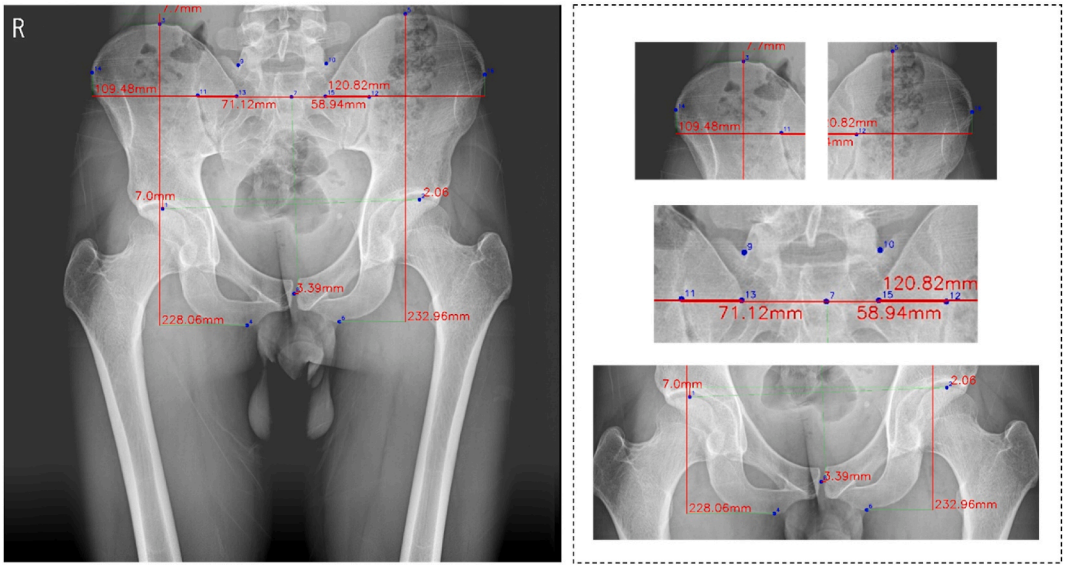
6. Conclusion

In conclusion, our proposed system for assessing pelvic displacement can accurately measure quantitative parameters, including angles and lengths. These parameters facilitate rapid and precise diagnosis through a staged execution of a deep learning model. However, the system did not incorporate radiographs containing implants for surgical treatment or severe skeletal abnormalities, nor did it analyze demographic variations or the degree of abnormalities. These factors are crucial in real medial settings and must be considered for the proposed system to operate more effectively. Therefore, for future research aiming to apply it in actual clinical settings, we plan to conduct performance analysis and verification from various perspectives, targeting a wide range of patients with different conditions, along with improving the system's performance by applying diverse methodologies.

In today's medical environment, the demand for computer-assisted diagnostic tools is increasing owing to lack of care providers



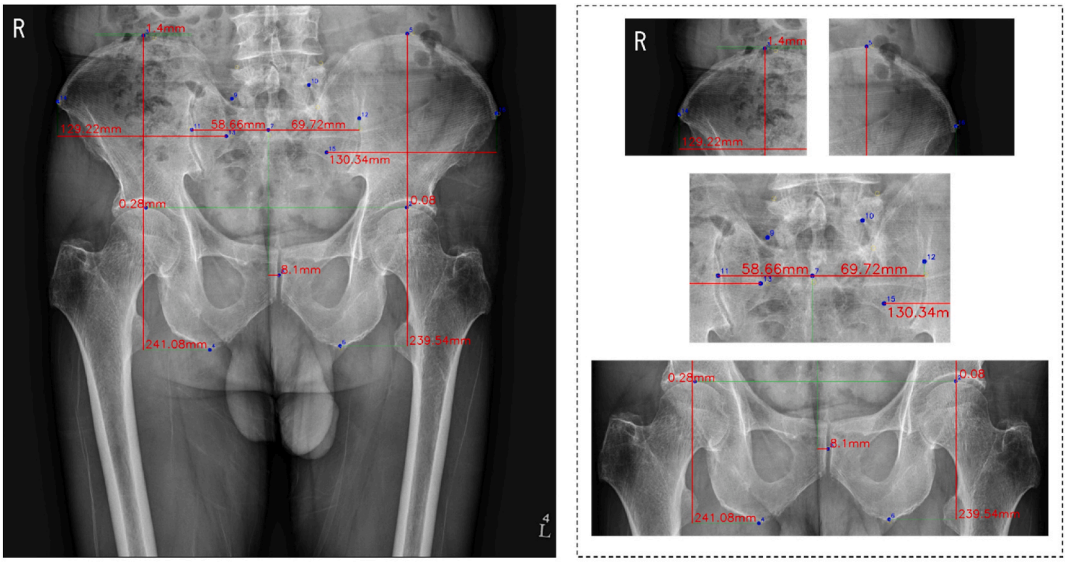
(a) Radiograph of a 70-year-old female patient



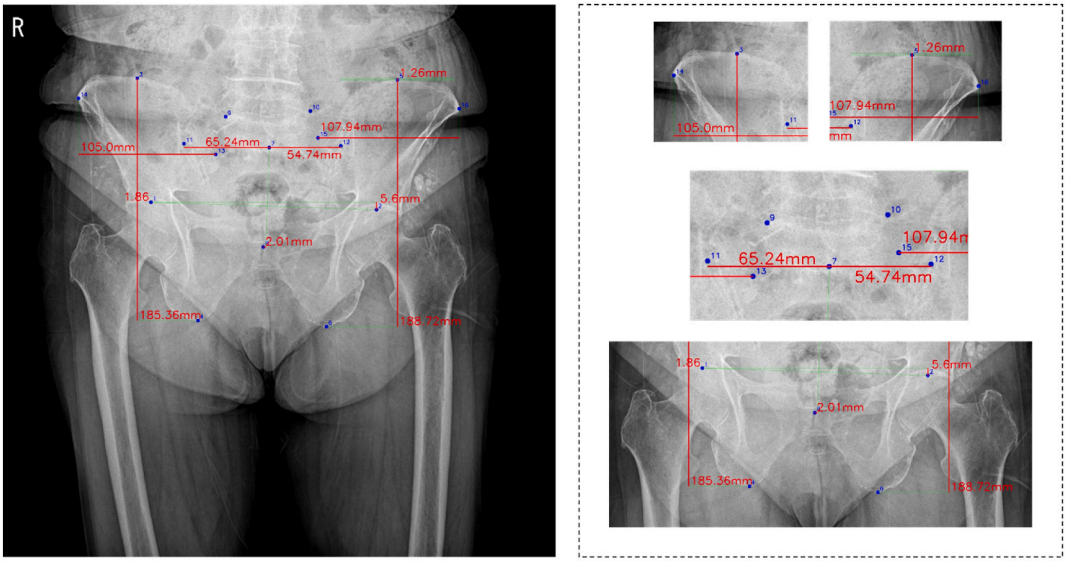
(b) Radiograph of a 25-year-old male patient

Fig. 5. Final results outputted by the system - precise measurement of landmarks necessary for assessing pelvic displacement. (a) Radiograph of a 70-year-old female patient. (b) Radiograph of a 25-year-old male patient.

and excessive workloads. However, no existing computer-aided software systems can measure such significant numbers of radiographic parameters on pelvic displacement. This study represents an important initial step in this field and has achieved promising results. Nevertheless, the inherent limitations of radiographic images projected onto a two-dimensional plane may impose constraints on elevating the performance of some landmark detections to the level of clinicians. This indicates the limited feasibility of an automatic measurement system without human intervention, thereby demonstrating how the automatic measurement system should be utilized in real clinical settings. We anticipate that in actual clinical practice, the system will serve as an adjunct tool, giving clinicians the flexibility to adjust and refine automatically measured results. This may alleviate the burden of the repetitive and time-consuming task of measuring various radiographic parameters required for surgical planning, diagnosis, and prognosis of relevant diseases, allowing clinicians to focus on effective treatment.



(a) An example of incorrectly detecting the landmark 9, 10 and 15



(b) An example of incorrectly detecting the landmark 9 and 15

Fig. 6. Final results outputted by the system - Example of failure to detect landmarks in the sacrum region among the necessary landmarks for pelvic displacement assessments. (a) An example of incorrectly detecting the landmark 9, 10 and 15. (b) An example of incorrectly detecting the landmark 9 and 15.

Ethic statement

This retrospective study was approved by the Institutional Review Board of Kyonggi University (IRB No. KGU-20230418-HR-103) and adhered to the ethical standards outlined in the Helsinki Declaration. Informed consent was waived by the Institutional Review Board of Kyonggi University as the data used were fully de-identified to protect patient confidentiality.

Funding statement

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (No. 2020R1A6A1A03040583).

Data availability statement

The datasets generated and/or analyzed during the current study are not publicly available due to specific institutional requirements governing privacy protection. However, they may be available from the corresponding author upon reasonable request.

CRediT authorship contribution statement

Ki-Ryum Moon: Writing – review & editing, Writing – original draft, Software, Formal analysis. **Shi Sub Byon:** Writing – review & editing, Writing – original draft, Formal analysis. **Sung Hyun Kim:** Writing – review & editing, Writing – original draft, Formal analysis. **Byoung-Dai Lee:** Writing – review & editing, Writing – original draft, Supervision, Software, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Byoung-Dai Lee reports article publishing charges and writing assistance were provided by Ministry of Education. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Vishy Mahadevan, Anatomy of the pelvis, Surgery 36 (7) (2018) 333–338.
[2] J.M. DeSilva, K.R. Rosenberg, Anatomy, development, and function of the human pelvis, Anat. Rec. 300 (4) (2017) 628–632, https://doi.org/10.1002/ar.23561.
[3] S.Y. Ma, A comparison of functional leg length inequality before and after manipulation of patients with low back pain, J. Korean Phys. Therapy Sci. 13 (1) (2006) 21–27.
[4] A. Delisle, M. Gagnon, C. Sicard, Effect of pelvic tilt on lumbar spine geometry, IEEE Trans. Rehabil. Eng. 5 (4) (1997) 360–366, https://doi.org/10.1109/86.650290.
[5] W. Xia, H. Fu, Z. Zhu, C. Liu, K. Wang, S. Xu, et al., Association between back muscle degeneration and spinal-pelvic parameters in patients with degenerative spinal kyphosis, BMC Muscoskel. Disord. 20 (2019) 454, https://doi.org/10.1186/s12891-019-2837-0.
[6] C.R. Reid, P.M. Bush, W. Karwowski, S.K. Durrani, Occupational postural activity and lower extremity discomfort: a review, Int. J. Ind. Ergon. 40 (3) (2010) 247–256, https://doi.org/10.1016/j.ergon.2010.01.003.
[7] P.K. Baker, Musculoskeletal origins of chronic pelvic pain: diagnosis and treatment, Obstet. Gynecol. Clin. N. Am. 20 (4) (1993) 719–742, https://doi.org/10.1016/S0889-8545(21)00389-2.
[8] Robert Cooperstein, Holzworth Madison, O'Brien Aidan, Intra- and Interexaminer reliability of compressive leg checking and correlation with the sit-stand test for anatomic leg length inequality, Chiropr. J. Aust. 45 (2) (2017) 184–195. Australian Chiropractors Association, 2017, https://search.informit.org/doi/10.3316/informit.903451209367000.
[9] T. Schwarz, A. Benditz, H.R. Springorum, J. Matussek, G. Heers, M. Weber, et al., Assessment of pelvic tilt in anteroposterior radiographs by means of tilt ratios, Arch. Orthop. Trauma Surg. 138 (8) (2018) 1045–1052, https://doi.org/10.1007/s00402-018-2931-z.
[10] M. Tannast, S.B. Murphy, F. Langlotz, S.E. Anderson, K.A. Siebenrock, Estimation of pelvic tilt on anteroposterior X-rays – a comparison of six parameters, Skeletal Radiol. 35 (3) (2006) 149–155, https://doi.org/10.1007/s00256-005-0050-8.
[11] J.-S. Lee, K.-W. Park, H.-T. Kim, S.-Y. Park, B.-C. Shin, Systematic review of the diagnosis of pelvic deviation for Chuna manual therapy, J. Korean Med. Rehabi. 32 (2) (2022) 83–94, https://doi.org/10.18325/jkmr.2022.32.2.83.
[12] P.P. Kaneriya, M.E. Schweitzer, C. Spettell, M.J. Cohen, D. Karasick, The cost-effectiveness of routine pelvic radiography in the evaluation of blunt trauma patients, Skeletal Radiol. 28 (1999) 271–273, https://doi.org/10.1007/s002560050514.
[13] D. Shen, G. Wu, H.-I. Suk, Deep learning in medical image analysis, Annu. Rev. Biomed. Eng. 19 (2017) 221–248, https://doi.org/10.1146/annurev-bioeng-071516-044442.
[14] Chuanbin Liu, et al., Misshapen pelvis landmark detection with local-global feature learning for diagnosing developmental dysplasia of the hip, IEEE Trans. Med. Imag. 39 (12) (2020) 3944–3954.
[15] Rebecca Smith, Kayvan Najarian, Automated segmentation of pelvic bone structure in x-ray radiographs using active shape models and directed hough transform, in: 2008 IEEE International Conference on Bioinformatics and Biomeidcine Workshops, IEEE, 2008.
[16] Suyog J. Pathare, Rutuja P. Solkar, Gajanan D. Nagare, Detection of fractures in long bones for trauma centre patients using hough transform, in: 2020 International Conference on Communication and Signal Processing (ICCSP), IEEE, 2020.
[17] Z. Xu, X. Baojie, W. Guoxin, Canny edge detection based on Open CV, in: 2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), Yangzhou, China, 2017, pp. 53–56, https://doi.org/10.1109/ICEMI.2017.8265710.
[18] V.F. Leavers, "Which Hough Transform?" CVGIP: Image Understanding 58.2, 1993, pp. 250–264.
[19] R. Smith, K. Najarian, K. Ward, A hierarchical method based on active shape models and directed Hough transform for segmentation of noisy biomedical images; application in segmentation of pelvic X-ray images, BMC Med. Inf. Decis. Making 9 (2009) 1–11, https://doi.org/10.1186/1472-6947-9-S1-S2.
[20] A. Jodeiri, R.A. Zoroofi, Y. Hiasa, M. Takao, N. Sugano, Y. Sato, et al., Region-based convolution neural network approach for accurate segmentation of pelvic radiograph, in: 2019 26th National and 4th International Iranian Conference on Biomedical Engineering (ICBME), Tehran, Iran, 2019, pp. 152–157, https://doi.org/10.1109/ICBME49163.2019.9030401.
[21] W. Yang, Q. Ye, S. Ming, X. Hu, Z. Jiang, Q. Shen, et al., Feasibility of automatic measurements of hip joints based on pelvic radiography and a deep learning algorithm, Eur. J. Radiol. 132 (2020) 109303, https://doi.org/10.1016/j.ejrad.2020.109303.
[22] J.A. Buckwalter, C. Saltzman, T. Brown, The impact of osteoarthritis: implications for research, Clin. Orthop. Relat. Res. 427 (Suppl) (2004) S6–S15, https://doi.org/10.1097/01.blo.0000143938.30681.9d, 1976-2007.
[23] C. Dezateux, K. Rosendahl, Developmental dysplasia of the hip, Lancet 369 (9572) (2007) 1541–1552, https://doi.org/10.1016/S0140-6736(07)60710-7.
[24] S.D. Trigg, D.S. Jeremy, C. Hulsopple, Femoroacetabular impingement syndrome, Curr. Sports Med. Rep. 19 (9) (2020) 360–366, https://doi.org/10.1249/JSR.0000000000000748.
[25] B. Bier, M. Unberath, J.-N. Zaech, J. Fotouhi, M. Armand, G. Osgood, et al., X-ray-transform invariant anatomical landmark detection for pelvic trauma surgery, in: A. Frangi, J. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger (Eds.), International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 11074, Springer International Publishing, Cham, 2018.
[26] C. Stotter, T. Klestil, C. Röder, P. Reuter, K. Chen, R. Emprechtinger, et al., Deep learning for fully automated radiographic measurements of the pelvis and hip, Diagnostics 13 (3) (2023) 497.

[27] Y. Pei, L. Mu, C. Xu, Q. Li, G. Sen, B. Sun, et al., Learning-based landmark detection in pelvis X-rays with attention mechanism: data from the osteoarthritis initiative, Biomed. Phys. Eng. Express 9 (2) (2023) 025001, https://doi.org/10.1088/2057-1976/ac8ffa.
[28] Y.-Q. Du, B. Zhang, J.-Y. Sun, H.-Y. Ma, J.-M. Shen, M. Ni, et al., The variation of the pelvis in unilateral Crowe type IV developmental dysplasia of the hip, Orthop. Surg. 13 (2) (2021) 546–552, https://doi.org/10.1111/os.12903.
[29] C. Yoon, C.H. Shin, D.O. Kim, M.S. Park, W.J. Yoo, C.Y. Chung, et al., Overgrowth of the lower limb after treatment of developmental dysplasia of the hip: incidence and risk factors in 101 children with a mean follow-up of 15 years, Acta Orthop. 91 (2) (2020) 197–202, https://doi.org/10.1080/17453674.2019.1688485.
[30] J. Park, et al., Analysis of the research trends of pelvic malposition in Korean clinical research, J. Korean Med. Rehab. (2015), https://doi.org/10.18325/jkmr.2015.25.2.081. The Society of Korean Medicine Rehabilitation.
[31] R.W. Herbst, Gonstead chiropractic science and art: the chiropractic methodology of clarence S. Gonstead, D.C, Today's Chiropractic 3 (4) (1974) 22–24. August-September.
[32] A. Krizhevsky, I. Sutskever, G.E. sHinton, Imagenet classification with deep convolutional neural networks, in: F. Pereira, C.J. Burges, L. Bottou, K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, vol. 25, Curran Associates, Inc., 2012.
[33] A.M. Reza, Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement, J. VLSI Sig. Proc. Syst. 38 (2004) 35–44.
[34] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Las Vegas, USA, 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.
[35] C. Runnan, Y. Ma, N. Chen, D. Lee, W. Wang, Cephalometric landmark detection by attentive feature pyramid fusion and regression-voting, in: D. Shen, et al. (Eds.), Medical Image Computing and Computer Assisted Intervention –MICCAI, Lecture Notes in Computer Science, vol. 11766, Springer, Cham, 2019, https://doi.org/10.1007/978-3-030-32248-9_97.
[36] M.A. Khalid, K. Zulfiqar, U. Bashir, A. Shaheen, R. Iqbal, Z. Rizwan, et al., Aariz: a benchmark dataset for automatic cephalometric landmark detection and CVM stage classification, arXiv preprint, arXiv:2302.07797 (2023), https://doi.org/10.6084/m9.figshare.13265471.v1.
[37] K. Takahashi, Y. Shimamura, C. Tachiki, Y. Nishii, M. Hagiwara, Cephalometric landmark detection without X-rays combining coordinate regression and heatmap regression, Res. Square (2023), https://doi.org/10.21203/rs.3.rs-2116273/v1.
[38] L. Lu, Wolterink Jelmer, B. Christoph, V. Raymond, Anatomy-aided deep learning for medical image segmentation: a review, Phys. Med. Biol. 66 (11) (2021).
[39] K. Uday, Z. Mohammad, N. Nusrat, H. Taufiq, Anatomy-XNet: an anatomy aware convolutional neural network for thoracic disease classification in chest X-rays, IEEE J. Biomed. Health Inform. 26 (11) (2022) 5518–5528.
[40] T. Tim, M. Philip, K. Georgios, R. Daniel, Interactive and explainable region-guided radiology report generation, in: 2023 Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Vancouver, Canada, 2016, pp. 7433–7442, https://doi.org/10.1109/CVPR52729.2023.00718.