

RESEARCH

Open Access



Automatic assessment of lower limb deformities using high-resolution X-ray images

Reyhaneh Rostamian¹, Masoud Shariat Panahi^{1*} , Morad Karimpour¹, Alireza Almasi Nokiani^{2,3}, Ramin Jafarzadeh Khaledi³ and Hadi Ghattan Kashani¹

Abstract

Background Planning an osteotomy or arthroplasty surgery on a lower limb requires prior classification/identification of its deformities. The detection of skeletal landmarks and the calculation of angles required to identify the deformities are traditionally done manually, with measurement accuracy relying considerably on the experience of the individual doing the measurements. We propose a novel, image pyramid-based approach to skeletal landmark detection.

Methods The proposed approach uses a Convolutional Neural Network (CNN) that receives the raw X-ray image as input and produces the coordinates of the landmarks. The landmark estimations are modified iteratively via the error feedback method to come closer to the target. Our clinically produced full-leg X-Rays dataset is made publically available and used to train and test the network. Angular quantities are calculated based on detected landmarks. Angles are then classified as lower than normal, normal or higher than normal according to predefined ranges for a normal condition.

Results The performance of our approach is evaluated at several levels: landmark coordinates accuracy, angles' measurement accuracy, and classification accuracy. The average absolute error (difference between automatically and manually determined coordinates) for landmarks was 0.79 ± 0.57 mm on test data, and the average absolute error (difference between automatically and manually calculated angles) for angles was $0.45 \pm 0.42^\circ$.

Conclusions Results from multiple case studies involving high-resolution images show that the proposed approach outperforms previous deep learning-based approaches in terms of accuracy and computational cost. It also enables the automatic detection of the lower limb misalignments in full-leg x-ray images.

Keywords Skeletal landmark detection, Lower limb deformities, Osteotomy, Arthroplasty surgery, Convolutional neural networks

Background

A lower limb deformity is commonly defined as the deviation of the limb's physiological axis from its normal condition. Those deformities may be congenital or constitutional in origin. Children may experience them due to a growth condition that causes the epiphyseal plate to close too soon. They could also be linked to traumatic events, metabolic disorders like rickets or such osteopathies as renal osteopathy. Neurologic diseases or systemic myopathies may also be connected [1].

*Correspondence:

Masoud Shariat Panahi
mshariatp@ut.ac.ir

¹ School of Mechanical Engineering, College of 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



The alignment of the lower limbs is assessed using a full-length standing anteroposterior (AP) projection radiograph [1]. Quantities of importance, such as the anatomical and mechanical axes of the femur and tibia, the femoral condyle tangent, and the tibia plateau tangent, are used to determine parameters related to the lower limb alignment. Following the stated quantities of interest, the typical parameters describing the lower limb alignment are calculated from lower limb frontal X-ray images (Fig. 1) [2–5]:

- Mechanical femoral mechanical tibial angle (MFMTA) is the angle between the mechanical axes of the femur and tibia with a normal range of $(-2^{\circ}, 4^{\circ})$.
- Mechanical lateral distal femoral angle (MLDFA) is the angle between the femoral condyle tangent and the mechanical axis of the femur with a normal range of $(85^{\circ}, 90^{\circ})$.

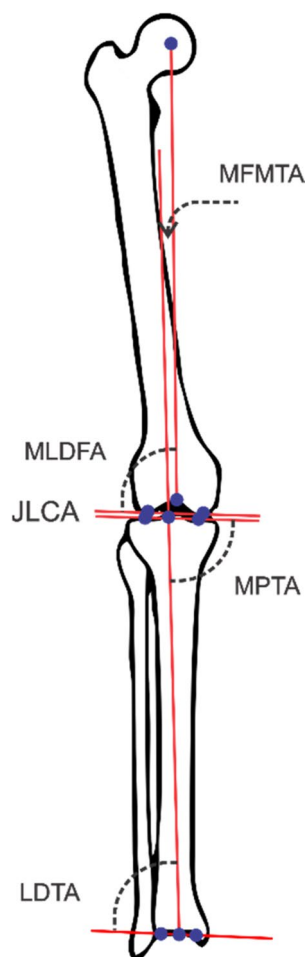


Fig. 1 Typical measurements describing the lower limb, obtained from a frontal X-ray image: MFMTA, MLDFA, MPTA, LDTA, and JLCA

- Mechanical proximal tibial angle (MPTA) is the angle between the tangent to the tibial plateau and the anatomical axis of the tibia with a normal range of $(85^{\circ}, 90^{\circ})$.
- Lateral distal tibial angle (LDTA) is the angle between the distal tibial articular surface and the anatomical axis of the tibia, which measures lateral distal tibial angle with a normal range of $(86^{\circ}, 92^{\circ})$.
- Joint line convergence angle (JLCA) is the angle formed between a line tangential to the distal femoral condyle and the tibial plateau with a normal range of $(0^{\circ}, 2^{\circ})$ [1, 6]

Orthopedic surgeons typically review radiography in AP projection with the individual standing up straight to support body weight to detect lower limb abnormalities in patients. However, modern technological advancements allow the employment of semi-automated and automatic procedures where accuracy and productivity are independent of the human experience [1, 7–11]. Recent works with a deep learning approach utilize two independent networks for a region of interest extraction and landmark detection, respectively [12]. Nguyen et al. suggested a decentralized deep learning system for computing the angles of angle between the femoral mechanical and anatomical axes (FAMA), hip-knee-ankle (HKA), MLDFA, and MPTA angle computation from full-leg radiographs. First, regions of interest around each joint are detected using a convolutional neural network (CNN). Then, by employing a second CNN, landmark coordinates are computed [8]. Pei et al. presented a U-net-based method to evaluate the HKA angle on unilateral lower limb X-rays. The different joints like hip, knee, and ankle were segmented using a U-net. The segmentations are utilized to identify the joints' centers. These points were used to estimate the HKA angle [13]. A method for determining HKA angles from full-leg radiographs was put out by Tack et al. The YOLOv4 was used in their method to identify regions of interest in joints. After that, each detected region of interest was regressed using ResNet to find landmark coordinates. Then, HKA angle was calculated using the identified landmarks [7].

Based on expert human landmarking patterns, we provide an image pyramid-based approach for landmark regression and angle measurement to assess lower limb deformities. This approach is inspired by the reinforcement learning strategy used for supervised learning tasks [14]. On our publicly accessible data from a local imaging center, we compare our estimates for landmark coordinates with those of human specialists to assess the accuracy of our approach. The primary contributions of the suggested approach are outlined below: I. Removing the stage involving the detection of regions

of interest by introducing an image-pyramid approach to produce many low-resolution feature maps around the goal points, II. Size-dependent logarithmic scaling of the input data that makes the approach applicable to high-resolution images. III. Employing a ResNet architecture to precisely regress landmark locations in each scale. IV. Using an error-feedback approach to modify patches in each scale and each estimated landmark. V. Diversifying the training data set to include cases with/without primary treatments. VI. Evaluating the accuracy of landmark detection on X-ray images of the lower limb VII. Simultaneous calculation of MFMTA, ML DFA, MPTA, and JLCA angles VIII. Automatic classification of those angles as lower than normal, normal or higher than normal according to predefined ranges for a normal condition. IX. presenting a publically available landmark-labeled full-leg X-ray dataset to support future works in the field.

Materials and methods

Full-leg X-rays dataset

We use our publicly available lower limb dataset to evaluate the performance of the proposed approach.

The dataset is including high resolution DICOM format for bilateral weight-bearing X- rays. All patients underwent imaging based on clinical indications of lower-limb malalignment. None of the data used in this study has a torsional deformation component and they underwent capturing for lower limb deformity in coronal plane, therefore in this model all the data was used to detect 2D landmarks. The data set contains various lower limb deformities; specifically, which related to varus valgus deformity with different origin along femur, tibia, and joint line. The data covers three categories of lower than normal, normal, and higher than normal for all 2D angular indicators (MFMTA, ML DFA, MPTA, LD TA, and JLCA). The dataset (named Rad lower limb Xray dataset), provided by a local imaging center, was generated using a radiography imaging device. All the cases have pixel size information in DICOM format metadata. All cases present a complete field of view (including pelvis, both hip joints, both knee joints, and both ankles). For this dataset, 426 subjects underwent the capturing. The participants included people with an osteotomy (54 subjects), total hip arthroplasty (2 subjects), total knee arthroplasty (73 subjects), miscellanies (6 subjects), and people with no primary treatment (291 subjects). The resolution of the images varies from 2548 to 3342 × 6995 to 10023 pixels according to the different settings in the imaging step. The imaging accuracy for each pixel is about 0.1 mm. In the annotation step, each x-ray is first labeled according to its primary treatment and then labeled with 26 landmarks (13 landmarks for each leg). The labeled landmarks

in the Rad dataset are shown in Fig. 2. Our full-leg Xray dataset as well as the landmarks' annotations will be made publicly available with this publication to support upcoming works in the field (https://aailab.ir/datasets/xray_rad_lower_limb.html). The work was approved by the Research Ethics Committees of Tehran University—Faculty of Sport Sciences and Health, University of Tehra, Tehran, Iran (Approval ID: IR.UT.SPORT.REC.1402.131).

Method

We first build a Gaussian pyramid with seven levels to produce smooth multi resolution pyramid (Fig. 3). Each pyramid level is a 64 × 64 matrix around the initial estimation for landmark coordinates referring to a specific image downsample [15, 16]. To retain images' scale and have a uniform dataset, we add black pixels around each x-ray to have the dataset with an image size of 10112 × 3584. The number of pyramid levels is set so that the size of the last down-sampled image (with two as a ratio size) would be approximately equal to the size of our desired patch (64 × 64). Therefore, levels of pyramids are referred to resolutions of 4096 × 4096 to 64 × 64 (in seven steps). For the training phase, the initial location of the landmark is initialized randomly from a normal landmark distribution for the training data. For the evaluation phase, however, it is initialized to the mean of the training labels for the landmark.

After building a 7-level pyramid, each level of the pyramid is processed by the pre-trained CNN, which is well suited for accurate regression of landmark coordinates in medical images [17]. For this purpose, 34-layer ResNet with some modifications is employed [18]. The structure of the modified ResNet is shown in Fig. 4. Each patch with the size of 64 × 64 is processed with CNN independently using the weights from the green channel. This allows for effective processing of the grayscale data while benefiting from the pre-trained weights, leading to faster convergence during training. The green channel was selected because it typically contains more informative features compared to the red or blue channels in many natural images. The stride of the first convolution was reduced from 2 to 1. A stride of 2 would halve the dimensions of the feature map, resulting in a considerable loss of spatial information. For landmark localization, maintaining higher spatial resolution is crucial to predict precise locations. Finally, the last 7 layers (including the fully-connected layer) of ResNet are removed. By truncating the network, its depth was effectively decreased, minimizing the downsampling applied to the feature maps throughout the architecture. This is especially important in landmark localization tasks, where fine spatial details are crucial. In this way, the output size of each

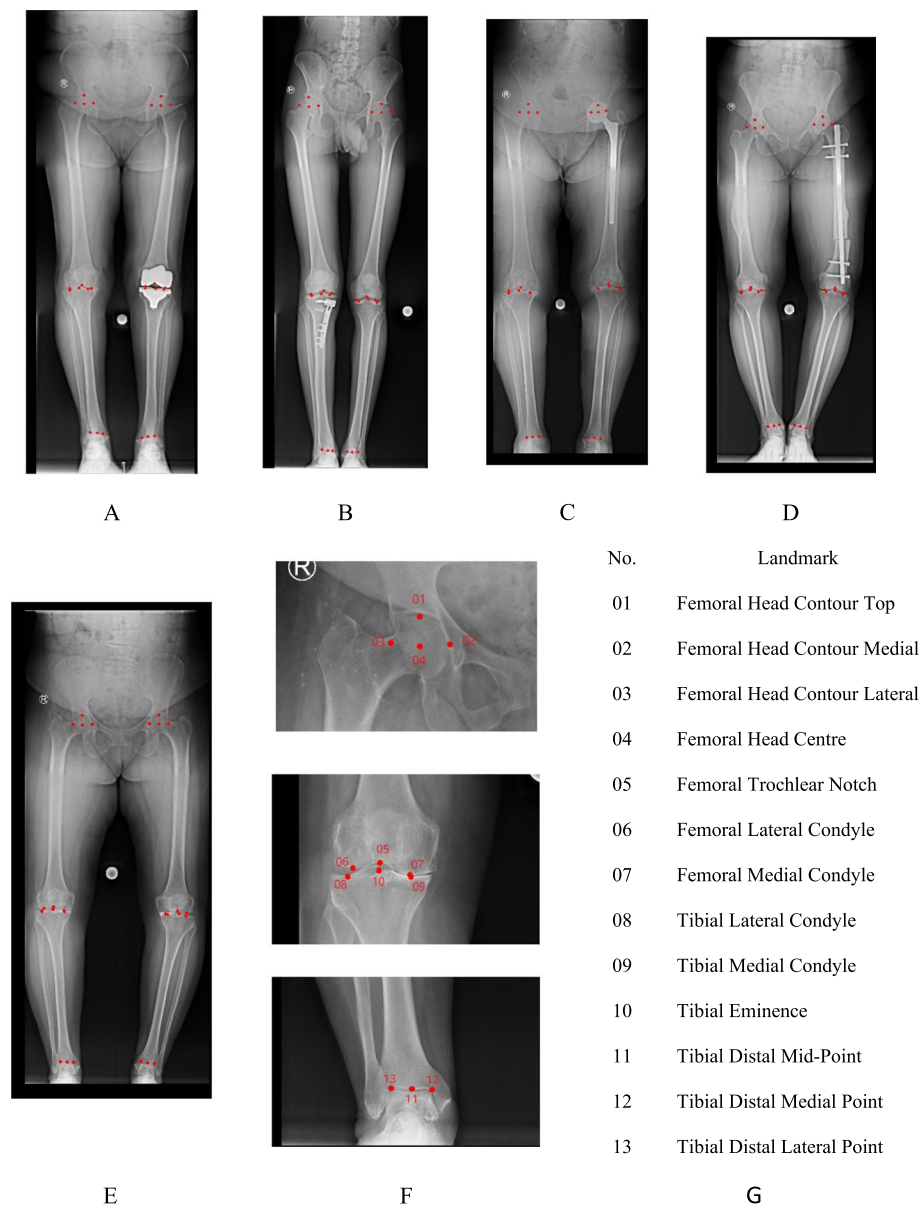


Fig. 2 Four types of cases in the dataset are presented: patients with **A** total knee arthroplasty, **B** osteotomy, **C** total hip arthroplasty, **D** miscellanies, and **E** no primary treatments. Parts F and G placement of landmarks for the right leg and landmark characteristics are shown, respectively

modified ResNet would be $256 \times 8 \times 8$. The output can be thought of as 256 low-resolution 8×8 heatmaps.

Heatmap regression has been utilized in several recent works for human pose estimation and 2D landmark detection [16, 20–22]. Given a learned heat map represents the probability of the location being the goal point, many works considered the location of the point with the maximum likelihood as the final point. Since this method is not end-to-end differentiable and suffers from quantization error, the integral regression approach was recommended [20]. To do this, each heatmap is processed

by Softmax to obtain probability distributions, and then each heatmap is reduced to a single point using Integral Regression. The third feature for each heatmap is produced using Softmax-weighted average pooling with an 8×8 kernel size to consider the relative importance of each point feature to others [16]. Now output with the size of 256×3 , which represents heatmap features, flatten for 7 levels of the pyramid and concatenate them to create a 1D vector as an input for the following network.

A 3-layer fully connected neural network is employed to regress the estimated offset for landmark location. 1D

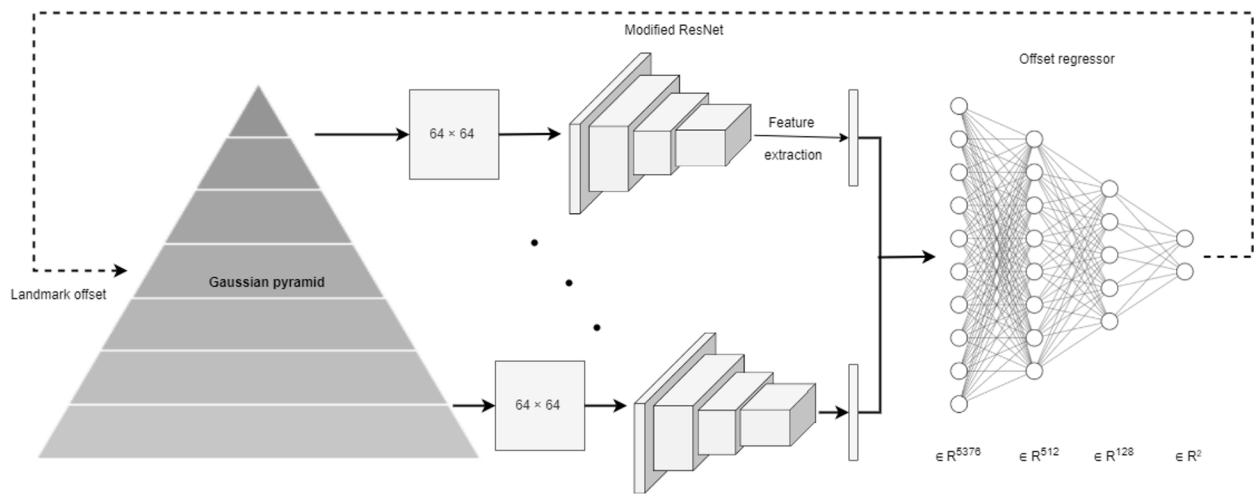


Fig. 3 Visual presentation of the proposed approach. The Gaussian pyramid, including different image resolutions is used to create 64×64 matrices around the estimated landmark. Then each patch is fed into a modified CNN to regress landmark location. The neural network is employed on features extracted from CNN's output for all pyramid levels to detect offset. This offset is used to update estimation for a landmark location

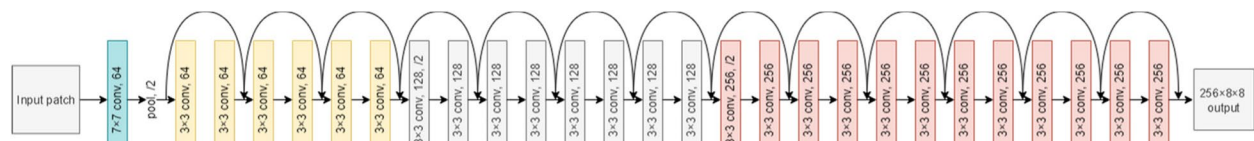


Fig. 4 Diagram of the modified ResNet architecture. In first layer the stride is changed to 1. The last 7 layers including dense layer are removed. [19]

features vector from heatmaps with the length of 5376 ($256 \times 3 \times 7$) is the network's input, and the output is the 1D vector with the length of 2 to use as model feedback error. 2 hidden layers of the network have 512 and 128 neurons with Relu activation function. This estimation for landmark offset is employed to modify landmark location. Based on this new estimation for a landmark, the pyramid's layers are recreated. This loop is repeated for a pre-defined number of iterations (e.g., 10); and in each iteration, the estimated landmark gets closer to the target (Pseudo code 1).

Training

For 30 epochs, we trained the network using the ADAM optimizer with learning rates of $1e-4$ for the first 20 epochs and $1e-5$ for the other 10. A batch size of 2 photos was employed. 85% of samples are utilized for training and 15% are used for validation. The procedure has only been trained for the right leg. The data is flipped from left to right for the left leg to double the number of training data and further reduce the variation in the data. As input for the network, all X-Ray images are resized to $5056 \times 10,112$ pixels by adding

black pixels around each image. Data augmentation improves prediction performance and handles wider variability within the provided image data. A patch was scaled by $\pm 5\%$ and randomly rotated by $\pm 15^\circ$ during training.

Pseudo code 1: Training step for a single landmark on single x-ray image

Require: X-ray image (I), mean (μ) and standard deviation (σ) for landmark position on training data

Initial landmark estimation: $x_0 \sim \mathcal{N}(\mu, \sigma)$

Initialize Gaussian pyramid with n level

For $i = 1, 10$ **do**

For $j = 1, n$ **do**

Create 64×64 pixel patch around x_0 for pyramid level j

Employ CNN on patch to get $C = 256 \times H = 8 \times W = 8$ output

Feature extraction from $H \times W$ output to create $C \times 3$ vector
Flat $C \times 3$ vector and concatenate into P

End for

MLP with P vector as an input and location offset ($\hat{x}$) as an output

Update landmark estimation $x_0 \leftarrow x_0 + \hat{x}$

Backpropagate $|x_0 - \hat{x}|$

End for

Results

Thirteen independent networks were developed, one per landmark. The performance of our suggested approach was evaluated at several levels: landmark coordinates accuracy in millimeters, angles accuracy in degrees, and class assignment accuracy. The results were reported for both primary cases (cases with no initial treatment), and all cases (including primary and revision cases).

Landmark coordinates accuracy

The estimated values achieved using our proposed approach were produced separately, and their distances from the 2D coordinates of landmarks achieved by human experts were calculated. The details of the results are presented in Table 1. The detection efficiency within the 3 mm error range was reported for each landmark. The median and maximum errors were reported to better comprehend the error magnitude for the unsuccessful cases. The accuracy was measured for successful detections using the mean, median, and standard deviation for both primary and all cases (including revision cases). The femoral trochlear notch point displayed the highest level of accuracy, with a mean L1 error of 0.33 ± 0.24 mm, and the tibial lateral condyle showed the lowest level of accuracy (1.18 ± 0.63 mm) for primary cases. The discovered landmarks had an overall mean inaccuracy of 0.79 ± 0.57 mm and 0.82 ± 0.60 mm for primary and all cases, respectively.

To determine the center of the femoral head, two scenarios are probable according to expert preference. The landmark can be determined directly within one point in the center of the head or calculated from the fitted circle using 3 points on the femoral head counter (femoral head contour top, femoral head contour medial, and femoral head contour lateral). The estimated values for the center of the femoral head are achieved using direct and indirect methods. The details of the results are presented in Table 2. The distance error between calculated landmarks using 3 points on the femoral counter for both target and detection, displayed the highest level of accuracy, with a mean error of 0.51 ± 0.36 mm and 0.61 ± 0.50 mm for primary and all cases, respectively.

To better understand error distribution for each landmark and compare model performance between each landmark, the boxplot on all test data and histogram on test data excluding failed cases are reported, respectively (Fig. 5A and B).

Angle accuracy

For a better model assessment, final angles' values (MFMTA, ML DFA, MPTA, LD TA, and JLCA) were acquired using our approach estimation for landmarks and compared with angles values using landmarks evaluated by human experts. Table 3 displays the results in detail. The detection performance within the error range of 2° was 95.6% for all angles. For the failed cases, the median and maximum errors were presented. The

Table 1 Results from various anatomical landmarks on the femur and tibia. The failed cases' columns display the proportion of unsuccessful cases, the median, and the maximum absolute error for this group. The success rates define the accuracy (excluding failed ones) for primary and all cases. The final row displays the results averaged over all landmarks. The absolute error (L1) is expressed in millimeters

Bone type	Landmark	Failed cases (error more than 3 mm)			Accuracy (excl. failed cases) for primary cases		Accuracy (excl. failed cases) for all cases	
		% Failure	Median	Max	Mean $\pm$ SD	Median	Mean $\pm$ SD	Median
Femur	Femoral Head Contour Top	0%	-	-	0.62 ± 0.34	0.57	0.65 ± 0.41	0.56
	Femoral Head Contour Medial	4%	4.59	5.65	1.04 ± 0.65	0.90	1.03 ± 0.64	0.84
	Femoral Head Contour Lateral	8%	3.64	4.58	1.04 ± 0.57	0.96	1.18 ± 0.64	1.10
	Femoral Head Center	0%	-	-	0.89 ± 0.41	0.93	0.96 ± 0.45	1.00
	Femoral Trochlear Notch	0%	-	-	0.33 ± 0.24	0.29	0.33 ± 0.25	0.30
	Femoral Lateral Condyle	10%	4.38	4.41	0.77 ± 0.63	0.52	0.91 ± 0.72	0.72
	Femoral Medial Condyle	4%	3.27	3.39	0.96 ± 0.82	0.59	0.98 ± 0.81	0.63
Tibia	Tibial Lateral Condyle	4%	3.10	3.17	1.18 ± 0.63	1.03	1.15 ± 0.73	1.03
	Tibial Medial Condyle	2%	3.24	3.24	1.03 ± 0.65	0.95	1.04 ± 0.76	0.95
	Tibial Eminence	0%	-	-	0.70 ± 0.43	0.60	0.66 ± 0.40	0.52
	Tibial Distal Mid-Point	0%	-	-	0.58 ± 0.50	0.42	0.59 ± 0.47	0.82
	Tibial Distal Medial Point	0%	-	-	0.49 ± 0.27	0.45	0.48 ± 0.27	0.46
	Tibial Distal Lateral Point	2%	3.23	3.23	0.68 ± 0.45	0.67	0.70 ± 0.43	0.64
Average	all	2.6%	3.39	4.58	0.79 ± 0.57	0.68	0.82 ± 0.60	0.67

Table 2 Results from various methods to detect the center of the femoral head. The failed cases' columns display the proportion of unsuccessful cases, the median, and the maximum absolute error for this group. The success rates define the accuracy (excluding failed ones) for primary and all cases. The absolute error (L1) is expressed in millimeters

Error type	Failed Cases (error more than 3 mm)			Accuracy (excl. failed cases) for primary cases		Accuracy (excl. failed cases) for all cases	
	% Failure	Median	Max	Mean $\pm$ SD	Median	Mean $\pm$ SD	Median
Direct femoral head center vs its estimation	0%	-	-	0.89 $\pm$ 0.41	0.93	0.96 $\pm$ 0.45	1.00
Direct femoral head center vs calculated center using three estimated points	0%	-	-	1.12 $\pm$ 0.63	1.03	1.19 $\pm$ 0.62	1.11
Calculated center using 3 points vs calculated center using estimated 3 points	2%	4.49	4.49	0.51 $\pm$ 0.36	0.41	0.61 $\pm$ 0.50	0.46

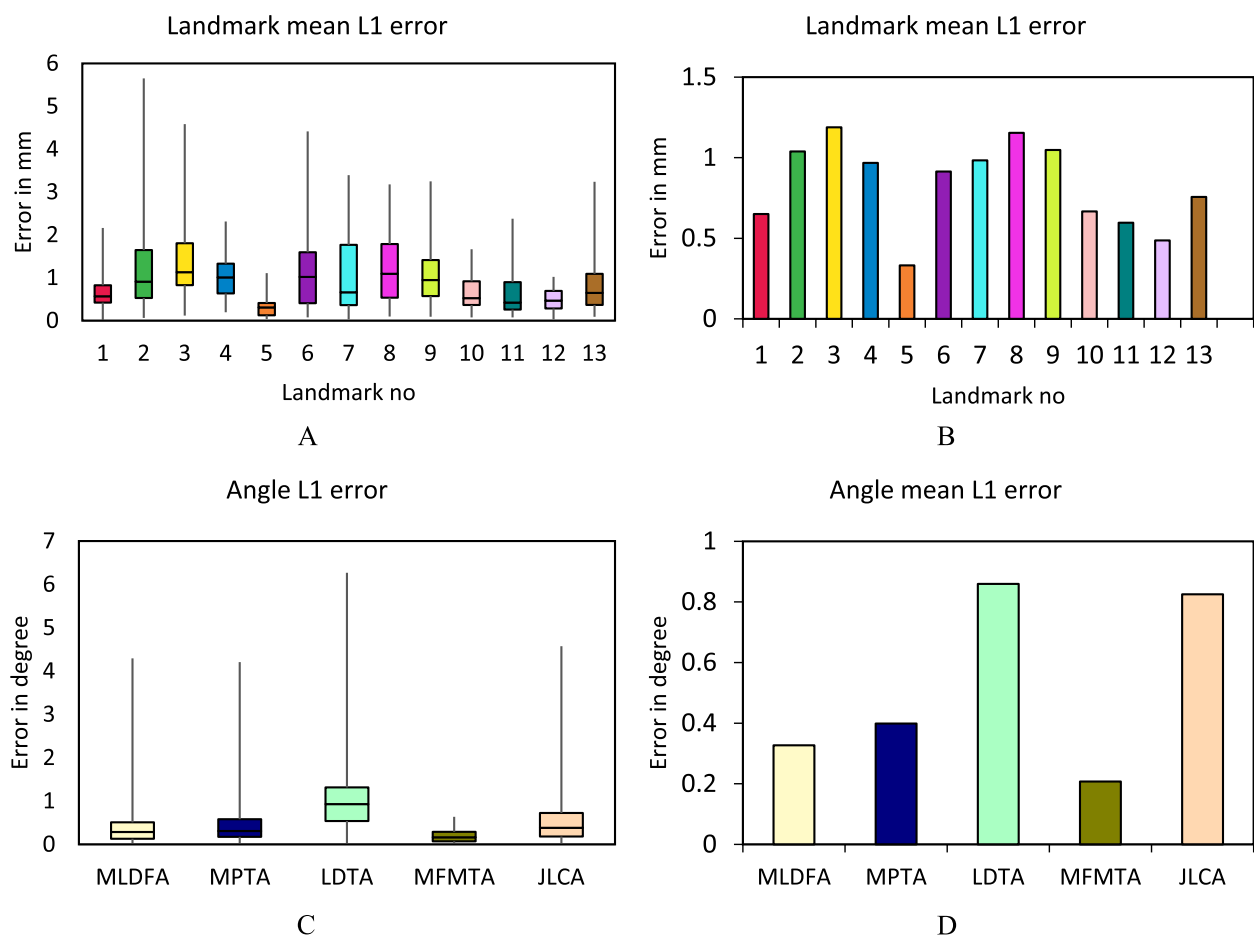


Fig. 5 Boxplot to describe **A** landmarks' L1 error and **CB** landmarks' mean L1 error and **D** angles' mean L1 error for all test data excluding failed cases

accuracy of successful detections was assessed using the mean, median, and standard deviation. With a mean error of $0.19 \pm 0.17^\circ$, the MFMTA displayed the highest accuracy level, while the LDTA displayed the lowest accuracy level ($0.91 \pm 0.51^\circ$). The discovered angles had an overall mean inaccuracy of $0.45 \pm 0.42^\circ$

and $0.43 \pm 0.39^\circ$ for primary and all cases, respectively. Error distribution for each calculated angle on all test data is shown using boxplots (Fig. 5C). The mean error value of the model for each angle on test data excluding failed cases is also presented using a histogram (Fig. 5D).

Table 3 Results from the MFMTA, ML DFA, MPTA, LD TA, and JL CA angles for both legs using discovered landmarks. The unsuccessful cases are identified using the absolute error (L1 loss) between our estimates and the expert measures. Within this category, the first four columns display the proportion of failed cases, absolute angle error, maximum error, and classification error (between normal, greater than normal, and lower than normal). The success rates define the accuracy (excluding failed ones) for primary and all cases. The final row displays the results averaged over all angles. The error is expressed in degrees

Angle type	Failed cases (error more than 2°)			Accuracy (excl. failed cases) for primary cases		Accuracy (excl. failed cases) for all cases	
	% Failure	median [Degree]	Max	angle error [Degree]	Classification error (accuracy)	angle error [Degree]	Classification error (accuracy)
ML DFA	2%	4.29	4.29	0.32 ± 0.22	100%	0.32 ± 0.22	100%
MPTA	4%	3.27	4.20	0.42 ± 0.38	100%	0.39 ± 0.33	100%
LD TA	8%	2.52	6.26	0.91 ± 0.51	80%	0.85 ± 0.49	82%
MFMTA	0%	-	-	0.19 ± 0.17	97%	0.20 ± 0.16	98%
JL CA	8%	2.81	4.57	0.44 ± 0.39	86%	0.42 ± 0.37	89%
Average	4.4%	3.08	6.26	0.45 ± 0.42	94%	0.43 ± 0.39	94.1%

Classification agreement

To assess the ultimate effect of landmark error on orthopedic diagnosis, the authors employed a classification method. The clinical classifications for angles are divided into three categories: normal, lower than normal, and higher than normal. According to Paley, the normal range for each indicator is specified through the average range of many healthy adults' indicators (described in section 1). In this study, each angular indicator was calculated using the estimated landmark positions for all cases. Subsequently, each indicator was assigned to one of the categories based on its magnitude relative to Paley's established thresholds. For ground truth validation, we computed the angular indicators using manually identified landmarks and categorized them according to the same thresholds. A classification was considered correct if the estimated category matched the ground truth category; otherwise, it was considered a misclassification (binary threshold). Selecting classification thresholds is crucial for optimizing diagnostic performance. If set too leniently, unhealthy conditions could be misclassified as normal (false negatives), while thresholds set too strictly could misidentify normal variations as abnormal (false positives). To assess classification efficiency, we defined accuracy as the ratio of the number of correctly classified cases to the total number of cases. The classification accuracy between the three classes for both failed, and successful cases are determined (Table 3). For all cases, the accuracy of angle class assignment between our model and expert measurements is about 94%. MPTA and ML DFA have the highest accuracy (100%) for both primary and all cases; LD TA has the lowest

accuracy of 80% and 82% for primary and all cases, respectively.

The accuracy levels of the test group participants were also assessed. While the patient (considering both legs) with the lowest accuracy (Fig. 6A) had a prediction error of 1.16 ± 0.70 mm, the person with the highest accuracy level (Fig. 6B) had a prediction error of 0.61 ± 0.43 mm. The figure provides the referred and estimated landmarks for these two instances to illustrate this prediction error level. Considering each case referring to one leg, the lowest and highest accuracy cases had a prediction error of 1.38 ± 0.69 mm and 0.55 ± 0.28 mm for landmark detection. For the angle detection, the case with the lowest and highest accuracy had a prediction error of $1.98 \pm 1.35^\circ$ and $0.12 \pm 0.43^\circ$, respectively.

The overall outcome of our model on full-leg input x-ray is shown in Fig. 7. Red spots indicate the estimated landmarks. The Red line indicates the locations of the LFDA, MPTA, LD TA, MFMTA, LMA, and JACL angles. Their values for both legs are also presented in white.

Discussion

A novel, pyramid-based approach is proposed to determine the type and severity of human lower limb deformities for osteotomy surgery. The approach employs a deep neural network of the CNN architecture to determine such characterizing angles as MFMTA, ML DFA, MPTA, LD TA, and JL CA using a set of landmarks that are automatically detected by the CNN. Each calculated angle is then categorized as normal, greater than normal, or smaller than normal based on the MD-predefined normal range for each angle.

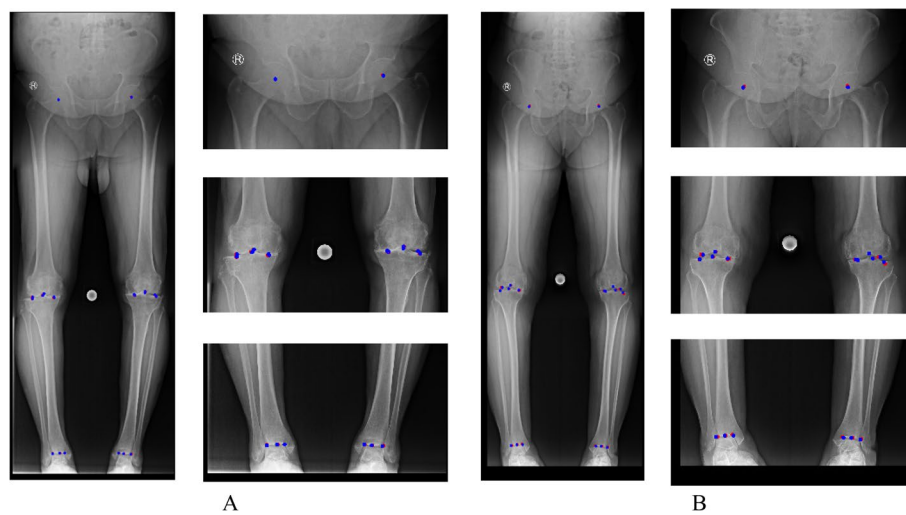


Fig. 6 The best and worst landmark predictions for the test group. **A** had the most accurate estimate (0.61 ± 0.43 mm), while **B** had the worst (1.16 ± 0.70 mm in prediction error). The experts produced the red points, while our model produced the blue ones

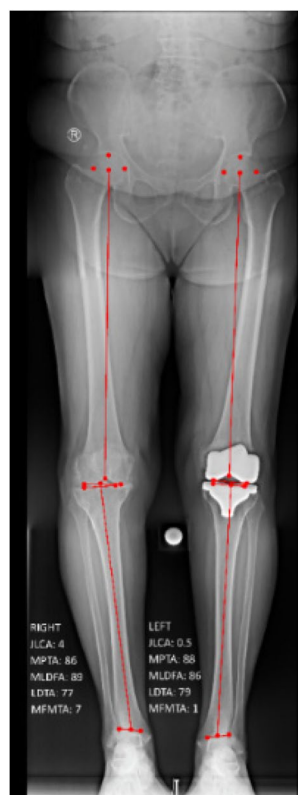


Fig. 7 A representation of the overall result. Red spots are estimated landmarks, and white lines are goal angles for both legs based on estimated landmarks

Using feedback mechanisms in multiscale feature extraction in high-resolution images are crucial for several reasons: I. Feedback mechanisms allow the model

to refine its feature extraction process based on initial outputs. This iterative refinement helps in capturing fine details in high-resolution images, leading to more precise landmark detection. II. High-resolution images often contain objects or landmarks at various scales. A feedback mechanism can help integrate features from multiple scales effectively, ensuring that both local and global landmarks are accurately detected. III. Feedback from higher-level features can extract lower-level features, helping the model understand the context in which landmarks appear. This is particularly useful in high-resolution images where both global and local features can influence the perception of a landmark. IV. High-resolution images may have noise or unwanted distortions. Feedback mechanisms identify and reduce the impact of these issues, leading to more reliable landmark detection. V. As new images are processed, feedback allows the model to adjust its strategies based on what it learns from the data. This means it can handle variations like changes in contrast, lighting or obstructions.

The most recent efforts use first- and second-order detection to find regions of interest at joints and landmark coordinates, respectively [7, 8, 13]. In comparison, our approach gets around high-resolution datasets by constructing multiresolution Gaussian pyramids to learn features in many scales in a single-step detection. This eliminates the need to extract each region of interest independently. Tack et al. used pre-trained CNN (ResNet) to identify the regions of joints. However, some failed cases were reported based on the different contrast of images. Additionally, certain outliers were displayed in landmark regression because of significant bone deformities and low image contrast. Even so, the total HKA

measurement and landmark detection error are acceptable for patients without primary treatment [7]. While Pei et al. used U-net to benefit from segmentation, Nguyen et al. used CNN to determine the location of the joints. However, Pei et al. reported a 0.5° average error between the automatic results and the manual measurements for the patient without primary treatments for the HKA angle. Nguyen et al. also reported 0.59°, 1.01°, and 1.08° average error for HKA, ML DFA, and MPTA, respectively [8, 13]. Our approach identifies the entire range of angles used in lower limb evaluations to precisely assess deformities. The approach easily deals with high-resolution images, cases with primary treatment like total hip arthroplasty, and more promising results are reported (Table 4).

For our high-resolution lower-limb dataset, the success of our approach is encouraging. Three levels of performance are shown for the suggested approach: I. determining the error between the landmarks' coordinates, which are automatically provided by the proposed approach, and the manually measured values used as standard references for the test images. II. the difference between the defined angles using manually measured values and automatically estimated landmarks. III. classification error for angles range between our estimated classification and standard references.

Some failed cases were identified due to high errors in landmark positioning and indicator evaluation. Those cases primarily relate to points that are inherently challenging to specify manually. To address this issue, an effective method is to define difficult landmarks in relation to others (the authors employed the same idea for the central point of the femoral head). Increasing the amount of training data is another potential solution. Utilizing more technicians for data labeling can also help reduce human error for manual labeling. The other primary reason for high errors, particularly in the joints, is the shadowing effect in imaging. A better imaging device with higher contrast that clearly defines the joint lines would help resolve this issue.

Table 4 Comparison with other methods: Average absolute error $\pm$ standard deviation (average results over all test data) is reported for angular indicators. ^aTop corresponds to the right and bottom to the left leg measurements

Method	JLCA [°]	MLDFA [°]	MPTA [°]	HKA [°]
Neguyen et al. ^a [8]	-	0.90 $\pm$ 0.80 1.14 $\pm$ 0.89	1.15 $\pm$ 0.94 1.03 $\pm$ 0.67	0.67 $\pm$ 0.42 0.54 $\pm$ 0.49
Jo et al. [23]	0.72 $\pm$ 0.64	0.52 $\pm$ 0.47	0.46 $\pm$ 0.45	0.22 $\pm$ 0.17
Sanchez et al. [24]	-	0.73 $\pm$ 0.69	1.09 $\pm$ 1.02	0.34 $\pm$ 0.43
Ours	0.44 $\pm$ 0.39	0.32 $\pm$ 0.22	0.42 $\pm$ 0.38	0.19 $\pm$ 0.17

3D abnormalities, such as rotational deformities, cannot obviously be detected through 2D X-ray imaging. Furthermore, in cases with rotational deformities, the coronal plane, which is employed to define all 2D landmarks and indicators, becomes misaligned. None of the data used in this study has a torsional deformation component, therefore in this model all the data was used to detect 2D landmarks. For general application of the model, the user needs to remove cases with a torsional deformation component. For these cases 3D imaging is recommended to get a better understanding of the patient's anatomy and deformities.

Conclusion

In this study, the pyramid-based approach proved quite effective results for automatic 2D lower-limb annotation for high-resolution x-ray images. The proposed approach involves constructing a pyramid using an adaptively sized 2D Gaussian patch over each scale and feeding those patches to a modified ResNet architecture to extract features at each individual resolution and finally estimating the coordinates of target landmarks using an integrating neural network. Results from multiple simulated experiments showed that the proposed approach can provide the required precision for direct clinical applications. Additionally, our system didn't need additional segmentation or region of interest detection processes. We anticipate that the benefits of our approach will make planning for osteotomies and lower-limb analysis more expedient.

The work reported here can be extended to utilize 3D data to evaluate 3D angular indicators like femoral and tibial rotations that do not appear in 2D images [25]. We envision analyzing the 3D data by estimating the landmark position using a multi-view technique. A pyramid-based network is utilized to estimate the landmark position for each view, followed by a 3-layer neural network that serves as a fusion strategy to determine the 3D locations of the landmarks [22]. This constitutes part of an ongoing research by the authors.

Abbreviations

AP	Anteroposterior
MFMTA	Mechanical Femoral Mechanical Tibial Angle
MLDFA	Mechanical Lateral Distal Femoral Angle
MPTA	Mechanical Proximal Tibial Angle
LDTA	Lateral Distal Tibial Angle
JLCA	Joint Line Convergence Angle
FAMA	Femoral Mechanical and Anatomical Axes
HKA	Hip-Knee-Ankle
CNN	Convolutional Neural Network

Acknowledgements

The authors declare that there are no acknowledgments to disclose.

Authors' contributions

All authors contributed to the study conception and design. Material preparation and analysis were performed by Reyhaneh Rostamian and Massoud Shariat-Panahi. Data collection were performed by Morad Karimpour, Alireza

Almasi Nokiani, and Ramin Jafarzadeh Khaledi. The first draft of the manuscript was written by Reyhaneh Rostamian. Hadi G. Kashani aided in interpreting the results and worked on the manuscript. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability

Full-leg X-ray dataset and the landmarks' annotations will be made publicly available with this publication to support upcoming works in the field (https://aialab.ir/datasets/xray_rad_lower_limb.html).

Declarations

Ethics approval and consent to participate

The study was approved by the Research Ethics Committees of Tehran University—Faculty of Sport Sciences and Health, University of Tehra, Tehran, Iran (Approval ID: IR.UT.SPORT.REC.1402.106). Informed consent was obtained from all individual participants included in the study.

Consent for publication

The authors affirm that human research participants provided informed consent for publication of the images.

Competing interests

The authors declare no competing interests.

Received: 8 April 2024 Accepted: 20 May 2025

Published online: 27 May 2025

References

1. Luís NM, Varatojo R. Radiological assessment of lower limb alignment. *Effort Open Rev.* 2021;6(6):487–94.
2. Paley D, Herzenberg JE, Tetsworth K, McKie J, Bhav A. Deformity planning for frontal and sagittal plane corrective osteotomies. *Orthop Clin North Am.* 1994;25(3):425–65.
3. Chao EY, Neluhani EV, Hsu RW, Paley D. Biomechanics of Malalignment. *Orthop Clin North Am.* 1994;25(3):379–86.
4. Cooke TDV, Li J, Scudamore RA. Radiographic assessment of bony contributions to knee deformity. *Orthop Clin North Am.* 1994;25(3):387–93. Available from: <https://europepmc.org/article/med/8028882>. Cited 2022 Jul 25.
5. Inman VT. The joints of the ankle. Baltimore: Williams & Wilkins; 1976.
6. Paley D. Principles of deformity correction. 3rd ed. Herzenberg m JE, editor. Springer-Verlag Berlin Heidelberg New York; 2005.
7. 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. <https://doi.org/10.1016/j.cmpb.2021.106080>.
8. Nguyen TP, Chae DS, Park SJ, Kang KY, Lee WS, Yoon J. Intelligent analysis of coronal alignment in lower limbs based on radiographic image with convolutional neural network. *Comput Biol Med.* 2020;120(March):103732. <https://doi.org/10.1016/j.combiomed.2020.103732>.
9. Cooke T, Scudamore R, Bryant J, Sorbie C, Siu D, Fisher B. A quantitative approach to radiography of the lower limb Principles and applications. *J Bone Joint Surg Br.* 1991;73-B(5):715–20.
10. Sled EA, Sheehy LM, Felson DT, Costigan PA, Lam M, Cooke TDV. Reliability of lower limb alignment measures using an established landmark-based method with a customized computer software program. *Rheumatol Int.* 2011;31(1):71–7.
11. 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.
12. Erne F, Grover P, Dreischarf M, Reumann MK, Saul D, Histing T, et al. Automated Artificial Intelligence-Based Assessment of Lower Limb Alignment Validated on Weight-Bearing Pre- and Postoperative Full-Leg Radiographs. *Diagnostics.* 2022;12(11):2679. Available from: <https://www.mdpi.com/2075-4418/12/11/2679/htm>. Cited 2023 Aug 8.
13. Pei Y, Yang W, Wei S, Cai R, Li J, Guo S, et al. Automated measurement of hip–knee–ankle angle on the unilateral lower limb X-rays using deep learning. *Phys Eng Sci Med.* 2021;44(1):53–62. <https://doi.org/10.1007/s13246-020-00951-7>.
14. Alansary A, Oktay O, Li Y, Le FL, Hou B, Vaillant G, et al. Evaluating reinforcement learning agents for anatomical landmark detection. *Med Image Anal.* 2019;1(53):156–64.
15. Szeliski R. *Computer Vision: Algorithms and Applications*. London: Springer, London; 2011. Available from: <http://link.springer.com/10.1007/978-1-84882-935-0>. Cited 2022 Jul 13.
16. Gilmour L, Ray N. Locating Cephalometric X-Ray Landmarks with Foveated Pyramid Attention. *Proc Mach Learn Res.* 2020;iii:262–74. Available from: <http://arxiv.org/abs/2008.04428>.
17. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016. p. 770–8.
18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit.* 2016;2016:770–8.
19. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Adv Neural Inf Process Syst.* 2019;8024–8035. Available from: <https://papers.nips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html>. Cited 2022 Jul 13.
20. Sun X, Xiao B, Wei F, Liang S, Wei Y. Integral Human Pose Regression. Available from: <https://github.com/JimmySuen/integral-human-pose>. Cited 2022 Jul 13.
21. Payer C, Štern D, Bischof H, Urschler M. Regressing heatmaps for multiple landmark localization using CNNs. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2016;9901 LNCS:230–8. Available from: https://link.springer.com/chapter/10.1007/978-3-319-46723-8_27. Cited 2022 Jul 13.
22. Rostamian R, Shariat Panahi M, Karimpour M, Kashani HG, Abi A. A deep learning-based multi-view approach to automatic 3D landmarking and deformity assessment of lower limb. *Sci Rep.* 2025;15(1):1–12.
23. Jo C, Hwang D, Ko S, Yang MH, Lee MC, Han HS, et al. Deep learning-based landmark recognition and angle measurement of full-leg plain radiographs can be adopted to assess lower extremity alignment. *Knee Surgery, Sport Traumatol Arthrosc.* 2022;(August). <https://doi.org/10.1007/s00167-022-07124-x>.
24. Sanchez SA, Vanoverschel P, Lebleu J, Pauwels A, Servaes W, Wiersinga W. Solutions for lower limb malalignment : a segmentation-guided coordinate regression approach for landmark detection and automatic measurement. *Epic Ser Heal Sci.* 2024;7:1–6.
25. Kuiper RJA, Seevinck PR, Viergever MA, Weinans H, Sakkars RJB. Automatic Assessment of Lower-Limb Alignment from Computed Tomography. *J Bone Jt Surg.* 2023;105(9):700–12.

Publisher's Note

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