



Computer-aided automatic measurement of leg length on full leg radiographs

Chan Su Lee¹ · Mu Sook Lee² · Shi Sub Byon¹ · Sung Hyun Kim³ · Byoung Il Lee¹ · Byoung-Dai Lee⁴ 

Received: 20 July 2021 / Revised: 25 September 2021 / Accepted: 26 September 2021 / Published online: 30 September 2021
© ISS 2021

Abstract

Objectives To develop and evaluate a deep learning (DL)-based system for measuring leg length on full leg radiographs of diverse patients, including those with orthopedic hardware implanted for surgical treatment.

Methods This study retrospectively assessed 2767 X-ray scanograms of 2767 patients who did or did not have orthopedic hardware implanted between January 2016 and December 2019. A cascaded DL model was developed to localize the relevant landmarks on the pelvis, knees, and ankles required for measuring leg length. Statistical analysis was performed using the correlation coefficient analysis and Bland–Altman plots to assess the agreement between the reference standard and DL-calculated lengths.

Results Testing data comprised 400 radiographs from 400 patients. Of these radiographs, 100 were from patients with orthopedic hardware implanted in their pelvis, knees, or ankles. For all testing data, leg lengths derived from the DL-based measurement system, with or without internal fixation devices, showed excellent agreement with the reference standard (femoral length, $r=0.99$ ($P<.001$); root mean square error (RMSE)=0.17 cm; mean difference, -0.01 ± 0.17 cm; 95% limit of agreement (LoA), -0.35 to 0.34 ; tibial length, $r=0.99$ ($P<.001$); RMSE=0.17 cm; mean difference, -0.02 ± 0.17 cm, 95% LoA, -0.34 to 0.31 ; and full leg length, $r=1.0$ ($P<.001$); RMSE=0.19 cm; mean difference, 0.05 ± 0.18 cm; 95% LoA, -0.31 to 0.40). The mean time for leg length measurement for each patient using the DL-based system was 8.68 ± 0.18 s.

Conclusion The DL-based leg length measurement system could provide similar performance to radiologists in terms of accuracy and reliability for a diverse group of patients.

Keywords Deep learning · Full leg radiography · Leg length measurement · Retrospective studies

Abbreviations

CCC Concordance correlation coefficient
DL Deep learning
LLD Leg length discrepancies

LoA Limit of agreement
MAE Mean absolute error
MSE Mean squared error
POI Point of interest
 r Pearson correlation coefficient
RMSE Root mean square error
ROI Region of interest

Chan Su Lee and Mu Sook Lee contributed equally to this work.

Key points: • Strong correlations ($P<.001$) were observed between the reference standard and lengths of the separated femurs, tibias, and full legs that were calculated through deep learning.

- The deep learning-based system measured each patient's leg length in approximately 8.68 ± 0.18 s.
- There was no significant difference in performance noted when the radiographs of patients with internal fixation devices implanted in the pelvis, knees, or ankles were compared with those without implants.

✉ Byoung-Dai Lee
blee@kgu.ac.kr

Extended author information available on the last page of the article

Introduction

Leg length discrepancies (LLDs), in which the lengths of the paired lower extremity limbs are unequal, are considered orthopedic deformities [1]. LLDs are relatively common and found in as many as 40–70% of adult and pediatric populations [1–3]. The causes of LLDs include idiopathic developmental abnormalities, scoliosis, fixed pelvic tilt, orthopedic degenerative disorders, and surgical disorders,

such as joint replacements [1–5]. In particular, studies have suggested that LLD is associated with several lower limb and lumbar biomechanical conditions and musculoskeletal disorders, such as foot pronation, low back pain, osteoarthritis of the knee and hip joints, and stress fractures [1, 6–10]. Therefore, it is crucial that the leg length be accurately and reliably measured to allow for proper diagnosis and treatment planning.

Direct and indirect clinical measurement of leg length relies on the palpation of bony landmarks, which is prone to errors [7]. Therefore, due to their accuracy and reliability, radiographs are considered the gold standard for measuring leg length [1, 7–10]. However, manual leg length measurements using radiographs, while cognitively simple for a radiologist, are time-consuming tasks [10]. Computer-assisted quantification programs may allow for automatic leg length measurements without these limitations. The recent development of a deep learning–based (DL-based) approach showed promising results for LLD measurement on limited datasets [10].

This study aimed to evaluate the accuracy and reliability of DL-based leg length measurements using full leg radiographs of a diverse group of patients, including those with internal fixation devices that were implanted for surgical treatment, by comparing the performance of a DL-based system to that of trained radiologists.

Materials and methods

Study population and dataset

The appropriate institutional review board approved this retrospective study, and the requirement for informed consent was waived due to the retrospective design of the study. We identified 2791 X-ray scanograms that were taken between January 2016 and December 2019 (Innovision; DongKang Medical Systems Co. Ltd., Seoul, South Korea), and a total of 2767 X-ray scanograms from 2767 patients were manually selected for this study. Of these scanograms, there are approximately twice as many female patients as male patients, and 328 were from patients with internal fixation devices implanted in the pelvis, knees, or ankles. Patient's radiographic studies were excluded from the study if they (a) overlapped ($n=0$); (b) had segmental views of the hips, knees, and ankles ($n=21$); (c) had skeletal dysplasia ($n=0$); (d) had external orthopedic hardware ($n=0$); or (e) had undergone radiography with a scale at their side ($n=3$). Full leg, standing, anteroposterior view radiographs that included both lower extremities from the hips to the ankles were obtained through an image acquisition system (HPACS software; HealthHub Co. Ltd., Seoul, South Korea). The image resolution was approximately 3000×8000 pixels in a 24-bit gray scale JPEG format. All images were de-identified before use. The patient demographic statistics are summarized in Table 1.

Table 1 Data and patient characteristics

Dataset source	Training data	Validation data	Testing data
No. of total patients	1895	472	400
No. of patients with internal fixation devices*	183 (915)	45 (225)	100
Pelvis only	16 (80)	3 (15)	20
Knees only	108 (540)	27 (135)	40
Ankles only	23 (115)	5 (25)	20
More than 2 places	36 (180)	10 (50)	20
Mean age (y)	55 ± 16	55 ± 16	66 ± 13
Age range (y)	8–91	10–89	11–88
Sex (n)			
Men	682	151	142
Women	1213	321	258
Leg length (cm)			
Femoral length	45.7 ± 4.1 (35.3–57.3)	45.2 ± 3.3 (36.1–53.8)	46.5 ± 3.6 (36.9–56.5)
Tibial length	36.3 ± 3.0 (30.0–46.9)	35.6 ± 2.7 (29.5–44.8)	35.7 ± 2.9 (28.9–45.3)
Full leg length	80.3 ± 7.5 (63.7–102.2)	80.5 ± 7.8 (63.7–101.5)	82.1 ± 6.3 (65.9–100.5)

Note—Except where indicated, data are expressed as the mean \pm standard deviation, with ranges in parentheses

*Numbers in parentheses are the total number after the data augmentation was applied

Patient studies were split into training, validation, and testing data without overlaps (Fig. 1). Images with internal fixation devices in the pelvis, knee, or ankle only accounted for 9.7% (183/1,895) and 9.5% (45/472) of the total number of training and validation data, respectively. To address the class imbalance, we applied a simple data augmentation scheme based on image cropping and horizontal flip. A given radiograph was manually cropped at all four corners, while ensuring that no body parts were lost. Training data were used to train DL models for detecting regions of interest (ROI) containing relevant landmarks for leg length measurement as well as segmenting bones, including implanted prostheses, from ROI images. ROIs in the individual radiographs and the femoral area and knee and ankle joint masks that were within corresponding ROIs were manually annotated by a board-certified radiologist. These annotations were used as the ground truth for ROI

detection and segmentation. Details of the training and validation datasets for deep learning models are summarized in Table E1 and E2.

Reference standard for leg length measurements

A board-certified radiologist performed the leg length measurements on full leg, standing, anteroposterior radiographs in the testing data. Here, femoral, tibial, and full leg lengths were measured based on the mechanical axis [11]. The point of interest (POI) for measurements in the femoral area was the top of the femoral head or prosthetic femoral head. The POI for the knee joint was the most distal end of the medial femoral condyle or the prosthesis that replaced it. The POI for the ankle joint was the center of the tibial plafond or the prosthesis that replaced it. Leg length measurements on the full leg, standing radiograph are presented in Fig. 2.

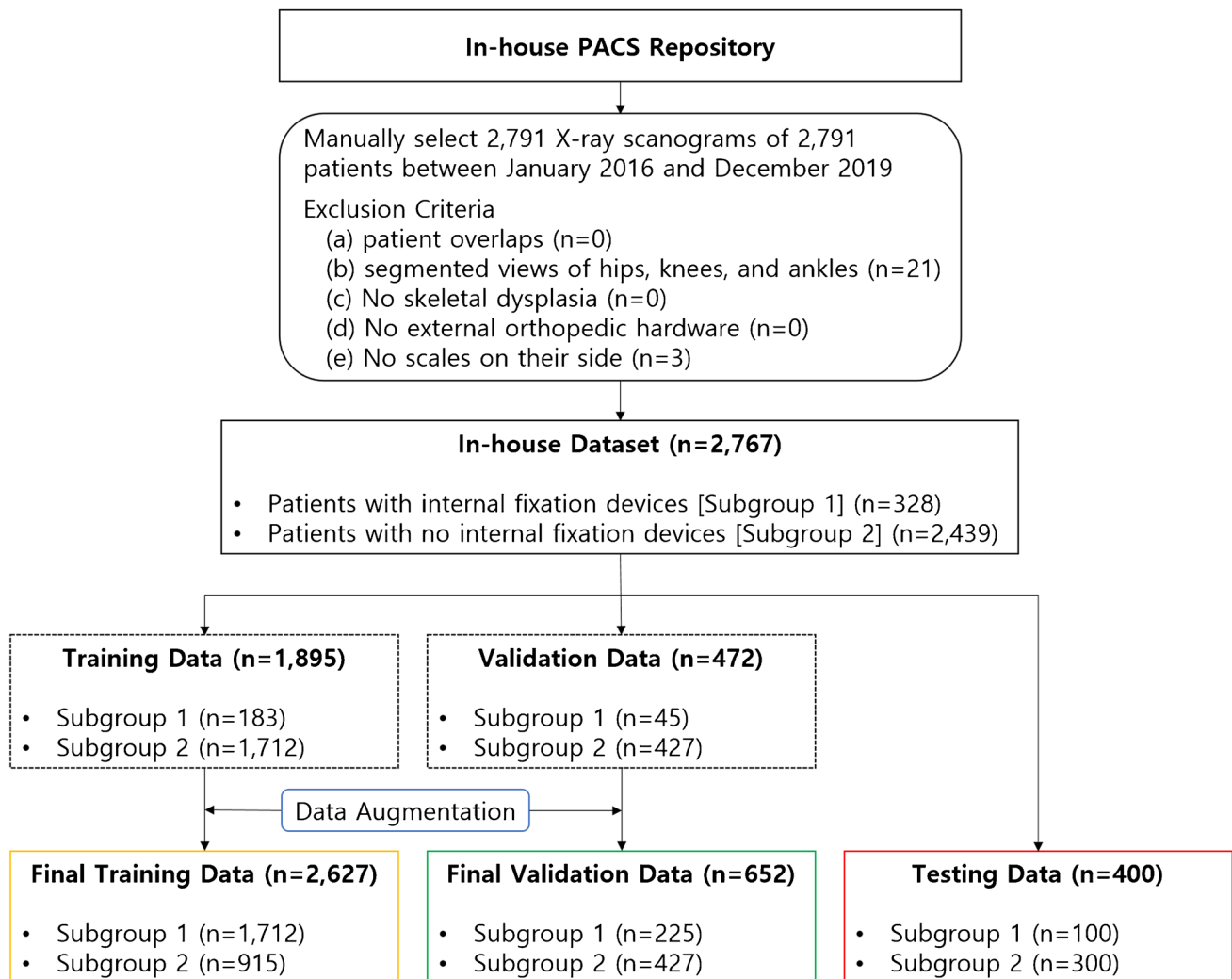


Fig. 1 Flow chart for the datasets. PACS, picture archiving and communication system

Development of the DL-based, fully automated, leg length measurement system

The system had a cascade architecture that comprised three stages (Fig. 2). The first stage involved locating six ROIs and identifying their classes. Based on the ROI's location (e.g., femoral area, knee joint and ankle joint on the left and right sides) and the presence of internal fixation devices in the ROI, the ROIs were classified into 12 different classes. We customized the single shot multibox detector [12] architecture for ROI detection. The second stage comprised 12 segmentation DL models, each of which was responsible for segmenting bones in the designated ROIs. All segmentation DL models shared the same model architecture; however, they were trained with different training data. For the segmentation DL model, we customized the standard U-Net [13] model. In the third stage, the bone contours in the segmentation masks were traversed using conventional image processing algorithms to find the accurate POI locations. Once all six POIs were identified, the final leg lengths on both sides were calculated. For training and testing, we used a single NVIDIA Titan-XP GPU (NVIDIA, Santa Clara, CA, USA). The code is accessible at <https://github.com/RTOSChansu/llm/>. See Appendix E1 for additional details of the model development.

The ROI detection performance was evaluated using the mean average precision that was the mean of the average precision scores [14] for the 12 different ROI classes. The ROI segmentation performance was evaluated using the Dice score that was calculated using the following formula: $(2 \times TP) / ((TP + FP) + (TP + FN))$, where *TP*, *FP*, and *FN* indicated the number of true positive, false positive, and false negative pixels, respectively.

Statistical analysis

The accuracy and reliability of the DL-based leg length measurements were evaluated using the concordance correlation coefficient (CCC) and Pearson correlation coefficient (*r*) in relation to a reference standard. A Bland–Altman plot was also used to assess the agreement between the reference leg lengths and the leg length that was calculated through DL. Additionally, the mean squared error (MSE), mean absolute error (MAE), and root mean square error (RMSE) were used to evaluate the accuracy of the leg length that was calculated through DL in relation to a reference standard. All statistical analyses were performed using Minitab software (Minitab 17.3.1; Minitab LLC, State College, PA, USA) and R-programming (version 3.1.2 [2014]; <https://www.r-project.org/>).

Besides analyzing all testing data, subgroup analysis was performed to determine if the implanted orthopedic hardware influenced the accuracy and reliability of the DL-based leg length measurement system. Testing data were grouped into two disjoint subgroups: subgroup 1, which included patients with any internal fixation devices implanted in the pelvis, knees, or ankles; and subgroup 2, which included patients who had no orthopedic hardware.

Results

Study participants

In total, 2767 patients (mean age \pm standard deviation, 56 ± 16 years; age range, 8–91 years; 975 men and 1972 women) were included in our study. Among those patients, 328 had internal fixation devices implanted in their pelvis,

Table 2 Mean average precision of the DL model for ROI detection and the Dice scores of the individual segmentation deep learning models

ROI Class		AP	Dice score
Left side	The femoral area with hardware (ROI class 1)	0.92	0.98
	The femoral area without hardware (ROI class 2)	0.99	0.99
	The knee joint with hardware (ROI class 3)	0.99	0.99
	The knee joint without hardware (ROI class 4)	0.99	0.99
	The ankle joint with hardware (ROI class 5)	0.97	0.99
	The ankle joint without hardware (ROI class 6)	0.99	0.98
Right side	The femoral area with hardware (ROI class 7)	0.94	0.99
	The femoral area without hardware (ROI class 8)	1.0	0.99
	The knee joint with hardware (ROI class 9)	0.99	0.99
	The knee joint without hardware (ROI class 10)	0.99	0.98
	The ankle joint with hardware (ROI class 11)	0.97	0.98
	The ankle joint without hardware (ROI class 12)	0.99	0.98
mAP		0.98	
Average of the 12 segmentation DL models		0.98	

Note—*DL* deep learning, *ROI* region of interest, *AP* average precision, *mAP* mean AP

knees, or ankles (Fig. 1). Patient studies were split into training, validation, and testing data (68%, 1895/2767; 17%, 472/2767; and 15%, 400/2767), each of which comprised patients who did or did not have implanted orthopedic hardware. The baseline characteristics of the study population are described in Table 1.

Performance of the DL models for ROI detection and segmentation

The DL models for ROI detection and segmentation performed excellently with the entire testing dataset. There were

no cases of failure with the ROI detection and the resultant mean average precision was 0.98. The subsequent segmentation masks generated by the DL models overlapped significantly with the corresponding ground truth. The average Dice score of the 12 segmentation DL models was 0.98. The DL models' performance in ROI detection and segmentation is summarized in Table 2.

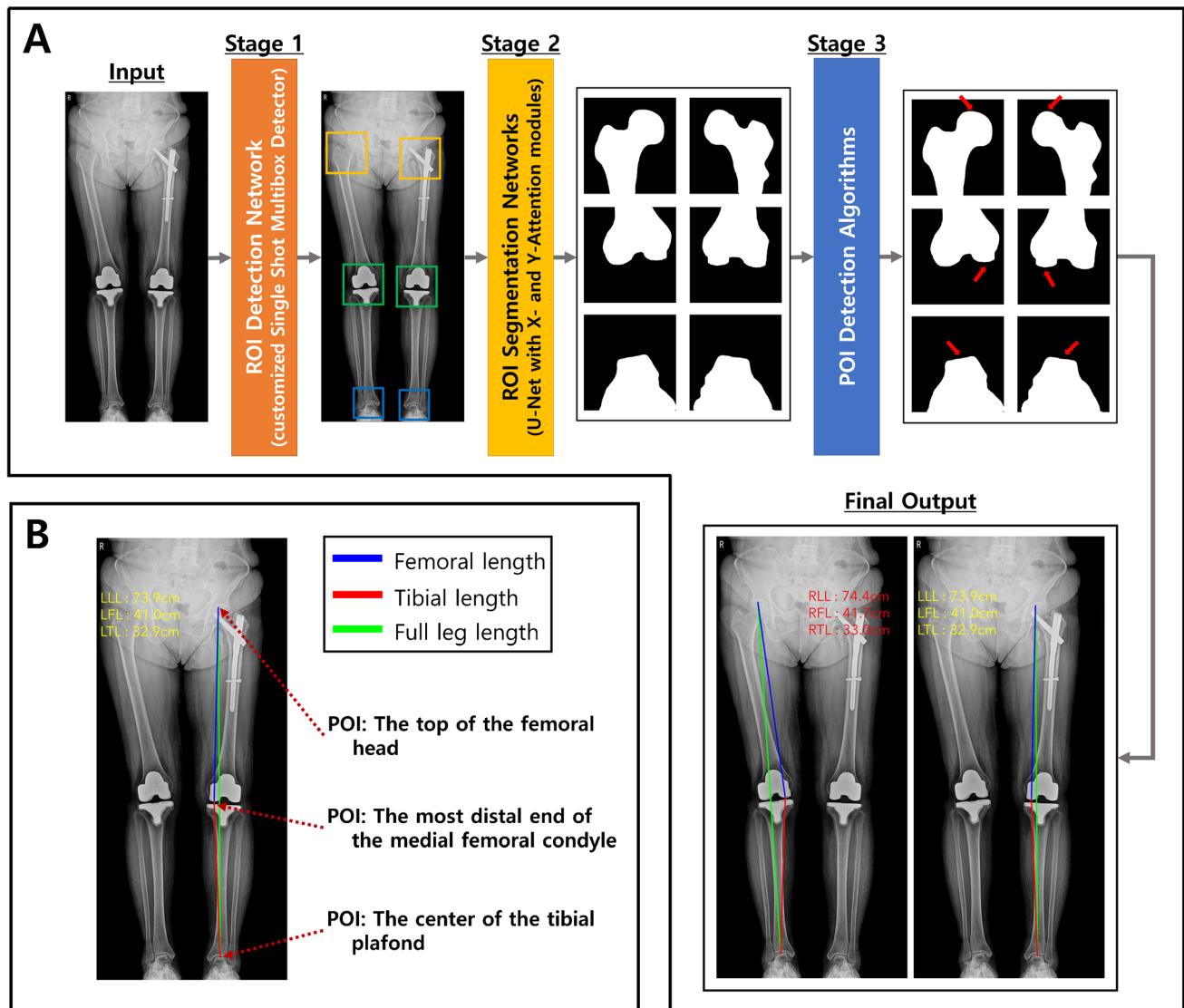


Fig. 2 An illustration of the leg length measurement process. **A** The deep learning-based system had a cascade architecture. The ROI detection network first detected six ROIs that contained the relevant landmarks that were needed for the measurements. The ROI segmentation networks then segmented the bones, including prostheses, within ROIs. The segmentation masks were processed using

image processing algorithms to identify the POIs that corresponded to the relevant landmarks in the pelvis, knees, and ankles. Finally, the leg length was measured by calculating the distance between the extracted pixels; **B** POIs for the leg length measurement based on mechanical axes. Best viewed in color. ROI, region of interest, POI, point of interest

Accuracy and reliability of DL-based leg length measurements

The mean time for leg length measurements by the DL-based system was 8.68 ± 0.18 s for each patient when using a single NVIDIA Titan-XP GPU (NVIDIA, Santa Clara, CA, USA). The leg length that was derived from the DL-based system showed excellent agreement with the reference standard (femoral length, CCC=0.99, $r=0.99$; tibial length, CCC=0.99, $r=0.99$; and full leg length, CCC=1.0, $r=1.0$; $P < 0.001$ for all correlations) (Table 3).

According to the Bland–Altman agreement analysis, there was no systematic bias between the reference standard and lengths that were calculated through the DL-based system. The mean differences \pm standard deviation between the two femoral, tibial, and full leg length measurements were -0.01 ± 0.17 cm, -0.02 ± 0.17 cm, and 0.05 ± 0.18 cm, respectively. The 95% limits of agreement (LoA) between the two femoral, tibial, and full leg length measurements ranged from -0.35 to 0.34 , -0.34 to 0.31 , and -0.31 to 0.40 , respectively (Fig. 3).

The DL-based system's high measurement accuracy was also observed with the MSE, MAE, and RMSE (femoral length, MSE=0.03 cm, MAE=0.14 cm, RMSE=0.17 cm; tibial length, MSE=0.03 cm, MAE=0.13 cm, RMSE=0.17 cm; and full leg length, MSE=0.04 cm, MAE=0.15 cm, RMSE=0.19 cm) (Table 3).

Subgroup analysis

The DL-based measurement system's performance was not influenced by the presence of internal fixation devices in the pelvis, knees, or ankles. For both subgroups, it was observed that the two measurements correlated significantly (subgroup 1, CCC=0.99, 0.99, and 1.0; $r=0.99$, 0.99, and 1.0 for femoral, tibial, full leg lengths, respectively; subgroup 2, CCC=0.99, 0.99, and 1.0; $r=0.99$, 0.99, and 1.0 for femoral, tibial, and full leg lengths, respectively; $P < 0.001$ for all correlations) (Table 3).

The Bland–Altman plots also demonstrated that the accuracy and reliability of the DL-based leg length measurements were not affected by the internal fixation devices that were implanted in patients' lower limbs (Fig. 3). For subgroup 1, the mean differences \pm standard deviation between the two femoral, tibial, and full leg length measurements were -0.01 ± 0.18 cm (95% LoA: -0.37 to 0.35); 0.06 ± 0.14 cm (95% LoA: -0.32 to 0.21); and -0.05 ± 0.21 cm (95% LoA: -0.47 to 0.36), respectively. For subgroup 2, the mean differences between the two femoral, tibial, and full leg length measurements were 0.00 ± 0.17 cm (95% LoA: -0.34 to 0.33); -0.01 ± 0.17 cm (95% LoA: -0.34 to 0.33); and 0.05 ± 0.18 cm (95% LoA: -0.31 to 0.40), respectively.

For subgroup 1, the RMSEs for the femoral, tibial, and full leg lengths were 0.18 cm, 0.15 cm, and 0.22 cm, respectively, whereas the corresponding errors of the

Table 3 Correlation between the leg lengths that were calculated using deep learning and the reference standard

Variables	Femoral length			Tibial length			Full leg length		
	Subgroup 1	Subgroup 2	Total	Subgroup 1	Subgroup 2	Total	Subgroup 1	Subgroup 2	Total
Concordance correlation coefficient (cm)	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00
Pearson correlation coefficient (cm)*	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00
Mean difference \pm SD (cm)	-0.01 ± 0.18	-0.00 ± 0.17	-0.01 ± 0.17	-0.06 ± 0.14	-0.01 ± 0.17	-0.02 ± 0.17	-0.05 ± 0.21	0.08 ± 0.16	0.05 ± 0.18
95% limit of agreement (cm)	($-0.37, 0.35$)	($-0.34, 0.33$)	($-0.35, 0.34$)	($-0.32, 0.21$)	($-0.34, 0.33$)	($-0.34, 0.31$)	($-0.47, 0.36$)	($-0.24, 0.39$)	($-0.31, 0.40$)
Mean square error (cm)	0.03	0.03	0.03	0.02	0.03	0.03	0.05	0.03	0.04
Mean absolute error (cm)	0.14	0.14	0.14	0.11	0.14	0.13	0.16	0.15	0.15
Root mean square error (cm)	0.18	0.17	0.17	0.15	0.17	0.17	0.22	0.18	0.19

* P values for correlations are $P < .001$

SD standard deviation

femoral, tibial, and full leg lengths in subgroup 2 were 0.17 cm, 0.17 cm, and 0.18 cm, respectively. Both MAE and MSE also showed similar levels of errors as RMSE. The results from the performance evaluation are summarized in Fig. 3 and Table 3, and examples cases of leg lengths that are calculated using the DL-based system are presented in Fig. 4.

Discussion

We developed and validated a deep learning–based fully automated leg length measurement system using full leg standing anteroposterior view radiographs that included both lower extremities from the hips to the ankles. For all the testing data that randomly included patients who

did or did not have internal fixation devices in the pelvis, knees, or ankles, the ability of the DL-based system for detecting and segmenting regions of interest was excellent (mean average precision = 0.98, average Dice score = 0.98). Additionally, strong correlations were observed between the reference standard and separated femur, tibia, and full leg lengths that were calculated using deep learning (femoral length, $r = 0.99$ ($P < 0.001$), root mean square error (RMSE) = 0.17 cm, mean absolute error (MAE) = 0.14 cm, mean difference -0.01 ± 0.17 cm, 95% limit of agreement (LoA) -0.35 to 0.34 ; tibial length, $r = 0.99$ ($P < 0.001$), RMSE = 0.17 cm, MAE = 0.13 cm, mean difference -0.02 ± 0.17 cm, 95% LoA -0.34 to 0.31 ; and full leg length, $r = 1.0$ ($P < 0.001$), RMSE = 0.19 cm, MAE = 0.15 cm, mean difference 0.05 ± 0.18 cm, 95% LoA -0.31 to 0.40). The average measurement time

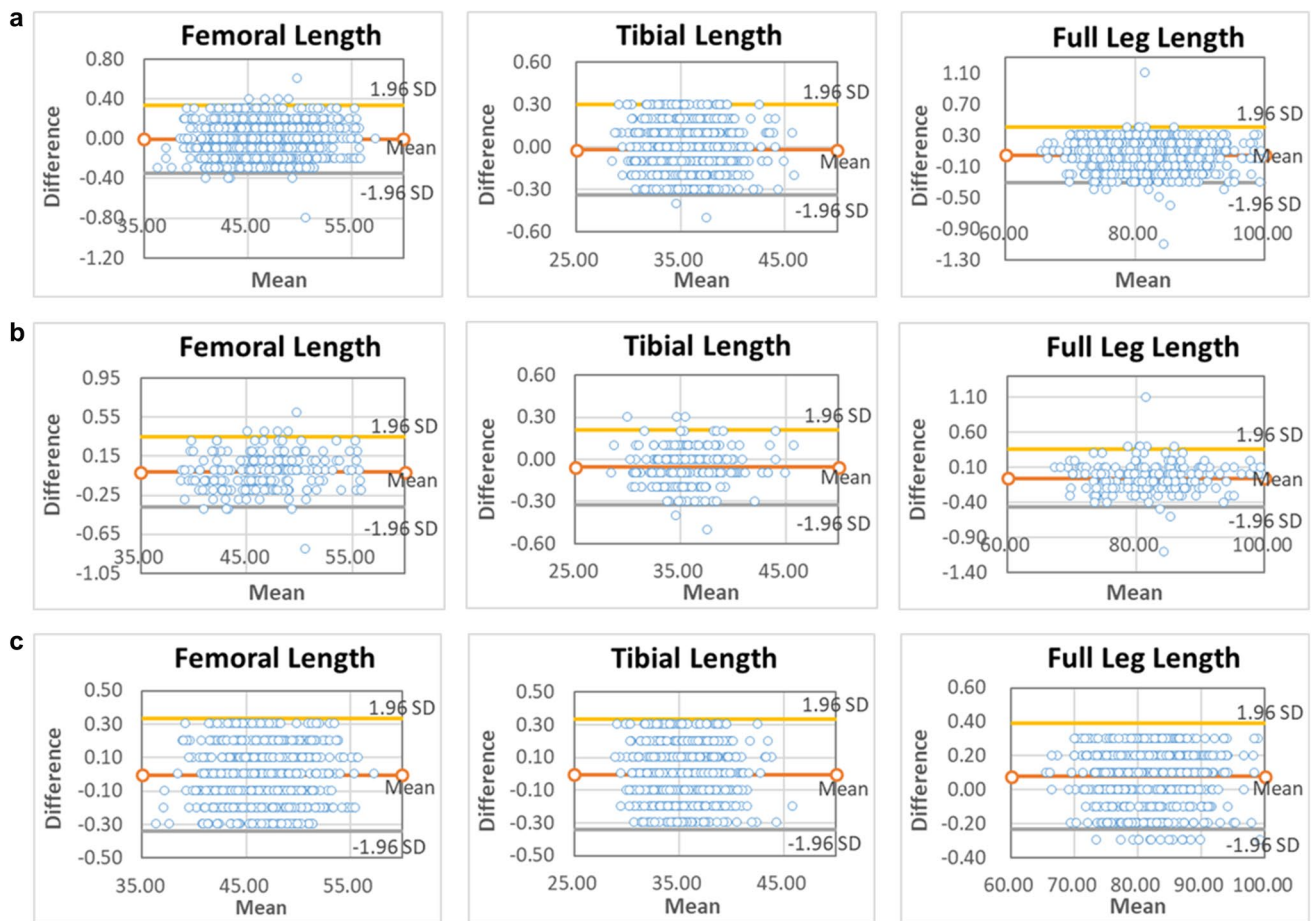
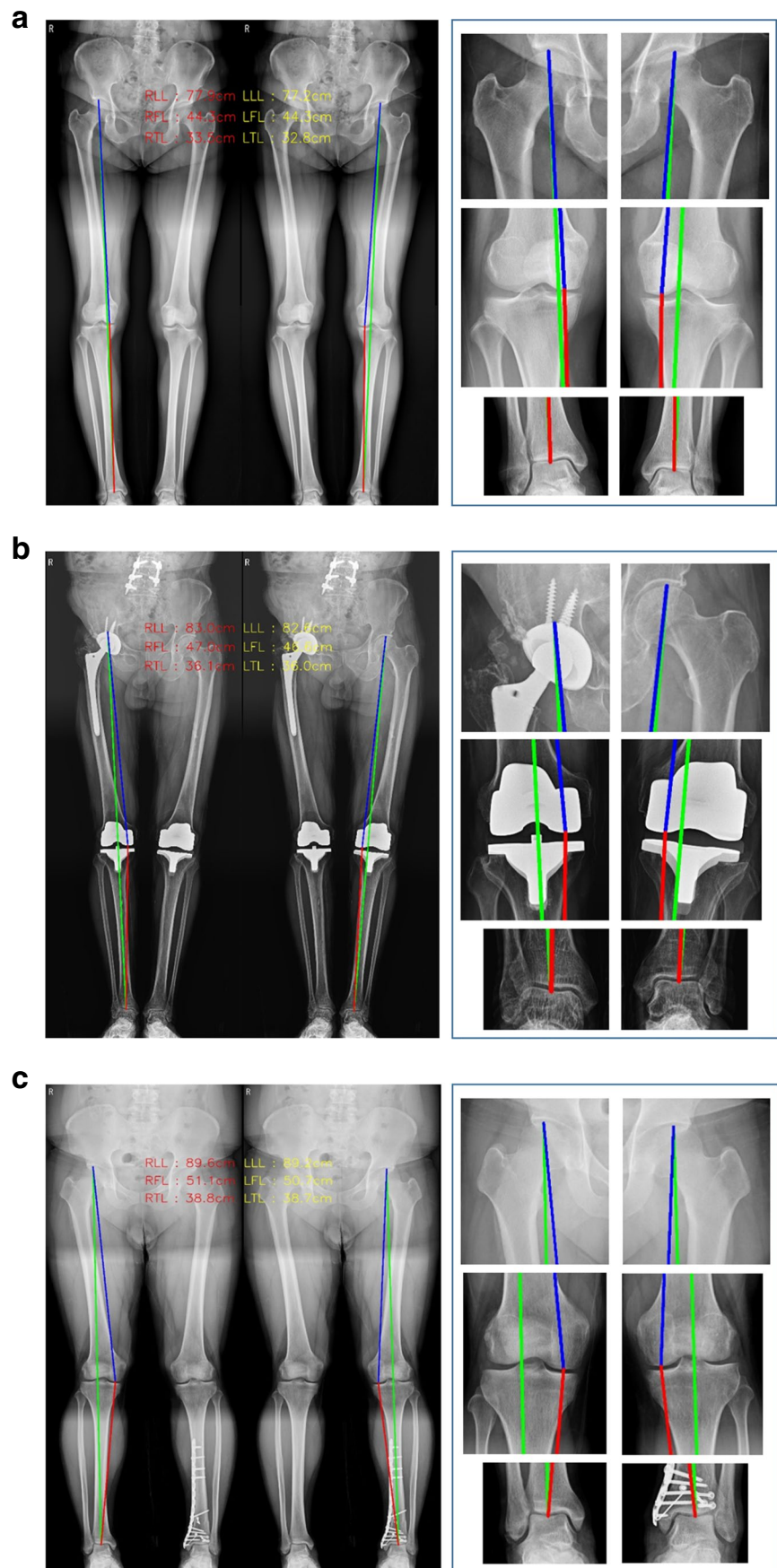


Fig. 3 Bland–Altman plots of the reference leg lengths and leg lengths that were calculated using DL. The x-axis represents the mean of the reference and DL-calculated lengths (cm), whereas the

y-axis represents the difference between the reference and DL-calculated lengths (cm). **A** All the testing data; **B** subgroup 1; **C** subgroup 2. *DL* deep learning, *SD* standard deviation

Fig. 4 Radiographic images with the DL-calculated leg lengths. The lines were generated by a DL-based automatic leg length measurement system. The blue line represents the femoral length, and the red line represents the tibial line. The green line represents the full leg length. Images on the right side demonstrated that the DL-based system was able to accurately and reliably identify relevant landmarks that were required to measure the leg length. **A** A patient with no orthopedic hardware implanted; **B** A patient with internal fixation devices implanted in the pelvis and knees; **C** A patient with internal fixation devices implanted in the ankle. *RLL* right full leg length, *RFL* right femoral length, *RTL* right tibial length, *LLL* left full leg length, *LFL* left femoral length, *LTL* left tibial length, *DL* deep learning



per radiograph for the deep learning–based system was 8.68 ± 0.18 s. There were no significant differences in performance when the subgroups were assessed according to the presence of internal fixation devices (subgroup 1, $r = 0.99$ ($P < 0.001$), 0.99 ($P < 0.001$), and 1.0 ($P < 0.001$); RMSE = 0.18 cm, 0.15 cm, and 0.22 cm; MAE = 0.14 cm, 0.11 cm, 0.16 cm for the femoral, tibial, and full leg lengths, respectively; subgroup 2, $r = 0.99$ ($P < 0.001$), 0.99 ($P < 0.001$), and 1.0 ($P < 0.001$); RMSE = 0.17 cm, 0.17 cm, and 0.18 cm; MAE = 0.14 cm, 0.14 cm, 0.15 cm for the femoral, tibial, full leg lengths, respectively). For subgroups 1 and 2, the mean differences between the two measurements were small with a 95% confidence interval in femoral, tibial, and full leg lengths (subgroup 1, -0.01 ± 0.18 cm, -0.06 ± 0.14 cm, and -0.05 ± 0.21 cm; 95% LoA -0.37 to 0.35 , -0.32 to 0.21 , and -0.47 to 0.36 , respectively; subgroup 2, 0.00 ± 0.17 cm, -0.01 ± 0.17 cm, and 0.08 ± 0.16 cm; 95% LoA -0.34 to 0.33 , -0.34 to 0.33 , and -0.24 to 0.39 , respectively). These findings indicated that there was no systematic bias between the two measurements.

Zheng et al. [10] demonstrated that the measurement of LLD in pediatric patients on radiographs could be automated and performed rapidly by deep learning algorithm (CCC = 0.99, $r = 0.99$ ($P < 0.001$), MAE = 0.45 cm for separated femur and tibia lengths; CCC = 0.99, $r = 0.99$ ($P < 0.001$), MAE = 0.45 cm for the full leg length). Our findings similarly suggest that automated and precise measuring of leg length on full leg standing radiographs of diverse patients, including those with orthopedic hardware implanted for surgical treatment, can be enabled by deep learning algorithms. As in our study, Zheng et al.'s approach employed a multi-step process for image segmentation, in which the DL model first roughly split radiographs into left and right leg images, then finely segmented femurs and tibias on the unilateral full-limb images. However, there were differences in our methodology. We believe that these differences were the key to our system's improved generalization capability. First, to accurately recognize and localize relevant landmarks on the hips, knees, and ankles, we included the radiographs of both pre- and postoperative patients who had internal orthopedic hardware implanted for training DL models. To the best of our knowledge, this is the first study to assess the effectiveness of DL-based leg length measurements for patients with internal fixation devices. Second, instead of exploring the entire image to segment the femurs and tibias, our approach only allowed ROIs that enclosed relevant landmarks (e.g., POIs) to be considered; thus, it was more effective in capturing discriminative local features. Third, we adopted state-of-the-art attention mechanisms to enable the fine segmentation of ROIs where prostheses were present.

This study had some limitations. First, we did not evaluate the DL models by using external datasets that are completely independent from those used for model training. Utilizing additional data from multiple institutions for model training is needed to improve the DL models' generalizability. Second, patients with skeletal dysplasia were not included in either training or testing. Third, we did not consider the angulation of the femoral and tibial axes. Fourth, we used only a single model of the equipment for X-ray scanogram acquisition.

In conclusion, we developed and evaluated a clinically relevant deep learning–based leg length measurement system that performed similar to radiologists in terms of accuracy and reliability. The performance of our system was not influenced by the orthopedic hardware implanted in lower extremity limbs of patients. This will be helpful in reducing repetitive and time-consuming task of leg length measurement required for surgical planning and for the diagnosis and prognosis of relevant disease. Further prospective studies should be performed to extend this deep learning–based system to provide comprehensive set of leg geometry measurements by increasing diversity of and numbers contained in the training data for deep learning models, as well as by enhancing feature learning capability of deep learning models.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00256-021-03928-z>.

References

1. Gurney B. Leg length discrepancy. *Gait Posture*. 2002;15:195–206.
2. Yong M, Park S. Leg length discrepancy to influence on kinematic changes of the pelvis and the hip during gait. *J Kor Phys Ther*. 2019;31:368–71.
3. Khamis S, Carmeli E. A new concept for measuring leg length discrepancy. *J Orthop*. 2017;14:276.
4. Meermans G, Malik A, Witt J, Haddad F. Preoperative radiographic assessment of limb-length discrepancy in total hip arthroplasty. *Clin Orthop Relat Res*. 2011;469:1677–82.
5. Blake RL, Ferguson H. Limb length discrepancies. *J Am Podiatr Med Assoc*. 1992;82:33–8.
6. Rannisto S, Okuloff A, Uitti J, et al. Leg-length discrepancy is associated with low back pain among those who must stand while working. *BMC Musculoskelet Disord*. 2015;16:110.
7. Ahmed K. Leg length discrepancy: assessment and secondary effects. *Ortho & Rheum Open Access*. 2017;5:555678.
8. Sabharwal S, Kumar A. Methods for assessing leg length discrepancy. *Clin Orthop Relat Res*. 2008;466:2910–22.
9. Murray KJ, Azari MF. Leg length discrepancy and osteoarthritis in the knee and lumbar spine. *J Can Chiropr Assoc*. 2015;59:226–37.

10. Zheng Q, Shellikeri S, Huang H, Hwang M, Sze RW. Deep learning measurement of leg length discrepancy in children based on radiographs. *Radiology*. 2020;296:152–8.
11. Rungprai C, Goetz JE, Arunakul M, et al. Validation and reproducibility with a biplanar imaging system versus conventional radiography. *Foot Ankle Int*. 2014;35:1166–75.
12. Liu W, Anguelov D, Erhan D et al (2016) SSD: Single shot multi-box detector. In: Leibe B., Matas J., Sebe N., Welling M. (eds) *Computer Vision – ECCV 2016*. ECCV 2016. Lecture Notes in Computer Science, Springer, Cham. 9905:21–37.
13. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015. Lecture Notes in Computer Science, Springer, Cham. 9351:234–241
14. Zhang P, Su W (2012) Statistical inference on recall, precision and average precision under random selection. *Proceedings of 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, Chongqing, China. 1348–1352

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

Authors and Affiliations

Chan Su Lee¹ · Mu Sook Lee² · Shi Sub Byon¹ · Sung Hyun Kim³ · Byoung Il Lee¹ · Byoung-Dai Lee⁴ 

¹ AI Lab, HealthHub, Co. Ltd, 623, Gangnam-daero, Seocho-gu, Seoul 06524, Republic of Korea

² Department of Radiology, Keimyung University Dongsan Hospital, 1035, Dalgubeol-daero, Sindang-dong, Daegu 42601, Republic of Korea

³ Human Medical Imaging and Intervention Center, 621, Gangnam-daero, Seocho-gu, Seoul 06524, Republic of Korea

⁴ Division of AI Computer Science and Engineering, Kyonggi University, 154-42, Gwanggyosan-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16227, Republic of Korea