Journal of Orthopaedic Surgery and Research

RESEARCH ARTICLE

Open Access

HKA-Net: clinically-adapted deep learning for automated measurement of hip-knee-ankle angle on lower limb radiography for knee osteoarthritis assessment

Young-Tak Kim[1,2†], Beom-Su Han[3†], Jung Bin Kim[4], Jason K. Sa[2,5], Je Hyeong Hong[6], Yunsik Son[7], Jae-Ho Han[8], Synho Do[1,9,10], Ji Seon Chae[11] and Jung-Kwon Bae[3*]

Abstract

Background Accurate measurement of the hip-knee-ankle (HKA) angle is essential for informed clinical decision-making in the management of knee osteoarthritis (OA). Knee OA is commonly associated with varus deformity, where the alignment of the knee shifts medially, leading to increased stress and deterioration of the medial compartment. The HKA angle, which quantifies this alignment, is a critical indicator of the severity of varus deformity and helps guide treatment strategies, including corrective surgeries. Current manual methods are labor-intensive, time-consuming, and prone to inter-observer variability. Developing an automated model for HKA angle measurement is challenging due to the elaborate process of generating handcrafted anatomical landmarks, which is more labor-intensive than the actual measurement. This study aims to develop a ResNet-based deep learning model that predicts the HKA angle without requiring explicit anatomical landmark annotations and to assess its accuracy and efficiency compared to conventional manual methods.

Methods We developed a deep learning model based on the variants of the ResNet architecture to process lower limb radiographs and predict HKA angles without explicit landmark annotations. The classification performance for the four stages of varus deformity (stage I: 0°–10°, stage II: 10°–20°, stage III: > 20°, others: genu valgum or normal alignment) was also evaluated. The model was trained and validated using a retrospective cohort of 300 knee OA patients (Kellgren-Lawrence grade 3 or higher), with horizontal flip augmentation applied to double the dataset to 600 samples, followed by fivefold cross-validation. An extended temporal validation was conducted on a separate cohort of 50 knee OA patients. The model's accuracy was assessed by calculating the mean absolute error between predicted and actual HKA angles. Additionally, the classification of varus deformity stages was conducted to evaluate the model's ability to provide clinically relevant categorizations. Time efficiency was compared between the automated model and manual measurements performed by an experienced orthopedic surgeon.

Results The ResNet-50 model achieved a bias of – 0.025° with a standard deviation of 1.422° in the retrospective cohort and a bias of – 0.008° with a standard deviation of 1.677° in the temporal validation cohort. Using

[†]Young-Tak Kim and Beom-Su Han have contributed equally to this manuscript.

*Correspondence:
Jung-Kwon Bae
roundnfirm@hanmail.net
Full list of author information is available at the end of the article

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

the ResNet-152 model, it accurately classified the four stages of varus deformity with weighted F1-score of 0.878 and 0.859 in the retrospective and temporal validation cohorts, respectively. The automated model was 126.7 times faster than manual measurements, reducing the total time from 49.8 min to 23.6 sec for the temporal validation cohort.

Conclusions The proposed ResNet-based model provides an efficient and accurate method for measuring HKA angles and classifying varus deformity stages without the need for extensive landmark annotations. Its high accuracy and significant improvement in time efficiency make it a valuable tool for clinical practice, potentially enhancing decision-making and workflow efficiency in the management of knee OA.

Keywords Hip-knee-ankle angle, Deep learning, Residual network, Varus deformity, Knee osteoarthritis, Lower limb scanogram

Background

Knee osteoarthritis (OA) is a pervasive and debilitating condition that significantly impairs joint function and diminishes the overall quality of life [1]. Knee OA is commonly associated with varus deformity [2], characterized by medial angulation of the knee joint, which increases the load on the medial compartment and accelerates joint degeneration. This abnormal load distribution leads to greater stress on the cartilage, causing accelerated wear and tear, pain, and functional impairment [3]. The continuous stress further exacerbates OA progression, ultimately altering the biomechanics of the lower limb [4]. The severity of varus deformity is quantified using the hip-knee-ankle (HKA) angle [5]. At the core of orthopedic assessment, the HKA angle serves as a pivotal metric, providing a quantitative measurement of the lower limb alignment by evaluating the angular relationships among the hip, knee, and ankle joints [6, 7]. The precise determination of the HKA angle is essential in the OA management, aiding in the diagnosis of medical conditions such as Genu Varum and Valgum [3, 8], as well as in the strategic planning and evaluation of surgical intervention [9, 10]. Accurate measurement of the HKA angle is critical for selecting appropriate treatment strategies, as it guides decisions ranging from conservative management to surgical interventions, such as high tibial osteotomy or total knee arthroplasty [11, 12]. Therefore, precise quantification of this angle is indispensable in managing knee OA, particularly in patients with varus deformity.

While the gold standard in orthopedic clinical practice relies on manual landmark measurements of the HKA angle, it proves to be labor-intensive and susceptible to both inter- and intra-observer variability [13, 14]. Furthermore, the overwhelming influx of radiographs in medical facilities, particularly in resource-constrained settings, constrains orthopedic surgeons from striving for timely and precise analyses [15]. Despite the promising artificial intelligence (AI) solutions in medical imaging [16], their development often mandates additional labor-intensive data collection beyond the HKA angle measurements, such as segmentation [14, 17] or detailed landmark annotation [18–22] for specific anatomical structures. For instance, C Jo, et al. trained their model using 15 anatomical landmarks within such joint areas [18]. Similarly, A Tack, et al. utilized 12 anatomical landmarks [19], and Y Pei, et al. focused on three regions of interest (ROI) at the femur and the tibia [14]. WP Gielis, et al. employed an extensive set of 111 anatomical landmarks across these regions [20], while TP Nguyen, et al. targeted 5 ROIs in the joint areas for their model training [17]. Likewise, J Yang, et al. proposed a framework involving 10 landmarks, including key points such as the center of the femoral head, the knee joint center, and the ankle joint center [21]. I Tanner, et al. also developed a deep learning model that detected key points such as the femoral head, knee, and ankle joints [22]. The latest current AI applications, while enhancing precision, have introduced time-consuming processes, limiting widespread adaptation within the clinical framework and requiring additional post-processing and interpretation efforts [23]. Models extensively trained on landmark annotations face challenges during continual fine-tuning, as updating these annotations for new data is resource-intensive and time-consuming. This can make the models less adaptable in clinical practice, particularly when faced with new data that requires re-annotation. To address this issue, there is a need for methods that reduce reliance on manual annotations, allowing for easier updates and quicker adaptability in real-world clinical workflows. Streamlining this process would help facilitate broader adoption in orthopedic settings by reducing the time and effort required for ongoing model maintenance.

To address these challenges, we have developed an approach to predict HKA angle measurements using the residual network (ResNet) architecture [24] with transfer learning. ResNet architecture have renowned for its capabilities in deep learning and image analysis

[25]. By applying transfer learning, we efficiently trained the model using pre-trained weights from large image datasets, allowing the model to fully exploit the feature extraction capabilities of the ResNet architecture. The aim of this study is to develop an automated, deep learning-based model that utilizes the ResNet architecture with transfer learning to accurately predict HKA angles from lower limb radiographs. This approach aims to reduce the reliance on labor-intensive manual landmark annotations, improving the efficiency and accuracy of clinical workflows in orthopedic practice, particularly in resource-constrained environments, thereby alleviating the workload for orthopedic surgeons and enhancing patient care.

Methods

Study design and setting

This retrospective study was conducted at Seoul Medical Center, Republic of Korea. The following narrative provides a comprehensive overview of this study (Fig. 1). The HKA angle is the primary measure being predicted in the study. For the determination of the HKA angles, the initial line extends from the femoral head to the intercondylar notch of the femur, while the subsequent line connects the tibial interspinous point to the tibial midplafond (Fig. 1A). This study proposes a model based on variants of the ResNet architecture (i.e., ResNet with 18, 34, 50, 101 and 152 layers) [24] that individually predicts the HKA angles of each lower limb through a unilateral pathway (Fig. 1B). This study utilized data from the lower limbs of 300 patients diagnosed with knee OA, which was augmented through horizontal flipping to double the dataset size, to construct a model that assessed the HKA angle and evaluated varus deformity classifications based on fivefold cross-validation (Fig. 1C). Varus deformity was classified into four stages based on the HKA angle: stage I (0°–10°), stage II (10°–20°), stage III (>20°), and others (genu valgum or normal alignment) [26]. Moreover, the five optimal models from each fold of the cross-validation were ensembled to minimize overfitting and tested on an additional cohort of 50 individuals for extended temporal validation.

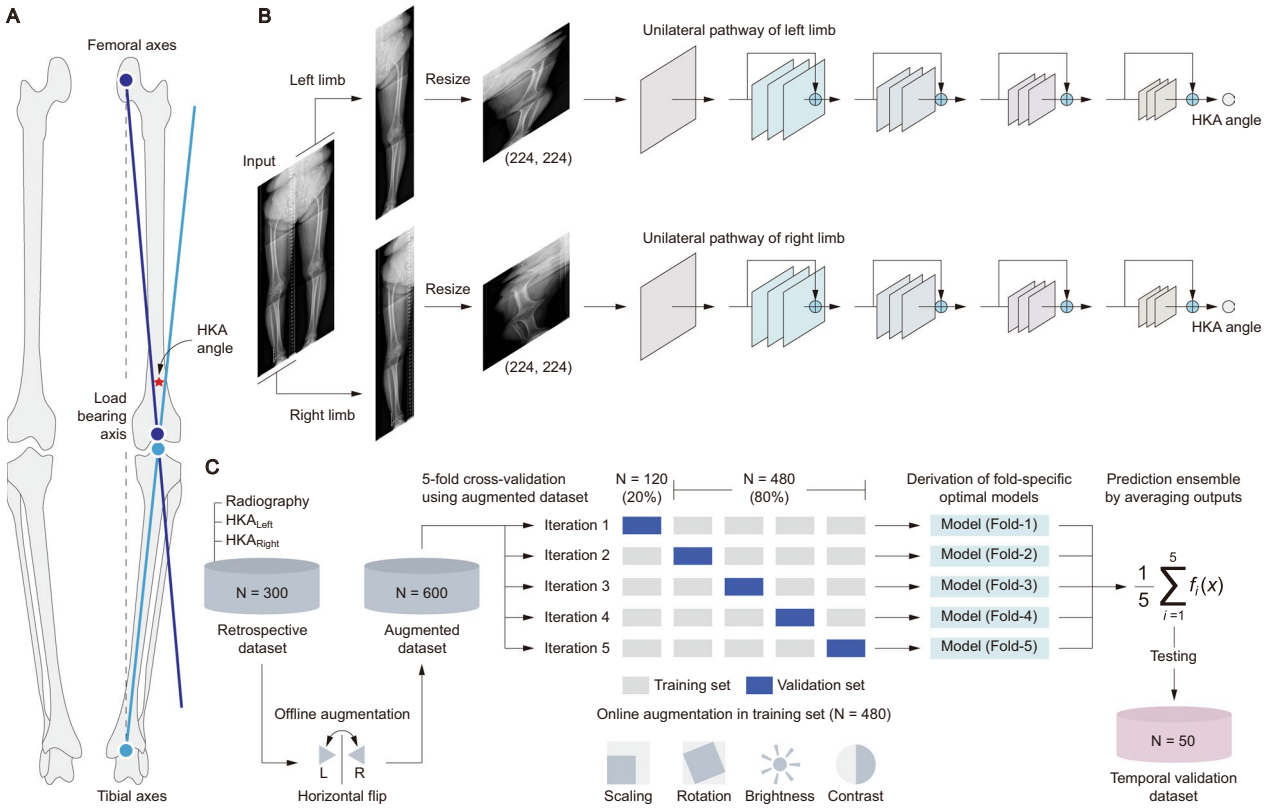


Fig. 1 Overview of deep learning approach for HKA assessment. **A** depicts the anatomical landmarks vital for HKA angle calculation. Here, the femoral axis (indigo) and the tibial axis (blue) converge at the knee, with the load-bearing axis (illustrated by a dotted line) indicating the weight distribution path. **B** outlines the architecture of the ResNet-based predictive model. This model processes images through separate pathways for each limb, each utilizing a ResNet architecture. **C** demonstrates the integrative validation approach, combining a retrospective fivefold cross-validation with data augmentation, followed by temporal validation using an ensemble of the five models

Participants

Three hundred lower limb radiographs of patients who had been diagnosed with knee OA and had undergone total knee replacement arthroplasty between January 2012 and December 2022 at the Seoul Medical Center were used to develop the deep learning models. Additionally, extended temporal validation data, comprising 50 patients, were acquired from January to October 2023 at the same institution to validate the model based on this retrospective cohort. The inclusion criteria were as follows: (1) adult patients aged 18 years or older; (2) patients diagnosed with knee OA of Kellgren-Lawrence (KL) grade 3 or higher; (3) patients who underwent primary total knee arthroplasty at Seoul Medical Center during the study period; and (4) availability of preoperative weight-bearing full-length lower limb radiographs taken in the standing position. The exclusion criteria included: (1) patients with tumors or other significant lesions in the lower limb; (2) patients with congenital deformities or neuromuscular disorders affecting lower limb alignment; and (3) cases where radiograph quality was insufficient for accurate analysis due to poor image resolution, improper positioning, or artifacts. All patients underwent bilateral leg scanning on their first admission. Ethical approval was obtained from the Institutional Research and Ethics Board (SEOUL-2023-12-002-001). The total cohort size was determined to ensure high-quality patient data, optimizing the available resources within practical constraints. This allowed us to prioritize reliable and well-curated patient data. Patient data were anonymized and included demographic and clinical histories (e.g., age, sex, height, weight, Body Mass Index (BMI), and American Society of Anesthesiologists (ASA) score [27]).

Measurement of HKA angle

The measurement of the HKA angle for both lower limbs was confirmed based on the consensus of two orthopedic surgeons. These measurements were performed using the Cobb angle measurement tool provided by the INFINITT PACS M6 software (Version 6.0.11.0, INFINITT Healthcare, Seoul, South Korea). The time taken for HKA angle measurement was measured as the interval from the initiation of the scanogram data load to the completion of the HKA angle measurement and was recorded concurrently when acquiring the HKA angle for the temporal validation cohort.

Data acquisition and preprocessing

Standard lower limb radiography was obtained using DigitalDiagnost Rel. 4.3 (Philips Medical Systems, Best, Netherlands) in the anterior–posterior view with patients in a standing, weight-bearing position. It has dimensions of 3040×7422 pixels, and a resolution of 7.35 pixels per mm. The window center and width were set to 2047 and 4095 pixels, respectively, to optimize the contrast and brightness of the images for enhanced clarity. These images were preprocessed to facilitate analysis. Each image was divided into two sections, representing the left and right limbs, by bisecting the image at its midpoint, resulting in two images of 1520×7422 pixels each. These were then resized to 224×224 pixels to suit the input requirements of the deep learning architecture based on ResNet (Fig. 1B).

Model construction

ResNet architectures with configurations of 18, 34, 50, 101 and 152 layers were analyzed to identify the most suitable model. ResNet features convolutional, pooling, and fully connected layers. This architecture is distinguished by its use of residual blocks with skipped connections. Each block consists of several convolutional layers, followed by batch normalization and rectified linear unit (ReLU) activation with a skip connection. To effectively process the grayscale knee X-ray images and utilize the pre-trained ResNet architecture with minimal alteration, we converted the grayscale images to RGB format by duplicating the single channel across the three RGB channels. Additionally, we applied the standard normalization used in ResNet models, based on the ImageNet dataset, by adjusting the RGB pixel values with the mean [0.485, 0.456, 0.406] and standard deviations [0.229, 0.224, 0.225]. This ensures that the images are compatible with the ResNet model while preserving the original pixel intensity values and anatomical details, making them accurate for analysis. To address the unique anatomical and positional differences of each limb in the knee OA images, we utilized separate but identical ResNet models to process the images of the left and right limbs.

For training, we employed transfer learning by initializing the ResNet architecture with pretrained weights to boost the initial feature recognition. The training process was optimized using the Adam optimizer with an initial learning rate of 0.001. We utilized the L1 loss for the regression and monitored the mean absolute error (MAE) as our primary performance metric.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y}_i\right|$$

where n denotes the total number of observations in the dataset. It averages the absolute differences between the actual (y_i) and predicted ($\widehat{y}_i$) values. These metrics provide a quantitative measure of the prediction error, indicating the average deviation of the predicted values from the actual values and the variability of these errors. The training procedure was performed for 100 epochs.

Data augmentation

To improve the robustness and generalizability of the model, we employed both online and offline augmentation techniques throughout the training process. Offline augmentation was applied before the training process began, where we performed horizontal flipping (mirroring) of all images. This effectively doubled the size of our dataset, increasing the number of samples from 300 to 600. By creating mirrored versions of the left and right leg images, the offline augmentation ensured that both lower limbs were well-represented in the training set.

During the training process, we applied online augmentation, which dynamically altered the images during each epoch. These transformations included random rotations within a 30-degree angle to simulate slight variations in limb positioning during radiographic imaging. We also applied random cropping, where between 90 and 100% of the original image was randomly selected and resized back to the input size required by the model. This ensured that the model encountered variability in image composition and patient size. Additionally, slight adjustments in brightness and contrast (within ± 10%) were applied to account for differences in exposure and imaging conditions.

Since these online augmentations were applied in real-time during training, the model encountered a varied set of inputs with each epoch. With 480 training samples per fold (after the offline augmentation), and training performed for 100 epochs, the model encountered a maximum of 48,000 augmented images during training for each fold (480 samples × 100 epochs). This augmentation process significantly improved the model's ability to generalize to unseen data by exposing it to a wide range of image variations. Off-line augmentation was applied to the validation set during cross-validation, while no online augmentation was used. For the temporal validation dataset, no augmentation of any kind was applied.

Validation strategy

A fivefold cross-validation approach was implemented during the training and validation phases to ensure the robustness and generalizability of the model (Fig. 1C). In this process, the dataset was randomly divided into 5 equal parts. In each validation fold, one part was used as the validation set and the remaining four parts were combined to form the training set. This procedure was repeated five times, with each part used as the validation set exactly once. Subsequently, the optimal models with the lowest error, selected from the fivefold cross-validation, were employed for a temporal validation dataset. These models were ensembled by averaging their outputs to minimize overfitting and were tested immediately without any additional training procedures. Furthermore, using gradient-weighted class activation mapping (Grad-CAM) [28], we explored the specific image regions where the model focused on estimating the HKA angle. The Grad-CAM results were averaged across the five optimal models from the fivefold cross-validation.

Performance assessments

After performing a retrospective fivefold cross-validation, the best result was aggregated from each validation set to evaluate the accuracy of the HKA angle prediction. This process employed each segment of the complete dataset as a test set:

$$A = \bigcup_{i=1}^{5} \{(y_i, \hat{y}_i) | y_i \in ground\ truth_i, \hat{y}_i \in prediction_i\}$$

where the aggregated result A represents the aggregation of the actual values (y_i) and their corresponding predicted values ($\hat{y}_i$) across a retrospective dataset. In this case, the extended temporal validation was identical to the operation, with $i = 1$. We used the mean error (bias) to assess the accuracy of HKA angle regression.

$$Bias = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)$$

Bias indicates the average bias in the predictions made by a model. In this calculation, y_i represents each actual value and $\hat{y}_i$ is the corresponding predicted value, with the differences between these pairs being averaged over n and the total number of observations. Furthermore, the standard deviation of error (SDE) was used to assess the variability or error:

$$SDE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} ((y_i - \hat{y}_i) - Bias)^2}$$

This involves calculating the difference between each actual value y_i and its predicted counterpart $\hat{y}_i$ subtracting the bias from these differences and squaring the results. The square root of the average of these squared differences, taken over all n observations, yielded the SDE. In addition, the coefficient of determination, or R^2, was evaluated to understand the linear relationship between the actual and predicted values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

This formula is used to calculate the coefficient of determination, or R^2, where n is the total number of observations, y_i is the actual value, $\hat{y}_i$ denotes the value predicted by the model, and $\bar{y}$ is the mean of the actual values. This metric, ranging from 0 to 1, reflects the proportion of variance in the dependent variable that is predictable from that of the independent variable.

Bland–Altman (BA) plots and scatter plots with regression lines were used.

The categorical predictions of varus deformity (no varus, stage I, stage II, and stage III) were presented in a 4×4 confusion matrix. Accuracy provides an overall indication of the performance of the model across all stages and is calculated as follows:

$$Accuracy = \frac{\sum True\ Positives}{Total\ Number\ of\ Samples}$$

However, accuracy alone may not fully reflect the model's performance, particularly in imbalanced dataset. To provide a more comprehensive evaluation, we calculated balanced accuracy. Balanced accuracy is the average of recall obtained for each class, ensuring that each class is equally considered, regardless of its representation in the dataset:

$$Balanced\ Accuracy = \frac{1}{k}\sum_{i=1}^{k}\frac{True\ Positive_i}{True\ Positive_i + False\ Negative_i}$$

where k is the number of classes, and i represents each class. Additionally, we calculated the weighted F1-score to account for class imbalance by assigning different weights to each class based on its support (i.e., the number of true instances for each class). The weighted F1-score is computed as:

$$Weighted\ F1\text{-}score = \frac{\sum_{i=1}^{k} F1\text{-}score_i \times Support_i}{\sum_{i=1}^{k} Support_i}$$

where i represents each class and $Support_i$ refers to the number of true instances in class i. This metric ensures that performance is evaluated proportionally to the number of samples in each class, providing a more balanced assessment in the presence of class imbalance.

The time elapsed for the HKA angle measurement using the model was determined from the initiation of the data load to the completion of the HKA angle measurement. Statistical analyses were performed using SciPy (Version 1.10.1). All development and evaluation procedures were executed using PyTorch (Version 2.0.1) and Scikit-learn (Version 1.3.1) in a Python (Version 3.9.0) environment. The computational environment used in the study included 6 NVIDIA Tesla V100 PCIe GPUs (32 GB each) and 2 NVIDIA GeForce RTX 4090 GPUs (24 GB each).

Results

Demographics

The study encompassed knee OA patients undergoing total knee arthroplasty, revealing distinct demographic and preoperative characteristics in retrospective and extended temporal validation cohorts (Table 1). Both groups demonstrated similarities in height, weight, and BMI. While no significant differences were observed in varus deformity and KL grade, a significant difference was noted in the HKA angle.

HKA angle prediction

The primary architectural variations of ResNet and their performance on the retrospective and temporal validation datasets are summarized in Table 2. Irrespective of the complexity of the model architecture, MAE values for the retrospective cohort ranged from 0.839° to 1.193° for the right limb and from 0.852° to 1.005° for the left limb. Bias values remained close to zero across both limbs, indicating minimal systematic error. In the temporal validation cohort, all models exhibited increased MAE compared to the retrospective cohort, with values ranging from 0.990° to 1.476° for the right limb and from 1.351° to 1.725° for the left limb. Bias values were more variable in the temporal cohort. R^2 values remained consistently high across both cohorts, ranging from 0.823 to 0.960, indicating strong correlations between predicted and actual HKA angles across models and limbs. Notably, the ResNet-50 model achieved a bias of −0.025° in the retrospective cohort and −0.008° in the temporal validation cohort when the results from both limbs were combined. The BA plot of the ResNet models was used to further illustrate the agreement between predicted and actual HKA angles (Fig. 2). Detailed performance of the HKA angle prediction based on sex and varus deformity is provided in the Supplemental Material.

Varus deformity classification

We further analyzed how the models predict the severity of varus deformity, extending from the HKA angle predictions (Fig. 3). While ResNet-50 performed the best in predicting the HKA angle, ResNet-18 demonstrated the best classification performance, achieving an accuracy of 0.917, balanced accuracy of 0.886, and weighted F1-score of 0.917 during cross-validation (Fig. 3A). In temporal validation, ResNet-18 maintained superior performances (Fig. 3B).

Comparison of time efficiency

For the temporal validation cohort, the elapsed time from data loading to HKA angle measurement in both lower

Table 1 Baseline characteristics

	Retrospective cohort (N = 300)		Temporal validation cohort (N = 50)		p-value
	Median (IQR)	N (%)	Median (IQR)	N (%)	
Age, years	72.0 (68.0 – 76.0)		76.0 (72.0 – 79.3)		0.001
Sex					0.310
Male		32 (10.7)		3 (6.0)	
Female		268 (89.3)		47 (94.0)	
Height, cm	151.0 (148.0 – 156.0)		151.0 (147.8 – 157.0)		0.650
Weight, kg	63.0 (57.0 – 71.0)		61.0 (54.0 – 67.3)		0.090
BMI	27.0 (25.0 – 30.1)		26.6 (23.8 – 29.2)		0.290
KL grade					0.288
3		137 (45.7)		27 (54.0)	
4		163 (54.3)		23 (46.0)	
ASA score					< .001
1		7 (2.3)		0 (0)	
2		268 (89.3)		36 (72.0)	
3		23 (7.7)		14 (28.0)	
4		4 (0.7)		0 (0)	
Preoperative HKA (left), °	−9.4 (−12.7 to −6.5)		−6.8 (−11.0 to −5.4)		0.013
Preoperative HKA (right), °	−8.9 (−12.3 to −5.7)		−7.7 (−10.2 to −4.1)		0.023
Varus deformity (left)					0.308
Normal		8 (2.7)		2 (4.0)	
Stage I		154 (51.3)		32 (64.0)	
Stage II		125 (41.7)		15 (30.0)	
Stage III		13 (4.3)		1 (2.0)	
Varus deformity (right)					0.312
Normal		16 (5.3)		2 (4.0)	
Stage I		167 (55.7)		34 (68.0)	
Stage II		108 (36.0)		14 (28.0)	
Stage III		9 (3.0)		0 (0)	

ASA: American Society of Anesthesiologists, BMI: Body Mass Index, IQR: interquartile range, KL: Kellgren-Lawrence

limbs was compared between the manual and automated measurements (Fig. 4). The average time required to manually measure the HKA angle was 59.78 s per patient, whereas automated measurement took 0.47 s.

Attention pattern in HKA angle prediction

The attention pattern was analyzed using an ensemble model from five models derived through fivefold cross-validation. Figure 5 shows representative cases from the temporal validation dataset. The model exhibited attention distributed across the entire lower limb, consistently focusing on both the skeletal structure and the surrounding spatial changes. In severe cases, the model emphasized the spatial regions most affected by the varus deformity, indicating that it considers not only the bone structure but also the changes in surrounding space as the deformity progresses.

Discussion

In this study, we developed a deep-learning model using lower limb radiography without relying on anatomical landmark annotations to improve HKA angle measurements in knee OA patients. Our approach relies solely on routinely acquired data in clinical practice, eliminating the need for additional labor-intensive anatomical annotation and preserving the current clinical workflow. Through retrospective analysis of lower limb radiography, we developed a predictive model focusing

Table 2 HKA angle prediction performance for retrospective and temporal validation cohorts across ResNet variants

Model	Retrospective cohort (cross-validation)			Temporal validation cohort (test)		
	MAE, °	Bias, °	R^2	MAE, °	Bias, °	R^2
Left limb						
ResNet-18	0.852 (0.664)	−0.005 (1.081)	0.959	1.531 (0.933)	0.971 (1.517)	0.888
ResNet-34	0.962 (0.751)	−0.057 (1.219)	0.948	1.633 (1.184)	1.129 (1.681)	0.859
ResNet-50	1.005 (0.812)	0.152 (1.284)	0.941	1.508 (1.330)	0.127 (2.018)	0.860
ResNet-101	0.992 (0.807)	−0.110 (1.275)	0.943	1.649 (1.447)	−0.033 (2.206)	0.833
ResNet-152	0.983 (0.774)	0.135 (1.245)	0.945	1.725 (1.462)	0.143 (2.270)	0.823
Right limb						
ResNet-18	0.839 (0.666)	−0.048 (1.071)	0.960	1.351 (0.826)	1.157 (1.086)	0.860
ResNet-34	0.850 (0.698)	0.066 (1.099)	0.957	1.476 (0.987)	1.281 (1.233)	0.824
ResNet-50	1.193 (0.977)	−0.201 (1.529)	0.917	0.990 (0.771)	−0.143 (1.255)	0.912
ResNet-101	1.115 (0.877)	−0.592 (1.289)	0.929	1.242 (0.988)	−0.292 (1.569)	0.860
ResNet-152	1.030 (0.845)	−0.325 (1.293)	0.938	1.369 (1.116)	0.476 (1.711)	0.826
Both limbs merged						
ResNet-18	0.845 (0.665)	−0.027 (1.075)	0.959	1.441 (0.882)	1.064 (1.316)	0.878
ResNet-34	0.906 (0.727)	0.004 (1.162)	0.953	1.554 (1.087)	1.205 (1.469)	0.846
ResNet-50	1.099 (0.903)	−0.025 (1.422)	0.929	1.249 (1.112)	−0.008 (1.677)	0.880
ResNet-101	1.053 (0.845)	−0.351 (1.304)	0.936	1.445 (1.249)	−0.162 (1.909)	0.844
ResNet-152	1.006 (0.810)	−0.095 (1.289)	0.941	1.547 (1.306)	0.309 (2.007)	0.825

Standard deviations for MAE and bias are provided in parentheses

on HKA angle assessments and the categorization of varus deformity stages. The robustness and applicability of the model were further evaluated through extended temporal validation to ensure a thorough and unbiased assessment [29, 30]. Notably, our framework processes each lower limb independently, enhancing accuracy and streamlining the analysis process. This method allows the model to focus on a single limb at a time, leading to more precise measurements without introducing unnecessary complexity.

The results showed a consistent performance across various ResNet architectures. Minimal biases of − 0.025 and − 0.008° were observed in the original and extended temporal validation cohorts, with a MAE of 1.099° and 1.249°, respectively. This bias does not significantly affect the management of patients with OA. In addition, the SDE was maintained within acceptable limits, at 1.422° in the retrospective cohort and 1.677° in the temporal validation cohort. The model also proved effective in classifying varus deformity stages, achieving a weighted F1-score of 88%. Remarkably, our method demonstrated a notable advancement in efficiency, being 126.7 times faster than the traditional gold standard method employed by orthopedic experts for HKA angle measurements. This substantial increase in speed significantly enhances the efficiency of the clinical workflow, particularly in time-constrained settings.

Orthopedics, with its demanding environmental requirements, has witnessed a revolutionary advancement in the field of automation efforts in measuring scales for various diagnoses [31] and treatment planning [32]. The HKA angle, a critical measurement in patient diagnosis, has been the primary focus of such attempts. Traditionally, an accuracy level of 3° in measuring the HKA angle has been deemed adequate for clinical applications in orthopedic surgery, as this reflects the precision of osteotomy correction [33]. From this perspective, previous anatomical landmark-based studies have retrospectively demonstrated a viable level of precision [14, 17–22]. Our research also showed that 99.1% (1189/1200 lower limbs) of cases in the retrospective analysis and 91% (91/100 lower limbs) in the temporal validation were within 3° of error, even without using landmarks. This ability to detect and interpret the intricate details of skeletal shapes in lower limb radiographs could have been a key factor in achieving a high level of accuracy in our measurements. The effectiveness of ResNet lies in its deep convolutional layers that are extremely adept at extracting hierarchical features from complex images [34]. These layers allow ResNet to progressively learn finer details from radiographs [35], starting from the basic edges and textures to more sophisticated skeletal patterns. In addition, ResNet's residual connections facilitate the training of these deeper networks by addressing

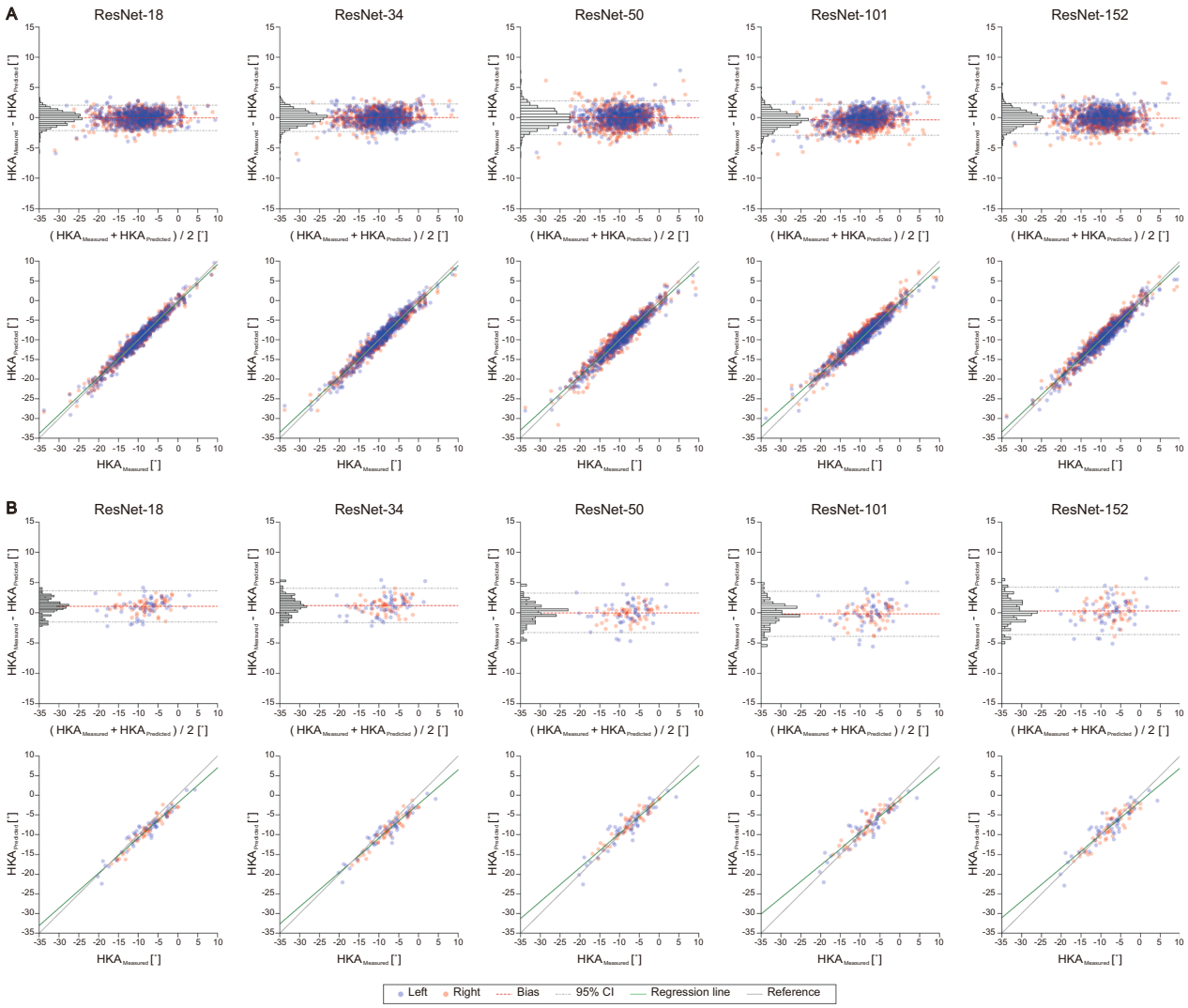


Fig. 2 Bland–Altman and scatter plot evaluations for HKA angle predictions, comparing the agreement and correlation between predicted and actual angles across five ResNet variants. **A** Cross-validation performance. **B** Temporal validation performance

the vanishing gradient problem [24] ensuring that even subtle features are not lost during the learning process.

Beyond accuracy, a crucial aspect of clinical tools is their effect on workflow efficiency, particularly in real-world practice [36]. In developing and fine-tuning models, utilizing routine clinical data is one of the most critical factors for successful AI deployment [37]. Relying on specialized data that requires extra effort, such as segmentation [14, 17] or detailed annotation [18–22], becomes highly impractical in busy orthopedic settings where time and resources are already limited. Models that demand additional manual inputs are unlikely to be adopted in environments where efficiency is paramount. The direct use of HKA angle data for model training integrates smoothly with current orthopedic practices, eliminating the need for additional time-consuming measurement or analysis steps in model development. Accordingly, a crucial aspect of our study lies in the direct use of HKA angle data for model training, seamlessly aligning with existing orthopedic workflows and eliminating the need for deviation from current standard procedures.

The repetitive nature of tasks in Picture Archiving and Communication Systems (PACS) environments can lead to reduced efficiency and an increased likelihood of errors [38, 39]. This issue may be further exacerbated in the demanding context of orthopedic departments [40]. Accordingly, there is a clear demand for technologies

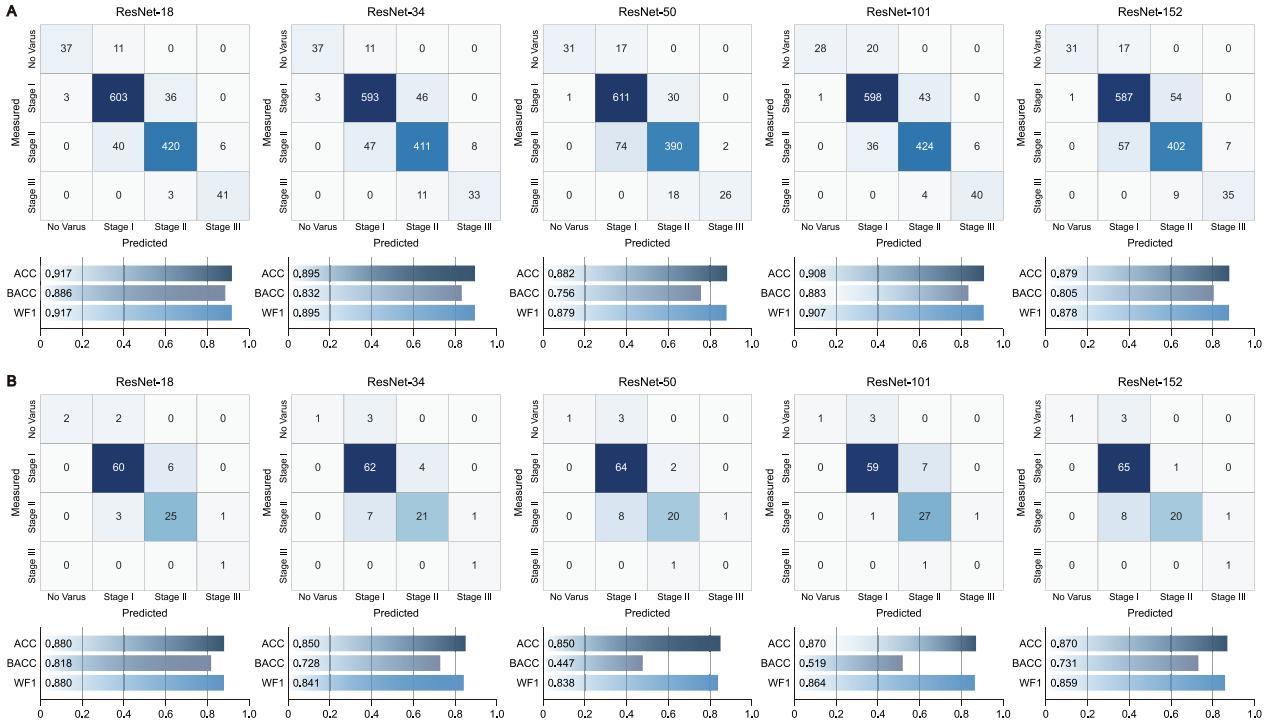


Fig. 3 Confusion matrices and performance metrics for varus deformity classification across both limbs using five ResNet variants. **A** Cross-validation results. **B** Temporal validation results. Performance metrics include accuracy (ACC), balanced accuracy (BACC) and weighted F1-score (WF1)

that support routine repetitive tasks in orthopedics. AI may serve as a strategic tool for addressing the underlying contributors to surgeon fatigue and mitigate the incidence of clinical errors [41]. Although previous studies have introduced models for automating HKA angle measurement, they have not demonstrated the extent of time efficiency improvements in actual clinical applications [14, 17–22]. The proposed model demonstrated its efficiency by processing the HKA angle measurement of 100 lower limbs, saving approximately 46 min compared to the time taken by an experienced orthopedic surgeon (Fig. 4). This signifies a shift towards more automated and less labor-intensive methods in orthopedics. Furthermore, the potential for seamless integration into clinical data systems by automated measurement of the HKA angle through AI during imaging transactions in PACS holds promise for optimized orthopedic practices.

The interpretative capability of our model offers insight into how AI can augment traditional clinical measurements. In standard practice, the HKA angle is measured by drawing a line from the head of the femur to the intercondylar notch and another from the tibia's interspinous point to the center of the ankle joint [42]. However, our bilateral model was trained to predict the HKA angle directly, resulting in a broader, more dynamic attention mechanism. Rather than focusing solely on specific bone landmarks, the model demonstrated an ability to distribute attention across the entire lower limb. This was particularly evident in more severe cases of varus deformity, where the model expanded its focus to include spatial changes around the bones, not just the skeletal structure itself. Such attention patterns indicate that the model is capturing a more holistic view of the limb's anatomical changes, suggesting an alternative approach to interpreting HKA angles that extends beyond traditional methods focused on fixed landmarks. By considering both the bones and their surrounding environment, the model may provide deeper insights into the progression of deformities like Genu Varum, offering a more comprehensive assessment of the lower limb's overall architecture.

Despite the promising outcomes, our study has limitations. The relatively limited sample size of the extended temporal validation cohort constrains the generalizability of our findings. In particular, the study population consisted predominantly of female patients, with the majority of varus deformity cases falling into Stage I and Stage II categories. Larger, multi-center and prospective studies involving diverse populations are imperative to validate and extend our conclusions. Future research warrants

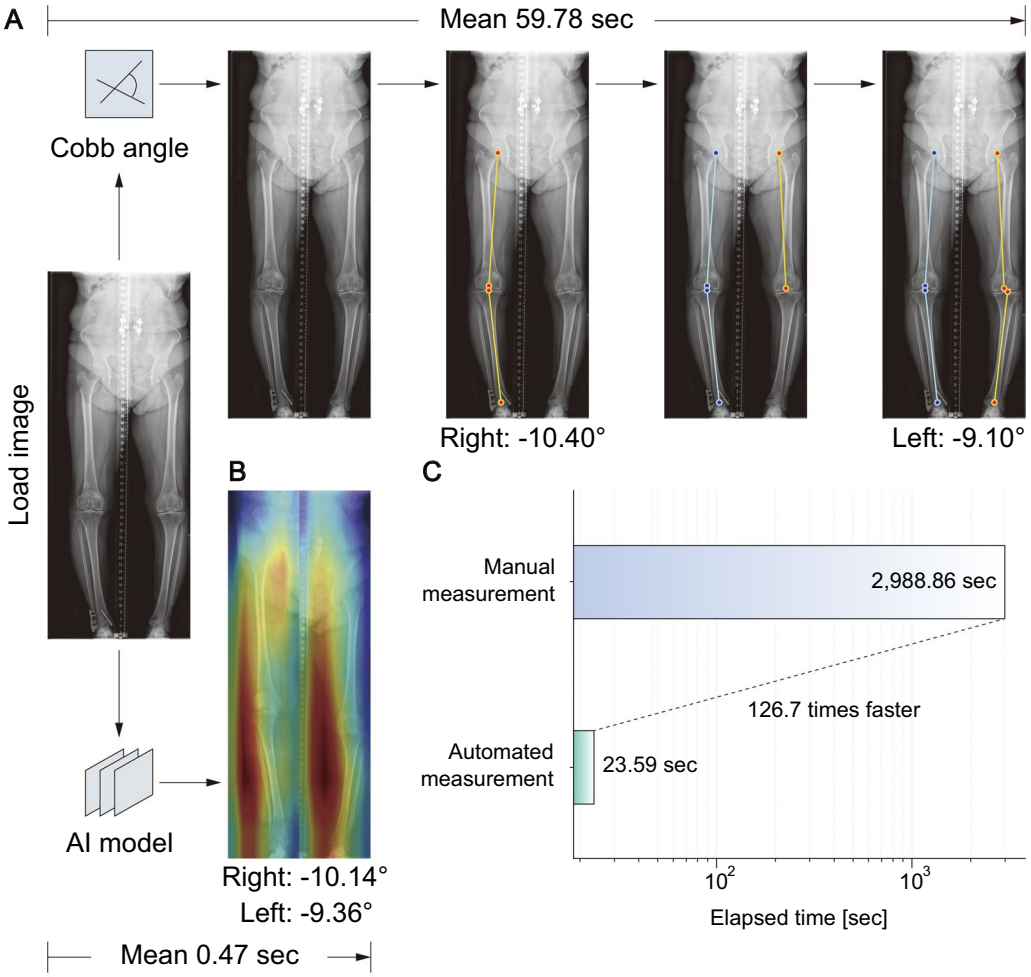


Fig. 4 Comparison of the time required for manual and automated measurements of the HKA angle. Procedure of manual HKA angle assessment (**A**), AI model-based assessment of HKA angle (**B**), and comparison of elapsed time between the manual and automated measurements (**C**)

further expansion of the model's scope with various etiologies and enhances its adaptability to diverse radiographic methods and resolutions. However, we did not directly assess whether the ResNet-based model could improve the accuracy of manual measurements due to the lack of a standard ground-truth dataset. Investigating inter-rater agreement in future studies could help determine whether the model reduces subjective bias among orthopedic surgeons, offering additional evidence of its reliability. Furthermore, longitudinal assessments of patient outcomes post-intervention, particularly evaluating the practical efficiency and reliability of AI-assisted clinical decision-support systems, constitute the basis for future investigation.

Conclusions

In conclusion, our study marks a notable advancement in merging AI with orthopedic practice, specifically in the accurate assessment of the HKA angle in knee OA patients. By employing the ResNet architecture coupled

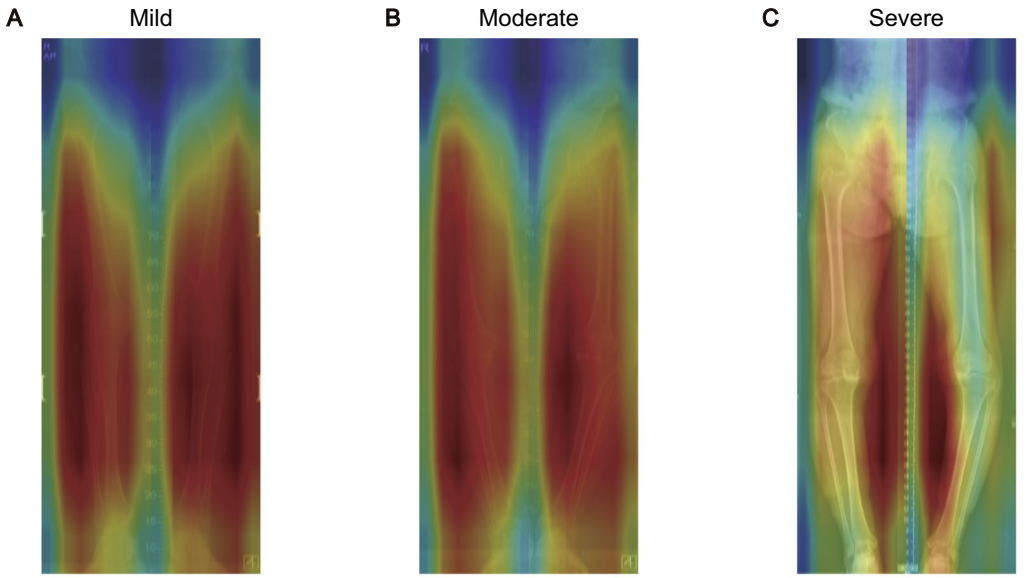


Fig. 5 Representative Grad-CAM visualizations of the ensemble model's HKA angle predictions for varus deformity stages in stage I (**A**), stage II (**B**), and a mixed case with stage II and III (**C**), with each panel showing the right limb on the left and the left limb on the right. 'G' stands for the ground truth HKA angle values, and 'P' for the model's predictions

with our innovative approach, we achieved clinically relevant accuracy without the burden of extensive landmark annotation. This approach holds the potential to substantially enhance patient care, particularly in resource-constrained environments.

Abbreviations

AI	Artificial Intelligence
ACC	Accuracy
ASA	American Society of Anesthesiologists
BA	Bland-Altman
BACC	Balanced Accuracy
BMI	Body Mass Index
Grad-CAM	Gradient-weighted Class Activation Mapping
HKA	Hip-Knee-Ankle
IQR	Interquartile Range
KL	Kellgren-Lawrence
MAE	Mean Absolute Error
OA	Osteoarthritis
PACS	Picture Archiving and Communication Systems
ReLU	Rectified Linear Unit
ResNet	Residual Network
ROI	Region of Interest
SDE	Standard Deviation of Error
WF1	Weighted F1-score

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13018-024-05265-y.

Supplementary file 1.

Acknowledgements
Not applicable.

Authors' contributions
YTK conceptualized the study, conducted the deep learning analysis, created the visualizations, drafted the original manuscript, and contributed to the review and editing. BSH contributed to conceptualization, data curation, drafting the original manuscript, and participated in the review and editing. JBK, JKS, JH-Hong, YS, JH-Han, SD, and JSC critically reviewed and edited the manuscript. JKB supported conceptualization, data curation, and supervision, and contributed to the manuscript's review and editing. All authors read and approved the final manuscript.

Funding
This work was supported by a National Research Foundation of Korea (NRF) Grant funded by the Korean government (Ministry of Science and ICT, MSIT) (RS-2023-00212574, RS-2024-00415812) and the MSIT, Korea, under the ITRC (Information Technology Research Center) support program (IITP-2024-RS-2022-00156225) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Availability of data and materials
The datasets generated and/or analysed during the current study are not publicly available due to institutional regulation but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Ethical approval was obtained from the Institutional Research and Ethics Board of Seoul Medical Center (SEOUL 2023-12-002-001), Republic of Korea. The study was conducted in accordance with the ethical standards of the 1964 Declaration of Helsinki. No patient-identifiable information was included

in the manuscript, and all data were anonymized to ensure confidentiality. As this was a retrospective study, informed consent was waived.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Radiology, Massachusetts General Hospital, Harvard Medical School, 125 Nashua Street, Boston, MA 02114, USA. [2]Department of Biomedical Sciences, Korea University College of Medicine, 73 Goryeodae-Ro, Seongbuk-Gu, Seoul 02841, Republic of Korea. [3]Department of Orthopedic Surgery, Seoul Medical Center, 156 Shinnae-Ro, Jungnang-Gu, Seoul 02053, Republic of Korea. [4]Department of Neurology, Korea University Anam Hospital, Korea University College of Medicine, 73 Goryeodae-Ro, Seongbuk-Gu, Seoul 02841, Republic of Korea. [5]Department of Biomedical Informatics, Korea University College of Medicine, 73 Goryeodae-Ro, Seongbuk-Gu, Seoul 02841, Republic of Korea. [6]Department of Electronic Engineering, Hanyang University, 222 Wangsimni-Ro, Seongdong-Gu, Seoul 04763, Republic of Korea. [7]Department of Computer Science and Engineering, Dongguk University, 30 Pildong-Ro 1-Gil, Jung-Gu, Seoul 04620, Republic of Korea. [8]Department of Brain and Cognitive Engineering, Korea University, 145 Anam-Ro, Seongbuk-Gu, Seoul 02841, Republic of Korea. [9]KU-KIST Graduate School of Converging Science and Technology, Korea University, 145 Anam-Ro, Seongbuk-Gu, Seoul 02841, Republic of Korea. [10]Kempner Institute, Harvard University, 150 Western Ave, Boston, MA 02134, USA. [11]Department of Anesthesiology and Pain Medicine, College of Medicine, Ewha Womans University, 260 Gonghang-Daero, Gangseo-Gu, Seoul 07804, Republic of Korea.

Received: 20 September 2024 Accepted: 12 November 2024
Published online: 20 November 2024

References

1. Vitaloni M, Botto-van Bemden A, Sciortino Contreras RM, Scotton D, Bibas M, Quintero M, et al. Global management of patients with knee osteoarthritis begins with quality of life assessment: a systematic review. BMC Musculoskelet Disord. 2019;20:1–12.
2. Brouwer G, Tol AV, Bergink A, Belo J, Bernsen R, Reijman M, et al. Association between valgus and varus alignment and the development and progression of radiographic osteoarthritis of the knee. Arthritis Rheum. 2007;56:1204–11.
3. Sharma L, Song J, Felson DT, Cahue S, Shamiyeh E. Dunlop DD The role of knee alignment in disease progression and functional decline in knee osteoarthritis. JAMA. 2001;286:188–95.
4. Felson DT. Osteoarthritis as a disease of mechanics. Osteoarthritis Cartilage. 2013;21:10–5.
5. Moreland JR, Bassett L. Hanker G Radiographic analysis of the axial alignment of the lower extremity. JBJS. 1987;69:745–9.
6. Chao E, Neluheni E, Hsu R. Paley D Biomechanics of malalignment. Orthop Clin North Am. 1994;25:379–86.
7. Cooke D, Scudamore A, Li J, Wyss U, Bryant T. Costigan P Axial lower-limb alignment: comparison of knee geometry in normal volunteers and osteoarthritis patients. Osteoarthritis Cartilage. 1997;5:39–47.
8. Tanamas S, Hanna FS, Cicuttini FM, Wluka AE, Berry P. Urquhart DM Does knee malalignment increase the risk of development and progression of knee osteoarthritis? A systematic review. Arthritis Care Res (Hoboken). 2009;61:459–67.
9. Ai S, Wu H, Wang L. Yan M HKA angle—a reliable planning parameter for high tibial osteotomy: a theoretical analysis using standing whole-leg radiographs. J Knee Surg. 2020;35:054–60.
10. Gao F, Ma J, Sun W, Guo W, Li Z. Wang W Radiographic assessment of knee–ankle alignment after total knee arthroplasty for varus and valgus knee osteoarthritis. Knee. 2017;24:107–15.
11. Jiang X, Xie K, Han X, Ai S, Wu H, Wang L, et al. HKA angle—a reliable planning parameter for high tibial osteotomy: a theoretical analysis using standing whole-leg radiographs. J Knee Surg. 2022;35:054–60.
12. Minoda Y. Alignment techniques in total knee arthroplasty. J Joint Surg Res. 2023;1:108–16.
13. Zampogna B, Vasta S, Amendola A, Marbach BU-E, Gao Y, Papalia R, et al. Assessing lower limb alignment: comparison of standard knee xray vs long leg view. Iowa Orthop J. 2015;35:49.
14. 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;44:53–62.
15. Côté MJ, Smith MA. Forecasting the demand for radiology services. Health Syst (Basingstoke). 2018;7:79–88.
16. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18:500–10.
17. Nguyen TP, Chae D-S, Park S-J, Kang K-Y, Lee W-S, Yoon J. Intelligent analysis of coronal alignment in lower limbs based on radiographic image with convolutional neural network. Comput Biol Med. 2020;120: 103732.
18. 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 Surg Sports Traumatol Arthrosc. 2023;31:1388–97.
19. 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.
20. Gielis WP, Rayegan H, Arbabi V, Ahmadi Brooghani SY, Lindner C, Cootes TF, et al. Predicting the mechanical hip–knee–ankle angle accurately from standard knee radiographs: a cross-validation experiment in 100 patients. Acta Orthop. 2020;91:732–7.
21. Yang J, Ren P, Xin P, Wang Y, Ma Y, Liu W, et al. Automatic measurement of lower limb alignment in portable devices based on deep learning for knee osteoarthritis. J Orthop Surg Res. 2024;19:232.
22. Tanner IL, Ye K, Moore MS, Rechenmacher AJ, Ramirez MM, George SZ, et al. Developing a computer vision model to automate quantitative measurement of hip-knee-ankle angle in total hip and knee arthroplasty patients. J Arthroplasty. 2024;39:2225–33.
23. Kwee TC. Kwee RM Workload of diagnostic radiologists in the foreseeable future based on recent scientific advances: growth expectations and role of artificial intelligence. Insights Imaging. 2021;12:1–12.
24. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
25. Wu Z, Shen C, Van Den Hengel A. Wider or deeper: Revisiting the resnet model for visual recognition. Pattern Recognit. 2019;90:119–33.
26. Bagaria V, Kulkarni RV, Sadigale OS, Sahu D, Parvizi J, Thienpont E. Varus knee deformity classification based on degree and extra-or intra-articular location of coronal deformity and osteoarthritis grade. JBJS Rev. 2021;9: e20.
27. Doyle DJ, Hendrix JM, Garmon EH American society of anesthesiologists classification. 2017.
28. Chattopadhay A, Sarkar A, Howlader P, Balasubramanian VN, editors. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. 2018 IEEE winter conference on applications of computer vision (WACV); 2018: IEEE.
29. de Hond AA, Shah VB, Kant IM, Van Calster B, Steyerberg EW. Hernandez-Boussard T Perspectives on validation of clinical predictive algorithms. NPJ Digit Med. 2023;6:86.
30. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J. 2021;14:49–58.
31. Federer SJ, Jones GG. Artificial intelligence in orthopaedics: a scoping review. PLoS ONE. 2021;16: e0260471.
32. St Mart J-P, Goh EL, Liew I, Shah Z. Sinha J Artificial intelligence in orthopaedics surgery: transforming technological innovation in patient care and surgical training. Postgrad Med J. 2023;99:687–94.
33. Odenbring S, Berggren A-M, Peil L Roentgenographic assessment of the hip-knee-ankle axis in medial gonarthrosis. A study of reproducibility. *Clin Orthop Relat Res*. 1993;195–6.
34. Xu W, Fu Y-L, Zhu D ResNet and its application to medical image processing: research progress and challenges. *Comput Methods Programs Biomed*. 2023;107660.

35. Wang Y, Li S, Zhao B, Zhang J, Yang Y, Li B. A ResNet-based approach for accurate radiographic diagnosis of knee osteoarthritis. CAAI Trans Intell Technol. 2022;7:512–21.
36. Zayas-Cabán T, Okubo TH. Posnack S Priorities to accelerate workflow automation in health care. J Am Med Inform Assoc. 2023;30:195–201.
37. Markowetz F. All models are wrong and yours are useless: making clinical prediction models impactful for patients. NPJ Precis Oncol. 2024;8:54.
38. Boiselle PM, Levine D, Horwich PJ, Barbaras L, Siegal D, Shillue K, et al. Repetitive stress symptoms in radiology: prevalence and response to ergonomic interventions. J Am Coll Radiol. 2008;5:919–23.
39. Alexander R, Waite S, Bruno MA, Krupinski EA, Berlin L, Macknik S, et al. Mandating limits on workload, duty, and speed in radiology. Radiology. 2022;304:274–82.
40. Epstein S, Sparer EH, Tran BN, Ruan QZ, Dennerlein JT, Singhal D, et al. Prevalence of work-related musculoskeletal disorders among surgeons and interventionalists: a systematic review and meta-analysis. JAMA Surg. 2018;153:e174947-e.
41. Myers TG, Ramkumar PN, Ricciardi BF, Urish KL, Kipper J, Ketonis C. Artificial intelligence and orthopaedics: an introduction for clinicians. J Bone Joint Surg Am. 2020;102:830.
42. Luís NM, Varatojo R. Radiological assessment of lower limb alignment. EFORT Open Rev. 2021;6:487–94.

Publisher's Note

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