Automated analysis of knee joint alignment using detailed angular values in long leg radiographs based on deep learning

Hong Seon Lee

The Graduate School
Yonsei University

Department of Medicine

Automated analysis of knee joint alignment using detailed angular values in long leg radiographs based on deep learning

A Dissertation Submitted
to the Department of Medicine
and the Graduate School of Yonsei University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Medical Science

Hong Seon Lee

December 2024

This certifies that the Dissertation of Hong Seon Lee is approved

Thesis Supervisor	Sungjun, Kim
Thesis Committee Member	Sung-Hwan Kim
Thesis Committee Member	Dosik Hwang
Thesis Committee Member	Young Han Lee
Thesis Committee Member	Hwiyoung Kim

The Graduate School
Yonsei University
December 2024

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Sang Chul Hwang from the Department of Integrated Medicine, Yonsei University College of Medicine, for his invaluable assistance in the development of deep learning algorithms and the application of these algorithms for measurements.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Automated analysis of knee joint alignment using detailed angular values in long leg radiographs based on deep learning

Malalignment in the lower limb structure can occur due to various causes, and accurately evaluating limb alignment is essential, especially when correction is necessary. To address this need, we developed an automated system to assess lower limb alignment by quantifying key parameters: mechanical tibiofemoral angle (mTFA), mechanical lateral distal femoral angle (mLDFA), medial proximal tibial angle (MPTA), and joint line convergence angle (JLCA) from full-length weight-bearing radiographs. In this retrospective study, we analyzed 404 radiographs for algorithm development and testing, with external validation using 30 radiographs from another hospital. The segmentation algorithm's performance was compared to manual segmentation using the Dice Similarity Coefficient (DSC), and agreement of alignment parameters was assessed using the Intraclass Correlation Coefficient (ICC). The time taken to measure the four alignment parameters was recorded. The algorithm demonstrated excellent agreement with human-annotated segmentation (89–97% similarity), with good to very good alignment parameter agreement (ICC: 0.7213–0.9865). The automated method was 3.44 times faster than manual measurement. In a larger clinical application, we conducted a retrospective cohort study with 17,080 long-leg radiographs from 34,160 legs taken between 2010 and 2023 at a tertiary hospital in South Korea. The deep learning model automatically analyzed mTFA, mLDFA, MPTA, and JLCA. Patients were stratified by age (≤55 and >55 years) and sex, and linear regression analysis was performed to evaluate knee alignment trends over time. While mTFA remained stable across the population ($p > 0.05$), in those under 55 years of age, there was a significant decrease (right: β=-0.05, 95% CI: -0.08 to -0.02, $p = 0.01$; left: β=-0.05, 95% CI: -0.07 to -0.03, $p = 0.01$), with younger females showing a steeper decline (right: β=-0.08, 95% CI: -0.14 to -0.03, $p = 0.01$; left: β=-0.06, 95% CI: -0.10 to -0.01, $p = 0.01$). CPAK classification trends showed an increase in classification 3 in females under 55 (right: β=0.01, 95% CI: 0.00 to 0.01, $p = 0.01$; left: β=0.01, 95% CI: 0.00 to 0.01, $p = 0.01$). As a result, classification 3 increased from 17.9% to 26.4% on the right side and from 21.4% to 28.0% on the left side between 2010 and 2023. This study reveals distinct shifts in knee alignment, particularly among younger females, with trends indicating a shift toward valgus alignment. Further research is needed to understand the long-term clinical impact of these trends and develop targeted strategies for preserving knee health across the lifespan.

Key words : Knee joint alignment, Radiograph, Deep learning, Osteoarthritis

1. INTRODUCTION

1.1. Algorithm development and validation

Malalignment of lower limb structure occurs due to congenital, developmental, or post-traumatic causes, leading to knee joint malalignment, causing joint degeneration, abnormal gait, pain, and asymmetric overloading of articular compartments[1]. Tibiofemoral malalignment is considered a risk factor for osteoarthritis (OA), with genu varum and genu valgum increasing the risk of medial and lateral OA progression, respectively. The severity of malalignment is directly related to knee joint function deterioration[2,3].

Accurate evaluation of limb alignment is necessary for situations where malalignment needs correction, such as limb realignment surgery or joint replacement surgery[4]. Full-length weight-bearing radiographs of the lower extremities in an upright posture are commonly used in clinical settings to evaluate lower limb alignment, joint orientation, and leg length discrepancy[5]. During imaging, the patient stands upright with bare feet together, fully extended knees, and forward-facing patellae to prevent rotation of the lower limbs.

Whole limb alignment is evaluated based on the mechanical tibiofemoral angle (mTFA), mechanical lateral distal femoral angle (mLDFA), medial proximal tibial angle (MPTA), and joint line convergence angle (JLCA). Accurately measuring these parameters is crucial to identify the main source of deformity. Micicoi et al. Reported a physiologic value of 85.8° for mLDFA and 85.6° for MPTA, indicating a 4° valgus and 4° varus of femoral and tibial bone morphology, respectively[6]. In patients with OA, varus deformity (hip-knee-ankle angle < 177°) is caused by distal femoral wear (mLDFA = 89°), tibial varus obliquity (MPTA = 87°), and lateral joint line opening (JLCA = 3°)[7]. However, compensating for any measurement abnormalities can achieve a balanced limb position. Therefore, measuring each parameter is vital for comprehending alignment abnormalities and identifying their primary cause[7]. However, this may be a laborious and time-consuming task for radiologists.

Therefore, there is a clinical need for a standardised and reproducible automatic analysis tool that measures lower limb alignment using full-length weight-bearing radiographs[8,9]. Moreover, developing a technical framework based on artificial intelligence applicable in clinical settings is potentially feasible[9]. Our objective was to create, train, and validate an automated support system to evaluate lower limb alignment by quantifying mTFA, mLDFA, MPTA, and JLCA on full-length weight-bearing radiographs of both lower extremities (Fig. 1).

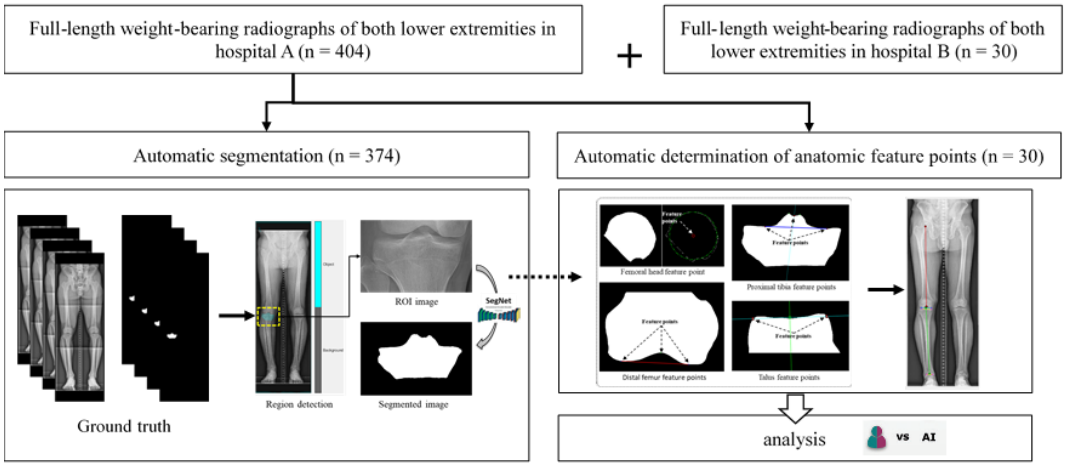

Fig. 1 Overview of patient enrollment, algorithm development, and analysis

1.2. Clinical application for large scale data

Knee joint alignment plays a pivotal role in the development and progression of degenerative OA[3,10-12]. Valgus alignment is associated with lateral compartment OA, while varus alignment is linked to medial compartment OA[13-15]. Various factors, including age, gender, body mass index (BMI), previous trauma, and conditions such as rheumatoid arthritis, significantly influence knee alignment and, consequently, the patterns of OA[16-19]. Additionally, knee joint biomechanics and muscle strength contribute to alignment and stress distribution, further impacting the onset and progression of OA[11,20-22]. Understanding these factors is crucial for early diagnosis, management, and prevention, as it can help mitigate OA's detrimental effects on joint function and quality of life[23-25].

In South Korea, the past decades have seen profound lifestyle changes, marked by increases in overnutrition and physical inactivity[26,27]. This has led to a rise in both adult and childhood obesity, which likely affects knee joint alignment not only after growth but also during critical developmental phases[26,27]. Research indicates that obesity is linked to greater valgus knee alignment in pubertal children, and higher BMI is correlated with increased variability in knee alignment, particularly among girls[28]. These alterations in knee alignment during key growth stages may influence future OA patterns. Moreover, the management of lateral and medial OA differs significantly in terms of treatment protocols, surgical complexity, and outcomes, highlighting the importance of early identification and management of knee alignment variations[29,30].

Recently, the Coronal Plane Alignment of the Knee (CPAK) classification system was developed to apply constitutional knee alignment patterns clinically. By evaluating constitutional knee alignment in non-arthritic or minimally arthritic conditions, CPAK classifies knee alignment based on the sum and difference of the mLDFA and MPTA[31-33]. This system is particularly useful for understanding individual variations in knee alignment, especially in the context of OA and knee replacement surgery.

Advances in deep learning have made it possible to automate knee alignment measurements, greatly enhancing the feasibility of large-scale studies[34-36]. These technologies provide precise, consistent, and efficient analysis of detailed knee alignment angles including mTFA, mLDFA, MPTA, and

JLCA from radiographic data, overcoming the limitations of traditional manual methods, which are often time-consuming and prone to variability. In our previous research, we developed an advanced algorithm for the automated measurement of detailed knee alignment angles, enabling more accurate assessments[37].

Given the lifestyle changes and the increasing prevalence of obesity, we hypothesize that there is a significant trend in knee alignment patterns over time. Furthermore, we propose that there will be notable differences in knee alignment between younger and older age groups, as well as between males and females, with younger individuals experiencing lifestyle changes during critical growth phases.

Therefore, the objective of this study is to analyze trends in knee alignment using deep learning-based automated measurements of detailed angular values. The focus will be on identifying age- and gender-specific differences in knee alignment patterns, particularly across time. By examining these longitudinal trends, we aim to provide insights into how changes in alignment patterns may influence future OA development and help guide early interventions in at-risk populations.

2. METHOD AND MATERIALS

2.1. Algorithm development and validation

2.1.1. Study participants and radiograph data

This retrospective study received approval from the institutional review boards of a tertiary hospital (A) (Yonsei University Gangnam Severance Hospital, Institutional Review Board, No 3-2020-0127) and a military hospital (B) (Armed Forces Capital Hospital, Institutional Review Board, 2023-02-002), and informed consent was waived because the data used in this retrospective study were fully de-identified to protect patient confidentiality. All methods were performed in accordance with the ethical standards of Helsinki Declaration. A total of 404 full-length weight-bearing radiographs of both lower extremities from 404 patients (mean age: 44.3 years, 188 men, 186 women) from hospital A were used to develop and test the algorithm. An external test set of 30 consecutive radiographs from 30 men (mean age: 30.2 years) from hospital B was included. The patients underwent long-leg radiography at the two institutions between March 2015 to January 2019 and between August 2022 and September 2022, respectively. Patients from hospital A with K-L grade 4, intra-articular fracture, deformity due to previous trauma, and knee arthroplasty, and those < 19 years were excluded (n = 426) (Fig. 2). The long-leg radiographs were obtained using two imaging acquisition systems and covered the whole lower limbs from the hips to the ankles under single anteroposterior exposure. Philips DigitalDiagnost (Philips, Best, The Netherlands) and Carestream DRX-Evolution (Carestream Health, Rochester, NY, USA) were used in hospitals A and B, respectively.

Next, 30 radiographs out of the 404 were used for clinical verification of the algorithm's anatomical feature points, chosen through stratified random splitting based on the K-L grade. The remaining 374 radiographs were used to develop and validate the automatic segmentation

algorithm. Cases with overlapping bones (n = 12), bones containing metal (n = 33), and unclear bone outline (n = 32) were excluded to ensure methodological consistency[38]. For the algorithm's development, 342 radiographs for the femoral head, 352 for the distal femur, 341 for the proximal tibia, 362 for the distal tibia, and 367 for the talus were used. The collected radiographs were divided into the training set (80%), validation set (10%), and test set (10%) (Fig. 2).

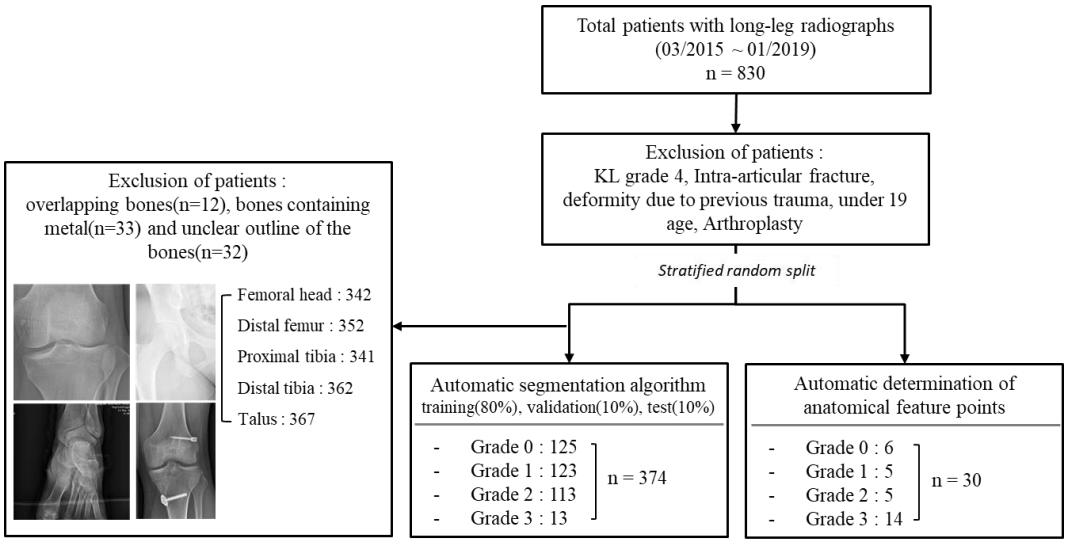

Fig. 2 Patient flowchart for algorithm development and clinical verification.

2.1.2. Manual segmentation

The femoral heads, knee joints, and ankle joints were manually segmented using Adobe Photoshop CC 2018 (Adobe Systems Inc., San Jose, CA, USA) to create masks, which served as the reference for comparison. A radiology technician, supervised by an experienced radiologist, labeled the masks.

2.1.3. Manual reference measurements

Lower limb alignment was evaluated based on the following anatomic feature points (Fig. 3): (1) the centre of the femoral head, (2) the centre of the femoral intercondylar notch, (3) centres of the medial and lateral tibial spines, (4) two most distal points of the medial and lateral femoral condyles, (5) two most proximal points of the medial and lateral tibial plateaus, and (6) mid-malleolar point (centre of the ankle).

The mechanical axis of the femur was defined as a line drawn from the centre of the femoral head to the centre of the femoral intercondylar notch. The mechanical and anatomical axes of the tibia were defined as the line connecting the centre of the tibial spines and the centre of the ankle. The distal femoral articular axis was defined by the line connecting the most distal points of the medial and lateral femoral condyles. The proximal tibial articular axis was defined as the line connecting the two most proximal points of the tibial plateaus. Four alignment parameters (mTFA,

mLDFA, MPTA, and JLCA) were measured using the aforementioned eight feature points and four lines.

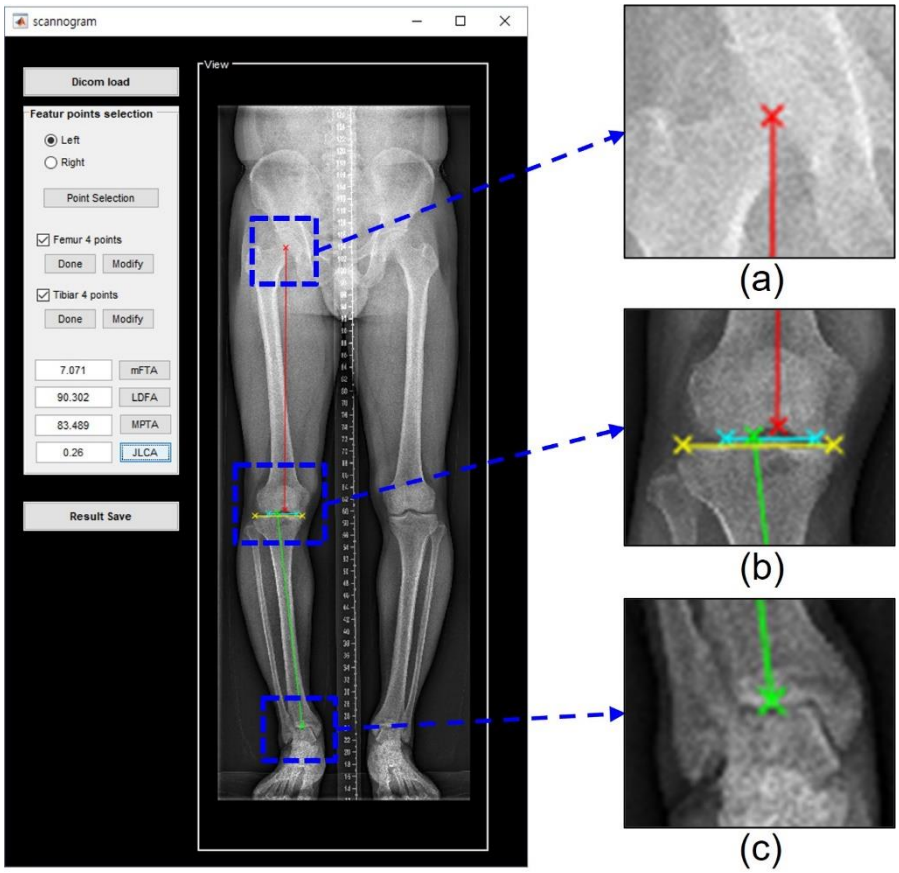

Fig. 3 Alignment parameter measurement tool by manually selecting 8 feature points and 4 lines (a) Femoral head centre, (b) centre of femoral intercondylar notch, centre of the tibial spines, two most distal points of medial and lateral femoral condyles, and two most proximal points of medial and lateral tibial plateaus and (c) mid-malleolar point.

We developed a tool for measuring alignment parameters using MATLAB's Graphical User Interface Development Environment (GUIDE) to create a Graphical User Interface (GUI) in MATLAB. This tool allows the designation of landmarks for angle measurement and calculates the angles using these points (Fig. 3). To assess the intraobserver and interobserver agreement of the measured values between the readers and algorithm, an orthopedic fellow measured the angles of the clinical verification data set (n = 30) twice, with a 2-week interval between the measurement sessions. Another radiology fellow measured the angles once. Regarding the test from the external institution, a fellowship-trained radiologist measured the angles twice. The time taken to load the data and measure the four alignment parameters using the tool was recorded.

2.1.4. Automated segmentation algorithm

Representative models of Semantic Segmentation include FCN (Fully Convolutional networks), U-Net, and SegNet. FCN needs to learn deconvolution when upsampling, so it needs weight parameters for learning, but in SegNet, this process is omitted, so the learning parameters are reduced. U-Net skip combines during the decoding process, but U-Net transfers the entire feature map information of the same layer from the encoder to the decoder and concats it. Therefore, it is heavier than SegNet, which only selects and uses some features of Max pooling indices.

For this reason, in this study, the outline of each bone was automatically segmented using SegNet. The SegNet architecture consists of a down sampling (encoding) path and a corresponding upsampling (decoding) path, followed by a final pixel-wise classification layer. In the encoder path, there are 13 convolutional layers that match the first 13 convolutional layers in the VGG16 network. Each encoder layer has a corresponding decoder layer; therefore, the decoder network also has 13 convolutional layers. The output of the final decoder layer is fed into a multi-class softmax classifier to produce class probabilities for each pixel independently[39].

To automatically segment the contours of each bone, we implemented a two-step segmentation algorithm (Fig. 4). In the initial step, we identified the region of interest containing the target bone, and subsequently, in the second step, we delineated the boundaries of the target bones within the identified image region. During the first step, the images were resized to 311 x 932 pixels, and the intensities were scaled to the range [0,1]. In the subsequent step, the images were resized to different pixel dimensions based on the size of each bone (Femoral head: 470 x 470, Distal femur: 740 x 540, Proximal tibia: 720 x 470, Distal tibia: 470 x 430, Talus: 370 x 220), and intensities were scaled to the range [0,1]. We used SGD (Stochastic Gradient Descent) Momentum as the solver to train the deep learning network. The maximum number of Epochs to train the SegNet model was set to 120, and a mini-batch with 4 observations was used for each iteration. And the momentum value was set to 0.9 and the learning rate to 1 x e^-2. The SegNet model was trained using the training and validation data and implemented with MATLAB R2018b on a GeForce GTX 1080Ti graphics processing unit.

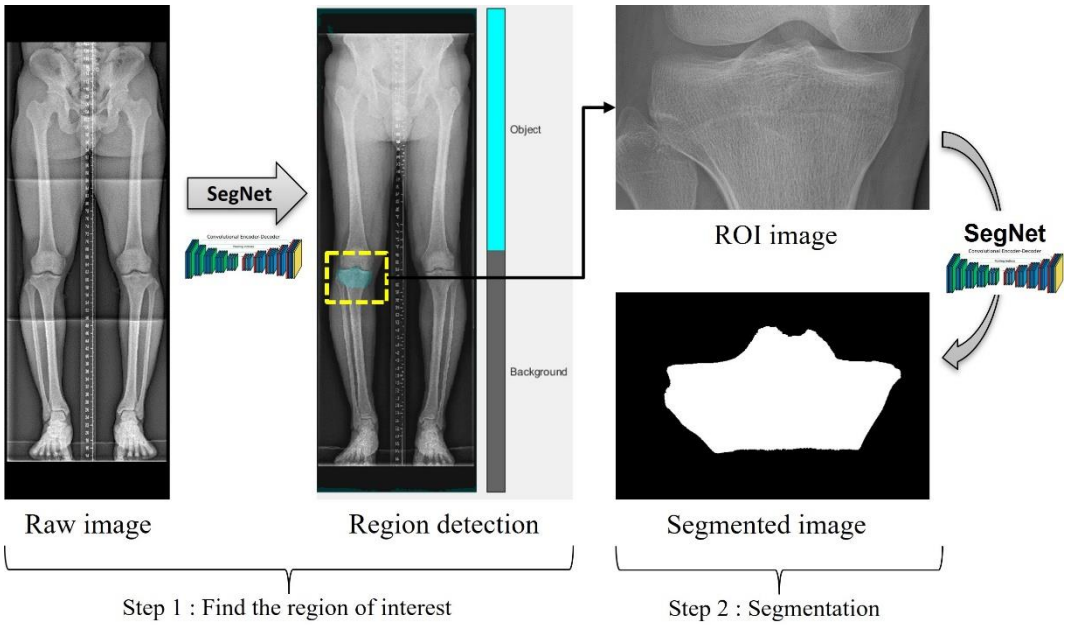

Fig. 4 Flowchart of the automatic segmentation algorithm.
The first step was performed on raw images. The second step was performed based on the region of interest (ROI) image created by cropping the raw image.

2.1.5. Automatic determination of anatomic feature points

The mechanical axes for lower limb alignment were automatically determined based on the segmentation masks (Fig. 5). The computer-aided automatic measurement times from image data loading to determining the four alignment parameters were recorded.

[The femoral head anatomic feature point]
A circle was fit to the segmentation outline of the femoral head to determine its centre.

[The distal femur anatomic feature point and the distal femur surface line]
The region comprising the distal femur surface line and the centroid of the segmentation outline was identified as the distal femur anatomic feature point. The distal femur surface line was determined by minimisig the distance between the bottom line of the bounding box and the segmentation outline, resulting in two points. The highest point within the defined area, encompassing the outline, was designated as the distal femur anatomic feature point.

[The proximal tibia anatomic feature points]
Two peaks were detected from the segmentation outline, and the midpoint between these two points was extracted to determine the proximal tibia anatomic feature points. Next, an orthogonal line was created by connecting the two points and the midpoint, and the position along the segmented outline

where the distance between the orthogonal line and the outline was minimised was defined as the proximal tibia anatomic feature point.

[The proximal tibia surface line]
The convex hull[40,41] and bounding box of the segmentation outline were calculated. To determine the feature points, candidate points were identified by selecting points above the centroid of the segmentation outline within the region defined by the convex hull. Next, the proximal tibia surface line was defined by identifying the two points closest to the upper corner points of the bounding box from the candidate points.

[Distal tibia anatomic feature points]
Two talus feature points were defined by applying the same method of defining the proximal tibia surface line. Next, an orthogonal line was constructed by connecting the midpoint of the two talus feature points, and the position where the distance between the orthogonal line and the segmented outline of the distal tibia was minimum was defined as the distal tibia anatomic feature point.

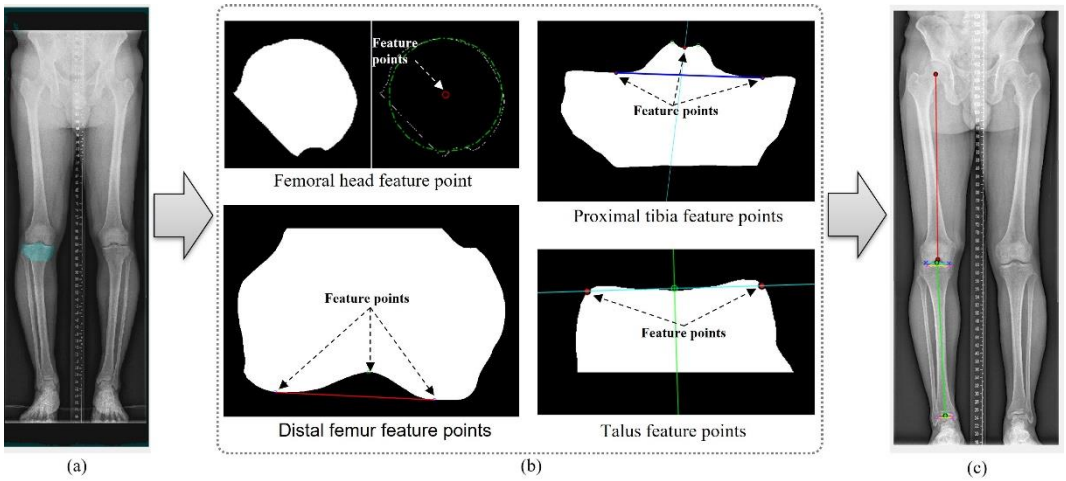

Fig. 5 Flowchart of automatic determination algorithm of anatomic feature points
(a) Segmented images. (b) Anatomic feature points automatically determined based on segmented images. (c) The mechanical axes for the lower limb alignment.

2.1.6. Statistical Analysis

We implemented global accuracy, mean accuracy, mean intersection over union (IoU), weighted IoU, and the dice similarity coefficient (DSC) to evaluate the segmentation algorithm's performance, which compares the similarity of the automated segmentation mask with the human-annotated segmentation mask. As a representative measurement, we considered a DSC ≥ 0.7 as indicative of excellent agreement between two segmented regions, following previous studies[42,43].

We confirmed normality in each group for mTFA, mLDFA, MPTA, and JLCA using the Shapiro–Wilk test and performed group-wise comparisons of their means and standard deviations (SDs) using repeated measures analysis of variance (ANOVA) between three groups or paired t-tests between two groups.

We evaluated the intraobserver and interobserver agreement of mTFA, mLDFA, MPTA, and JLCA between the readers and algorithm using the intraclass correlation coefficient (ICC) to assess measurement reproducibility. Altman considered an ICC of 0.81–1 as very good, 0.61–0.8 as good, and 0.41–0.6 as moderate. In the interobserver agreement test, we used the result of the second session for comparison when a reader performed two measurements.

Statistical significance was set at $p < 0.05$. We performed all statistical analyses using Medcalc software (version 20.114; MedCalc Software Ltd., Ostend, Belgium).

2.2. Clinical application for large scale data

2.2.1. Study Design and Population

This retrospective study received approval from the institutional review board of a tertiary hospital (Yonsei University Gangnam Severance Hospital, Institutional Review Board, No 3-2024-0133). Patient consent was waived due to the retrospective nature of the study. The study utilized long-leg radiographs obtained from a large cohort of patients who underwent long-leg radiography at a single tertiary hospital in South Korea between 2010 and 2023. The inclusion criteria required participants to have available long-leg radiographs (Fig. 6). Exclusion criteria included individuals under 18 years of age (n=3901), follow-up studies (n=14450), previous surgery or severe deformity (n=3972), no BMI record (n=170), and algorithm errors (n=1890). Surgeries included total knee arthroplasty (TKA), ACL reconstruction, correctional osteotomy, and internal fixation for previous fractures. Severe deformities included limb amputation, hereditary multiple exostoses, polio sequelae, and deformities from previous fractures. Algorithm errors occurred due to issues with the DICOM file or malfunctioning algorithms, with data unavailable for both legs in 491 cases, the right leg in 375 cases, and the left leg in 1024 cases. Due to unexplained DICOM file corruption that occurred frequently in 2019, we were only able to analyze fewer than 200 images from that year. As a result, data from 2019 were excluded from the analysis. Ultimately, 17,080 radiographs of 34,160 legs taken between 2010 and 2023 were included in the study.

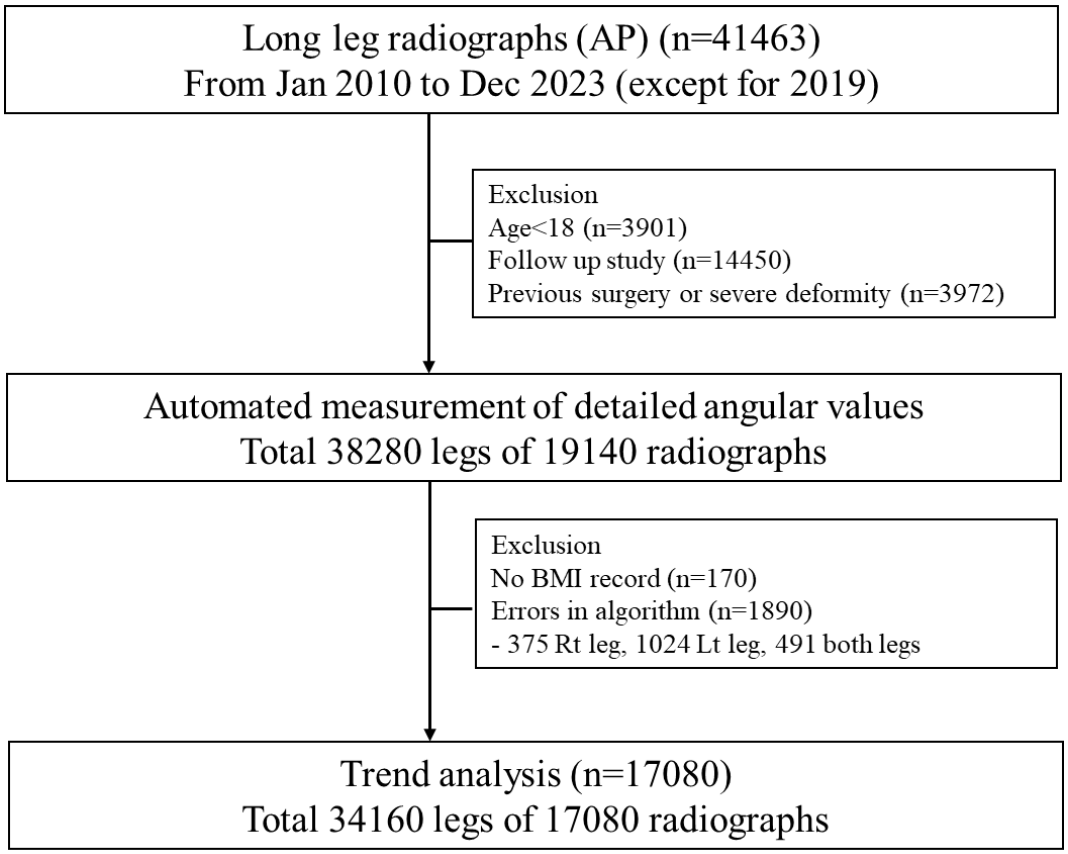

Fig. 6 Patient flowchart

2.2.2. Radiographic Data Acquisition

Long-leg radiographs were obtained using standardized protocols to ensure consistent and accurate measurement of knee alignment. The radiographs included weight-bearing anteroposterior (AP) views, capturing the entire lower limb from the hip to the ankle. The radiographs were obtained using an imaging acquisition system that covered the entire lower limbs from the hips to the ankles under single anteroposterior exposure (Philips DigitalDiagnost, Philips, Best, The Netherlands). This imaging technique allows for accurate assessment of the detailed angular values related to knee joint alignment.

2.2.3. Deep Learning Model

A deep learning model was developed and trained to automatically detect and measure angular values associated with knee joint alignment from the radiographs. The measurements extracted by the model included the MTFA, LDFA, MPTA, and JLCA. The model was trained using a subset of radiographs that had been manually annotated by radiology technicians under the supervision of

radiologists. Data augmentation techniques were applied to enhance the robustness of the model and prevent overfitting.

2.2.4. Data Analysis

The study examined changes in MTFA, LDFA, MPTA, and JLCA for both the right and left legs across age groups (18-29, 30-39, 40-49, 50-59, 60-69, and 70+ years). Additionally, BMI was categorized into underweight, normal weight, overweight, and obese based on the world health organization (WHO) BMI classification for Asian populations: underweight (BMI <18.5), normal weight (BMI 18.5 to <23), overweight (BMI 23 to <25), and obese (BMI ≥25).
To analyze temporal trends, the study evaluated changes in MTFA, LDFA, MPTA, and JLCA for the entire cohort, as well as stratified by sex and age groups (≤55 years and >55 years). The 55-year age threshold was chosen because the median age of the entire cohort was 56 years. Furthermore, to precisely analyze alignment patterns, the study used the CPAK classification to assess the proportion of each classification within the entire cohort, females under 55, males under 55, females over 55, and males over 55. Linear regression analysis was conducted to evaluate trends over time and across BMI categories.

2.2.5. Statistical Analysis

All statistical analyses were performed using Python (version 3.12.3) developed by the Python Software Foundation (Wilmington, DE). Jupyter Notebook (version 6.5.4), a product of Project Jupyter (San Diego, CA), was employed for interactive coding and documentation. A p-value of <0.05 was considered statistically significant. Since age, sex, and angular values, excluding BMI, did not follow a normal distribution, non-parametric statistical methods were applied. Descriptive statistics included the median, interquartile range, and proportions for categorical variables. To analyze trends in measurements across different years or groups, linear regression was performed.

3. RESULTS

3.1. Algorithm development and validation

3.1.1. Segmentation performance

As indicated in Table 1, we assessed the segmentation performance using metrics including global accuracy, mean accuracy, mean IoU, weighted IoU, and DSC to thoroughly analyze the results obtained in segmentation problems. The segmentation algorithm demonstrated excellent agreement with the human-annotated segmentation for all the anatomical regions, with an average DSC of 93% for the femoral head, 95% for the distal femur, 95% for the proximal tibia, 89% for the distal tibia, and 97% for the talus. Other values ranged from 96% to 98% for the femoral head, 95% to 96% for

the distal femur, 96% to 98% for the proximal tibia, 93% to 96% for the distal tibia, and 94% to 98% for the talus.

Table 1 Segmentation accuracy measured using various evaluation metrics

	Global accuracy	**Mean accuracy**	**Mean IoU**	**Weighted IoU**	**Mean DSC**
Femoral head	0.98	0.98	0.96	0.96	0.93
Distal femur	0.98	0.98	0.95	0.96	0.95
Proximal tibia	0.98	0.98	0.96	0.96	0.95
Distal tibia	0.98	0.98	0.93	0.96	0.89
Talus	0.98	0.98	0.94	0.96	0.97

IoU, Intersection over union; DSC, Dice similarity coefficients.

3.1.2. Assessment of measurement comparisons to algorithms

Measurements of the lower limb alignment did not significantly differ between the readers and algorithm in the internal institution test set, as shown in Table 2 (mTFA: Reader 1, 181.82°±3.39; Reader 2, 181.78°±3.33; Algorithm, 181.79°±3.48; mLDFA: Reader 1, 87.51°±1.96; Reader 2, 87.71°±1.8; Algorithm, 87.73°±1.86; MPTA: Reader 1, 86.76°±3.19; Reader 2, 86.41°±3.08; Algorithm, 86.99°±3.29; JLCA: Reader 1, 1.79°±1.43; Reader 2, 1.73°±1.07; Algorithm, 1.67°±1.41) (all $p > 0.05$). The average angle differences between the readers and algorithm in the internal and external institutions are shown in Fig. 7. The mean differences in mTFA, mLDFA, MPTA, and JLCA between the two readers were 0.04°±0.30, 0.20°±0.88, 0.35°±1.10, and 0.36°±1.08, respectively. The mean differences between Reader 1 and the algorithm and Reader 2 and the algorithm were 0.03°±0.79 and 0.01°±0.83 for mTFA, 0.23°±0.60 and 0.03°±0.84 for mLFDA, 0.23°±1.27 and 0.59°±1.66 for MPTA, and 0.12°±0.68 and 0.24°±1.17 for JLCA, respectively. based on a mechanical tibiofemoral angle. The intraobserver correlations (ICC range, 0.9836–0.9991) between sessions 1 and 2 for Reader 1 and the interobserver correlations (ICC range, 0.7751–0.9981) between Readers 1 and 2 were good to very good, as shown in Table 3. The ICC scores of angles measured by Reader 1, Reader 2, and the algorithm indicated good to very good agreement, as shown in Table 4 (ICC ranges: 0.9848–0.9865 for mTFA, 0.9443–0.9746 for mLDFA, 0.9273–0.9604 for MPTA, and 0.7213–0.9393 for JLCA).

Table 2 Details of manual and automatic measurements of lower limb alignment

	Group	Mean	SD	P-value
mTFA	Reader1	181.82	3.39	0.998
	Reader 2	181.78	3.33	
	Algorithm	181.79	3.48	
mLDFA	Reader1	87.51	1.96	0.765
	Reader 2	87.71	1.8	
	Algorithm	87.73	1.86	
MPTA	Reader1	86.76	3.19	0.598
	Reader 2	86.41	3.08	
	Algorithm	86.99	3.29	
JLCA	Reader1	1.79	1.43	0.315
	Reader 2	1.43	1.07	
	Algorithm	1.67	1.41	

SD, standard deviation; mTFA, mechanical tibiofemoral angle; mLDFA, mechanical lateral distal femoral articular angle; MPTA, medial proximal tibial angle; JLCA, joint line convergence angle

Table 3 Details of intraobserver and interobserver agreement of lower limb alignment between readers

	ICC		95% CI		P-value	
	R1 vs R1 (intraobserver)	R1 vs R2 (interobserver)	R1 vs R1 (intraobserver)	R1 vs R2 (interobserver)	R1 vs R1 (intraobserver)	R1 vs R2 (interobserver)
mTFA	0.9991	0.9981	0.9984 ~ 0.9995	0.9968 ~ 0.9988	<0.0001	<0.0001
mLDFA	0.9900	0.9420	0.9831 ~ 0.9940	0.9030 ~ 0.9654	<0.0001	<0.0001
MPTA	0.9949	0.9683	0.9914 ~ 0.9970	0.9470 ~ 0.9811	<0.0001	<0.0001
JLCA	0.9836	0.7751	0.9725 ~ 0.9902	0.6234 ~ 0.8656	<0.0001	<0.0001

ICC, in-class correlation coefficient; CI, confidence interval; R1, reader 1; R2, reader 2; mTFA, mechanical tibiofemoral angle; mLDFA, mechanical lateral distal femoral articular angle; MPTA, medial proximal tibial angle; JLCA, joint line convergence angle

Table 4 Details of intraobserver and interobserver agreement of lower limb alignment between the readers and algorithm

	ICC		**95% CI**		**P-value**	
	R1 vs Al	R2 vs Al	R1 vs Al	R2 vs Al	R1 vs Al	R2 vs Al
mTFA	0.9865	0.9848	0.9773 ~ 0.9919	0.9745 ~ 0.9909	<0.0001	<0.0001
mLDFA	0.9746	0.9443	0.9574 ~ 0.9848	0.9068 ~ 0.9668	<0.0001	<0.0001
MPTA	0.9604	0.9273	0.9336 ~ 0.9763	0.8782 ~ 0.9566	<0.0001	<0.0001
JLCA	0.9393	0.7213	0.8984 ~ 0.9638	0.5333 ~ 0.8335	<0.0001	<0.0001

ICC, in-class correlation coefficient; CI, confidence interval; R1, reader 1; R2, reader 2; mTFA, mechanical tibiofemoral angle; mLDFA, mechanical lateral distal femoral articular angle; MPTA, medial proximal tibial angle; JLCA, joint line convergence angle; R1, reader 1; R2, reader 2; AI, artificial intelligence

3.1.3. Measurement Times

The time taken for the manual measurements of lower limb alignment from the internal institution test set (n = 30) by the two readers averaged 86 min (average of 172 s/patient). In contrast, the time taken for computer-aided automatic measurements was 25 min, including the loading time for training data (average of 50 s/patient), which was 3.44 times faster than that for manual measurement. The processing time taken after data loading averaged 20 s/patient.

3.1.4. External validation of the algorithm

External validation included 30 long-leg radiographs from consecutive patients at an external hospital. Intraobserver correlations (ICC ranges: 0.9393-0.9979) between sessions 1 and 2 for Reader 3 and the interobserver correlations (ICC ranges, 0.7126–0.9695) between the manual and automatic measurements were good to very good, as shown in Table 5. There was no statistically significant difference between the measurements of the lower limb alignment by the reader and algorithm in the external validation, as shown in Table 6 (mTFA: Reader 3, 181.37°±2.26; Algorithm, 181.26°±2.56; mLDFA: Reader 3, 86.92°±2.03; Algorithm, 86.80°±2.01; MPTA: Reader, 86.20°±1.65; Algorithm, 86.55°±1.66; JLCA: Reader 3, 0.40°±1.74; Algorithm 0.49°±1.58) (all $p > 0.05$). The average angle differences between the reader and algorithm are shown in Fig. 7.

Table 5 Details of intraobserver and interobserver agreement of lower limb alignment between the manual and automatic measurement on external validation

	ICC		95% CI		P-value	
	R3 vs R3 (intraobserver)	R3 vs AI (interobserver)	R3 vs R3 (intraobserver)	R3 vs AI (interobserver)	R3 vs R3 (intraobserver)	R3 vs AI (interobserver)
mTFA	0.9979	0.9695	0.9965 ~ 0.9988	0.9489 ~ 0.9818	<0.0001	<0.0001
mLDFA	0.9830	0.9218	0.9716 ~ 0.9899	0.8692 ~ 0.9533	<0.0001	<0.0001
MPTA	0.9748	0.9199	0.9579 ~ 0.9850	0.8658 ~ 0.9521	<0.0001	<0.0001
JLCA	0.9353	0.7126	0.8916 ~ 0.9613	0.5189 ~ 0.8283	<0.0001	<0.0001

* ICC, in-class correlation coefficient; CI, confidence interval; R3, reader 3; mTFA, mechanical tibiofemoral angle; mLDFA, mechanical lateral distal femoral articular angle; MPTA, medial proximal tibial angle; JLCA, joint line convergence angle; AI, artificial intelligence

Fig. 7 Comparative evaluation of reader and algorithm
mTFA, mechanical tibiofemoral angle; mLDFA, mechanical lateral distal femoral angle; MPTFA, medial proximal tibial angle; JLCA, joint line convergence angle

3.2. Clinical application for large scale data

3.2.1. Patient Demographic Characteristics

The patient demographic characteristics are summarized in Table 6. A total of 17,080 images were analyzed across different years, with the percentage of female participants ranging from 57.2% to 69.3% per year. The overall percentage of female images was 62.0% (10,590/17,080). The median age of the entire population was 56.0 years (Q1-Q3: 42.0-66.0 years), with a female median age of 59.0 years (Q1-Q3: 50.0-68.0 years) and a male median age of 48.0 years (Q1-Q3: 29.0-61.0 years). The median age showed a general increase over the years, with the female population consistently older than the male population. The median BMI across all participants was 24.0 (Q1-Q3: 21.8-26.3), with a female median BMI of 23.6 (Q1-Q3: 21.4-26.0) and a male median BMI of 24.5 (Q1-Q3: 22.6-26.6). The median BMI remained relatively stable across the years, with males generally having a higher median BMI compared to females.

Table 6. Patient Demographic Characteristics

Year	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
No. of images	345	469	883	1115	1142	1786	1897
% of female*	65.2 (225/345)	61.6 (289/469)	65.6 (579/883)	69.3 (773/1115)	65.7 (750/1142)	58.7 (1048/1786)	57.2 (1086/1897)
Age (total)	52.0 (28.0-63.0)	52.0 (36.0-63.0)	56.0 (43.0-65.0)	58.0 (50.0-67.0)	57.0 (50.0-64.0)	52.0 (33.0-60.0)	52.0 (35.0-62.0)
Age (female)	56.0 (47.0-65.0)	57.0 (46.0-66.0)	58.0 (49.0-66.0)	60.0 (52.0-68.0)	59.0 (53.0-66.0)	55.0 (43.0-62.0)	56.0 (47.0-64.0)
Age (male)	28.0 (20.0-49.0)	41.0 (24.8-53.3)	50.0 (27.0-62.3)	54.0 (37.3-63.0)	53.0 (36.0-60.0)	40.0 (25.0-57.0)	42.0 (25.0-56.0)
BMI (total)	23.7 (22.2-25.4)	23.8 (22.2-25.7)	23.9 (21.9-25.9)	24.1 (22.0-26.3)	24.1 (22.1-26.2)	23.4 (21.4-26.0)	23.6 (21.6-26.4)
BMI (female)	23.9 (22.0-25.4)	23.6 (21.8-25.8)	23.7 (21.8-26.1)	24.0 (21.8-26.3)	23.8 (22.0-25.9)	22.9 (20.7-25.6)	23.4 (21.1-26.2)
BMI (male)	23.6 (22.2-25.3)	24.2 (22.9-25.5)	24.3 (22.2-25.7)	24.2 (22.5-26.2)	24.6 (22.3-26.3)	24.2 (22.3-26.4)	24.2 (22.1-26.5)
Year	**2017**	**2018**	**2020**	**2021**	**2022**	**2023**	**Overall**
No. of images	1452	1330	1391	1605	1791	1874	17080
% of female*	59.8 (869/1452)	59.5 (791/1330)	59.7 (830/1391)	63.7 (1023/1605)	62.6 (1121/1791)	64.4 (1206/1874)	62.0 (10590/17080)
Age (total)	53.0 (37.0-63.0)	55.0 (39.0-63.0)	58.0 (43.0-67.0)	61.0 (47.0-69.0)	60.0 (46.0-69.0)	61.0 (48.0-69.0)	56.0 (42.0-66.0)
Age (female)	57.0 (47.0-66.0)	58.0 (48.0-65.0)	61.0 (51.0-69.0)	63.0 (52.0-71.0)	64.0 (52.0-71.0)	63.0 (53.0-71.0)	59.0 (50.0-68.0)
Age (male)	43.0 (25.0-56.0)	47.0 (29.0-59.0)	50.0 (33.0-64.0)	56.0 (38.0-66.0)	53.0 (35.0-65.8)	53.0 (38.0-66.0)	48.0 (29.0-61.0)
BMI (total)	23.6 (21.6-26.2)	24.3 (21.8-26.5)	24.0 (22.0-26.6)	24.3 (22.0-26.4)	24.0 (21.6-26.4)	24.0 (21.9-26.1)	24.0 (21.8-26.3)
BMI (female)	23.4 (21.3-26.0)	23.7 (21.5-26.3)	23.8 (21.4-26.4)	23.9 (21.3-26.0)	23.4 (21.2-26.1)	23.6 (21.3-25.6)	23.6 (21.4-26.0)
BMI (male)	24.3 (22.4-26.5)	24.6 (22.6-26.7)	24.5 (22.8-26.8)	25.0 (23.0-27.0)	24.7 (22.9-26.8)	24.9 (22.9-26.8)	24.5 (22.6-26.6)

The data represents percentages (%), with the numbers in parentheses indicating the count of female images versus the total number of images.
Other data points represent median values, with the numbers in parentheses indicating the interquartile range (Q1-Q3). BMI, body mass index.

3.2.2. Correlation of detailed angular values with age groups and BMI classification

The correlation of detailed angular values with age groups and BMI classification is depicted in Fig. 8 and 9. The data (Fig. 8) show that as age increases, there is a general trend of increasing MTFA on both sides (right, β=0.73, 95% CI: 0.44 to 1.01; left, β=0.73, 95% CI: 0.63 to 0.84), LDFA (right, β=0.37, 95% CI: 0.33 to 0.39; left, β=0.39, 95% CI: 0.28 to 0.50), and JLCA (right, β=0.30, 95% CI: 0.05 to 0.54; left, β=0.23, 95% CI: 0.02 to 0.44) across all patients (all $p < 0.05$). As age increased, there was a trend toward a decrease in MPTA (right, β=-0.09, 95% CI: -0.14 to -0.04; left, β=-0.08, 95% CI: -0.17 to 0.01), but this was only statistically significant on the right side ($p < 0.05$), with no significance on the left side ($p = 0.067$). MTFA and LDFA exhibited similar trends in both males and females. However, in females, the decrease in MPTA (right, β=-0.19, 95% CI: -0.26 to -0.11; left, β=-0.19, 95% CI: -0.28 to -0.11) was statistically significant on both sides ($p < 0.05$), whereas in males, the increase in JLCA (right, β=0.09, 95% CI: -0.03 to 0.22; left, β=0.07, 95% CI: -0.05 to 0.19) was not statistically significant on either side (right, $p = 0.107$; left, $p = 0.159$).

As BMI increased, there was a general trend of increasing MTFA (right, β=0.47, 95% CI: -0.32 to 1.25; left, β=0.32, 95% CI: -0.32 to 0.96), LDFA (right, β=0.31, 95% CI: -0.21 to 0.83; left, β=0.22, 95% CI: -0.34 to 0.78), and JLCA (right, β=0.10, 95% CI: -0.09 to 0.29; left, β=0.11, 95% CI: -0.06 to 0.28), while MPTA (right, β=-0.06, 95% CI: -0.25 to 0.13; left, β=0.03, 95% CI: -0.14 to 0.20) remained relatively stable (Fig. 9). However, there was a slight decrease in Lt MTFA (β=-0.09, 95% CI: -0.19 to 0.00) in males. Except for this decrease in Lt MTFA in males ($p < 0.05$), none of the other trend lines were statistically significant (all $p > 0.05$). Thus, in this analysis, there were no consistent or statistically significant changes in detailed angular values on either side based on BMI.

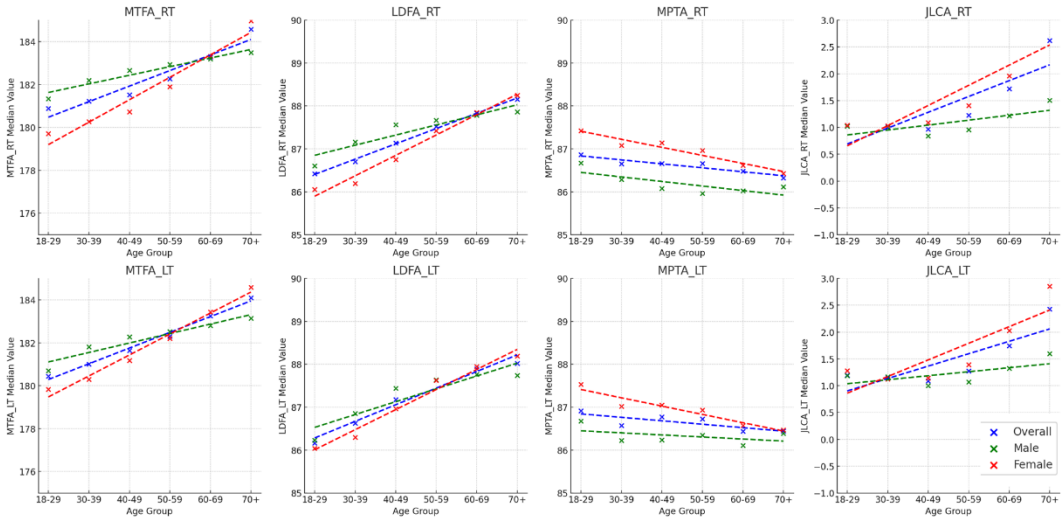

Fig. 8 Correlation of detailed angular values with age groups
The plot shows the median values of detailed angular values for each age group. The blue line represents the total population, the green line represents males, and the red line represents females. The dotted lines indicate the trend of the median values for each group.
MTFA, medial tibiofemoral angle; LDFA, lateral distal femoral angle; MPTA, medial proximal tibial angle; JLCA, joint line congruence angle; RT, right; LT, left

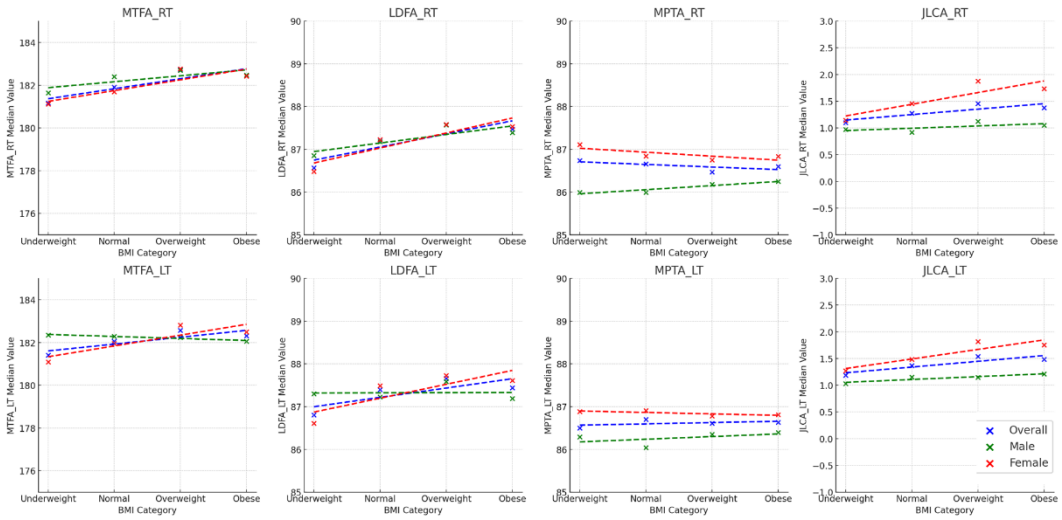

Fig. 9 Correlation of detailed angular values with BMI classification
The plot shows the median values of detailed angular values for each BMI group. The blue line represents the total population, the green line represents males, and the red line represents females. The dotted lines indicate the trend of the median values for each group.
BMI, body mass index; MTFA, medial tibiofemoral angle; LDFA, lateral distal femoral angle; MPTA, medial proximal tibial angle; JLCA, joint line congruence angle; RT, right; LT, left

3.2.3. Annual trends of detailed angular values of knee alignment

The annual trends of detailed angular values of knee alignment are depicted in Fig. 10. In the overall population, all females, all males, and the population over 55 years of age, MTFA, MPTA, and JLCA remained consistent each year without significant changes (all $p > 0.05$). However, there was an exception with the bilateral LDFA in the overall population (right, β=0.02, 95% CI: 0.00 to 0.04; left, β=0.03, 95% CI: 0.01 to 0.06) and in all males (right, β=0.04, 95% CI: 0.01 to 0.06; left, β=0.04, 95% CI: 0.01 to 0.08), both of which showed an annual increase (all $p < 0.05$). However, this increase did not lead to a significant rise in MTFA (right, β=-0.01, 95% CI: -0.05 to 0.04; left, β=-0.01, 95% CI: -0.04 to 0.02) (both $p > 0.05$).
In contrast, for individuals under 55 years of age, MTFA (right, β=-0.05, 95% CI: -0.08 to -0.02; left, β=-0.05, 95% CI: -0.07 to -0.03) showed a decreasing trend on both sides, and this trend was more pronounced in females under 55 (right, β=-0.08, 95% CI: -0.14 to -0.03; left, β=-0.06, 95% CI: -0.10 to -0.01), where the decline was steeper compared to the overall population under 55 (all $p < 0.05$).

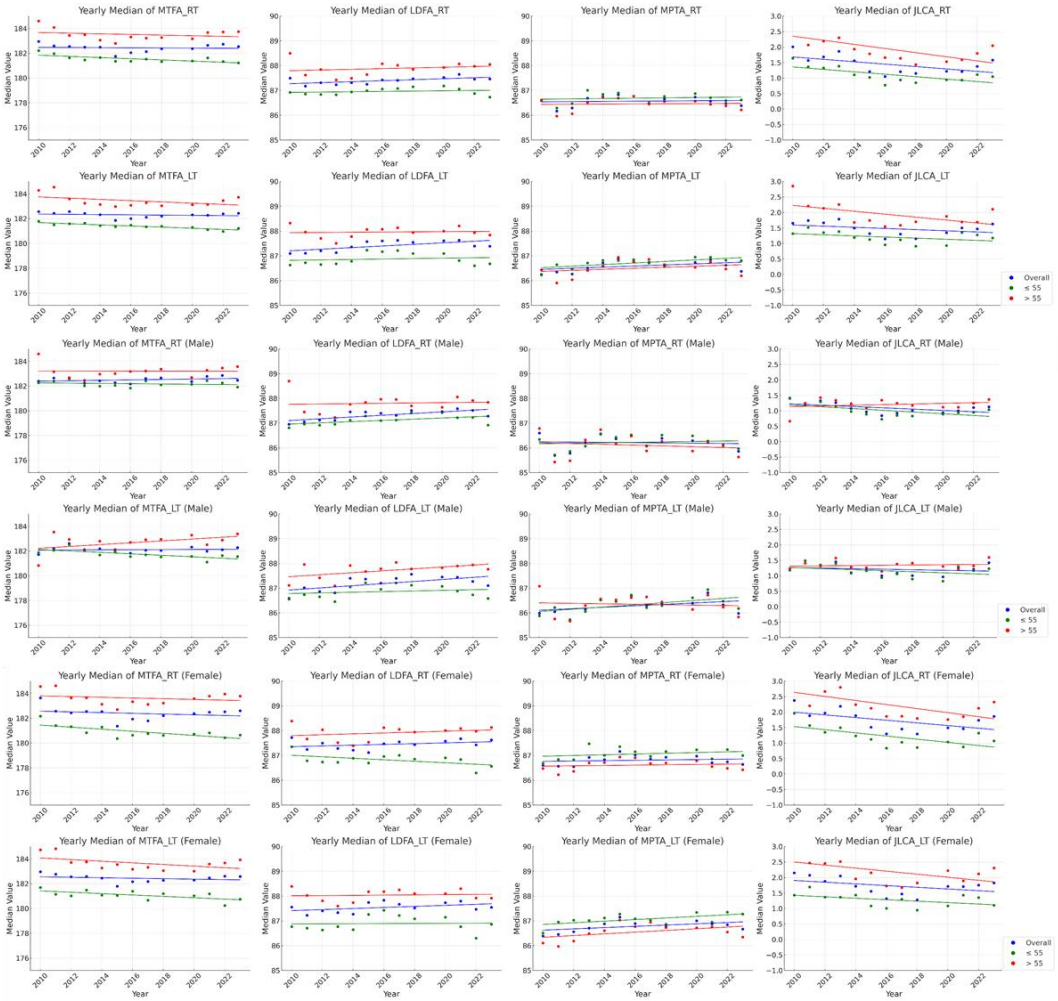

Fig. 10 Annual trends of detailed angular values of knee alignment
Each point represents the median values of detailed angular values for each year. The top 8 graphs show the values for the total population, the middle 8 graphs represent males, and the bottom 8 graphs represent females. The blue line indicates the total population, the green line represents individuals under 55 years old, and the red line represents individuals over 55 years old.
MTFA: medial tibiofemoral angle; LDFA: lateral distal femoral angle; MPTA: medial proximal tibial angle; JLCA: joint line congruence angle; RT: right; LT: left.

3.2.4. Annual trends of CPAK classification

The annual trends of CPAK classification are depicted in Fig. 11. Except for females under 55 years of age, there were no statistically significant trends in the proportions of each classification (all $p > 0.05$). However, in females under 55, there was a significant decrease in classification 1 on

both sides (right, β=-0.01, 95% CI: -0.01 to 0.00; left, β=0.00, 95% CI: -0.01 to 0.00) and a significant increase in classification 3 (right, β=0.01, 95% CI: 0.00 to 0.01; left, β=0.01, 95% CI: 0.00 to 0.01).

As a result, for females under 55, the proportion of classification 1 decreased from 24.4% in 2010 to 16.5% in 2023 on the right side, and from 22.6% to 14.1% on the left side. In contrast, the proportion of classification 3 increased from 17.9% in 2010 to 26.4% in 2023 on the right side, and from 21.4% to 28.0% on the left side.

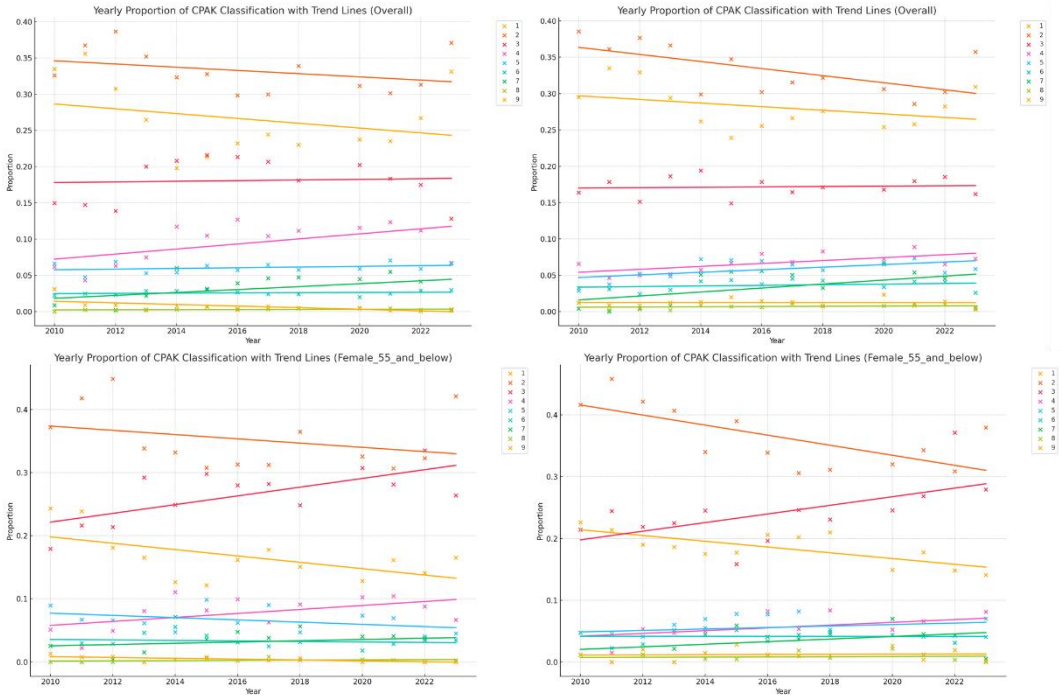

Fig. 11 Annual trends of CPAK classification
The top two graphs represent the total population, and the bottom two graphs represent females under 55 years old, showing the yearly proportion of CPAK classification. The colors correspond to CPAK groups 1 through 9, as indicated in the legend, with trend lines drawn for each group.
CPAK, coronal plane alignment of the knee.

4. DISCUSSION

4.1. Algorithm development and validation

The variability of conventional alignment measurement causes controversy. Surgeons have reported inconsistencies and discordance between conventional radiographic measurements and intraoperative navigation measurements[44,45]. Wright et al. reported three sources of measurement inconsistency: physiological variations, procedure variability (inconsistent positioning), and intra- and interobserver variability[46]. The mean interobserver difference was 1.4° (SD =1.1), and the mean intra-observer difference was 0.7° (SD = 0.9). Laskin et al. reported up to 7° variability in tibiofemoral angle measurements among 50 surgeons[47]. Automated measurement reduces these errors by minimising subjectivity.

We proposed a time-efficient system that automatically measures mTFA, mLDFA, MPTA, and JLCA from full-length leg weight-bearing radiographs. The system strongly correlated with the manual measurements in the internal and external institution tests.

Accurate segmentation is required for the automatic measurement of lower limb alignment. Previous studies performed femoral and tibial segmentation using a traditional spectral clustering and active shape model[48] or unsupervised or atlas-guided approaches[49-51]. Deep-learning methods have been applied in image segmentation, with UNet being popular in the medical field. However, UNet may not be the most efficient option for relatively simple data (images with fewer large objects) as it may require more resources. In this study, a SegNet model was used for image segmentation.

There have been studies utilizing long leg radiographs to investigate detailed angular values related to coronal alignment[52-56]. However, these papers commonly employ a method where landmarks are directly annotated by humans, and algorithms are subsequently trained based on this annotated data. This approach inherently introduces a potential bias to the reference values, as the ground truth is produced by humans marking points manually. In contrast, our approach involves segmentation followed by the identification of landmarks using a predetermined rule-based system. This method has the potential to reduce interobserver agreement on ground truth, as it eliminates the reliance on manual point annotation by humans. Moreover, the segmentation mask generated by the algorithm can be used to identify new geometric landmarks.

Zheng et al. proposed a method for automatically measuring leg length discrepancy in a pediatric population using deep learning[57]. The method demonstrated a high concordance rate between manual and automatic segmentation of the pediatric leg, with a Dice value of 0.94. However, their study employed a wide exclusion criteria. In contrast, Schock et al. achieved a high level of concordance rate across a wide range of clinical and pathologic indications, with an average Sørensen-Dice coefficient of 0.97 for the femur and 0.96 for the tibia[38].

In our internal validation, the readers and algorithm demonstrated a high concordance rate. The algorithm required 1 min/patient, in contrast to the manual measurement time of up to 3 min. In the external validation, the algorithm results significantly correlated with the manual measurements. However, the validation population consisted of young soldiers aged 20–30 years from a military hospital and may not represent the general population. JLCA values tended to be lower in military hospital patients than in those from the other included hospital. Nevertheless, these findings suggest that our algorithm may be useful in other populations.

Our study had several limitations. First, the training data did not include images from patients with skeletal dysplasia or hardware, limiting the clinical variability of the images. Second, several cases showed a large absolute error (> 5 degrees) between manual and automated measurement results. Future studies should include a wider variety and number of training data to reduce these errors. Third, our study included a total of 374 images from 374 patients for algorithm development, which may be considered too few compared to those in larger studies. However, studies by Zheng et al. and Schock et al. enrolled 179 and 255 patients, respectively, and showed convincing results

in their analyses, indicating that the number of cases analysed in our study (n = 374) was sufficient to demonstrate excellent performance[38,57].

4.2. Clinical application for large scale data

A key finding is that younger individuals, particularly females under 55, exhibited notable changes in knee alignment over time. MTFA showed a decreasing trend on both sides in individuals under 55, with the trend being more pronounced in younger females (Figure 3). This steeper decline in MTFA may reflect the influence of rising pediatric and adolescent obesity, known to affect valgus knee alignment. Previous studies[28] have linked higher BMI during growth phases to greater variability in knee alignment, especially in females, and the reduction in MTFA could increase the risk of lateral compartment OA later in life, given its association with valgus alignment.

The CPAK classification trends further highlighted gender-specific differences (Figure 4), with younger females showing a significant decrease in classification 1 and an increase in classification 3. This shift suggests that knee alignment in younger females is moving from traditional varus patterns toward more neutral or valgus alignments, potentially increasing the risk of lateral compartment OA. These findings underscore the need for early identification and intervention in populations at risk for alignment-related pathologies.

Age-related angular values generally showed a trend toward varization (Figure 1). This is likely due to degenerative changes that occur with aging, including remodeling of the medial femoral condyle and medial tibial plateau, as well as increased medial joint space narrowing. These findings are consistent with previous studies[58-60]. However, BMI had a more modest impact on knee alignment, with trends showing an increase in MTFA, LDFA, and JLCA as BMI increased (Figure 2). Nonetheless, these changes were not statistically significant, except for a slight decrease in left MTFA among males. The differences in obesity thresholds between Asian and Caucasian populations, with Caucasians typically having higher BMI values, may also contribute to variations in knee alignment patterns[61,62]. This underscores the importance of targeted research to better understand how body weight affects knee alignment over time.

One strength of the study is the use of a large cohort spanning over a decade, which allowed us to analyze long-term trends in knee alignment across various age groups and genders. This large sample size enhances the statistical power and generalizability of the findings, especially within the South Korean population. Additionally, advanced deep learning algorithms were applied for the automated measurement of angular values, reducing the potential for human error. The study also incorporated CPAK classification alongside detailed angular measurements, providing a comprehensive understanding of knee alignment patterns that are crucial for identifying populations at risk for OA.

Despite its strengths, this study has several limitations. First, as a retrospective analysis, it relies on previously collected data, limiting control over confounding variables and the ability to establish causal relationships. Second, the data were collected from a single tertiary hospital in South Korea, which may limit the generalizability of the findings to other populations. Third, the study lacks longitudinal follow-up for individual patients, making it difficult to track how knee alignment evolves over time. The exclusion of 2019 data due to DICOM file corruption may also introduce gaps in trend analysis. Additionally, the study focuses solely on radiographic knee alignment without correlating these findings to functional outcomes like pain, mobility, or quality of life. The reliance

on deep learning algorithms, while advantageous, introduces potential errors in measurement, which could affect the results. Lastly, important factors such as physical activity, occupation, and genetic predisposition, known to influence knee alignment and OA progression, were not accounted for.

5. CONCLUSION

Our deep-learning-based automated measurement algorithm accurately quantified lower limb alignment from long-leg radiographs and was significantly faster than manual measurements, making it well-suited for clinical application across various patient groups and conditions. Furthermore, this study emphasizes the importance of understanding knee alignment trends, particularly in younger females, who showed significant valgization in CPAK classification and mTFA. These trends suggest a potential rise in lateral osteoarthritis in females, highlighting the need for early, gender- and age-specific interventions to prevent OA. As treatment strategies for lateral OA are less standardized compared to medial OA, prevention and management approaches must be adapted accordingly. Further research is necessary to explore the long-term effects of these alignment changes and to develop targeted strategies for maintaining knee health throughout the lifespan.

References

1.Burghardt RD, Hinterwimmer S, Burklein D, Baumgart R. Lower limb alignment in the frontal plane: analysis from long standing radiographs and computer tomography scout views: an experimental study. Arch Orthop Trauma Surg 2013;133:29-36.
2.Zampogna B, Vasta S, Amendola A, Uribe-Echevarria Marbach B, Gao Y, Papalia R, et al. Assessing Lower Limb Alignment: Comparison of Standard Knee Xray vs Long Leg View. Iowa Orthop J 2015;35:49-54.
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, Niu J, Gross KD, Englund M, Sharma L, Cooke TDV, et al. Valgus malalignment is a risk factor for lateral knee osteoarthritis incidence and progression: findings from the Multicenter Osteoarthritis Study and the Osteoarthritis Initiative. Arthritis & Rheumatism 2013;65:355-62.
5.Sharma L, Chmiel JS, Almagor O, Felson D, Guermazi A, Roemer F, et al. The role of varus and valgus alignment in the initial development of knee cartilage damage by MRI: the MOST study. Ann Rheum Dis 2013;72:235-40.
6.Micicoi G, Jacquet C, Sharma A, LiArno S, Faizan A, Kley K, et al. Neutral alignment resulting from tibial vara and opposite femoral valgus is the main morphologic pattern in healthy middle-aged patients: an exploration of a 3D-CT database. Knee Surgery, Sports Traumatology, Arthroscopy 2021;29:849-58.
7.Thienpont E, Schwab PE, Cornu O, Bellemans J, Victor J. Bone morphotypes of the varus and valgus knee. Arch Orthop Trauma Surg 2017;137:393-400.
8.Kijowski R, Liu F, Caliva F, Pedoia V. Deep Learning for Lesion Detection, Progression, and Prediction of Musculoskeletal Disease. J Magn Reson Imaging 2020;52:1607-19.
9.Gyftopoulos S, Lin D, Knoll F, Doshi AM, Rodrigues TC, Recht MP. Artificial Intelligence in Musculoskeletal Imaging: Current Status and Future Directions. AJR Am J Roentgenol 2019;213:506-13.
10.Cerejo R, Dunlop DD, Cahue S, Channin D, Song J, Sharma L. The influence of alignment on risk of knee osteoarthritis progression according to baseline stage of disease. Arthritis & rheumatism 2002;46:2632-6.
11.Englund M. The role of biomechanics in the initiation and progression of OA of the knee. Best practice & research Clinical rheumatology 2010;24:39-46.
12.Vincent KR, Conrad BP, Fregly BJ, Vincent HK. The pathophysiology of osteoarthritis: a mechanical perspective on the knee joint. PM&R 2012;4:S3-S9.
13.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 & rheumatism 2007;56:1204-11.
14.Andriacchi TP. Valgus alignment and lateral compartment knee OA: A biomechanical paradox or new insight into knee OA? Arthritis and rheumatism 2013;65:310.
15.Sharma L, Song J, Dunlop D, Felson D, Lewis CE, Segal N, et al. Varus and valgus alignment and incident and progressive knee osteoarthritis. Annals of the rheumatic diseases 2010;69:1940-5.
16.Chen L, Zheng JJY, Li G, Yuan J, Ebert JR, Li H, et al. Pathogenesis and clinical management of obesity-related knee osteoarthritis: Impact of mechanical loading. Journal of orthopaedic translation 2020;24:66-75.
17.Musumeci G, Aiello FC, Szychlinska MA, Di Rosa M, Castrogiovanni P, Mobasheri A. Osteoarthritis in the XXIst century: risk factors and behaviours that influence disease onset and

progression. International journal of molecular sciences 2015;16:6093-112.
18.Silverwood V, Blagojevic-Bucknall M, Jinks C, Jordan J, Protheroe J, Jordan K. Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. Osteoarthritis and cartilage 2015;23:507-15.
19.Batushansky A, Zhu S, Komaravolu RK, South S, Mehta-D'souza P, Griffin TM. Fundamentals of OA. An initiative of Osteoarthritis and Cartilage. Obesity and metabolic factors in OA. Osteoarthritis and cartilage 2022;30:501-15.
20.Jackson B, Wluka A, Teichtahl A, Morris ME, Cicuttini FM. Reviewing knee osteoarthritis—a biomechanical perspective. Journal of Science and Medicine in Sport 2004;7:347-57.
21.Egloff C, Hügle T, Valderrabano V. Biomechanics and pathomechanisms of osteoarthritis. Swiss medical weekly 2012;142:w13583-w.
22.Heijink A, Gomoll AH, Madry H, Drobnič M, Filardo G, Espregueira-Mendes J, et al. Biomechanical considerations in the pathogenesis of osteoarthritis of the knee. Knee Surgery, Sports Traumatology, Arthroscopy 2012;20:423-35.
23.Bijlsma J, Knahr K. Strategies for the prevention and management of osteoarthritis of the hip and knee. Best practice & research Clinical rheumatology 2007;21:59-76.
24.Roy Davis Altman M. Early management of osteoarthritis. Am J Manag Care 2010;16:S41-S7.
25.Mahmoudian A, Lohmander LS, Mobasheri A, Englund M, Luyten FP. Early-stage symptomatic osteoarthritis of the knee—time for action. Nature Reviews Rheumatology 2021;17:621-32.
26.Lee C-H. Changes in the Food and Nutritional Status of Koreans over the Last Century. Korean Food and Foodways: The Root of Health Functional Food: Springer; 2022. p.263-88.
27.Kim S, Moon S, Popkin BM. Nutrition transition in the Republic of Korea. Asia Pacific Journal of Clinical Nutrition 2001;10:S48-S56.
28.Bout-Tabaku S, Shults J, Zemel BS, Leonard MB, Berkowitz RI, Stettler N, et al. Obesity is associated with greater valgus knee alignment in pubertal children, and higher body mass index is associated with greater variability in knee alignment in girls. The Journal of rheumatology 2015;42:126-33.
29.Fibel KH, Hillstrom HJ, Halpern BC. State-of-the-Art management of knee osteoarthritis. World Journal of Clinical Cases: WJCC 2015;3:89.
30.Uivaraseanu B, Vesa CM, Tit DM, Abid A, Maghiar O, Maghiar TA, et al. Therapeutic approaches in the management of knee osteoarthritis. Experimental and therapeutic medicine 2022;23:1-6.
31.MacDessi SJ, Griffiths-Jones W, Harris IA, Bellemans J, Chen DB. Coronal plane alignment of the knee (CPAK) classification: a new system for describing knee phenotypes. The bone & joint journal 2021;103:329-37.
32.Pagan CA, Karasavvidis T, Lebrun DG, Jang SJ, MacDessi SJ, Vigdorchik JM. Geographic variation in knee phenotypes based on the coronal plane alignment of the knee classification: a systematic review. The Journal of Arthroplasty 2023;38:1892-9. e1.
33.Huber S, Mitterer JA, Vallant SM, Simon S, Hanak-Hammerl F, Schwarz GM, et al. Gender-specific distribution of knee morphology according to CPAK and functional phenotype classification: analysis of 8739 osteoarthritic knees prior to total knee arthroplasty using artificial intelligence. Knee Surgery, Sports Traumatology, Arthroscopy 2023;31:4220-30.
34.Yeoh PSQ, Lai KW, Goh SL, Hasikin K, Hum YC, Tee YK, et al. Emergence of deep learning in knee osteoarthritis diagnosis. Computational intelligence and neuroscience 2021;2021:4931437.
35.Moon K-R, Lee B-D, Lee MS. A deep learning approach for fully automated measurements of lower extremity alignment in radiographic images. Scientific Reports 2023;13:14692.
36.Schock J, Truhn D, Abrar DB, Merhof D, Conrad S, Post M, et al. Automated analysis of

alignment in long-leg radiographs by using a fully automated support system based on artificial intelligence. Radiology: Artificial Intelligence 2020;3:e200198.
37.Lee HS, Hwang S, Kim S-H, Joon NB, Kim H, Hong YS, et al. Automated analysis of knee joint alignment using detailed angular values in long leg radiographs based on deep learning. Scientific Reports 2024;14:7226.
38.Schock J, Truhn D, Abrar DB, Merhof D, Conrad S, Post M, et al. Automated Analysis of Alignment in Long-Leg Radiographs by Using a Fully Automated Support System Based on Artificial Intelligence. Radiol Artif Intell 2021;3:e200198.
39.Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:2481-95.
40.Barber CB, Dobkin DP, Huhdanpaa H. The quickhull algorithm for convex hulls. ACM Transactions on Mathematical Software (TOMS) 1996;22:469-83.
41.O'Rourke J. Finding minimal enclosing boxes. International journal of computer & information sciences 1985;14:183-99.
42.Zou KH, Wells WM, 3rd, Kikinis R, Warfield SK. Three validation metrics for automated probabilistic image segmentation of brain tumours. Stat Med 2004;23:1259-82.
43.Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:297-302.
44.Yaffe MA, Koo SS, Stulberg SD. Radiographic and navigation measurements of TKA limb alignment do not correlate. Clinical orthopaedics and related research 2008;466:2736-44.
45.Han SB, Kim HJ, Lee DH. Effect of Computer Navigation on Accuracy and Reliability of Limb Alignment Correction following Open-Wedge High Tibial Osteotomy: A Meta-Analysis. Biomed Res Int 2017;2017:3803457.
46.Wright JG, Treble N, Feinstein AR. Measurement of lower limb alignment using long radiographs. J Bone Joint Surg Br 1991;73:721-3.
47.Laskin RS. Alignment of total knee components. SLACK Incorporated Thorofare, NJ; 1984. p.62-72.
48.Wu J, Mahfouz MR. Robust x-ray image segmentation by spectral clustering and active shape model. J Med Imaging (Bellingham) 2016;3:034005.
49.Gandhamal A, Talbar S, Gajre S, Razak R, Hani AFM, Kumar D. Fully automated subchondral bone segmentation from knee MR images: Data from the Osteoarthritis Initiative. Comput Biol Med 2017;88:110-25.
50.Gandhamal A, Talbar S, Gajre S, Hani A, Kumar D. Automatic and unsupervised femur and tibia segmentation using magnetic resonance images. Osteoarthritis and Cartilage 2017;25:S258.
51.Aprovitola A, Gallo L. Knee bone segmentation from MRI: A classification and literature review. Biocybernetics and Biomedical Engineering 2016;36:437-49.
52.Kim SE, Nam JW, Kim JI, Kim J-K, Ro DH. Enhanced deep learning model enables accurate alignment measurement across diverse institutional imaging protocols. Knee Surgery & Related Research 2024;36:4.
53.Jo C, Hwang D, Ko S, Yang MH, Lee MC, Han H-S, et al. Deep learning-based landmark recognition and angle measurement of full-leg plain radiographs can be adopted to assess lower extremity alignment. Knee Surgery, Sports Traumatology, Arthroscopy 2023;31:1795.
54.Nam HS, Park SH, Ho JPY, Park SY, Cho JH, Lee YS. Key-Point Detection Algorithm of Deep Learning Can Predict Lower Limb Alignment with Simple Knee Radiographs. Journal of Clinical Medicine 2023;12:1455.
55.Simon S, Schwarz GM, Aichmair A, Frank BJH, Hummer A, DiFranco MD, et al. Fully automated deep learning for knee alignment assessment in lower extremity radiographs: a cross-

sectional diagnostic study. Skeletal Radiology 2022;51:1249-59.
56.Meng X, Wang Z, Ma X, Liu X, Ji H, Cheng J-z, et al. Fully automated measurement on coronal alignment of lower limbs using deep convolutional neural networks on radiographic images. BMC Musculoskeletal Disorders 2022;23:869.
57.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.
58.Hayami T, Pickarski M, Wesolowski GA, Mclane J, Bone A, Destefano J, et al. The role of subchondral bone remodeling in osteoarthritis: reduction of cartilage degeneration and prevention of osteophyte formation by alendronate in the rat anterior cruciate ligament transection model. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology 2004;50:1193-206.
59.Gensburger D, Arlot M, Sornay-Rendu E, Roux Jp, Delmas P. Radiologic assessment of age-related knee joint space changes in women: A 4-year longitudinal study. Arthritis Care & Research 2009;61:336-43.
60.Fife RS, Brandt KD, Braunstein EM, Katz BP, Shelbourne KD, Kalasinski LA, et al. Relationship between arthroscopic evidence of cartilage damage and radiographic evidence of joint space narrowing in early osteoarthritis of the knee. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology 1991;34:377-82.
61.Organization WH. The Asia-Pacific perspective: redefining obesity and its treatment. Health Communications Australia 2000.
62.Consultation W. Obesity: preventing and managing the global epidemic. World Health Organization technical report series 2000;894:1-253.

Abstract in Korean

하지 방사선 사진에서 세부 각도 값을 사용한 무릎 관절 정렬의 딥러닝 기반 자동 분석

하지 구조에서의 변형은 여러 가지 원인에 의해 발생할 수 있으며, 특히 교정이 필요한 경우 하지 정렬을 정확하게 평가하는 것이 필수적이다. 이를 해결하기 위해, 우리는 하지 정렬을 평가하는 자동화 시스템을 개발하였으며, 전신 체중 부하 방사선 사진에서 기계적 경골대퇴각(mTFA), 기계적 외측 원위 대퇴각(mLDFA), 내측 근위 경골각(MPTA), 관절선 일치각(JLCA)과 같은 주요 매개변수를 정량화하였다. 본 후향적 연구에서는 알고리즘 개발 및 테스트를 위해 404개의 방사선 사진을 분석하였으며, 다른 병원의 30개 방사선 사진을 사용하여 외부 검증을 진행하였다. 분절화 알고리즘의 성능은 Dice 유사 계수(DSC)를 이용해 수동 분절화와 비교하였으며, 정렬 매개변수의 일치는 클래스 간 상관 계수(ICC)를 통해 평가하였다. 또한, 네 가지 정렬 매개변수를 측정하는 데 소요된 시간을 기록하였다. 알고리즘은 사람에 의해 그려진 분절화와 89~97%의 유사도를 보여 우수한 일치율을 보였으며, 정렬 매개변수에 대한 일치는 양호에서 매우 양호(ICC: 0.7213–0.9865)한 수준으로 평가되었다. 자동화된 측정 방식은 수동 측정 대비 약 3.44배 더 신속하였다.

더 큰 임상적 적용을 위해, 우리는 2010년부터 2023년까지 한국의 3차 병원에서 촬영된 34,160개의 다리를 대상으로 총 17,080장의 하지 전장 방사선 사진을 사용한 후향적 코호트 연구를 진행하였다. 딥러닝 모델을 사용하여 mTFA, mLDFA, MPTA, JLCA 를 자동 분석하였고, 환자는 연령(≤55세 및 >55세)과 성별에 따라 분류하였다. 무릎 정렬의 연도별 경향을 평가하기 위해 선형 회귀 분석을 시행하였습니다. mTFA 는 전체 인구에서 일관된 양상을 보였으나($p > 0.05$), 55세 이하 군에서는 유의미한 감소가 나타났다(우측: β=-0.05, 95% CI: -0.08 to -0.02, p = 0.01; 좌측: β=-0.05, 95% CI: -0.07 to -0.03, p = 0.01). 특히 젊은 여성 집단에서 더욱 뚜렷한 감소 경향이 관찰되었다(우측: β=-0.08, 95% CI: -0.14 to -0.03, p = 0.01; 좌측: β=-0.06, 95% CI: -0.10 to -0.01, p = 0.01). CPAK 분류의 경향 분석에서는 55세 이하 여성에서 3번 분류의 비율이 증가하는 것으로 나타났다(우측: β=0.01, 95% CI: 0.00 to 0.01, p = 0.01; 좌측: β=0.01, 95% CI: 0.00 to 0.01, p = 0.01). 이에 따라, 3번 분류는 2010년에서 2023년 사이에 우측에서는 17.9%에서 26.4%로, 좌측에서는 21.4%에서 28.0%로 증가하였다. 본 연구는 특히 젊은 여성 집단에서 무릎 정렬의 변화가 뚜렷하며, 외반 변형(Valgus alignment)으로의 전환 경향을 보여주고 있다. 이러한 경향의 장기적인 임상적 영향을 규명하고, 전 생애에 걸쳐 무릎 건강을 유지하기 위한 맞춤형 전략을 개발하기 위한 추가 연구가 필요할 것으로 생각된다.

핵심되는 말 : 무릎 관절 정렬, 방사선 촬영, 딥러닝, 골관절염

PUBLICATION LIST

Lee HS, Hwang S, Kim S-H, Joon NB, Kim H, Hong YS, et al. Automated analysis of knee joint alignment using detailed angular values in long leg radiographs based on deep learning. Scientific Reports 2024;14:7226.