Master's Thesis of Park Sang Hyun

Key-Point Detection Algorithm of Deep Learning Can Predict Lower Limb Alignment with Plain Knee Radiographs

단순 슬관절 방사선 사진으로 하지 정렬을 예측할 수 있는 딥 러닝의 키 포인트 감지 알고리즘

August 2023

Graduate School of Medicine
Seoul National University
Orthopedic surgery

Park Sang Hyun

Key-Point Detection Algorithm of Deep Learning Can Predict Lower Limb Alignment with Plain Knee Radiographs

Pf. Lee Yong Seuk

Submitting a master's thesis of
Orthopedic surgery

April 2023

Graduate School of Medicine
Seoul National University
Orthopedic surgery

Park Sang Hyun

Confirming the master's thesis written by
Park Sang Hyun
July 2023

Chair _____(Seal)
Vice Chair _____(Seal)
Examiner _____(Seal)

Abstract

Background: There have been many attempts to predict the weight-bearing line (WBL) ratio using plain knee radiographs. Using a convolutional neural network (CNN), we focused on predicting the WBL ratio quantitatively.

Methods: From March 2003 to December 2021, 2410 patients with 4790 knee AP radiographs were randomly selected using stratified random sampling. Our dataset was cropped by four points annotated by a specialist with a 10-pixel margin. The model predicted our interest points, which were both plateau points, i.e., starting WBL point and exit WBL point. The resulting value of the model was analyzed in two ways: pixel units and WBL error values.

Results: The mean accuracy (MA) was increased from around 0.5 using a 2-pixel unit to around 0.8 using 6 pixels in both the validation and the test sets. When the tibial plateau length was taken as 100%, the MA was increased from approximately 0.1, using 1%, to approximately 0.5, using 5% in both the validation and the test sets.

Conclusions: The DL-based key-point detection algorithm for predicting lower limb alignment through labeling using plain knee AP radiographs demonstrated comparable accuracy to that of the direct measurement using whole leg radiographs. Using this algorithm, the WBL ratio prediction with plain knee AP radiographs could be useful to diagnose lower limb alignment in osteoarthritis patients in primary care.

Keyword : knee; weight-bearing line; machine learning; convolutional neural network; prediction
Student Number : 2018-24084

Table of Contents

Chapter 1. Introduction

1.1. Study Background

Osteoarthritis (OA) is the most common form of arthritis, affecting millions of people worldwide [1]. It occurs when the protective cartilage that cushions the ends of the bones wears down [2]. Therefore, weight-bearing joints, such as the knee joint, are more vulnerable. The medial compartment of the knee joint is the most commonly affected site in OA, and the medial joint space narrows as OA progresses. This induces varus deformity of the lower limb. Consequently, the adduction moment, which is the magnitude of the ground reaction force, moves medially from the center of the knee joint during ambulation, and the moment arm of the ground reaction force increases [3]. In this manner, the varus deformity enters a vicious cycle. Therefore, it is important to intervene through treatment before such vicious cycle develops [4,5].

The evaluation of the weight-bearing axis of the lower limb is a fundamental step in the identification, classification, and treatment of lower limb deformities, which may result from degeneration, trauma, inflammation, or congenital diseases. When deciding between treatment options, such as conservative treatment, osteotomy, and arthroplasty, the weight-bearing axis should be considered in addition to the patient's basic demographics [6]. Several methods assess the weight-bearing axis of the lower limbs. Techniques to determine the hip-knee-ankle angle, mechanical axis, and weight-bearing line (WBL) ratio are commonly used methods [7,8]. These parameters are usually measured on whole leg radiographs (WLR) and in institutions, such as large hospitals or community clinics [9]. Therefore, they may not be readily available because of the high costs involved. Moreover, primary physicians may face difficulties in identifying such deformities, and several attempts have been made to predict the WBL ratio using

plain standing knee radiographs

The application of artificial intelligence in medicine has gained popularity in recent years because of its ability to improve the efficiency of healthcare delivery and patient diagnosis. Karnuta et al. used machine learning to identify knee arthroplasty implants from X-rays [11]. In addition, Joseph et al. employed machine learning to forecast the development of osteoarthritis over 8 years using combined MR imaging features, demographics, and clinical factors as input [12]. Convolutional neural networks (CNNs) are a subtype of deep learning (DL) that have shown impressive outcomes in image classification and recognition [13,14]. A CNN was used in a previous study that predicted the WBL ratio as a parameter for lower limb alignment in plain knee radiographs [15]. However, a limitation of this study was that the prediction was only possible within intervals. Therefore, a quantitative assessment that can predict the WBL ratio and translate it to an accurate point on the tibial plateau may be more intuitive for clinical use.

1.2. Purpose of Research

This study aimed to develop a DL algorithm to predict the point at which the weight bearing axis of the lower limb crosses the tibial plateau. The hypothesis of this study was that the WBL ratio obtained from weight-bearing WLR could be predicted by a specially designed DL model using standing plain knee anteroposterior (AP) radiographs with high predictive value.

Chapter 2. Body

2.1. Materials and Methods

With the assumption that standing plain knee radiographs are a part of WLR and may be related to the lower limb alignment, the most appropriate and accurate model to predict the WBL ratio after learning the WBL ratio was designed using plain knee radiographs. This was based on the key-point detection model [16,17]. The key-point detection model involves locating the key object parts that represent the underlying object in a feature-rich manner. After directly labeling the WLR picture according to how we draw the WBL in a real clinical situation, we cut it into a plain knee radiograph picture and trained the DL model. Subsequently, the accuracy of the learned DL algorithm for measuring the WBL ratio was investigated. For the analysis of the accuracy of the DL model, its mean absolute error (MAE) and intra-class correlation coefficients (ICC) were evaluated. In addition, the accuracy values of the DL model were compared with the ICCs of the rater, using real measurements of the WBL ratio in WLR to check whether the DL algorithm was reliable as the measurements of the rater. All procedures involving human participants were performed following the ethical standards of the institutional review board (IRB No. B-2210-784-101) and the Helsinki Declaration (1964) and its later amendments. Consent was not sought because this study retrospectively reviewed the medical record image data of patients who underwent X-ray examination at our hospital, and personal identification information was not included in the data analysis process. Therefore, it was difficult to evaluate if the risk for the patients included in this study was increased compared to that of other patients. There is no reason to presume the refusal of consent.

2.1.1 Data Set

From March 2003 to December 2021, 89,709 patients with knee pain and standing knee AP radiographs were obtained from the clinical database of our hospital. Among them, 3515 patients (3.9% of standing knee AP acquisition) who underwent weight-bearing WLR were included. The exclusion criteria were as follows: (1) previous ipsilateral surgery (hip, knee, or ankle joint); (2) children with remaining growth plates; and (3) patients with deformity due to previous trauma or congenital diseases. After excluding patients who met the exclusion criteria, finally, AP radiographs of 4790 knees in 2410 patients were randomly selected using stratified random sampling. To avoid the cluster effect between multiple radiographs in a single patient, only the initial knee AP radiograph was used.

2.1.2 WBL Ratio Measurement and Labeling

Our dataset was created by one orthopedic surgeon with five years of experience, and the data were cropped by four points annotated (tibial plateau at both ends, WBL starting and exit points) with a 10-pixel margin. Although four points varied on the cropped data, and the data size was variable, the data were resized uniformly in the training phase. Therefore, our model was invariant in relation to the image size to some degree. The model's robustness of the data size was evaluated with various random margins of 5 to 10 pixels. This experiment showed that the WBL ratio did not affect the data size, and our model predicted the line's tendency. This revealed that even if the users crop the image data abnormally, the model's prediction result will not decline.

The WBL ratio was measured using the weight-bearing WLR of all 4790 lower limbs in 2410 patients for labeling the training set and analysis of prediction accuracy in the validation and test sets. The WBL was drawn from the center of the femoral head to the

center of the superior articular surface of the talus. The WBL ratio was calculated as the ratio of the crossing point of the mechanical axis, from the medial edge to the entire width of the tibial plateau..

2.1.3 Image Processing

A standing knee AP radiograph was chosen as the research object. The PyDicom library (version 1.3.0) was used for the preprocessing of DICOM images. The right or left knee was cropped in a knee radiograph that included both knees. Strong augmentation, such as sheer, distortion, and high rate of random brightness, contrast, equalization, and hue saturation for robustness, was used to improve the performance of the algorithm.

2.1.4 DL Algorithm

The DL algorithm consists of two stages. In the first stage, key-points heat maps were predicted using the WBL prediction model. Our WBL prediction model is similar to conventional pose estimation, which predicts some key-points through heat maps and is composed of a feature extractor and a simple convolutional decoder. The model predicts our points of interest, which are both plateau points, i.e., the starting WBL point and the exit WBL point. Using the logits calculated from the model, we trained our model in an end-to-end manner with the Adam W optimizer and binary cross entropy loss. The other setting was similar to that of the conventional DL. In the second stage, the WBL ratio was calculated using the four predicted points. Two lines from four points were drawn to determine the intersection of the two lines and calculate the WBL ratio (Figure 1).

2.1.5 Experiment

In the experiment, our model for 10,000 iterations using Adam W with $\beta 1 = 0.9$, $\beta 2 = 0.999$ was trained; learning rate = 1×10^{-3}, weight decay = 1×10^{-2}, cosine decaying scheduler, and binary cross entropy loss. In the training phase, the cropped image data were transformed using random brightness, contrast, equalization, and hue saturation for robustness. Our model was on four 2080 ti GPU with an 8-batch size per GPU. The resulting value of the model was analyzed in two ways: WBL error values and pixel units. When analyzing the results with the WBL error value, a value of 1 was assigned if the DL prediction value was within each error value of the tibial plateau, and a value of 0 was assigned otherwise. When analyzing the results in pixel units, the accuracy was calculated by assigning a value of 1 if the DL prediction was received in each pixel unit, and a value of 0 otherwise. The algorithm was first trained by making as many landmarks as possible, i.e., 27 dots, around the knee joint. According to the traditional WBL calculation method, radiological landmarks that can symbolize the bony anatomy of both femur and tibia were used, so there were 27 of them, as follows. The starting point and exit point of the WBL line, both endpoints of the femur cortex in X-ray images, both endpoints of the tibia cortex in X-ray images, both endpoints of femur and tibia at the tibiofemoral joint-the most distant points medial and lateral from femur and tibia and transepicondyle points at the tibiofemoral joint-trochlear notch center-midpoint of both tibial spines and tibial spines, intersection of tibia and fibula, inflection points at femur and tibia. The marking was reduced by identifying the most appropriate dots that showed the best performance (Figure 1).

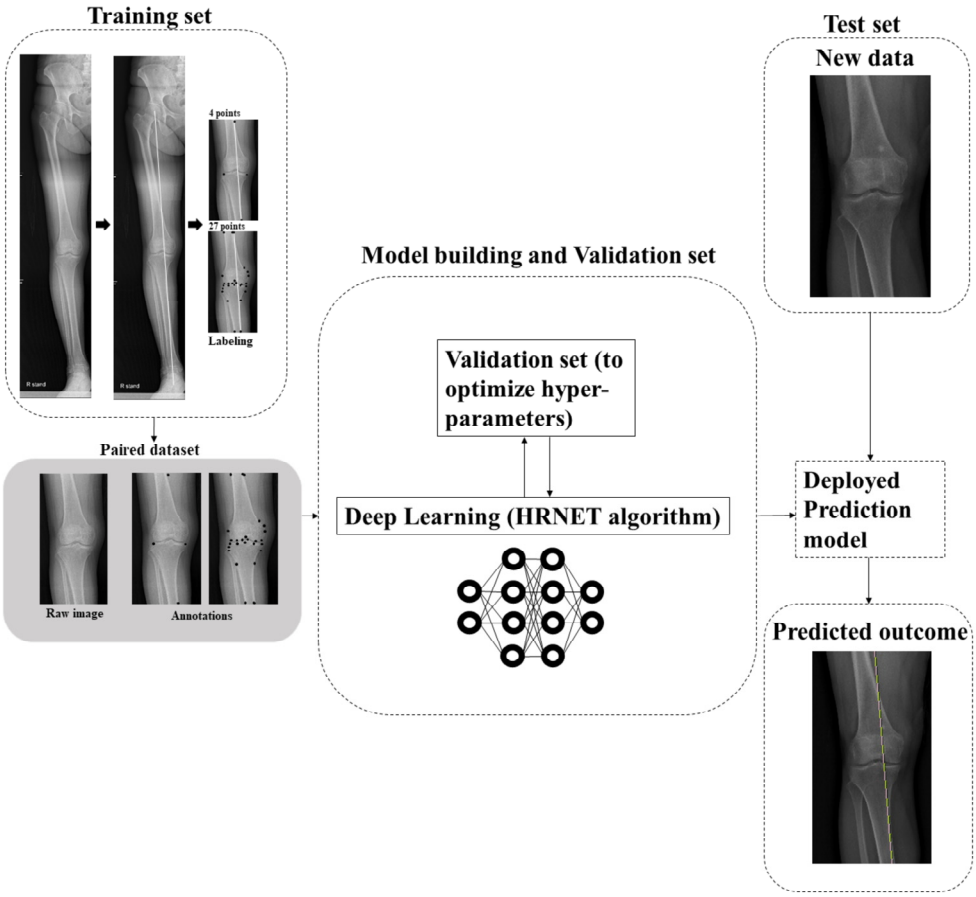

Figure 1. Pipeline of the study. Annotations: labeled image.
Training set: The data sample used to fit the model, Validation set: The data sample used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. Test set: The data sample used to provide an unbiased evaluation of a final model fit on the training dataset.

2.1.6 Statistical Anaylsis

Data are presented as means and standard deviations for continuous variables. One way analysis of variance was performed to compare the quantitative variables (i.e., age, body mass index (BMI), and WBL ratio). Pearson's chi-squared test or Fisher's exact test was used to compare the qualitative variables (i.e., gender). Statistical significance was set at $p < 0.05$. The data were analyzed using SPSS 25.0 (IBM, Armonk, NY, USA). To examine the reproducibility of the calculation of the WBL ratio using WLR,

two observers were chosen: A, an orthopedic surgeon with 5 years of experience; B. an orthopedic surgeon with 20 years of experience. Independent measurements obtained by each of the two raters for each data set (raters A, B) and two independent measurements (A1 and A2) obtained by a rater A for each data set at a different time were compared. The mean difference between the independent measurements obtained by the raters was evaluated. The inter- and intra-observer reliabilities of the measurements were analyzed using ICC, with ICC < 0.40 indicating poor agreement, in the range 0.40-0.75 indicating fair to good (moderate) agreement, and in the range 0.76-1.00 indicating excellent agreement. MAE was used as a measure to determine how well the CNN fit the WBL ratio [13–15]. MAE is a measure that indicate the difference between the actual labeled WBL ratio by A (AL) using WLR and the WBL ratio predicted by the CNN using plain knee radiographs, $MAE=\frac{1}{N}\sum_{i=1}^{N}\left|\widehat{y_i}-y_i\right|$, with $\widehat{y_i}$: estimated WBL ratio of ith data, y_i : ground-truth WBL ratio of ith data [16].

2.2. Results

The baseline characteristics of the patients and the distribution of the labels in the training, validation, and test sets are summarized in Table 1. Age, sex, BMI, and WBL ratio were not significantly different between the datasets. The performance was improved through simple labeling that marked both ends of the tibia and the starting and exit points of the WBL line. Four points indicated the best performance among the trials. A comparison of the WBL mean accuracy at 4 and 27 points is shown in Figure 2. As the threshold of the WBL error percentage value increased, the accuracy of taking four points approached 0.6, whereas when 27 points were taken, a steady state was reached with an accuracy not exceeding 0.1. These results were also similarly obtained in the

pixel unit, and as the threshold increased, the accuracy when taking four points was close to 0.8 in 6 pixels, while that achieved when taking 27 points was approximately 0.2 in 6 pixels. A representative example of a patient's difference between the predicted and the correct points on an X-ray image is shown in Figure 3. The differences between the actual measured WBL ratio and the WBL ratio predicted by the CNN in this patient were 0.03 in the right knee and 0.05 in the left knee.

Table 1. Baseline characteristics of dataset

	Training set	Validation set	Test set	Total	P-value
Age (year)	65.1 ± 9.31	65.0 ± 9.42	65.4 ± 9.27	65.2 ± 12.2	0.521
Gender (M/F)	391/1928	46/241	49/241	486/2410	0.215
BMI (kg/m^2)	25.7 ± 3.22	25.8 ± 2.51	25.3 ± 2.37	25.7 ± 3.21	0.333
WBL ratio	0.33 ± 0.17	0.31 ± 0.15	0.32 ± 0.12	0.32 ± 0.15	0.087

Values are presented as number or mean ± standard deviation. M, men; F, female; BMI, body mass index; WBL, weight-bearing line; Training set, The sample of data used to fit the model; Validation set, The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper-parameters; Test set, The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset

The prediction of the algorithm implemented by learning is shown in Table 2 using units of pixels and WBL error values. The mean accuracy (MA) was increased from around 0.5 using a 2-pixel unit to around 0.8 using 6 pixels in both the validation and the test sets. The probability of the prediction and target values entering within 6 pixels was close to 0.8. When the tibial plateau length was taken as 100%, the MA was increased from approximately 0.1, using 1%, to approximately 0.5, using 5%, in both the validation and the test sets. The probability of obtaining a value within 5% exceeded 0.5.

The mean difference, ICC, and MAE value are shown in Table 3. The mean difference of the validation and test sets between intra-observer and inter-observer measurements of the WBL ratio using a WLR ranged from 0.023 to 0.036. The MAE of the validation and test sets, with the DL model measuring the WLB

ratio using plain knee radiographs were 0.064 (95% CI, 0.057-0.071) and 0.051 (95% CI, 0.044-0.058), respectively. The ICCs of the validation and test sets were all over 0.8, which indicated excellent agreement. The distributions of the WBL predictions in the validation and test sets are shown in Figure 4. The distribution of the DL model showed less difference in the high incidence area, where the WBL ratio was between 0.25 and 0.50. However, higher differences in the section that showed a lower incidence of the WBL ratio were noted. These values were similar for both the validation and the test sets. The MA of the WBL ratio for the validation and test sets is shown in Figure 5. As the thresholds of the WBL error percentage value and pixel units increased, the accuracy increased in both the validation and the test sets. The accuracy increased from around 0.1 to 0.5 with the increase in the WBL error percentage value thresholds, and from around 0.5 to 0.8 with the increase in the pixel threshold.

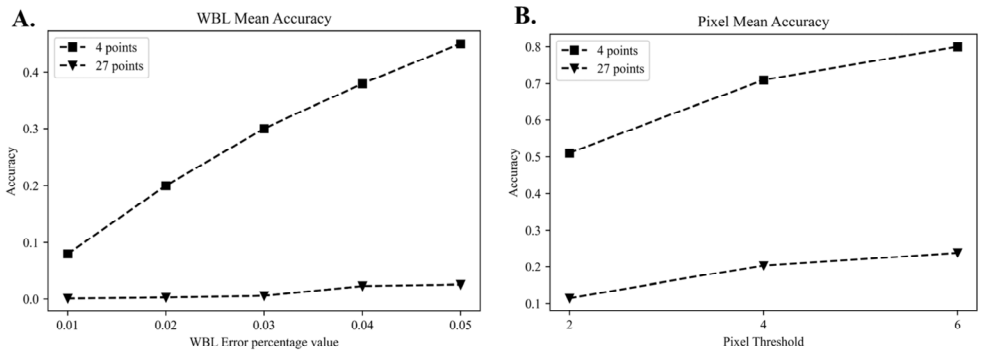

Figure 2. Comparison of WBL mean accuracy in 4 and 27 points.

(A) Comparison graph of 4 and 27 points in WBL error percentage. (B) Comparison graph of 4 and 27 points in pixel threshold, Validation set: The data sample used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. Test set: The data sample used to provide an unbiased evaluation of a final model fit on the training dataset.

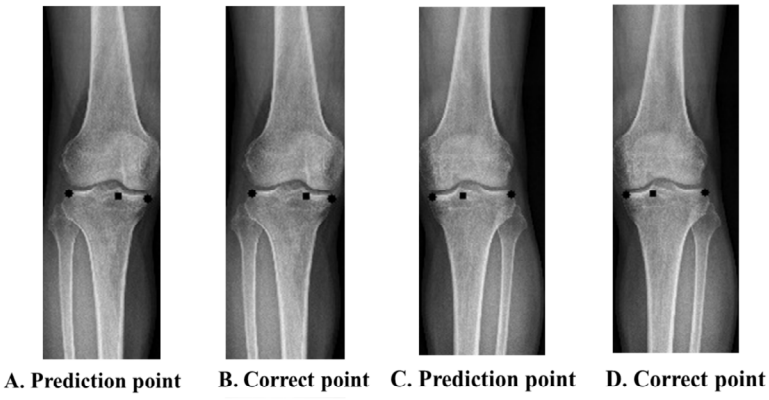

Figure 3. Prediction and correct point.
(A, B, C, D) Prediction point vs correct point. (A) Prediction point in Rt knee. (B) Correct point in Rt knee. (C) Prediction point in Lt knee. (D) Correct point in Lt knee. Prediction point: point in which deep learning algorithm is predicted, Correct point: labeled point.

Table 2. Results for prediction of deep learning algorithm

			Validation set		
Pixel	**2 pixels**		**4 pixels**		**6 pixels**
MA	0.51 ± 0.07		0.71 ± 0.06		0.80 ± 0.08
WEV	**1 %**	**2 %**	**3 %**	**4 %**	**5 %**
MA	0.08 ± 0.04	0.20 ± 0.05	0.30 ± 0.06	0.38 ± 0.07	0.45± 0.07

			Test set		
Pixel	**2 pixels**		**4 pixels**		**6 pixels**
MA	0.52 ± 0.06		0.73 ± 0.06		0.84 ± 0.04
WEV	**1 %**	**2 %**	**3 %**	**4 %**	**5 %**
MA	0.10 ± 0.05	0.22 ± 0.07	0.37 ± 0.09	0.46 ± 0.11	0.53 ± 0.11

Values are presented as number or mean ± standard deviation. MA, Mean accuracy; WEV, WBL Error value; Validation set, The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper-parameters; Test set, The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.

Table 3. Mean difference and mean absolute error and ICC in validation and test sets

Validation set	Mean difference	ICC
A1 and A2	0.023 ± 0.005	0.96 ± 0.02
A1 and B	0.034 ± 0.003	0.94 ± 0.05
Test set	**Mean difference**	**ICC**
A1 and A2	0.024 ± 0.006	0.97 ± 0.03
A1 and B	0.036 ± 0.004	0.95 ± 0.05
Validation set	**Mean absolute error**	**ICC**
AL and DL	0.064 ± 0.007	0.89 ± 0.09
Validation set	**Mean absolute error**	**ICC**
AL and DL	0.051 ± 0.007	0.88 ± 0.08

Values are presented as number or mean ± standard deviation. ICC, Intra-class correlation coefficients; Validation set, The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper-parameters; Test set, The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset; A1, rater 1, A2: rater 1 at different times; B, rater 2; AL, rater 1 with labeling on the WLR; DL, deep learning

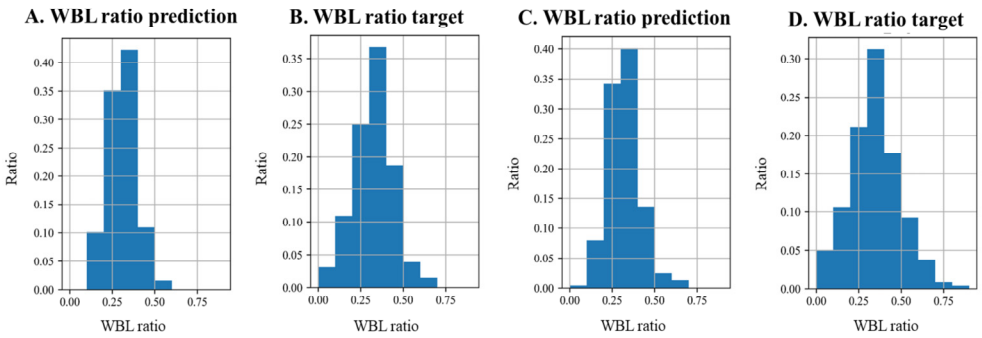

Figure 4. Histogram of WBL ratio distribution in validation and test set.

(A, B, C, D) Prediction distribution vs target distribution. (A) WBL ratio prediction distribution in validation set. (B) WBL ratio target distribution in validation set. (C) WBL ratio prediction distribution in test set. (D) WBL ratio target distribution in test set. WBL ratio prediction: predicted value of WBL, WBL ratio target: calculated WBL target.

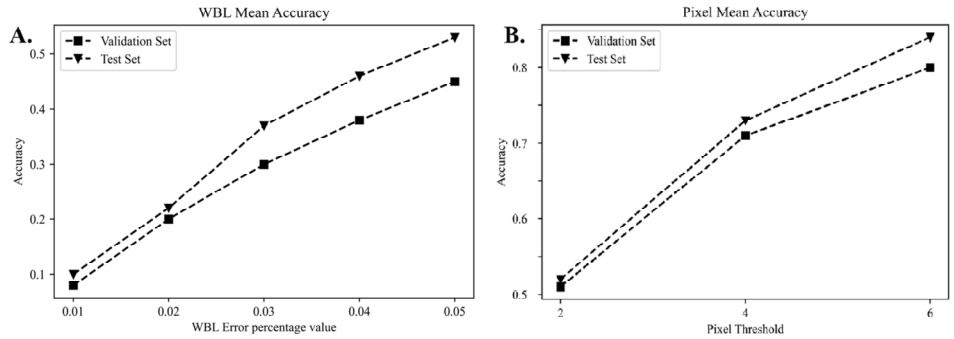

Figure 5. Graph of WBL mean accuracy.

(A) Accuracy graph of WBL error percentage value, (B) Accuracy graph of pixel threshold, validation set: The data sample used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters. Test set: The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.

2.3. Discussion

The principal finding of this study is that the novel approach using the four-point marked key-point detection algorithm could predict the alignment of the lower limb using standing knee AP radiographs with high accuracy, comparable to that achieved with the direct measurement of the WLR. As one pixel was 0.265 mm, approximately 80% of the test and validation set prediction values were entered into 6 pixels. Therefore, it was assumed that approximately 80% of patients could be correctly evaluated within approximately 1 mm intervals.

The mean difference, MAE, and ICC values were used to test the accuracy and reliability of this study [19]. The accuracy of DL was indirectly estimated by comparing the mean difference between the values measured by the raters using WLR and the MAE of the values of DL predicting the WBL ratio using plain knee radiographs. The WBL ratio showed a slightly increase in the MAE when compared with the mean difference, because the WLR was not provided. In addition, the ICC of DL was also lower than the ICC of the values measured by the raters using WLR, but we observed

that it provided a relatively good prediction at 0.8 or more.

Several trials to predict lower limb alignment using plain knee radiographs by linear regression analysis exist [10,20]. However, the results are not satisfactory, and to obtain satisfactory results, an X-ray image approximately 20 cm in length above and below the knee joint was required [21]. To solve these problems, a new prediction method based on DL using a key-point detection algorithm is proposed in our study. The key-point detection algorithm was considered the most suitable model because of the ability of the algorithm to find a specific point where the WBL passes through the tibia plateau [22,23]. Key-point detection algorithms are often used for pose estimation, face detection, and object detection [17,23]. Interestingly, the prediction accuracy decreased when the marking increased. This seems to be due to the characteristics of the DL model. DL is so automatic and high dimensional that the process of calculating the output by extracting features from the input is represented as a black box [13,24,25]. Therefore, it can be understood that if more labeling is performed, the labeling error increases, and the automatic feature extraction process is hindered; thus, the prediction accuracy could be lowered, as occurred in our study.

From the patient's point of view, visiting a tertiary hospital for WLR is time-consuming, expensive, and leads to high radiation exposure [26]. If the WBL ratio can be predicted through plain knee AP radiographs using this algorithm in primary care, it will be possible to easily determine the treatment process, as well as the degree of arthritis in more detail. This will also be useful for follow-up evaluation of patients who underwent re-alignment procedures such as osteotomy [19,27]. Expecting lower limb alignment using only plain knee radiographs has a lot of pros in the decision of patient-specific treatment protocols in various kinds of institutions. A DL model for predicting the WBL ratio through a plain knee radiograph was not attempted in the past, but it will become an essential medical technique in the future society characterized by the use of artificial intelligence [28].

The strength of this study is that a more accurate prediction of lower limb alignment can be obtained using the DL key-point detection model. Our study is meaningful in that it not only uses a key-point detection model, but also takes a significant point and trains machine learning to make a more precise prediction using this model. In addition, this study has significance as it allows predicting the WBL ratio using a plain knee AP radiograph in a situation where WLR imaging is limited. Primary care physicians can properly diagnose patients with knee OA using the DL model. In addition, it is expected that planning the realignment procedure would be possible with high accuracy.

This study has several limitations. First, it was difficult to interpret the developed CNN model itself; therefore, it was hard to determine whether the CNN model focused on the WBL prediction. Second, the number of test sets was small, although the ratio between the validation and the test sets was adequate. Third, the WBL predictions of our model were relatively distributed around the center compared with the target WBL value. Thus, our model was trained on general WBL values and has limitations in predicting infrequent WBL values. This is because of the small amount of WBL data corresponding to outliers, as shown in Figure 4. This phenomenon is expected to decrease as the amount of data increases. Fourth, the prediction would be inaccurate if there is a deformity in the proximal femur or distal tibia, because this cannot be checked on standing plain knee radiographs. Fifth, since the study was conducted on patients who presented to tertiary medical institutions, there may be a selection bias in the patient group. This is because these patients were referred from primary and secondary medical institutions. The severity of disease in patients visiting tertiary care may be higher than that of patients visiting primary or secondary health care institutions.

Chapter 3. Conclusions

This DL-based key-point detection algorithm for predicting lower limb alignment makes primary physicians possible to predict lower limb alignment of patients with only plain knee AP radiographs in considerable amount of accuracy. It will help us to determine the treatment process, as well as the degree of arthritis in more detail.

Bibliography

1. Neogi, T. (2013). The epidemiology and impact of pain in osteoarthritis. Osteoarthritis and Cartilage, 21(9), 1145–1153.
2. Lee, D. H., Kim, S. J., Kim, S. A., & Ju, G.-I. (2022). Past, present, and future of cartilage restoration: From localized defect to arthritis. Knee Surgery & Related Research, 34(1), 1.
3. Costello, K. E., Felson, D. T., Neogi, T., Segal, N. A., Lewis, C. E., Gross, K. D., Nevitt, M. C., Lewis, C. L., & Kumar, D. (2021). Ground reaction force patterns in knees with and without radiographic osteoarthritis and pain: Descriptive analyses of a large cohort (the Multicenter Osteoarthritis Study). Osteoarthritis and Cartilage, 29(8), 1138–1146.
4. Fürmetz, J., Patzler, S., Wolf, F., Degen, N., Prall, W. C., Soo, C., Böcker, W., & Thaller, P. II. (2020). Tibial and femoral osteotomies in varus deformities—Radiological and clinical outcome. BMC Musculoskeletal Disorders, 21(1), 201.
5. Matsumoto, T., Hashimura, M., Takayama, K., Ishida, K., Kawakami, Y., Matsuzaki, T., Nakano, N., Matsushita, T., Kuroda, R., & Kurosaka, M. (2015). A radiographic analysis of alignment of the lower extremities—Initiation and progression of varus-type knee osteoarthritis. Osteoarthritis and Cartilage, 23(2), 217–223.
6. Wang, J. H., Shin, J. M., Kim, H. H., Kang, S.-H., & Lee, B. H. (2017). Discrepancy of alignment in different weight bearing conditions before and after high tibial osteotomy. International Orthopaedics, 41(1), 85–92.
7. Lin, Y.-II., Chang, F.-S., Chen, K.-II., Huang, K.-C., & Su, K.-C. (2018). Mismatch between femur and tibia coronal alignment in the knee joint: Classification of five lower limb types according to femoral and tibial mechanical alignment. BMC Musculoskeletal Disorders, 19(1), 411.

8. Zampogna, B., Vasta, S., Amendola, A., Uribe-Echevarria Marbach, B., Gao, Y., Papalia, R., & Denaro, V. (2015). Assessing Lower Limb Alignment: Comparison of Standard Knee Xray vs Long Leg View. The Iowa Orthopaedic Journal, 35, 49–54.
9. Choi, H.-U., Kim, D.-H., Lee, S.-W., Choi, B.-C., & Bae, K.-C. (2022). Comparison of Lower-Limb Alignment in Patients with Advanced Knee Osteoarthritis: EOS Biplanar Stereoradiography versus Conventional Scanography. Clinics in Orthopedic Surgery, 14(3), 370–376.
10. Pan, S., Huang, C., Zhang, X., Ruan, R., Yan, Z., Li, Z., Pang, Y., Guo, K., & Zheng, X. (2022). Non-weight-bearing short knee radiographs to evaluate coronal alignment before total knee arthroplasty. Quantitative Imaging in Medicine and Surgery, 12(2), 1214–1222.
11. Karnuta, J. M., Luu, B. C., Roth, A. L., Haeberle, H. S., Chen, A. F., Iorio, R., Schaffer, J. L., Mont, M. A., Patterson, B. M., Krebs, V. E., & Ramkumar, P. N. (2021). Artificial Intelligence to Identify Arthroplasty Implants From Radiographs of the Knee. The Journal of Arthroplasty, 36(3), 935–940.
12. Joseph, G. B., McCulloch, C. E., Nevitt, M. C., Link, T. M., & Sohn, J. H. (2022). Machine learning to predict incident radiographic knee osteoarthritis over 8 Years using combined MR imaging features, demographics, and clinical factors: Data from the Osteoarthritis Initiative. Osteoarthritis and Cartilage, 30(2), 270–279.
13. Ko, S., Pareek, A., Ro, D. H., Lu, Y., Camp, C. L., Martin, R. K., & Krych, A. J. (2022). Artificial intelligence in orthopedics: Three strategies for deep learning with orthopedic specific imaging. Knee Surgery, Sports Traumatology, Arthroscopy: Official Journal of the ESSKA, 30(3), 758–761.
14. Ye, Q., Shen, Q., Yang, W., Huang, S., Jiang, Z., He, L., & Gong, X. (2020). Development of automatic measurement for patellar height based on deep learning and knee radiographs. European Radiology, 30(9), 4974–4984.

15. Moon, H.-D., Choi, H.-G., Lee, K.-J., Choi, D.-J., Yoo, H.-J., & Lee, Y.-S. (2021). Can Deep Learning Using Weight Bearing Knee Anterio-Posterior Radiograph Alone Replace a Whole-Leg Radiograph in the Interpretation of Weight Bearing Line Ratio? Journal of Clinical Medicine, 10(8), 1772.
16. Qorchi, S., Vray, D., & Orkisz, M. (2021). Estimating Arterial Wall Deformations from Automatic Key-Point Detection and Matching. Ultrasound in Medicine & Biology, 47(5), 1367–1376.
17. Gao, J., & Yang, T. (2022). Research on Real-Time Face Key Point Detection Algorithm Based on Attention Mechanism. Computational Intelligence and Neuroscience, 2022, 6205108.
18. Qi, J., Du, J., Siniscalchi, S. M., Ma, X., & Lee, C.-H. (2020). On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression. IEEE Signal Processing Letters, 27, 1485–1489.
19. Lee, Y. S., Lee, B. K., Kwon, J. H., Kim, J. I., Reyes, F. J. V., Suh, D. W., & Nha, K.-W. (2014). Serial assessment of weight-bearing lower extremity alignment radiographs after open-wedge high tibial osteotomy. Arthroscopy: The Journal of Arthroscopic & Related Surgery: Official Publication of the Arthroscopy Association of North America and the International Arthroscopy Association, 30(3), 319–325.
20. Oh, S.-M., Bin, S.-I., Kim, J.-Y., Lee, B.-S., & Kim, J.-M. (2020). Short knee radiographs can be inadequate for estimating TKA alignment in knees with bowing. Knee Surgery & Related Research, 32(1), 9.
21. Sgroi, M., Faschingbauer, M., Reichel, H., & Kappe, T. (2016). Can the frontal tibiofemoral alignment be assessed on anteroposterior knee radiographs? Journal of Orthopaedics and Traumatology: Official Journal of the Italian Society of Orthopaedics and Traumatology, 17(4), 339–343.
22. Wang, R., Zeng, L., Wu, S., Cao, W., & Wong, K. (2020). Illumination-Invariant Feature Point Detection Based on Neighborhood Information. Sensors (Basel, Switzerland), 20(22), 6630.

23. Sun, Y., Xing, Y., Zhao, Z., Meng, X., Xu, G., & Hai, Y. (2022). Comparison of manual versus automated measurement of Cobb angle in idiopathic scoliosis based on a deep learning keypoint detection technology. European Spine Journal: Official Publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society, 31(8), 1969–1978.
24. Hsu, W., & Elmore, J. G. (2019). Shining Light Into the Black Box of Machine Learning. Journal of the National Cancer Institute, 111(9), 877–879.
25. Yi, P. H., Wei, J., Kim, T. K., Sair, H. I., Hui, F. K., Hager, G. D., Fritz, J., & Oni, J. K. (2020). Automated detection & classification of knee arthroplasty using deep learning. The Knee, 27(2), 535–542.
26. Perka, N., Kopf, S., & Hommel, H. (2019). A whole leg radiograph is not necessary for postoperative determination of the mechanical leg axis after total knee arthroplasty. Archives of Orthopaedic and Trauma Surgery, 139(10), 1455–1460.
27. Yoon, T.-H., Choi, C. H., Kim, S.-J., Kim, S.-H., Kim, N.-H., & Jung, M. (2019). Effect of Medial Open-Wedge High Tibial Osteotomy on the Patellofemoral Joint According to Postoperative Realignment. The American Journal of Sports Medicine, 47(8), 1863–1873.
28. Richardson, J. P., Smith, C., Curtis, S., Watson, S., Zhu, X., Barry, B., & Sharp, R. R. (2021). Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digital Medicine, 4(1), 140.

Abstract

배경: 단순 슬관절 방사선 사진을 사용하여 체중 지지선(WBL) 비율을 예측하려는 시도는 이전부터 있었다. 이번 연구에서는 컨볼루션 신경망(CNN)을 사용하여 WBL 비율을 정량적으로 예측하는 데 중점을 두었다.
방법: 2003년 3월부터 2021년 12월까지 2410명의 환자에 대해 시행한 4790건의 슬관절 전후방 방사선 사진을 계층화된 무작위 추출을 사용하여 무작위로 선택했다. 데이터는 전문가에 의해 주석이 달린 4점으로 10픽셀 마진을 두고 처리하였다. 이 모델은 경골 고평부의 양측 말단, 및 체중 지지선의 양 말단을 예측하였다. 모델의 결과 값은 픽셀 단위와 WBL 오류 비율의 두 가지 방법으로 분석하였다.
결과: 평균 정확도(MA)는 유효성 검사 및 테스트 세트 모두에서 2픽셀 단위를 사용했을 때 약 0.5에서 6픽셀을 사용하여 약 0.8로 확인되었다. 경골 고평부 길이를 100%로 했을 때 평균 정확도(MA)는 검증 및 테스트 세트 모두에서 1%를 오류비율에서 약 0.1에서 5% 시 약 0.5로 증가했다.
결론: 단순 슬관절 전후방 방사선 사진을 이용한 특정 점 학습을 통해 하지 정렬을 예측하는 딥러닝(Deep learning) 기반 키 포인트 감지 알고리즘은 하지 전장 방사선 사진을 사용한 직접 측정의 정확도와 유사한 정확도를 보여주었다. 이 알고리즘을 사용하여 단순 슬관절 전후방 방사선 사진을 사용한 WBL 비율 예측은 일차의료기관에서 진료 중인 골관절염 환자의 하지 정렬을 진단하는 데 유용하게 사용될 수 있다.