

RESEARCH ARTICLE

Open Access



Development of an artificial intelligence model for predicting implant size in total knee arthroplasty using simple X-ray images

Yeuni Yu^{1†}, Yoon Jae Cho^{2,3†}, Sohee Park^{4,5}, Yun Hak Kim^{6,7*} and Tae Sik Goh^{2,3*}

Abstract

Background Accurate estimation of implant size before surgery is crucial in preparing for total knee arthroplasty. However, this task is time-consuming and labor-intensive. To alleviate this burden on surgeons, we developed a reliable artificial intelligence (AI) model to predict implant size.

Methods We enrolled 714 patients with knee osteoarthritis who underwent total knee arthroplasty from March 2010 to February 2014. All surgeries were performed by the same surgeon using implants from the same manufacturer. We collected 1412 knee anteroposterior (AP) and lateral view x-ray images and retrospectively investigated the implant size. We trained the AI model using both AP and lateral images without any clinical or demographic information and performed data augmentation to resolve issues of uneven distribution and insufficient data. Using data augmentation techniques, we generated 500 images for each size of the femur and tibia, which were then used to train the model. Using data augmentation techniques, we generated 500 images for each size of the femur and tibia, which were then used to train the model. We used ResNet-101 and optimized the model with the aim of minimizing the cross-entropy loss function using both the Stochastic Gradient Descent (SGD) and Adam optimizer.

Results The SGD optimizer achieved the best performance in internal validation. The model showed micro F1-score 0.91 for femur and 0.87 for tibia. For predicting within \pm one size, micro F1-score was 0.99 for femur and 0.98 for tibia.

Conclusion We developed a deep learning model with high predictive power for implant size using only simple x-ray images. This could help surgeons reduce the time and labor required for preoperative preparation in total knee arthroplasty. While similar studies have been conducted, our work is unique in its use of simple x-ray images without any other data, like demographic features, to achieve a model with strong predictive power.

Keywords Knee arthroplasty, Radiographic analysis, Deep learning, Predictive modeling

[†]Yeuni Yu and Yoon Jae Cho contributed equally to this work as first authors.

*Correspondence:

Yun Hak Kim
yunhak10510@pusan.ac.kr
Tae Sik Goh
taesikgoh@pusan.ac.kr

¹Biomedical Research Institute, School of Medicine, Pusan National University, Yangsan, Republic of Korea

²Department of Orthopaedic Surgery, School of Medicine, Pusan National University, Busan, Republic of Korea

³Biomedical Research Institute, Pusan National University Hospital, Busan, Republic of Korea

⁴Convergence Medical Sciences, Pusan National University, Yangsan, Republic of Korea

⁵Data Science Center, Insilicogen, Inc, Yongin-si, Korea

⁶Department of Biomedical Informatics, School of Medicine, Pusan National University, Yangsan, Republic of Korea

⁷Department of Anatomy, School of Medicine, Pusan National University, Yangsan, Republic of Korea



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

Introduction

Preoperative templating in total knee arthroplasty is essential for achieving satisfactory outcomes. In particular, accurate size prediction allows surgeons to avoid size mismatch or tibial overhang [1]. Appropriate size prediction can also improve preparedness for unexpected situations during surgery (e.g., implant or equipment contamination) through the preparation of additional implants and equipment. Furthermore, it could allow hospitals to manage their inventory more efficiently [2, 3]. Despite the numerous benefits of preoperative templating for both surgeons and patients, it is a time-consuming task, and the results can vary depending on the individual performing it. Therefore, numerous efforts have been made to carry out preoperative templating more accurately and conveniently [4–6].

To improve the accuracy of templating, three-dimensional (3D) methods have been explored, but conventional templating using radiographs is still more commonly used. This is because 3D templating is more labor-intensive, and the difference in accuracy between 3D and conventional templating is not statistically significant [7]. Instead, the researches using artificial intelligence (AI) technology to predict the size of the implants more quickly and accurately have been conducted. However, most of these studies have relied solely on demographic features (age, sex, BMI, weight, height etc.) to predict implant size [8–10]. And only one research has used radiographs and demographic features simultaneously [11]. Demographic features such as gender, height, and weight provide quantitative information sufficient to estimate an individual's bone size, and models using these features for predicting implant size have reported accuracies exceeding 80%. However, to date, there have been no reports of a reliable implant size prediction model developed using only X-rays, without any demographic or scaling information provided. Can an artificial neural network successfully predict the appropriate implant size using only X-rays in an environment with minimal scaling information, such as demographic features? Motivated by this curiosity, we initiated our study and created a reliable AI model using a convolutional neural network (CNN), which yielded satisfactory results [12].

Methods

Data acquisition and preparation

From March 2010 to February 2014, a total of 714 patients with knee osteoarthritis who underwent total knee arthroplasty were enrolled in the study. All surgeries were performed by the same surgeon and exclusively utilized the NexGen® product from Zimmer Biomet, specifically the posterior stabilized (PS) type. Anteroposterior (AP) and lateral view x-ray images of the patients' knees were collected, and the inserted implant sizes

were investigated retrospectively. During the surgery, the implant size was determined using the sizing trials provided by Zimmer Biomet. For both the femur and tibia, trial implants were applied to assess factors such as rotation and overhang, ultimately guiding the decision on the appropriate implant size. Cases requiring size adjustments due to rotation or gap balance issues were excluded from the study at the outset.

All images were converted into the grayscale Joint Photographic Experts Group (JPEG) format and resized each image to a resolution of 224×224 pixels to make it compatible with the original ResNet network architecture [13]. 20% of the total images were randomly extracted for the test images.

Deep learning model development

In this study, to predict the implant size, ResNet-101 was used to predict the implant size. The ResNet architecture, which involves repeated convolutional and pooling layers, was trained using preprocessed images as input training data. The first layer with max pooling about input images was extracted to feature map through convolution of 7×7 size kernel. The kernel of size 1×1, 3×3, and 1×1 performs 64, 128, 256, and 512 convolution operations, and finally average pooling (Fig. 1). Femur and tibia dataset were divided into 80% as a training dataset and 20% as a test dataset.

To increase the number of training images, the images were augmented using geometric transformations. The images of class which has the most numerous images, like femur insert size D and tibia insert size 1, were only horizontally flipped. ImageFilter in Pillow packages was applied to the other images, respectively: blur, contour, detail, edge enhance, edge enhance more, emboss, find edges, sharpen, smooth, and smooth more. Images were rotated to 45° to fulfill 500 images.

Using images generated through data augmentation, we trained the femur model with 2,000 images and the tibia model with 2,500 images, utilizing the 101-layer ResNet architecture (Fig. 2). The 101-layer Resnet models were evaluated using 5-fold cross validation. The learning rate of the 101-layer ResNet that has been observed as 0.0001 and batch size is set to 20. In this work, the Stochastic Gradient Descent (SGD) optimizer used to reduce loss performance. All images were resized according to the selected model requirements and the results were collected. Similarly, all the images are resized by 224×224 before fed into the neural network ResNet.

Based on the training dataset, the model was optimized with a target of minimizing the cross-entropy loss function using the SGD optimizer. The final output score was calculated as the sum value of the model. Cross entropy loss was defined as:

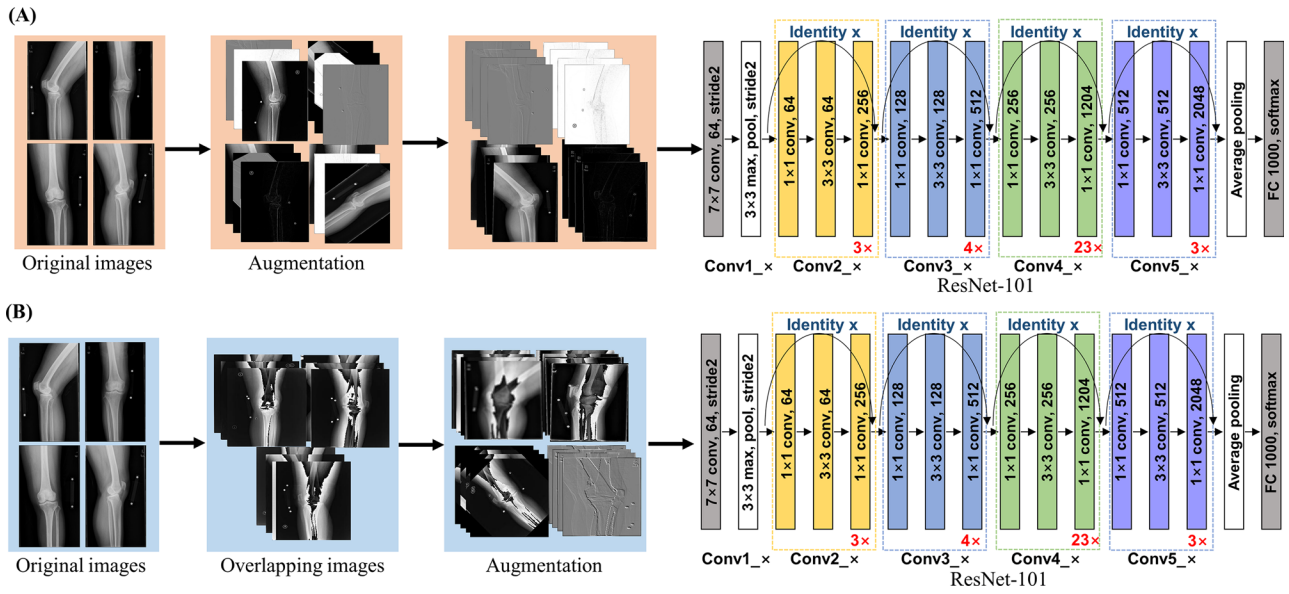


Fig. 1 Implant size detection framework using ResNet-101 (A) Training workflow based on 101-layer residual network. Inputted front or lateral X-ray images independently after augmentation. (B) Overlapped X-ray images in pairs as input images then augmentation

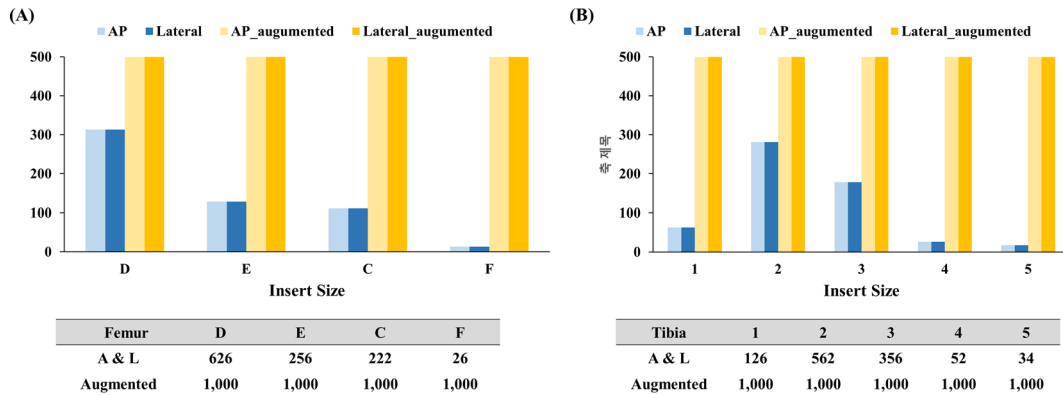


Fig. 2 Graphical representation illustrating the alteration in image quantity after augmentation. (A) 1,130 each image both AP and lateral combined and amplified to 2,000 for femur dataset. (B) Same images were augmented to 2,500 for tibia dataset

$$l_n = \sum_{c=1}^C w_c \log \frac{\exp(x_{n,c})}{\sum_{i=1}^C \exp(x_{n,i})} y_{n,c}$$

$$l(x, y) = \begin{cases} \frac{\sum_{n=1}^N l_n}{N}, & \text{if reduction} = 'mean' \\ \sum_{n=1}^N l_n, & \text{if reduction} = 'sum' \end{cases}$$

Where C is the number of classes, x is the input, y is the target, w is the weight and N spans the minibatch dimension. The maximal number of training epochs was set to be 50. The initial learning rate was set to 0.001 and would be decayed by 0.1 every 10 epochs. Continuous variables were described as the mean ± standard deviation. The performance of the classifier is calculated using test samples. The confusion matrix is calculated for evaluating classifier performances. It is generated by comparing the responses of the classification algorithm to the test

set with the actual values in the data set. Micro-averaging F1 score is based on the cumulative True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) of the dataset.

Micro-Average Precision (P_{micro}) is calculated as:

$$P_{micro} = \frac{TP_{total}}{TP_{total} + FP_{total}}$$

Micro-Average Recall (R_{micro}) is calculated as:

$$R_{micro} = \frac{TP_{total}}{TP_{total} + FN_{total}}$$

Micro F1 score ($F1_{micro}$) is calculated as:

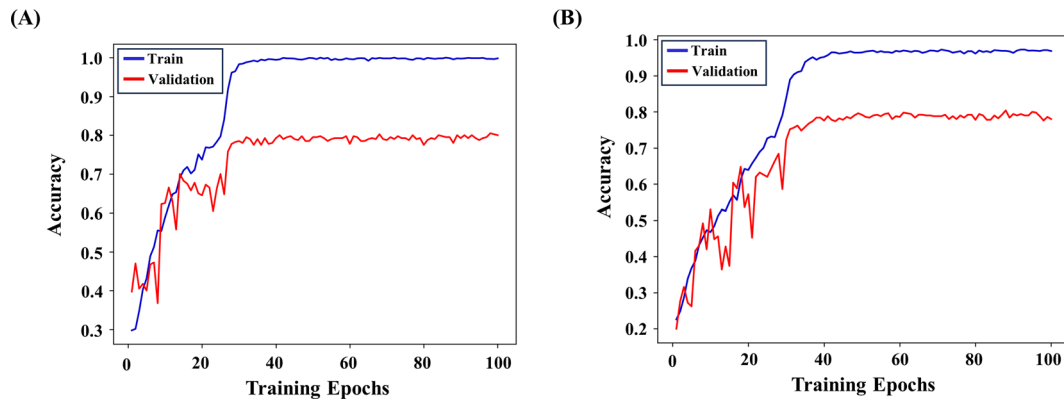


Fig. 3 Accuracy curve based on 100 epochs. (A) Training and validation accuracy with 101-layer ResNet for femur classification model. (B) Training and validation accuracy with 101-layer ResNet for tibia classification model

Table 1 The prediction accuracy of each model used in this study

	AP	Lateral	Overall
Femur	90.7%	90.78%	90.78%
Tibia	85.11%	87.94%	86.17%

AP=Anteroposterior

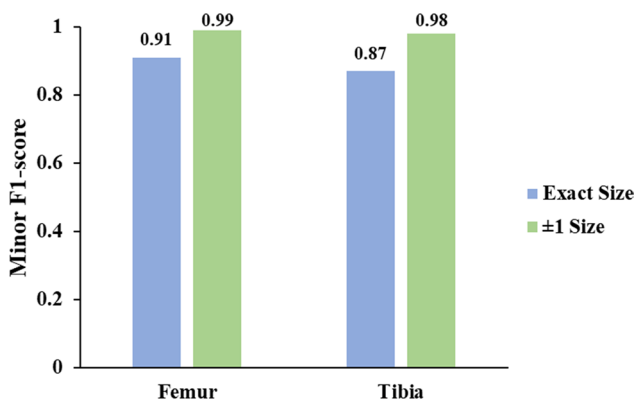


Fig. 4 The graph of micro F1-score. The blue chart graph is score of predicting exact size. The orange chart graph is score of predicting within ±one size

$$F1_{micro} = 2 \times \frac{P_{micro} \times R_{micro}}{P_{micro} + R_{micro}}$$

System requirement

The information of the computer used in this study was as follows: Intel Xeon 4144 (2.20 GHz), 128 gigabytes of RAM, and t GeForce RTX 3090 Ti Ubuntu 18.04 LTS. Model development was performed by Python (version 3.7.11) with libraries of torch (version 1.8.0), and statistical analyses were performed with a R software.

Results

Three versions of the predictive model were developed. First, models using only AP or lateral images were created. Second, a model using both AP and lateral images without discrimination was developed. The accuracy of these models was then compared.

Training accuracy of all models was around 80% (Fig. 3). The model using the AP images showed the lowest accuracy in both femur (90.7%) and tibia (85.11%) size prediction. (Table 1) Although it was a slight difference, the model using the lateral images had the highest accuracy (femur 90.78%, tibia 87.84%). The model using overall images showed good accuracy in femur prediction same of lateral model. But, in tibia prediction, it showed slight lower accuracy (86.17%) than lateral model.

The final model for predicting exact implant size demonstrated a micro F1-score 0.91 for femur and 0.87 for tibia. For predicting within ±one size, the micro F1-score was 0.99 for femur and 0.98 for tibia (Fig. 4).

Discussion

The AI model, which uses only simple radiographs to figure out the implant model, has already been extensively researched and produces high performance [9, 14–16]. However, there is currently no report on a model using only radiographs to predict component size. Most reports used demographic features only or combination with other information. The significant correlation between demographic feature and implant size has been reported, and there has also been a size prediction model using the Baysien model [17]. In 2022, Kunze KN et al. [9]. showed more than 95% accuracy in predicting ±one size of the implant using age, sex, height, weight and BMI of 11,777 patients. However, the accuracy of predicting exact size was only 42.2%. Only 1 study, which used radiographs, Yue Y et al. [11]. developed an error correct output coding model using demographic features and radiographs

of 308 patients in 2022 and reported accuracy of 86.27% and 88.23% in prediction of femur and tibia size.

During the study, prediction accuracies of 70.59% and 68.72% were achieved using only simple radiographs. Notably, our prediction model demonstrated exceptional performance with a micro F1-score exceeding 0.98 when allowing for a margin of \pm one size, solely based on simple X-rays without any additional information. Even the model for predicting exact implant size have shown micro F1-score 0.91 for femur and 0.87 for tibia. The difference between study of Yue Y and this study are (1) the number of our patients was larger and (2) only data augmentation was used without applying transfer learning. It was believed that this approach could result in a model with better accuracy. In 2023, According to a report of Riechelmann F et al [18], even if surgeons conduct 2D size template, the exact size match was observed only 34% of cases and size match within \pm one size in 57.5%. Considering the report by Riechelmann F and the results from other AI studies, these findings are notably impressive. Additionally, the model demonstrated the highest performance when using lateral images, highlighting the effectiveness of this approach. The superior performance of the model using lateral images is likely due to the increased sensitivity of surgeons performing TKA to anteroposterior size mismatch compared to mediolateral size mismatch.

This study showed satisfactory results in predicting the component size of TKA patients performed by a single institution, one surgeon. It could be considered the model incorporates surgeon's implant preference. It could be considered as a lack of diversity or data volume in this study, but it also means that each individual who follow our method can get a more suitable predictive model for their own. Moreover, this study is made with a relatively simple algorithm using only simple radiographs, so everyone can easily follow it and can get a reliable size prediction model. This is expected to relieve the surgeon's labor and time for preoperative planning.

As mentioned above, the limitation of this study is that it has been performed in one institution and one surgeon. The radiographs of about 700 patients gathered from a single institution couldn't represent the population sufficiently. (1) Since the current model was trained only by small Asians, size mismatch would occur when it has to predict the size out of the learning range. (2) The prediction accuracy would be affected by the characteristics of the imaging device and the imaging method, so additional algorithm modifications are required to apply the current model to a multicenter study.

Conclusion

A deep learning model with high predictive power for implant size was developed using a small number of patient groups. This model is expected to contribute significantly to preoperative planning and implant preparation for total knee arthroplasty. While similar studies have been conducted, this model uniquely achieves sufficient predictive power with only radiographs.

Acknowledgements

Not applicable.

Author contributions

The manuscript has been read and approved by all the Authors. YY contributed to the analysis and interpretation of data and drafting the article. YJC contributed to the interpretation of data and drafting the article. SHP contributed to the analysis and interpretation of data. YHK contributed to the conception and design of the study and revising the article critically for important intellectual content. TSG contributed to the conception and design of the study, acquisition of data, and revising the article critically for important intellectual content.

Funding

This work was supported by Biomedical Research Institute Grant (20210017), Pusan National University Hospital.

Data availability

The datasets generated or analyzed during the study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

This work has not been published previously, it is not under consideration for publication elsewhere, and, if accepted, it will not be published elsewhere in the same form, in English or in any other language, without the written consent of the Publisher.

Competing interests

The authors declare no competing interests.

Received: 9 July 2024 / Accepted: 19 August 2024

Published online: 27 August 2024

References

1. Blackley HR, Howell GE, Rorabeck CH. Planning and management of the difficult primary hip replacement: preoperative planning and technical considerations. *Instr Course Lect*. 2000;49:3–11.
2. Cichos KH, Hyde ZB, Mabry SE, Ghanem ES, Brabston EW, Hayes LW, McGwin G Jr, Ponce BA. Optimization of Orthopedic Surgical Instrument trays: lean principles to reduce fixed operating room expenses. *J Arthroplasty*. 2019;34:2834–40.
3. Dyas AR, Lovell KM, Balentine CJ, Wang TN, Porterfield JR Jr, Chen H, Lindeman BM. Reducing cost and improving operating room efficiency: examination of surgical instrument processing. *J Surg Res*. 2018;229:15–9.
4. Karnuta JM, Luu BC, Roth AL, Haeberle HS, Chen AF, Iorio R, Schaffer JL, Mont MA, Patterson BM, Krebs VE, Ramkumar PN. Artificial Intelligence to identify arthroplasty implants from radiographs of the knee. *J Arthroplasty*. 2021;36:935–40.
5. Morsy AM, Elbana EG, Mostafa AG, Edward MA, Hafez MA. Comparison of Functional Outcome of Total and Unicompartamental Knee Arthroplasty Using Computer-Assisted Patient-Specific Templating. *Adv Orthop*. 2021 (2021) 5524713.

6. Levine B, Fabi D, Deirmengian C. Digital templating in primary total hip and knee arthroplasty. *Orthopedics*. 2010;33:797.
7. Kobayashi A, Ishii Y, Takeda M, Noguchi H, Higuchi H, Toyabe S. Comparison of analog 2D and digital 3D preoperative templating for predicting implant size in total knee arthroplasty. *Comput Aided Surg*. 2012;17:96–101.
8. Kunze KN, Polce EM, Patel A, Courtney PM, Levine BR. Validation and performance of a machine-learning derived prediction guide for total knee arthroplasty component sizing. *Arch Orthop Trauma Surg*. 2021;141:2235–44.
9. Kunze KN, Polce EM, Patel A, Courtney PM, Sporer SM, Levine BR. Machine learning algorithms predict within one size of the final implant ultimately used in total knee arthroplasty with good-to-excellent accuracy. *Knee Surg Sports Traumatol Arthrosc*. 2022;30:2565–72.
10. Sershon RA, Li J, Calkins TE, Courtney PM, Nam D, Gerlinger TL, Sporer SM, Levine BR. Prospective validation of a demographically based primary total knee arthroplasty size calculator. *J Arthroplasty*. 2019;34:1369–73.
11. Yue Y, Gao Q, Zhao M, Li D, Tian H. Prediction of knee prosthesis using patient gender and BMI with non-marked X-Ray by Deep Learning. *Front Surg*. 2022;9:798761.
12. Zhang Y, Hong D, McClement D, Oladosu O, Pridham G, Slaney G. Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging. *J Neurosci Methods*. 2021;353:109098.
13. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
14. Lambrechts A, Wirix-Speetjens R, Maes F, Van Huffel S. Artificial Intelligence Based Patient-Specific Preoperative Planning Algorithm for Total Knee Arthroplasty. *Front Robot AI*. 2022;9:840282.
15. Yi PH, Wei J, Kim TK, Sair HI, Hui FK, Hager GD, Fritz J, Oni JK. Automated detection & classification of knee arthroplasty using deep learning. *Knee*. 2020;27:535–42.
16. Patel R, Thong EHE, Batta V, Bharath AA, Francis D, Howard J. Automated identification of orthopedic implants on radiographs using deep learning. *Radiol Artif Intell*. 2021;3:e200183.
17. Blevins JL, Rao V, Chiu YF, Lyman S, Westrich GH. Predicting implant size in total knee arthroplasty using demographic variables. *Bone Joint J*, 102–b (2020) 85–90.
18. Riechelmann F, Lettner H, Mayr R, Tandogan R, Dammerer D, Liebensteiner M. Imprecise prediction of implant sizes with preoperative 2D digital templating in total knee arthroplasty. *Arch Orthop Trauma Surg*. 2023;143:4705–11.

Publisher's note

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