BMC Musculoskeletal Disorders

RESEARCH

Open Access

Development and accuracy of an artificial intelligence model for predicting the progression of hip osteoarthritis using plain radiographs and clinical data: a retrospective study

Ryo Hidaka[1*], Kenta Matsuda[1], Takashi Igari[2], Shu Takeuchi[2], Yuichi Imoto[1], Satoshi Yagi[2] and Hirotaka Kawano[1]

Abstract

Background Predicting the progression of hip osteoarthritis (OA) remains challenging, and no reliable predictive method has been established. This study aimed to develop an artificial intelligence (AI) model to predict hip OA progression via plain radiographs and patient data and to determine its accuracy.

Methods This retrospective study utilized anteroposterior pelvic radiographs of consecutive patients with hip OA who underwent primary unilateral total hip arthroplasty. Radiographs diagnosed with Kellgren–Lawrence (KL) grade 0–2 were extracted from 361 patients and 1697 images. This AI model was developed to predict whether OA would progress from KL grade 0–2 to KL grade ≥ 3 within *n* years ($n = 3, 4, 5$). A gradient-boosting decision tree approach was utilized according to feature extractions obtained by a convolutional neural network from radiographs and patient data (height, body weight, sex, age, and KL grade given by an orthopedic surgeon) with five-fold cross-validation. The model performance was assessed using accuracy, specificity, sensitivity, and the area under the receiver operating characteristic curve (AUC).

Results The mean accuracy, specificity, sensitivity, and AUC of our prediction model were, respectively, 81.8%, 88.0%, 66.7%, and 0.836 for 3 years; 79.8%, 85.0%, 71.6%, and 0.836 for 4 years; and 78.5%, 80.4%, 76.9%, and 0.846 for 5 years.

Conclusions The proposed AI model performed adequately in predicting hip OA progression and may be clinically applicable with additional datasets and validation.

Keywords Hip, Osteoarthritis, Artificial intelligence, Kellgren–Lawrence classification, Convolutional neural network, Deep learning, Machine learning

*Correspondence:
Ryo Hidaka
qqdr59qd@yahoo.co.jp
[1]Department of Orthopaedic Surgery, Teikyo University School of Medicine, 2-11-1 Kaga, Itahashi-ku, Tokyo 173-8606, Japan
[2]Technology Administration Supervision, FUJISOFT INCORPORATED, 1-1 Sakuragi-cho, Naka-ku, Yokohama-shi, Kanagawa 231-8008, Japan

Background

An estimated 242 million people worldwide face daily limitations due to symptomatic hip and/or knee osteoarthritis (OA) [1]. Hip OA prevalence in individuals over 45 years of age is approximately 10% [2, 3]. Conservative OA treatment includes education, weight loss, and exercise supplemented by non-steroidal anti-inflammatory drugs, corticosteroid injections, and adjunctive medications. For advanced symptomatic OA with structural damage, total joint arthroplasty effectively alleviates pain; however, it is costly, invasive, and affects a patient's quality of life [4]. A cohort study linked functional limitations due to hip and knee OA to increased mortality and serious cardiovascular events [5]. Predicting OA onset and progression is crucial, with studies showing older age, female sex, and overweight as risk factors for both hip and knee OA [1, 6]. Additionally, congenital and acquired anatomic hip morphologies (e.g., hip dysplasia and femoroacetabular impingement syndrome) are risk factors for hip OA [7, 8]. However, the overall etiology of OA remains unclear, and knowledge regarding factors contributing to hip OA progression is lacking.

In recent years, the use of artificial intelligence (AI) to reduce the variability and human error in interpreting complex data has become increasingly widespread. Plain radiographs were used for the initial radiographic evaluation to diagnose or assess hip OA severity. The Kellgren–Lawrence (KL) grade [9] is widely used to standardize radiographic hip OA severity. However, the semi-quantitative KL grade is highly dependent on practitioner subjectivity, and several studies have reported that the KL grade can be automated via deep learning in knee OA [10–12].

Machine learning, based on radiographs, magnetic resonance imaging (MRI), and biomarkers, has been widely used in early OA diagnosis [13]. A multitask deep learning model for grading radiographic hip OA features has been reported in hip OA [14]. Their prospective observational study developed a deep learning model utilizing standing pelvic radiographs of 4368 participants in the Osteoarthritis Initiative to evaluate femoral osteophytes, acetabular osteophytes, joint crevice narrowing at four levels (normal, mild, moderate, and severe), and subchondral bone sclerosis and subchondral cyst formation at two levels (with and without).

AI models in knee OA have been reported to predict OA progression and the risk of total knee arthroplasty [15, 16]. The deep learning model on knee radiographs better predicted the risk of total knee arthroplasty than did binary outcome models using standard grading systems [15]. A multimodal machine learning model (deep learning and a gradient boosting decision tree), based on knee OA progression prediction from radiographs and clinical data, showed better results than did a logistic regression model [16]. The AI model using machine learning, and not deep learning, for hip OA was developed based on the original Shape-Score automatically derived from radiographs in the Cohort Hip and Cohort Knee (CHECK) study; the model predicts OA progression or progression to total hip arthroplasty (THA) after 8 years [17]. A deep learning algorithm predicting the necessity for THA within 3 months has been developed for hip OA [18]. Compared to knee OA, there are few studies on AI, and no AI model using deep learning has been utilized to predict hip OA progression.

This study aimed to develop an AI model for predicting hip OA progression using hip radiographs and patient demographics and determine its accuracy.

Methods

This study was approved by the Institutional Review Board of Teikyo University and conducted in accordance with the World Medical Association Declaration of Helsinki (IRB approval No. 19-152-5). All individuals included in this study provided written informed consent.

This retrospective study enrolled 500 patients with hip OA who underwent primary unilateral THA at our institution between October 2011 and March 2022. Anteroposterior (AP) pelvic radiographs in a non-weight-bearing (supine) position obtained from the first until the last visit were used for each patient, excluding images of hip fractures, post-periacetabular osteotomy, and pediatric skeletal diseases. All images were evaluated by a single hip orthopedic surgeon (RH) to determine the KL grade. Images diagnosed with KL grade 0–2 were extracted from 361 patients, and 1697 images were obtained. Among these, in 249 patients, a KL grade of 3 or higher within 3 years could be predicted, with a dataset of 915. The number of patients for whom a KL grade of 3 or higher within 4 years could be predicted was 223, with a dataset of 792. The number of patients for whom a KL grade of 3 or higher within 5 years could be predicted was 203, with a dataset of 703. To assess the intra- and inter-observer reliability of the KL grade, 40 hips were randomly selected and assessed by two orthopedic surgeons (RH and KM). The intra- and inter-observer reliability values were 0.94 (95% confidence interval [CI], 0.92–0.95) and 0.93 (95% CI, 0.91–0.95), respectively.

Demographics

Height, body weight, and sex were recorded during THA, whereas age was obtained from Digital Imaging and Communications in Medicine (DICOM) radiographic image files. The KL grades were classified on a scale of 0–4. (Table 1)

Table 1 Patient demographics

Variables	Values
Height, m	1.54 (0.075)
Body weight, kg	58.3 (11.7)
Female sex, %	89.3%
Age, years	67.3 (11.1)
KL grade, n	
0	880/1697
1	364/1697
2	453/1697

Values are presented as means (SD)

KL, Kellgren–Lawrence; SD, standard deviation

Algorithm design

The workflow of the algorithm is illustrated in Fig. 1. An NVIDIA Tesla V100 GPU with CUDA 11.3 was utilized to run the deep learning models. A combination of Python [19] (version 3.8.10, http://www.python.org), OpenCV [20] (version 4.6.0, https://opencv.org/), and Pydicom [21] (version 2.3.1, https://pydicom.github.io) was utilized to crop and resize the DICOM radiographic images.

A pretrained Faster R-CNN implemented in MMDetection [22] (https://github.com/open-mmlab/mmdetection) was trained to identify the hip regions of interest (ROI) from hip AP radiographs. Bounding boxes were placed at the femoral head center, around the anterior superior iliac spine on the upper side, around the pubic symphysis on the medial side, and distal to the lesser trochanter on the lower side. The ROI annotators used these rules to annotate datasets. A total of 499 hip AP radiographs (349 for training, 50 for validation, and 100 for testing) were used to develop the object-detection model. The left and right hip ROIs were detected and cropped, and incorrect ROIs were manually modified. Afterward, the left hip ROI was flipped to the right. The square hip ROI was resized to 224 pixels × 224 pixels.

EfficientNet-B4 [23] (version 0.7.1, https://github.com/lukemelas/EfficientNet-PyTorch), pretrained on the ImageNet dataset, was used as a feature extractor of the images. Cropped radiographs were normalized using the mean and standard deviation of the ImageNet dataset. The extracted features were the outputs of the Global Average Pooling layer, resulting in 1792-dimensional real-valued vectors.

Demographic data, which were preprocessed into a six-dimensional (four dimensions for height, body weight, age, and KL grade and two dimensions for sex, which is a one-hot encoded) real vector, were finally concatenated with the extracted image features, yielding 1798-dimensional real-valued vectors and inputs. These data were inputted into a gradient boosting decision tree, using the GradientBoostingClassifier class in the scikit-learn library [24], to predict OA progression to KL grade≥3 within n years. (n=3,4,5). In the initial stages of the study, the performance of four different algorithms—linear regression, random forest, gradient boosting decision tree, and support vector machine—was evaluated for

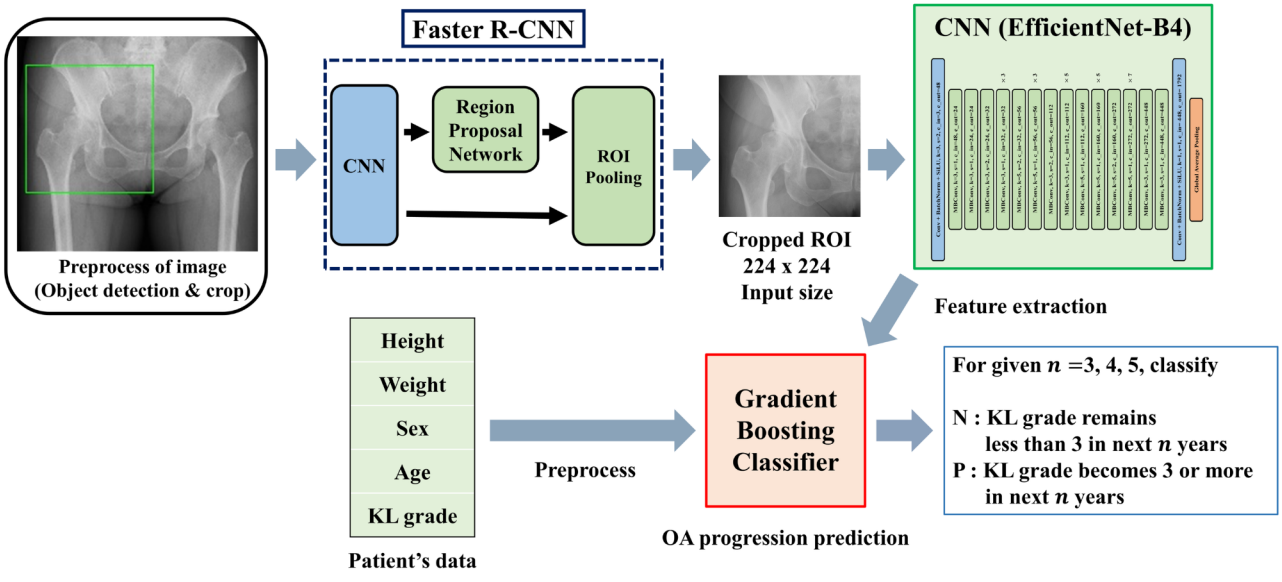

Fig. 1 Schematic representation of our artificial intelligence modal pipeline, predicting the progression of hip OA. A pretrained Faster R-CNN was trained to depict hip ROI from pelvic anteroposterior radiographs. First, input images detected the hip ROI, and the cropped the hip area as our target ROI was resized to 224 × 224 pixels. Afterwards, these cropped images were inputted into EfficientNet-B4 pretrained by ImageNet dataset. Feature extractions obtained by EfficientNet-B4 and patient data (height, body weight, sex, age, and KL grade given by an orthopedic surgeon) were inputted into a Gradient Boosting Classifier for predicting whether OA would progress to KL grade ≥ 3 within *n* years (*n* = 3, 4, 5). OA, osteoarthritis; ROI, regions of interest; KL, Kellgren-Lawrence

similar tasks. The gradient boosting decision tree was selected for this study due to its superior performance.

A five-fold cross-validation was used to train and evaluate our model. To prevent patients from joining both training and test data, the StratifiedGroupKFold class in the scikit-learn library was utilized to split the entire dataset and fix the random seed used for division to ensure reproducibility. Training data were normalized using the StandardScaler class in the scikit-learn library. The gradient boosting decision tree was trained using normalized training data and fixed random seeds to reproduce the training results. Test data were normalized using StandardScaler and fitted to the training data in the scikit-learn library.

Performance assessment

To assess the model's performance, the accuracy, specificity, sensitivity, and area under the receiver operating characteristic curve (AUC) of the model were calculated. Statistical analyses were performed using the Python library scikit-learn. The 95% CIs in the AUC were calculated using methods outlined in a previous study [25]. We retrieved the feature importance of the gradient boosting decision tree to identify the factors that contribute to the prediction of OA progression.

Results

The important demographics of patients were as follows: mean height, 1.54 m (SD 0.075); mean body weight, 58.3 kg (SD 11.7); sex, female (89.3%); and mean age, 67.3 years (SD 11.1). Of the 1697 images diagnosed as KL 0–2 at the first visit, 880 were KL grade 0, 364 were KL grade 1, and 453 were KL grade 2. (Table 1)

The mean accuracy, specificity, sensitivity, and AUC of our prediction model for OA progression to KL grade ≥ 3 within *n* years (*n* = 3,4,5) were, respectively, as follows: 81.8%, 88.0%, 66.7%, and 0.836 for 3 years, 79.8%, 85.0%, 71.6%, and 0.836 for 4 years, and 78.5%, 80.4%, 76.9%, and 0.846 for 5 years. (Table 2) The receiver-operating characteristic curves for each fold of cross-validation in our prediction model predicting OA progression to KL grade ≥ 3 within *n* years (*n* = 3,4,5) are summarized in Fig. 2. The analysis of feature importance showed a strong impact of KL grade in each fold of cross-validation in our prediction model for OA progression to KL grade ≥ 3 within n years (*n* = 3, 4, 5).

Discussion

This study developed an AI model to predict whether hip OA would progress from KL grade 0–2 to ≥ 3 within *n* years (*n* = 3,4,5), according to AP pelvic radiographs and patient data (height, body weight, sex, age, and KL grade). All prediction models for hip OA progression within 3–5 years were highly accurate, with an AUC of 0.83 or higher.

A cohort study reported an AI prediction model for hip OA [17]. They applied machine learning algorithms to provide the Shape-Score, a single value describing the risk for progression of KL grade ≥ 2 or THA at 8-year follow-up based solely on joint shape from hip radiographs. Their prediction model containing only the Shape-Score was comparable to a model combining patient

Table 2 Overview of our prediction model for OA progression to KL grade ≥ 3 within *n* years (*n* = 3,4,5), with performance results in each Fold of cross-validation and average values

		Accuracy (%)	Specificity (%)	Sensitivity (%)	AUC	95% CI
3 years	0	76.1	86.3	58.9	0.776	0.705–0.847
	1	82.9	89.4	69.0	0.887	0.829–0.946
	2	81.9	86.9	62.3	0.831	0.754–0.908
	3	85.1	90.2	70.8	0.859	0.789–0.929
	4	83.0	87.4	72.5	0.829	0.745–0.913
	average	81.8	88.0	66.7	0.836	
4 years	0	83.8	92.4	62.8	0.843	0.764–0.920
	1	74.5	82.1	68.3	0.803	0.741–0.865
	2	86.2	90.6	77.4	0.910	0.854–0.966
	3	75.4	84.0	64.4	0.787	0.716–0.859
	4	79.2	75.9	85.1	0.837	0.759–0.914
	average	79.8	85.0	71.6	0.836	
5 years	0	73.4	78.4	66.7	0.815	0.737–0.893
	1	79.7	88.4	74.7	0.874	0.813–0.936
	2	79.7	78.8	80.8	0.886	0.831–0.941
	3	75.6	71.4	78.9	0.784	0.706–0.863
	4	84.2	85.1	83.3	0.873	0.820–0.925
	average	78.5	80.4	76.9	0.846	

OA, osteoarthritis; KL, Kellgren–Lawrence; AUC, area under the receiver operating characteristic curve; CI, confidence interval

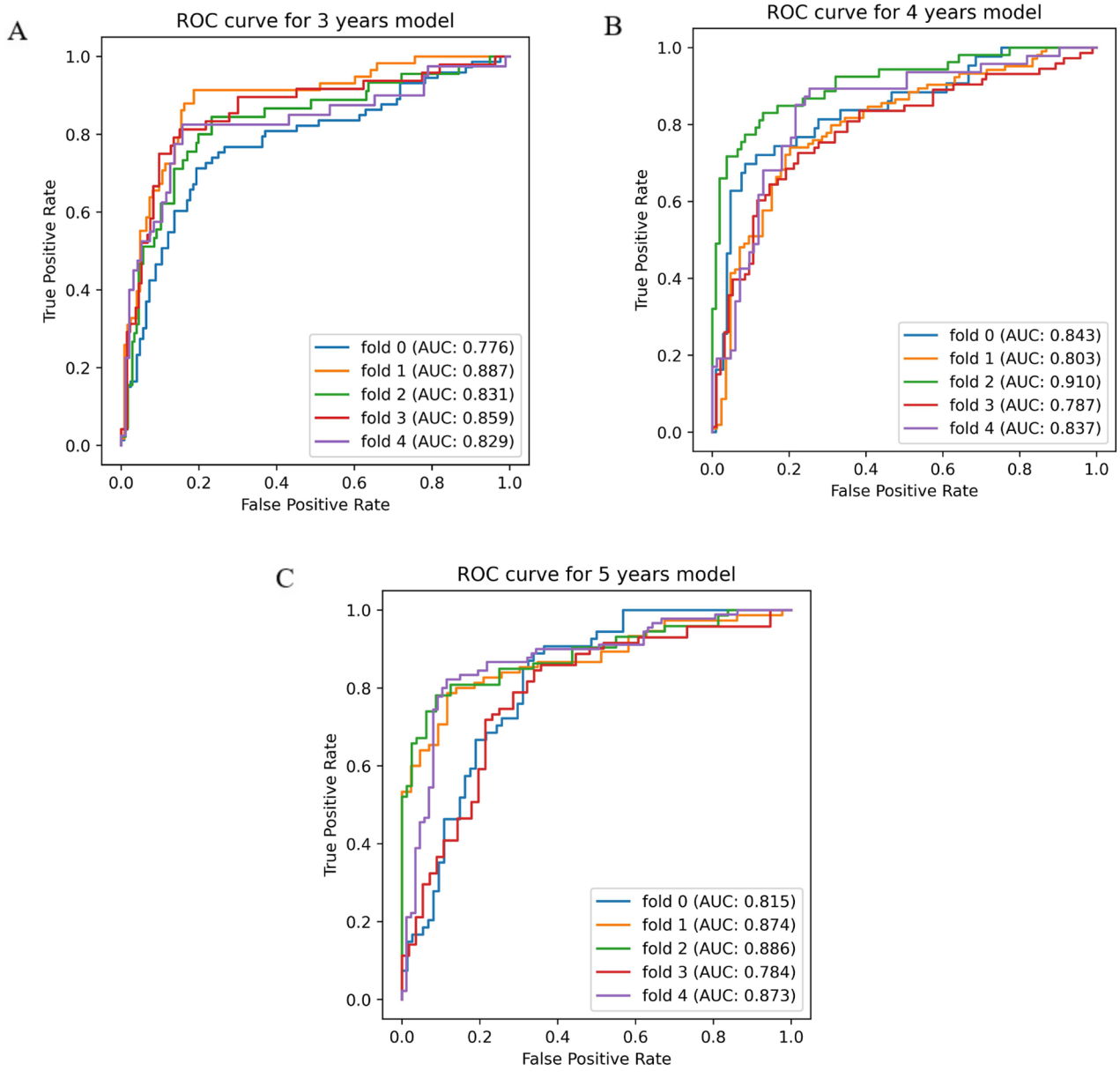

Fig. 2 Receiver operating curves for each fold of cross-validation in osteoarthritis prediction (progression of KL grade ≥ 3) models. KL, Kellgren–Lawrence; AUC, area under the receiver operating characteristics curve. **A**: Prediction model to determine the progression of KL-grade ≥ 3 within 3 years. **B**: Prediction model to determine the progression of KL-grade ≥ 3 within 4 years. **C**: Prediction model to determine the progression of KL-grade ≥ 3 within 5 years

demographics, clinical examination, and basic radiographic parameters (AUC 0.798 vs. 0.795). Adding the Shape-Score to the latter model improved the discriminative ability from an AUC of 0.795–0.863. These results suggested the significance of developing an AI model that includes both medical information and imaging data. The AUC of 0.836–0.846 in our prediction model was higher than that of the prediction model containing only the Shape-Score and comparable to the prediction model combining the Shape-Score and detailed patient information. Notably, the prediction model in this study provided sufficient model performance, according to hip radiographs and brief patient data.

The Shape-Score was developed based on hip joint shape with the risk of femoroacetabular impingement syndrome, such as cam morphology and increased acetabular coverage. However, the etiology of hip OA in Japan was acetabular dysplasia in most patients [26], and the Shape-Score may not be suitable for use in Japan. Additionally, the etiology and prevalence of hip OA differ among Asians, North Americans, and Europeans [3]. A systematic review of prediction models to estimate the future risk of osteoarthritis in the general

population indicated a reliance on data from a small pool of appropriate cohort datasets and concerns over general population applicability [27]. The AI model in this study was developed for patients with hip OA who were all Japanese. Because the OA population in this study does not necessarily reflect the epidemiology of the Japanese OA population, future studies are warranted to expand the data and validate this model in OA populations worldwide.

Two studies in the Rotterdam cohort-1 have reported the prediction model for the progression of KL grade ≥ 2 or THA without using AI. One study showed that the hip geometry from baseline pelvic radiographs has a moderate ability in predicting hip OA progression, which is largely independent of the predictive value of clinical risk factors (AUC = 0.67 [hip geometry], 0.66 [sex, age, body mass index] and combining both: AUC = 0.73) [28]. Another study developed a prediction model using demographics, urinary C-terminal cross-linked telopeptide of type II collagen levels, and radiographic parameters, including the Wiberg and alpha angles, to quantify acetabular coverage and cam morphology, respectively [29]. Their model showed an AUC of 0.82 in the Rotterdam cohort-1, 0.75 when validated in the Rotterdam cohort-2, and 0.69 when validated in the CHECK cohort. Their risk prediction models provided reasonable calibration in the second cohort of the Rotterdam study; however, the calibrations were poor for the CHECK cohort. The AUC of our prediction model using AI was comparable with that of the prediction model developed without AI. However, our model was not validated externally. In future studies, external validation should be performed to verify its applicability across large populations.

This study developed a prediction model to determine whether OA would progress to KL grade ≥ 3, excluding THA. In a cohort study from Research on Osteoarthritis/Osteoporosis Against Disability, the percentage of participants with hip pain was less than 1% among participants with KL grade 0–2, whereas it was more than 30% among those with KL grade ≥ 3 hip OA, with an odds ratio of approximately 80 for KL grade ≥ 3 hip OA and hip pain [30]. Previous predictive models defined OA progression as KL grade ≥ 2 [17], whereas this study defined it as KL grade ≥ 3. An association between radiographic and clinical OA has been found in the knee joint, whereas an inconsistent association exists in the hip joint [31]. Decisions for THA may be influenced by clinical presentations such as rest pain, range of motion, underlying conditions [32], and economic and environmental conditions of the patient [33]. Therefore, THA may not have been included in the disease outcomes.

In the treatment of patients with hip pain, orthopedic clinicians often encounter cases in which OA does not progress despite anatomical abnormalities, such as acetabular dysplasia and cam morphology on radiographs, or cases in which OA progresses despite the absence of obvious anatomical abnormalities. Hence, predicting hip OA remains challenging. Our prediction model may assist orthopedic clinicians to provide patients with optimal information regarding their individualized future OA progression risk and select appropriate treatments (intensity), potentially increasing treatment adherence. In the future, the prediction model could be integrated into a software package linked to electronic medical records (including picture archiving and communication systems), aiding clinicians to estimate the future progression risk of hip OA.

This study has several limitations. First, the dataset size was limited to training the AI model. Due to the limited dataset size, the experiments were performed using a five-fold cross-validation. Second, this study included only a dataset from a single center and lacked external validation, thus making its generalizability uncertain. Future studies are needed to evaluate the generalizability of this method across multiple datasets from diverse centers and populations. External validation with data from different populations is a crucial step in implementing this machine-learning-based approach in primary care. Third, the retrospective nature of this study may have introduced bias. Additionally, our patient group had a sex bias, with a predominance of female patients. All participants in this study were Japanese patients with hip OA, which aligns with previous studies indicating that most hip OA patients in Japan were female [34]. Fourth, our AI prediction model only utilized a single AP pelvic radiograph for its assessment, even though in clinical practice, additional views, such as detected hip radiographs, are acquired for OA evaluation. A review of machine learning in early OA diagnosis reported models using MRI for knee OA [13]. Although deep learning models using MRI have been applied to osteonecrosis of the femoral head [35, 36], they have not been used for hip OA. Therefore, future studies are needed to develop AI models for hip OA using MRI. Fifth, we could not collect clinical parameters such as pain score and limitation of activity to train our algorithm, and incorporating clinical parameters could enhance its performance.

Conclusions

We developed an AI model to predict the progression of OA from KL grade 0–2 to KL grade ≥ 3 within 3–5 years based on AP pelvic radiographs, height, body weight, sex, age, and KL grade. This model performed well and

may be clinically applicable with additional datasets and validation. Furthermore, it may serve as a pilot study for research on AI models to predict hip OA progression.

Abbreviations

OA	Osteoarthritis
AI	Artificial intelligence
KL	Kellgren–Lawrence
CHECK	Cohort Hip and Cohort Knee
THA	Total hip arthroplasty
AP	Anteroposterior
DICOM	Digital Imaging and Communications in Medicine
ROI	Region of interest
AUC	Area under the receiver operating characteristic curve
CI	Confidence interval

Acknowledgements
We would like to thank Editage (www.editage.jp) for English language editing.

Author contributions
The authors made the following contributions: (1) conception and design (RH, KM, TI, SY, HK); analysis and interpretation of the data (RH, TI, ST); (2) drafting of the article (RH); critical revision of the article for important intellectual content (KM, TI, ST); and (3) final approval of the article (RH, KM, TI, ST, YI, SY, HK).

Funding
This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability
The datasets used and analyzed in the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
This study was approved by the Institutional Review Board of Teikyo University (ethical approval number: 19-152-5) and conducted in accordance with the World Medical Association Declaration of Helsinki. All participants provided written informed consent for their data to be used for research purposes.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 21 February 2024 / Accepted: 6 November 2024
Published online: 08 November 2024

References

1. Hawker GA. Osteoarthritis is a serious disease. Clin Exp Rheumatol. 2019;37:3–6.
2. Jordan JM, Helmick CG, Renner JB, Luta G, Dragomir AD, Woodard J, et al. Prevalence of hip symptoms and radiographic and symptomatic hip osteoarthritis in African americans and caucasians: the Johnston County Osteoarthritis Project. J Rheumatol. 2009;36:809–15. https://doi.org/10.3899/jrheum.080677.
3. Fan Z, Yan L, Liu H, Li X, Fan K, Liu Q, et al. The prevalence of hip osteoarthritis: a systematic review and meta-analysis. Arthritis Res Ther. 2023;25:51. https://doi.org/10.1186/s13075-023-03033-7.
4. Katz JN, Arant KR, Loeser RF. Diagnosis and treatment of hip and knee osteoarthritis: a review. JAMA. 2021;325:568–78. https://doi.org/10.1001/jama.2020.22171.
5. Hawker GA, Croxford R, Bierman AS, Harvey PJ, Ravi B, Stanaitis I, et al. All-cause mortality and serious cardiovascular events in people with hip and knee osteoarthritis: a population based cohort study. PLoS ONE. 2014;9:e91286. https://doi.org/10.1371/journal.pone.0091286.
6. Runhaar J, van Berkel AC, Agricola R, van Meurs J, Bierma-Zeinstra SMA. Risk factors and population-attributable fractions for incident hip osteoarthritis. HSS J. 2023;19:407–12. https://doi.org/10.1177/15563316231192461.
7. Shapira J, Chen JW, Bheem R, Lall AC, Rosinsky PJ, Maldonado DR, et al. Radiographic factors associated with hip osteoarthritis: a systematic review. J Hip Preserv Surg. 2020;7:4–13. https://doi.org/10.1093/jhps/hnz073.
8. Casartelli NC, Maffiuletti NA, Valenzuela PL, Grassi A, Ferrari E, van Buuren MMA, et al. Is hip morphology a risk factor for developing hip osteoarthritis? A systematic review with meta-analysis. Osteoarthr Cartil. 2021;29:1252–64. https://doi.org/10.1016/j.joca.2021.06.007.
9. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis. 1957;16:494–502.
10. Chen P, Gao L, Shi X, Allen K, Yang L. Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. Comput Med Imaging Graph. 2019;75:84–92. https://doi.org/10.1016/j.compmedimag.2019.06.002.
11. Yeoh PSQ, Lai KW, Goh SL, Hasikin K, Hum YC, Tee YK, et al. Emergence of deep learning in knee osteoarthritis diagnosis. Comput Intell Neurosci. 2021;2021:4931437. https://doi.org/10.1155/2021/4931437.
12. Yoon JS, Yon CJ, Lee D, Lee JJ, Kang CH, Kang SB, et al. Assessment of a novel deep learning-based software developed for automatic feature extraction and grading of radiographic knee osteoarthritis. BMC Musculoskelet Disord. 2023;24:869. https://doi.org/10.1186/s12891-023-06951-4.
13. Xuan A, Chen H, Chen T, Li J, Lu S, Fan T, et al. The application of machine learning in early diagnosis of osteoarthritis: a narrative review. Ther Adv Musculoskelet Dis. 2023;15:1759720X231158198. https://doi.org/10.1177/1759720X231158198.
14. von Schacky CE, Sohn JH, Liu F, Ozhinsky E, Jungmann PM, Nardo L, et al. Development and validation of a multitask deep learning model for severity grading of hip osteoarthritis features on radiographs. Radiology. 2020;295:136–45. https://doi.org/10.1148/radiol.2020190925.
15. Leung K, Zhang B, Tan J, Shen Y, Geras KJ, Babb JS, et al. Prediction of total knee replacement and diagnosis of osteoarthritis by using deep learning on knee radiographs: data from the osteoarthritis initiative. Radiology. 2020;296:584–93. https://doi.org/10.1148/radiol.2020192091.
16. Tiulpin A, Klein S, Bierma-Zeinstra SMA, Thevenot J, Rahtu E, Meurs JV, et al. Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data. Sci Rep. 2019;9:20038. https://doi.org/10.1038/s41598-019-56527-3.
17. Gielis WP, Weinans H, Welsing PMJ, van Spil WE, Agricola R, Cootes TF, et al. An automated workflow based on hip shape improves personalized risk prediction for hip osteoarthritis in the CHECK study. Osteoarthr Cartil. 2020;28:62–70. https://doi.org/10.1016/j.joca.2019.09.005.
18. Chen CC, Huang JF, Lin WC, Cheng CT, Chen SC, Fu CY, et al. The feasibility and performance of total hip replacement prediction deep learning algorithm with real world data. Bioeng (Basel). 2023;10:458. https://doi.org/10.3390/bioengineering10040458.
19. Van Rossum G, Drake FL. Python 3 reference Manual. Scotts Valley, CA: CreateSpace; 2009.
20. Bradski G. The OpenCV Library. Dr Dobb J Softw Tool. 2000;120:122–5.
21. Mason D et al. scaramallion, mrbean-bremen, rhaxton, Suever J, Vanessasaurus,. pydicom/pydicom: pydicom 2.3.1 (v2.3.1). Zenodo; 2022. https://doi.org/10.5281/zenodo.7319790
22. Chen K, Wang J, Pang J, Cao Y, Xiong Y, Li X et al. MMDetection: Open mmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155; 2019. https://doi.org/10.48550/arXiv.1906.07155
23. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. ArXiv abs/1905.11946. 2019.
24. Pedregosa F, Varoquaux G, Michel V, Thirion B, Grisel O, Blondel M, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
25. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29–36. https://doi.org/10.1148/radiology.143.1.7063747.
26. Jingushi S, Ohfuji S, Sofue M, Hirota Y, Itoman M, Matsumoto T, et al. Multiinstitutional epidemiological study regarding osteoarthritis of the hip in Japan. J Orthop Sci. 2010;15:626–31. https://doi.org/10.1007/s00776-010-1507-8.
27. Appleyard T, Thomas MJ, Antcliff D, Peat G. Prediction models to estimate the future risk of osteoarthritis in the general population: a systematic review. Arthritis Care Res (Hoboken). 2023;75:1481–93. https://doi.org/10.1002/acr.25035.
28. Castaño-Betancourt MC, Van Meurs JB, Bierma-Zeinstra S, Rivadeneira F, Hofman A, Weinans H, et al. The contribution of hip geometry to the prediction

of hip osteoarthritis. Osteoarthr Cartil. 2013;21:1530–6. https://doi.org/10.1016/j.joca.2013.06.012.
29. Saberi Hosnijeh F, Kavousi M, Boer CG, Uitterlinden AG, Hofman A, Reijman M, et al. Development of a prediction model for future risk of radiographic hip osteoarthritis. Osteoarthr Cartil. 2018;26:540–6. https://doi.org/10.1016/j.joca.2018.01.015.
30. Iidaka T, Muraki S, Akune T, Oka H, Kodama R, Tanaka S, et al. Prevalence of radiographic hip osteoarthritis and its association with hip pain in Japanese men and women: the ROAD study. Osteoarthr Cartil. 2016;24:117–23. https://doi.org/10.1016/j.joca.2015.07.017.
31. Kinds MB, Welsing PM, Vignon EP, Bijlsma JW, Viergever MA, Marijnissen AC, et al. A systematic review of the association between radiographic and clinical osteoarthritis of hip and knee. Osteoarthr Cartil. 2011;19:768–78. https://doi.org/10.1016/j.joca.2011.01.015.
32. Dreinhöfer KE, Dieppe P, Stürmer T, Gröber-Grätz D, Flören M, Günther KP, et al. Indications for total hip replacement: comparison of assessments of orthopaedic surgeons and referring physicians. Ann Rheum Dis. 2006;65:1346–50. https://doi.org/10.1136/ard.2005.047811.
33. Youm J, Chan V, Belkora J, Bozic KJ. Impact of socioeconomic factors on informed decision making and treatment choice in patients with hip and knee OA. J Arthroplasty. 2015;30:171–5. https://doi.org/10.1016/j.arth.2014.09.006.
34. Sato T, Yamate S, Utsunomiya T, Inaba Y, Ike H, Kinoshita K, et al. Life course epidemiology of hip osteoarthritis in Japan: a multicenter, cross-sectional study. J Bone Joint Surg Am. 2024;106:966–75. https://doi.org/10.2106/JBJS.23.01044.
35. Shen X, Luo J, Tang X, Chen B, Qin Y, Zhou Y, et al. Deep learning approach for diagnosing early osteonecrosis of the femoral head based on magnetic resonance imaging. J Arthroplasty. 2023;38:2044–50. https://doi.org/10.1016/j.arth.2022.10.003.
36. Shen X, He Z, Shi Y, Liu T, Yang Y, Luo J, et al. Development and validation of an automated classification system for osteonecrosis of the femoral head using deep learning approach: a multicenter study. J Arthroplasty. 2024;39:379–e862. https://doi.org/10.1016/j.arth.2023.08.018.

Publisher's note