

Artificial intelligence assistance in radiographic detection and classification of knee osteoarthritis and its severity: a cross-sectional diagnostic study

N. PONGSAKONPRUTTIKUL¹, C. ANGTHONG¹, V. KITTIKHA¹, S. CHUWONGIN², P. PUENGPATTRAKUL¹, P. THONGPAT¹, S. BOONSANG³, T. TONGLOY²

¹Faculty of Medicine, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

²College of Advanced Manufacturing Innovation, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

³Department of Electrical Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

Abstract. – **OBJECTIVE:** Radiographic interpretation suffers from an ever-increasing workload in orthopedic and radiology departments. The present study applied and assessed the performance of a convolutional neural network designed to assist orthopedists and radiologists in the detection and classification of knee osteoarthritis from early to severe degrees in accordance with the Kellgren-Lawrence (KL) classification system.

MATERIALS AND METHODS: In total, 1650 knee joint radiographs (anteroposterior view) were collected from the Osteoarthritis Initiative public resource. Two models were developed: one distinguished normal (KL 0-I) from osteoarthritic knees (KL II-IV), and the other classified the severity as normal (KL 0-I), non-severe (KL II), or severe (KL III-IV). The regions of interest were labeled under the supervision of experts. Our artificial intelligence (AI) models were trained using the You Only Look Once version 3 (YOLOv3) detection algorithm.

RESULTS: Our first AI model using YOLOv3 tiny could detect and classify normal and osteoarthritic knees on plain knee joint radiographs with 85% accuracy and 81% mean average precision. The second AI model for classifying severity achieved a total accuracy of 86.7% and mean average precision of 70.6%.

CONCLUSIONS: Our proposed deep learning models provided high accuracy and satisfactory precision for the detection and classification of early to severe knee osteoarthritis on anteroposterior radiographs. These models may be used as diagnostic aids by interpreting knee radiographs and guiding the treatment options via each osteoarthritic stage for related physicians and specialists.

Key Words:

Artificial intelligence, Knee osteoarthritis, Radiograph, Deep learning.

Introduction

Knee osteoarthritis (OA) has a detrimental effect on the quality of life of people worldwide¹. It is characterized by progressive destruction of the articular cartilage and formation of osteophytes, subchondral cysts, and subchondral sclerosis of the synovial joints, which can ultimately lead to functional disability². Many patients with knee OA require early diagnosis and treatment³. Healthcare providers can diagnose knee OA based on clinical evaluations and imaging investigations. The first-line investigation conventionally used in the diagnosis is plain radiography of the knee¹. However, because there are usually many patients with knee OA in daily practice, radiographic interpretation in orthopedic and radiology departments requires an ever-increasing workload. Approximately 15% of cases of early knee OA are reportedly missed in routine practice⁴. In addition, the diagnostic accuracy is highly dependent on the physician's knowledge and experience. Thus, OA detection can be missed by a less experienced reader. Multiple previous studies⁴⁻⁸ have shown highly inconsistent sensitivity and specificity ranging from 3.0% to 95.0% and from 60.0% to 98.0%, respectively, in detecting knee OA on posteroanterior standing knee X-ray images. These are issues for

which artificial intelligence (AI) may play an integral role in assisting orthopedists, radiologists, and related specialists in interpreting radiographs in a shorter time and with higher accuracy. However, minimal information is available regarding the role of AI in the diagnosis of knee OA via radiography. The application of AI has been emerging in the field of medical imaging, especially using deep learning approaches. Previous studies have applied both shallow machine learning and deep learning algorithms for the detection and classification of knee OA. Brahim et al⁹ applied a shallow machine learning algorithm that included naive Bayes and random forest classifiers to detect early knee OA based on the Kellgren-Lawrence (KL) classification¹⁰ of KLO as normal and KL2 as early knee OA. They achieved an accuracy rate of approximately 83% with their model using 1024 knee X-ray images from the public database of the Osteoarthritis Initiative (OAI). Tiulpin et al¹¹ adopted a deep Siamese convolutional neural network (CNN) to automatically classify knee OA severity based on the KL classification. Their model patched the medial and lateral sides of the knee joint, with a convolutional network applied to each side. Each output was concatenated and used in the final fully connected layer to make a prediction. This model produced a multiclass accuracy of 66.7%. Thomas et al¹² used DenseNet, a 169-layer CNN with a dense convolutional network architecture, to propose a model that could predict the KL grading scale using the OAI dataset with an average accuracy of 0.71.

Multiple-step deep learning approaches were also formulated in previous studies. In 2017, Antony et al¹³ proposed a combination of object detection and classification techniques to classify the severity of knee OA. A fully CNN combined with a fine-tuned CNN with four convolutional layers and one fully connected layer was trained on 4446 and 2920 knee X-ray images from the OAI and MOST datasets. Detection of the knee joint was performed by the fully CNN, whereby the fine-tuned CNN with four convolutional layers and one fully connected layer was used to classify the severity according to the KL classification. A multiclass classification accuracy of 63.4% was achieved using this approach. Later, Chen et al¹⁴ adopted a similar protocol by first using the You Only Look Once version 2 (YOLO v2) algorithm for knee joint detection. Various renowned CNN models were then used to classify the severity. The best performing combination was YOLO v2 with the fine-tuned VGG-19 model, which achieved a classification accuracy of 69.7%.

In this paper, we propose an object detection method using a modern state-of-the-art object detection CNN architecture called YOLOv3 tiny. We developed two models: one was used to automatically distinguish a normal knee from knee OA, and the other classified the severity based on further grouping by the KL classification. This CNN model is designed to assist orthopedists, radiologists, and related specialists by automatically detecting and classifying knee OA from early to severe degrees in accordance with the KL classification.

Materials and Methods

Dataset and Preprocessing

In total, 1650 cropped right and left knee radiographs (anteroposterior view) with a resolution of 224×224 pixels were collected from public resources distributed by Chen¹⁵ in that author's "Knee Osteoarthritis Severity Grading Dataset" and published article¹⁴. This dataset was taken from the OAI and modified by cropping only the areas of the knee joint that are split into right and left joints^{14,15}. The OAI is a multicenter, longitudinal, prospective observational study of knee OA funded by the US National Institutes of Health. It provides a publicly accessible database of the natural history of knee OA, containing resources such as knee X-rays and magnetic resonance images. The present study involved 4796 men and women ranging in age from 45 to 79 years. Only trained and certified readers in each participating medical center were involved in classifying the radiographic images. The patients' radiographic manifestations were described mainly according to the following features in accordance with the KL classification^{1,10}: osteophytes, narrowing of the joint space, and subchondral bone changes such as subchondral bone sclerosis and subchondral bone cysts. The KL classification is defined as follows^{1,11,16,17}:

- Grade 0: no radiological findings of OA
- Grade I: doubtful narrowing of the joint space and possible osteophytic lipping
- Grade II: definite osteophytes and possible narrowing of the joint space
- Grade III: moderate multiple osteophytes, definite narrowing of the joint space, small pseudocystic areas with sclerotic walls, and possible bone contour deformity
- Grade IV: large osteophytes, marked narrowing of the joint space, severe sclerosis, and definite deformity of the bone contour

Table I. Two datasets used in the present study.

Dataset	Classification	Training set (Knee X-ray images)	Testing set (Knee X-ray images)
I	Normal (KL 0-I)	500	50
	OA (KL II-IV)	500	50
II	Normal (KL 0-I)	500	50
	Non-severe (KL II)	500	50
	Severe (KL III-IV)	500	50

Dataset I contains Normal (KL 0-I) and OA (KL II-IV) classes with 500 training and 50 testing samples each. Dataset II contains Normal (KL 0-I), Non-severe (KL II), and Severe (KL II-IV) classes with 500 training and 50 testing samples each. Abbreviations: OA, osteoarthritis; KL, Kellgren-Lawrence.

Because of the ambiguity of the KL classification, its intra-rater reliability is relatively unreliable, ranging only from 0.67 to 0.73^{14,18}. Two systems with which to categorize the subgroups of radiographic findings using the KL classification have been suggested. The first system divides all KL grades into two subgroups, where KL grade 0 to I is considered “normal” and KL grade \geq II is designated “definite OA”^{1,10,19}. The second system divides all KL grades into three subgroups of the severity of osteoarthritic changes, where KL grade 0 is “normal,” KL grade I to II indicates “non-severe OA,” and KL grade III to IV is considered “severe OA”²⁰. We propose a modified system in which KL grade 0 indicates that the knee joint is “normal,” KL grade 0 to I indicates “non-severe OA,” and KL grade III to IV indicates “severe OA.” The logic behind this classification is based on the cutoff point of OA being KL grade \geq II in the original study and various other studies^{1,10,19} and the response to treatment^{18,21}. In this study, the first system and our proposed classification system were adopted, resulting in two datasets as shown in Table I.

All regions of interest with features in accordance with the first and our proposed systems

were manually labeled with rectangular bounding boxes to create ground truth via CiRA-CORE, an in-house deep learning platform ([https:// git.cira-lab.com/cira/cira-core](https://git.cira-lab.com/cira/cira-core)), under the guidance of experts (Figure 1). Each class of each system was then randomly split into a training set and testing set at a ratio of 9:1 to train and test the model, respectively. After splitting, the testing set for validation contained 1500 knee X-ray images with 500 images in each class for both datasets.

Data Augmentation

Prior to training the model, multiple data augmentation techniques were implemented to improve the generalization capabilities of the model. This was conducted to increase the variation of the dataset in an attempt to imitate the real-life scenario in which the qualities and parameters, such as exposure and orientation, are not consistent. The techniques applied were as follows:

- Rotation of the image with the value ranging from -3 to 3 degrees, varying at every 45 degrees
- Adjustment of the brightness and contrast by multiplying all pixel red, green, and blue values by seven steps ranging from 0.6 to 1.0

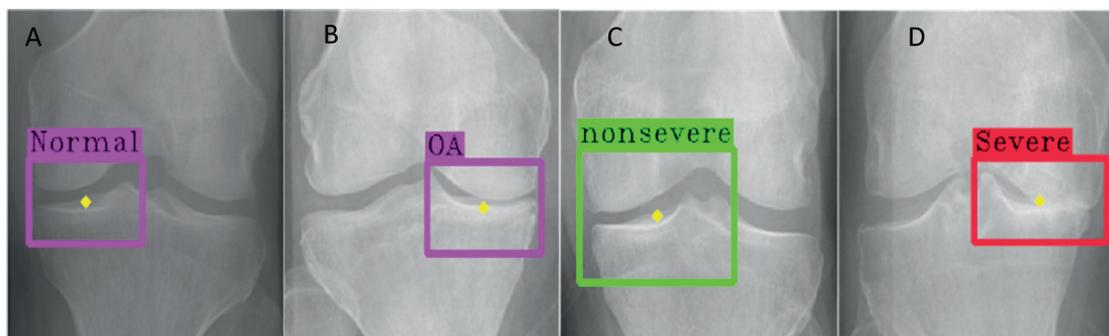


Figure 1. ROI labeling. ROIs are shown in X-rays of a (A) normal knee, (B) knee with OA, (C) knee with non-severe OA, and (D) knee with severe OA. Abbreviations: ROI, region of interest; OA, osteoarthritis

Deep Learning Model Configuration and Training

Modified YOLO algorithms were employed under the CiRA-CORE platform^{22,23} as the proposed methodology for developing models to detect and classify knee OA. These models were then assessed and evaluated based on their performance in detecting and classifying knee OA using relevant learning methods. Introduced earlier, YOLO is a state-of-the-art object detector that combines object detection and classification via a deep CNN. It uses features from the whole image to predict and split the image into multiple bounding boxes and then predicts the classification²⁴. YOLO is well known for its balance of speed and accuracy, especially for real-time detection, and for processing 45 to 155 frames per second^{22,24}; these features make it suitable for on-demand detection in the clinical setting. Furthermore, previous studies have suggested that YOLO exhibits impressive performance in detecting small objects²⁵, which is critical in detecting small lesions such as miniature osteophytes in non-severe OA. However, some versions of YOLO, such as YOLOv3, require a high level of computing power. For this reason, we employed a more simplified and optimized version dubbed the “tiny” variant

of YOLO. This version requires less computing power for comparable accuracy and is suitable as an embedded AI detector system²⁶.

We fed 1000 labeled knee X-ray images, with 500 annotated as “Normal” and 500 as “Osteoarthritis,” into the model to test the first system. We then fed 1500 labeled knee X-ray images, with 500 annotated as “Normal,” 500 as “Non-severe,” and 500 as “Severe,” into the model to test the second system. This study was performed using a dedicated deep learning server with a 64-bit Ubuntu 16.04 operating system and a library based on the CUDA 10.1 toolkit and cuDNN v7.5.0. All experiments were conducted on a server with the following configuration: CPU i7-8700 (3.70 GHz), RAM 16 × 2 GB, GPU NVIDIA GTX 1070Ti (8 GB), and the C++ programming language. A modified version of YOLOv3 tiny was employed under CiRA-CORE with a mini-batch size of 64 and 8 subdivisions, while the momentum and decay were set at 0.9 and 0.0005, respectively. The model learning rate was 0.001.

The model automatically located the knee lesions and regions of interest of each knee in the radiographic images, depicting a rectangular bounding box during the analysis. An intersection of union of 0.5 between the predicted detection

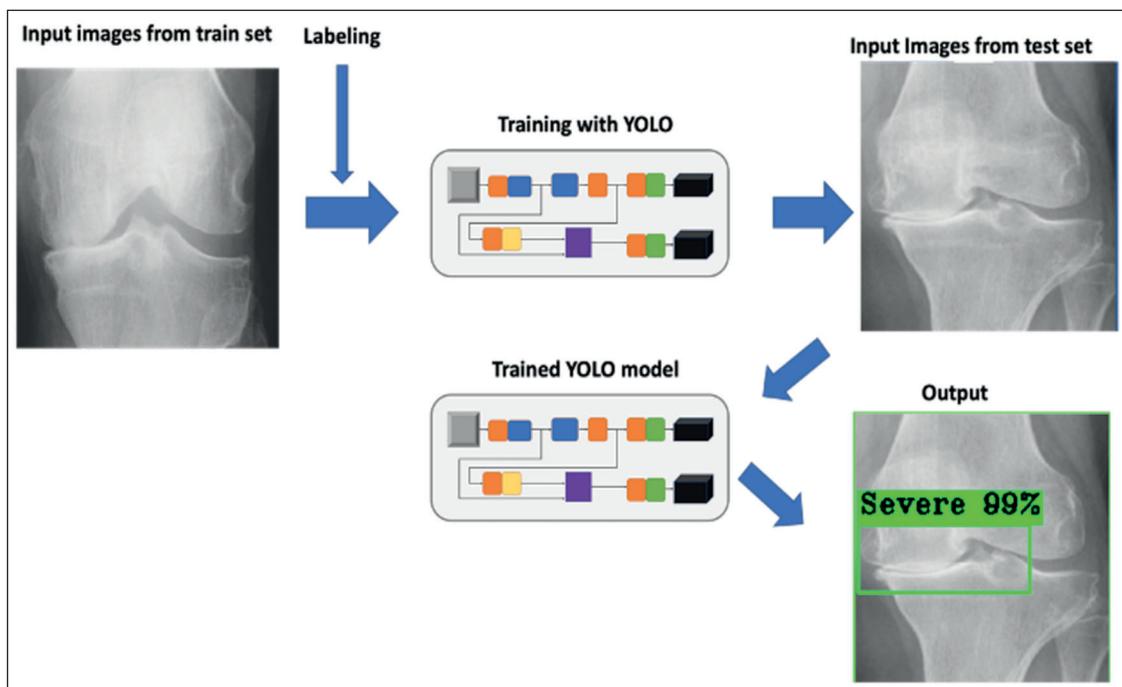


Figure 2. In total, 1000 samples and 1500 samples of plain knee radiographs in the first dataset and second dataset, respectively, were labeled and fed to the YOLO algorithm to train the model. After training, images from the test set were used to test and validate the trained YOLO model. Here, the output is a bounding box with a predicted class, such as “Severe.”

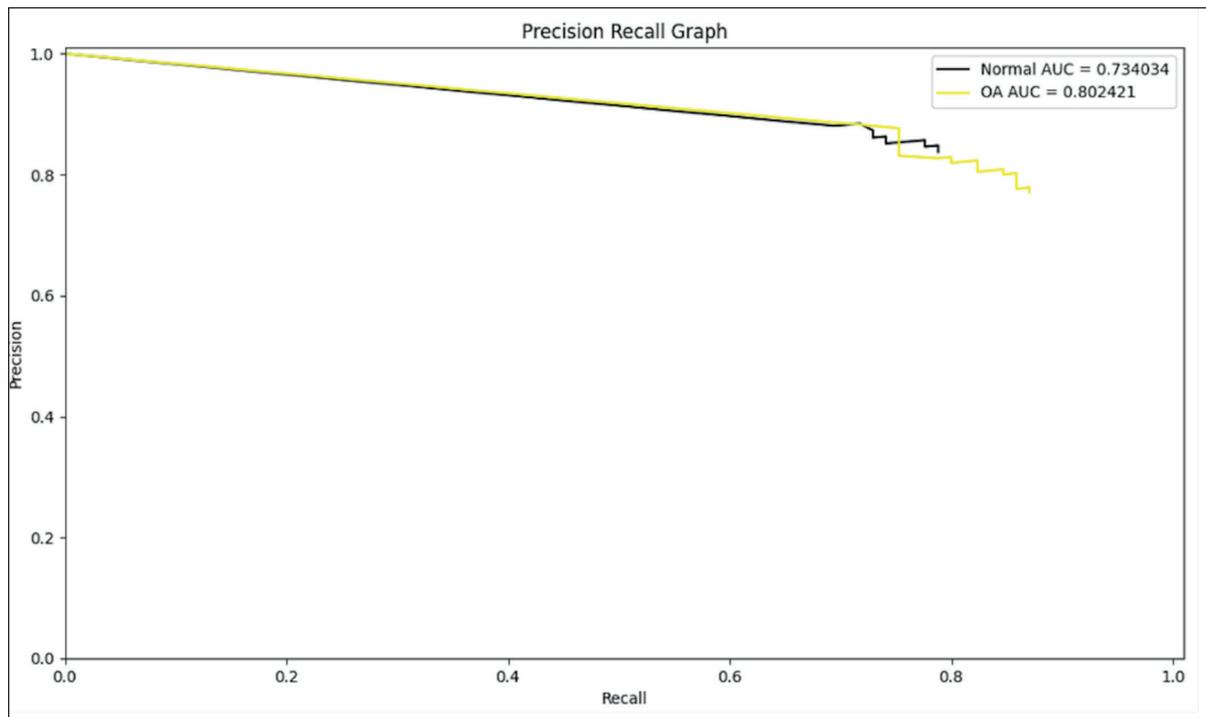


Figure 3. Precision-recall curve with AUC of the first proposed model using YOLOv3 tiny. The first proposed model using YOLOv3 tiny detected normal knees as KL 0-I and knees with OA as KL II-IV with an AUC of 0.73 and 0.80, respectively. *Abbreviations:* AUC, area under the curve; OA, osteoarthritis; KL, Kellgren-Lawrence

and manually labeled bounding box was used as the threshold to determine whether the predicted bounding box represented the actual class. A value less than the threshold was considered a false positive. The workflow of the methodology is shown in Figure 2.

Model Evaluation

To evaluate the performance of our models, we first validated the model using the testing set containing 50 normal and 50 OA knee X-ray images in the first dataset and 50 images each for the normal, non-severe, and severe classes for the second dataset. Next, a performance assessment was carried out by constructing a confusion matrix. Aside from the computation of four performance metrics [sensitivity (recall), specificity, accuracy, and precision (positive predictive value)], the true positive, true negative, false positive, and false negative were also analyzed using the confusion matrix. The formulas used to calculate these parameters are listed in Table II. The other parameter, the F1 score, was calculated using the following formula:

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Furthermore, we used a precision-recall curve (PRC) to evaluate the object detection model. A PRC is a two-dimensional graph in which the X-axis is the recall (sensitivity) and the y-axis is the precision. It depicts the relationship between the precision and recall at each threshold. Generally, a model is deemed satisfactory if the precision remains at a high level as the recall increases. The area under the curve (AUC) is the area under the PRC that summarizes the model performance in a single number. It is used to measure the usefulness of the tested model; a larger AUC is associated with better performance of the model. The mean average precision (mAP) was also computed.

Results

The present study included varying degrees of OA severity in accordance with the KL classification on the collected images. For the first system, 500 normal radiographs (KL grades 0-I) and 500 radiographs with osteoarthritic findings (KL grades II-IV) were used for training of the deep learning model via YOLOv3 tiny. Testing of the model was performed on 50 normal radiographs (KL grades 0-I) and 50 radiographs with osteoar-

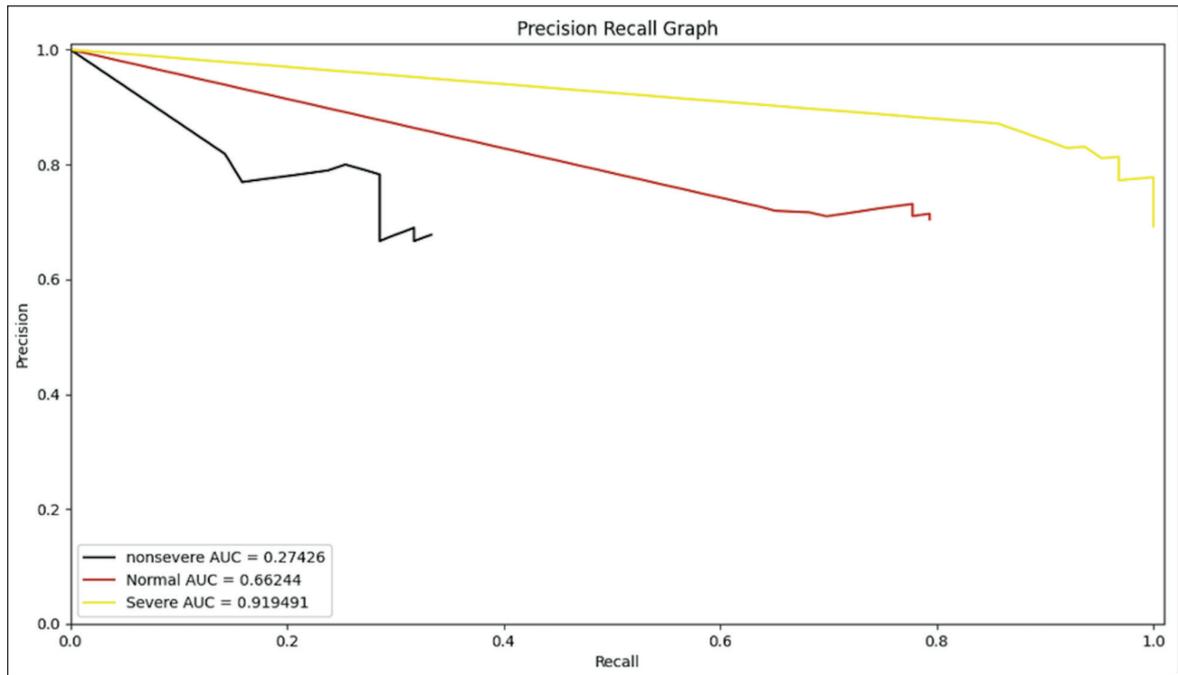


Figure 4. Precision-recall curve with AUC of the second proposed model using YOLOv3 tiny. The second proposed model using YOLOv3 tiny detected normal knees as KL 0-I, knees with non-severe OA as KL II, and knees with severe OA as KL III-IV with an AUC of 0.66, 0.27, and 0.92, respectively. *Abbreviations:* AUC, area under the curve; OA, osteoarthritis; KL, Kellgren-Lawrence.

thritic findings (KL grades II-IV). The analytical results of the tested model were promising with respect to the accuracy, sensitivity, and specificity, as shown in Table III. The overall accuracy of the first model using YOLOv3 tiny was 85% in detecting normal and osteoarthritic knee joints, while the overall sensitivity and specificity were the same at 85%. The mAP was also satisfactory at 81%. Individually, as shown in Table IV, similar performance was also observed in classifying normal and osteoarthritic knees (sensitivity of 86% and 84%, respectively). However, specificity of identifying normal knees was slightly lower

than that of osteoarthritic knees (84% and 86%, respectively).

For the second proposed model, there were 500 normal radiographs (KL grades 0-I), 500 radiographs with non-severe osteoarthritic findings (KL grade II), and 500 radiographs with severe osteoarthritic findings (KL grades III-IV) using YOLOv3 tiny. Testing of the model was performed using 50 unobserved normal radiographs (KL grade 0), 50 radiographs with non-severe osteoarthritic findings (KL grades I-II), and 50 radiographs with osteoarthritic findings (KL grades III-IV). Using the second system, our second

Table II. Confusion matrix with sensitivity, specificity, precision, negative predictive value, and accuracy calculation.

		Predicted class		
		Positive	Negative	
Actual class	Positive	TP	FN	Sensitivity (recall) TP/(TP + FN)
	Negative	FP	TN	Specificity TN/(TN + FP)
		Precision TP/(TP + FP)	Negative predictive value TN/(TN + FN)	Accuracy TP + TN/(TP + TN + FP + FN)

Abbreviations: TP, true positive; FP, false positive; FN, false negative; TN, true negative.

Table III. Model performance evaluation between two classification systems using YOLO network.

Model	Accuracy	Sensitivity	Specificity	F1 score	Mean average precision
First proposed model with YOLOv3 tiny	85.0%	85.0%	85.0%	85.0%	81.0%
Second proposed model with YOLOv3 tiny	86.7%	55.1%	85.9%	61.1%	70.6%

The first proposed model with YOLOv3 tiny defined the knee joint as normal if KL 0-I and OA if KL II-IV. The second proposed model with YOLOv3 tiny detected normal knees as KL 0-I, non-severe osteoarthritic knees as KL II, and severe osteoarthritic knees as KL III-IV. Abbreviations: KL, Kellgren-Lawrence; OA, osteoarthritis.

trained model demonstrated highly satisfactory results in terms of the overall performance, as shown in Table III, with accuracy and specificity of 86.7% and 85.9%, respectively. However, only fair overall sensitivity and mAP were achieved for this system (Table III). The individual performances for each class are shown in Table V. The PRCs with the AUCs of the first and second proposed models using YOLOv3 tiny are shown in Figures 3 and 4, respectively.

Discussion

OA affects more than 250 million people worldwide (approximately 4% of the world population) and is considered one of the 50 most common sequelae of diseases and injuries^{1,27}. By 2010, OA was responsible for approximately 17 million years lost to disability, and knee OA alone constituted 83% of this number^{1,27}. Currently, a combination of clinical presentation and plain radiography is considered the mainstay in the diagnosis of OA. In 1957, Kellgren and Lawrence were among the first to formalize the radiographic classification of OA^{1,10}. The present study highlights the benefits of AI in the interpretation and classification of knee OA findings

on radiographs. In the first proposed model, the accuracy, sensitivity, specificity, and mAP were high for the detection of the normal, OA, and overall groups. These levels were higher than the results of interpretation and classification by orthopedic surgeons in a previous study²⁸. Thus, the herein proposed deep learning model may be used to help not only orthopedic surgeons and related specialists but also general practitioners in interpreting radiographic findings and diagnosing OA in daily practice.

For our second proposed deep learning approach, the results demonstrated high accuracy and specificity but only fair sensitivity and mAP in the overall detection. The rationale of this system is related to the recommendation of a treatment option for knee OA that depends on the KL grade. Regarding non-severe OA as KL grade II, several studies^{20,29,30} have indicated that nonoperative treatments or joint-sparing surgeries could be the first-choice treatments. However, regarding severe OA as KL grades III-IV, joint replacement surgery may be the treatment of choice because of unsatisfactory results of joint-sparing surgeries^{30,31}. The second proposed model in this study provided excellent specificity and accuracy and satisfactory sensitivity for the detection of severe OA (KL grades III-IV).

Table IV. Evaluation results for first proposed model trained with 1000 samples using YOLOv3 tiny.

	Normal	OA	Average	Total
TPR (Sen/Rec)	0.860	0.840	0.850	0.850
FNR	0.140	0.160	0.150	0.150
TNR (Spec)	0.840	0.860	0.850	0.850
FPR	0.160	0.140	0.150	0.150
Precision	0.843	0.857	0.850	0.850
Accuracy	0.850	0.850	0.850	0.850
Misclassification rate	0.150	0.150	0.150	0.150
F1 score	0.851	0.848	0.850	0.850

Normal: KL 0-I, OA: KL II-IV. Abbreviations: TPR, true positive rate; FNR, false negative rate; TNR, true negative rate; FPR, false positive rate; Sen/Rec, sensitivity/recall; Spec, specificity; KL, Kellgren-Lawrence; OA, osteoarthritis.

Table V. Evaluation results for second proposed model trained with 1500 samples using YOLOv3 tiny.

	Normal	Non-severe OA	Severe OA	Average	Total
TPR (Recall)	0.540	0.425	0.780	0.582	0.551
FNR	0.460	0.575	0.220	0.418	0.449
TNR (Specificity)	0.890	0.746	0.980	0.872	0.859
FPR	0.110	0.254	0.020	0.128	0.141
Precision	0.711	0.521	0.951	0.728	0.687
Accuracy	0.773	0.913	0.913	0.867	0.867
Misclassification rate	0.227	0.560	0.087	0.291	0.291
F1 score	0.614	0.468	0.857	0.646	0.611

Normal: KL 0-I, non-severe OA: KL II, severe OA: KL III-IV. Abbreviations: TPR, true positive rate; FNR, false negative rate; TNR, true negative rate; FPR, false positive rate; KL, Kellgren-Lawrence; OA, osteoarthritis.

In addition, it provided excellent accuracy and specificity, albeit low sensitivity, for the detection of non-severe OA (KL grade II). The low sensitivity, especially in the normal and non-severe groups, might stem from the similarity and minimal differences stated in the KL classification criteria. Although low sensitivity was shown in this approach, it still outperformed orthopedic surgeons in detecting knee OA from plain-film knee X-rays in a previous study²⁸.

Limitations

This study had some limitations. The sample size was relatively small. A larger sample size might improve the overall performances of both models, especially for sensitivity and mAP. Additionally, other classification systems may require testing because the KL classification has been documented as ambiguous, and interpretation may vary based on the reader^{15,17}. However, despite these limitations, these models can aid orthopedic surgeons, related specialists, and even general practitioners in obtaining and differentiating normal findings from pathological findings and assessing the severity of knee OA. This AI assistance may help to develop treatment options that correspond to the severity grade.

Conclusions

Our proposed model provided highly satisfactory to excellent sensitivity, specificity, accuracy, and mAP for the detection and classification of normal (KL 0-I) and osteoarthritic findings (KL II-IV) on anteroposterior radiographs. Furthermore, the second proposed model based on our modified classification system assisted in the differentiation of severe grades (KL III-IV) from a non-severe grade (KL II) or normal findings (KL

0-I). The proposed AI technology could help orthopedic surgeons and related specialists interpret knee radiographs for the diagnosis of knee OA. It can also assist in classifying the disease severity for determining the appropriate treatment option.

Acknowledgements

We thank Angela Morben, DVM, ELS, from Edanz (www.edanz.com/ac) for editing a draft of this manuscript.

Funding

The authors did not receive any financial support for the research, authorship, and publication of this study.

Author Contributions

In this study, Chayanin Angthong, who is the corresponding author, has provided substantial contributions to the conception and design of the study. He was also responsible for drafting and making critical revision of the manuscript while providing supervision, validation, and the final approval of the version of the article to be published. Napat Pongsakonpruttikul, the first author, has made significant contributions to the study design, data acquisition, data analysis and interpretation of the result of study. He was also largely involved in the drafting of the article, making critical revisions, and producing the final version of manuscript. Furthermore, Veerayuth Kittichai has made great contribution in aiding in the study design, data analysis, result interpretation, and revision of the manuscript. Other authors including Santhad Chuwongin, Paisal Puengpipattrakul, Siridech Boonsang, and Teerawat Tongloy have contributed largely regarding the study's methodology and data analysis. Lastly, Pongphak Thongpat has contributed greatly to the acquisition and analysis of data in this study.

Conflicts of Interest

The authors declare no conflicts of interest.

ORCID ID

Name: Chayanin Anghthong, ORCID number: 0000-0002-1104-8945

Data Availability Statement

The original dataset was taken from a public resource distributed by Pingjun Chen¹⁵ titled “Knee Osteoarthritis Severity Grading Dataset” Version 1 which was published on 4 Sep 2018 in Mendeley Data accessed via doi: 10.17632/56rmx5bjcr.1 (<https://data.mendeley.com/datasets/56rmx5bjcr/1>). The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

References

- 1) Kohn MD, Sassoon AA, Fernando ND. Classifications in Brief: Kellgren-Lawrence Classification of Osteoarthritis. *Clin Orthop Relat Res* 2016; 474: 1886-1893.
- 2) Vaishya R, Pariyo GB, Agarwal AK, Vijay V. Non-operative management of osteoarthritis of the knee joint. *J Clin Orthop Trauma* 2016; 7: 170-176.
- 3) Cui A, Li H, Wang D, Zhong J, Chen Y, Lu H. Global. Regional prevalence, incidence, and risk factors of knee osteoarthritis in population-based studies. *EClinicalMedicine* 2020; 29: 100587.
- 4) Chaisson CE, Gale DR, Gale E, Kazis L, Skinner K, Felson DT. Detecting radiographic knee osteoarthritis: What combination of views is optimal? *Rheumatology (Oxford)* 2000; 39: 1218-1221.
- 5) Muñoz-García N, Cordero-Ampuero J, Madero-Jarabo R. Diagnostic accuracy of magnetic resonance images and weight-bearing radiographs in patients with arthroscopic-proven medial osteoarthritis of the knee. *Clin Med Insights Arthritis Musculoskelet Disord* 2020; 13: 1179544120938369.
- 6) Kijowski R, Blankenbaker DG, Stanton PT, Fine JP, De Smet AA. Radiographic findings of osteoarthritis versus arthroscopic findings of articular cartilage degeneration in the tibiofemoral joint. *Radiology* 2006; 239: 818-824.
- 7) Duncan ST, Khazzam MS, Burnham JM, Spindler KP, Dunn WR, Wright RW. Sensitivity of standing radiographs to detect knee arthritis: A systematic review of level I studies. *Arthroscopy* 2015; 31: 321-328.
- 8) Lysholm J, Hamberg P, Gillquist J. The correlation between osteoarthrosis as seen on radiographs and on arthroscopy. *Arthroscopy* 1987; 3: 161-165.
- 9) Brahim A, Jennane R, Riad R, Janvier T, Khedher L, Toumi H, Lespessailles E. A decision support tool for early detection of knee osteoarthritis using X-ray imaging and machine learning: Data from the OsteoArthritis Initiative. *Comput Med Imaging Graph* 2019; 73: 11-18.
- 10) Kellgren J, Lawrence J. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957; 16: 494-502.
- 11) Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Sarakkala S. Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach. *Sci Rep* 2018; 8.
- 12) Thomas K, Kidziński Ł, Halilaj E, Fleming S, Venkataraman G, Oei E, Gold G, Delp S. Automated classification of radiographic knee osteoarthritis severity using deep neural networks. *Radiol Artif Intell* 2020; 2: e190065.
- 13) Antony J, McGuinness K, Moran K, O'Connor NE. Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks. *MLDM* 2017; 376-390.
- 14) 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.
- 15) Chen P. Knee osteoarthritis severity grading dataset. Mendeley Data 2018.
- 16) Fife RS, Brandt KD, Braunstein EM, Katz BP, Shelbourne KD, Kalasinski LA, Ryan S. Relationship between arthroscopic evidence of cartilage damage and radiographic evidence of joint space narrowing in early osteoarthritis of the knee. *Arthritis Rheum* 1991; 34: 377-382.
- 17) Audrey HX, Abd Razak HR, Andrew TH. The truth behind subchondral cysts in osteoarthritis of the knee. *Open Orthop J* 2014; 8: 7-10.
- 18) Culvenor AG, Engen CN, Øiestad BE, Engebretsen L, Risberg MA. Defining the presence of radiographic knee osteoarthritis: A comparison between the Kellgren and Lawrence system and OARSI atlas criteria. *Knee Surg Sports Traumatol Arthrosc* 2015; 23: 3532-3539.
- 19) Rehman Y, Lindberg MF, Arnljot K, Gay CL, Lerdal A, Aamodt A. More Severe radiographic osteoarthritis is associated with increased improvement in patients' health state following a total knee arthroplasty. *J Arthroplasty* 2020; 35: 3131-3137.
- 20) Ekeland A, Nerhus TK, Dimmen S, Heir S. Better functional results of opening wedge HTO for varus knees with medial osteoarthritis than opening wedge LFO for valgus knees with lateral osteoarthritis. *Bone Jt Open* 2020; 1: 346-354.
- 21) Hoorntje A, Witjes S, Koenraadt KLM, Aarts R, Weert T, van Geenen RCI. More severe preoperative Kellgren-Lawrence grades of knee osteoarthritis were partially associated with better postoperative patient-reported outcomes in TKA patients. *J Knee Surg* 2019; 32: 211-217.
- 22) Kittichai V, Pongsakul T, Chumchuen K, Samung Y, Sriwichai P, Phatthamolrat N, Tongloy T, Jaksukam K, Chuwongin S, Boonsang S. Deep learning approaches for challenging species and gender identification of mosquito vectors. *Sci Rep* 2021; 11: 4838.

- 23) Kittichai V, Kaewthamasorn M, Thanee S, Jomtarak R, Klanboot K, Naing KM, Tongloy T, Chuwongin S, Boonsang S. Classification for avian malaria parasite *Plasmodium gallinaceum* blood stages by using deep convolutional neural networks. *Sci Rep* 2021; 11: 16919
- 24) Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. *CVPR* 2016; 779-788.
- 25) Nguyen N, Do T, Ngo T, Le D. An evaluation of deep learning methods for small object detection. *J Electr Eng Comput Sci* 2020; 2020: 1-18.
- 26) Mazzia V, Khaliq A, Salvetti F, Chiaberge M. Real-time Apple detection system using embedded systems with hardware accelerators: An edge AI application. *IEEE Access* 2020; 8: 9102-9114.
- 27) Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V, Abraham J, Ackerman I, Aggarwal R, Ahn SY, Ali MK, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Bahalim AN, Barker-Collo S, Barretero LH, Bartels DH, Basáñez MG, Baxter A, Bell ML, Benjamin EJ, Bennett D, Bernabé E, Bhalla K, Bhandari B, Bikbov B, Bin Abdulhak A, Birbeck G, Black JA, Blencowe H, Blore JD, Blyth F, Bolliger I, Bonaventure A, Boufous S, Bourne R, Boussinesq M, Braithwaite T, Brayne C, Bridgett L, Brooker S, Brooks P, Brugha TS, Bryan-Hancock C, Bucello C, Buchbinder R, Buckle G, Budke CM, Burch M, Burney P, Burstein R, Calabria B, Campbell B, Canter CE, Carabin H, Carapetis J, Carmona L, Cella C, Charlson F, Chen H, Cheng AT, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahiya M, Dahodwala N, Damere-Derry J, Danaei G, Davis A, De Leo D, Degenhardt L, Dellavalle R, Delossantos A, Denenberg J, Derrett S, Des Jarlais DC, Dharmaratne SD, Dherani M, Diaz-Torne C, Dolk H, Dorsey ER, Driscoll T, Duber H, Ebel B, Edmond K, Elbaz A, Ali SE, Erskine H, Erwin PJ, Espindola P, Ewingsokhan SE, Farzadfar F, Feigin V, Felson DT, Ferrari A, Ferri CP, Fèvre EM, Finucane MM, Flaxman S, Flood L, Foreman K, Forouzanfar MH, Fowkes FG, Franklin R, Fransen M, Freeman MK, Gabbe BJ, Gabriel SE, Gakidou E, Ganatra HA, Garcia B, Gaspari F, Gillum RF, Gmel G, Gosselin R, Grainger R, Groeger J, Guillemin F, Gunnell D, Gupta R, Haagsma J, Hagan H, Halasa YA, Hall W, Haring D, Haro JM, Harrison JE, Havmoeller R, Hay RJ, Higashi H, Hill C, Hoen B, Hoffman H, Hotez PJ, Hoy D, Huang JJ, Ibeanusi SE, Jacobsen KH, James SL, Jarvis D, Jasrasaria R, Jayaraman S, Johns N, Jonas JB, Karthikeyan G, Kassebaum N, Kawakami N, Keren A, Khoo JP, King CH, Knowlton LM, Kobusingye O, Koranteng A, Krishnamurthi R, Laloo R, Laslett LL, Lathlean T, Leasher JL, Lee YY, Leigh J, Lim SS, Limb E, Lin JK, Lipnick M, Lipshultz SE, Liu W, Loane M, Ohno SL, Lyons R, Ma J, Mabwijano J, MacIntyre MF, Malekzadeh R, Mallinger L, Manivannan S, Marcenes W, March L, Margolis DJ, Marks GB, Marks R, Matsumori A, Matzopoulos R, Mayosi BM, McAnulty JH, McDermott MM, McGill N, McGrath J, Medina-Mora ME, Meltzer M, Mensah GA, Merriman TR, Meyer AC, Miglioli V, Miller M, Miller TR, Mitchell PB, Mocumbi AO, Moffitt TE, Mokdad AA, Monasta L, Montico M, Moradi-Lakeh M, Moran A, Morawska L, Mori R, Murdoch ME, Mwaniki MK, Naidoo K, Nair MN, Naldi L, Narayan KM, Nelson PK, Nelson RG, Neuvitt MC, Newton CR, Nolte S, Norman P, Norman R, O'Donnell M, O'Hanlon S, Olives C, Omer SB, Ortblad K, Osborne R, Ozgediz D, Page A, Pahari B, Pandian JD, Rivero AP, Patten SB, Pearce N, Padilla RP, Perez-Ruiz F, Perico N, Pesudovs K, Phillips D, Phillips MR, Pierce K, Pion S, Polanczyk GV, Polinder S, Pope CA 3rd, Popova S, Porrini E, Pourmalek F, Prince M, Pullan RL, Ramaiah KD, Ranganathan D, Razavi H, Regan M, Rehm JT, Rein DB, Remuzzi G, Richardson K, Rivara FP, Roberts T, Robinson C, De Leòn FR, Ronfani L, Room R, Rosenfeld LC, Rushton L, Sacco RL, Saha S, Sampson U, Sanchez-Riera L, Sanman E, Schwebel DC, Scott JG, Segui-Gomez M, Shahraz S, Shepard DS, Shin H, Shivalokoti R, Singh D, Singh GM, Singh JA, Singleton J, Sleet DA, Sliwa K, Smith E, Smith JL, Stapelberg NJ, Steer A, Steiner T, Stolk WA, Stovner LJ, Sudfeld C, Syed S, Tamburlini G, Tavakkoli M, Taylor HR, Taylor JA, Taylor WJ, Thomas B, Thomson WM, Thurston GD, Tleyjeh IM, Tonelli M, Towbin JA, Truelsen T, Tsilimbaris MK, Ubeda C, Undurraga EA, van der Werf MJ, van Os J, Vavilala MS, Venketasubramanian N, Wang M, Wang W, Watt K, Weatherall DJ, Weinstock MA, Weintraub R, Weisskopf MG, Weissman MM, White RA, Whiteford H, Wiersma ST, Wilkinson JD, Williams HC, Williams SR, Witt E, Wolfe F, Woolf AD, Wulf S, Yeh PH, Zaidi AK, Zheng ZJ, Zonies D, Lopez AD, Murray CJ, AlMazroa MA, Memish ZA. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; 380: 2163-2196.
- 28) Muñoz-García N, Cordero-Ampuero J, Madero-Jarabo R. Diagnostic accuracy of magnetic resonance images and weight-bearing radiographs in patients with arthroscopic-proven medial osteoarthritis of the knee. *Clin Med Insights Arthritis Musculoskelet Disord* 2020; 13: 1179544120938369.
- 29) Zhou X, Liu Q, Liang T, Xu P, Liu Y, Fu S, Wang G, Zhang L. Arthroscopy combined with high tibial osteotomy for the treatment of knee medial compartment osteoarthritis and its influence on cartilage injury. *Chin J Reparative Reconstr Surg* 2021; 35: 690-696.
- 30) Li HB, Nie S, Lan M, Tang ZM, Liao XG. The risk factors for early conversion to total knee arthroplasty after high tibial osteotomy. *Comput Methods Biomech Biomed Engin* 2020; 19: 1-7.
- 31) Steadman JR, Briggs KK, Matheny LM, Ellis HB. Ten-year survivorship after knee arthroscopy in patients with Kellgren-Lawrence grade 3 and grade 4 osteoarthritis of the knee. *Arthroscopy* 2013; 29: 220-225.