

Automated Deep Learning Framework for Radiographic View Classification in 17,958 Total Knee Arthroplasty Patients: Enabling Scalable Clinical Imaging Analysis

Yuxuan Jin, Mingrui Yang, Xiaojuan Li, Carl S Winalski, Daniel M Rotroff¹, Nicolas Santiago Piuze

¹Department of Quantitative Health Sciences

INTRODUCTION:

Accurate identification of standardized knee radiograph views is essential for evaluating preoperative alignment and monitoring outcomes after total knee arthroplasty (TKA). Deep learning has received growing attention over the past few decades, with promising applications in orthopedics. Several deep learning architectures developed for general computer vision tasks have achieved strong performance; however, their effectiveness in screening and identifying knee radiographs of specific view from large-scale, real-world clinical encounters remains less well understood. In routine clinical workflows, radiographs are often captured in multi-view series (e.g., AP, PA, merchant, lateral, fixed-flexion), yet inconsistencies in Digital Imaging and Communications in Medicine (DICOM) metadata hinder large-scale automated analysis. This study introduces a clinically applicable, two-stage deep learning (DL) framework designed to screen and classify knee radiographs from 17,958 TKA patients. Our goal is to enable downstream tasks such as implant planning, surgical outcome prediction, and alignment measurement at scale.

METHODS:

A total of 81,156 unique accession numbers were identified from 17,958 unique patients undergoing a primary elective unilateral TKA at a large tertiary academic center in the United States between 2016-2022. We selected accession numbers from the knee radiograph session immediately before each patient's first primary TKA surgery, as well as from all subsequent radiograph sessions following the surgery. A total of 78,464 unique accession numbers, resulting in 396,579 DICOM images of various views, were pulled from Picture Archiving and Communication System (PACS). To distinguish bilateral AP/PA knee radiographs from those of other views (merchant, lateral, unilateral AP/PA), a set of 433 images was manually labeled as either bilateral or other view. The images were then randomly split at the accession number level into training, validation and test sets according to a 60%/20%/20% ratio. Standard image pre-processing steps were applied. The models were trained using ResNet-18, a convolutional neural network (CNN), to classify images as bilateral or other view. The models were initialized using three different seeds, each trained with early stopping (if validation loss did not improve for 5 epochs) and without early stopping (30 epochs). All model parameters were trained using the Adam optimizer with a learning rate of 0.001, cross-entropy loss, and a batch size of 8. Results were reported using accuracy, Area under the ROC curve (AUC), precision, recall and F-1-score. A similar approach was used to train deep learning models to distinguish bilateral fixed-flexion view versus bilateral standing view. To screen the images at scale, three models without early stopping were applied, and majority voting was used to obtain the final prediction. The models were trained using the PyTorch library (version 2.3.1) in Python (version 3.9) on a CPU-based Linux workstation.

RESULTS: A total of 433 images of various views from 131 accession numbers were manually labeled, with 272 images (107 bilateral, 165 others) in training set, 82 images (31 bilateral, 51 others) in validation set, and 79 images (30 bilateral, 49 others) in test set. Figure 1 shows example images. The performance of the six models trained to distinguish between bilateral vs other view is shown in Table 1. Overall, the models have achieved good performance on test set with 0.99 AUC. The model results based on majority voting are further evaluated using a random sample of 250 images predicted as bilateral and another 250 images predicted as other view; all predictions are found to be correct. Similarly, the models trained with another 477 manually labeled images (338 in training, 72 in validation, 67 in test) to classify bilateral fixed-flexion versus bilateral standing view achieved perfect results on test set, with 100% accuracy and 1.00 AUC on all of them. After applying these models to the 396,579 images pulled from PACS, we identify that 87,696 are predicted as bilateral images, and 13,968 are predicted as bilateral fixed-flexion views. Based on this, we conclude that approximately 8.5% of the patients did not have preoperative bilateral images, and only 42% of them had preoperative bilateral fixed-flexion view images.

DISCUSSION AND CONCLUSION:

This study demonstrates the feasibility and accuracy of using deep learning models to automatically classify clinically relevant radiographic views in a large cohort of TKA patients. Our two-stage framework reliably identified bilateral and fixed-flexion knee radiographs, overcoming the limitations of inconsistent DICOM metadata and significantly reducing the work required to obtain specific views from radiographs in the absence of good quality series description in the DICOM header while working with large-scale imaging data. This capability enables scalable triage of high-volume imaging datasets and lays the groundwork for integrating radiographs into predictive models for alignment correction goals, surgical planning, and outcomes after TKA. By facilitating image curation at scale, this approach supports downstream applications such as AI-driven segmentation, deformity assessment, and personalized surgical planning. Ultimately, it provides a practical and scalable tool for advancing data-driven orthopaedic care and optimizing surgical decision-making in

TKA.

Stage One Model



Stage Two Model



Table 1. Model performance on the test data for distinguishing bilateral AP/PA knee radiographs from other views (79 images in total; 30 bilateral, 49 other views). The final model (not shown in this table) was developed as an ensemble of the three models without early stopping described below, using majority voting.

Seed	Early Stopping (Yes/No)	Accuracy	AUC	Precision	Recall	F-1 Score
142	Yes	98.73%	0.99	Bilateral: 0.97 Other: 1.00	Bilateral: 1.00 Other: 0.98	Bilateral: 0.98 Other: 0.99
142	No	98.73%	0.99	Bilateral: 0.97 Other: 1.00	Bilateral: 1.00 Other: 0.98	Bilateral: 0.98 Other: 0.99
456	Yes	97.47%	0.99	Bilateral: 0.97 Other: 0.98	Bilateral: 0.97 Other: 0.98	Bilateral: 0.97 Other: 0.98
456	No	97.47%	0.99	Bilateral: 0.97 Other: 0.98	Bilateral: 0.97 Other: 0.98	Bilateral: 0.97 Other: 0.98
789	Yes	94.94%	0.99	Bilateral: 0.93 Other: 0.96	Bilateral: 0.93 Other: 0.96	Bilateral: 0.93 Other: 0.96
789	No	94.94%	0.99	Bilateral: 0.93 Other: 0.96	Bilateral: 0.93 Other: 0.96	Bilateral: 0.93 Other: 0.96