Multimodal Deep Learning Models Integrating Preoperative Imaging and Clinical Data to Predict One-Year Patient-Reported Outcomes After Total Knee Arthroplasty

Yuxuan Jin, Mingrui Yang, Xiaojuan Li, Carl S Winalski, Daniel M Rotroff[1], Nicolas Santiago Piuzzi
[1]Department of Quantitative Health Sciences

INTRODUCTION: Artificial intelligence offers major potential in orthopedics by enabling predictive modeling for surgical outcomes. Historically, most models have relied on structured clinical variables from electronic health records. However, radiographic imaging may offer additional predictive insight if effectively integrated. This study developed and evaluated multimodal deep learning (DL) models combining clinically available preoperative bilateral knee radiographs and clinical variables to predict one-year patient-reported outcomes (PROMs) after total knee arthroplasty (TKA).

METHODS: A total of 17,958 unique patients undergoing a primary elective unilateral TKA at a large tertiary academic center in the United States between 2016-2022 were included. Based on a previously developed deep learning framework, we screened and selected the preoperative bilateral AP/PA knee radiographs from the session immediately before each patient's first primary TKA surgery. If multiple bilateral images existed from the same session, one image was selected at random. We also obtained preoperative clinical variables such as age, sex, BMI, race, years of education, smoking status, insurance, patient-reported outcomes (PROM) phenotype, ADI, CCI, diagnosis, anesthesia, baseline KOOS JR, NARX score; and 1-year PROMs including KOOS Pain and PS from electronic health records. Stratified sampling based on outcome was applied to split the dataset at the patient level into training, validation and test sets according to a 70%/15%/15% ratio (6958 patients in training, 1491 in validation, and 1492 in test). We followed standard image pre-processing steps to handle raw pixel data extracted from radiograph images, including resizing images to 224×224, normalizing using the mean and standard deviation from the training set, applying augmentation methods such as random rotations up to 10 degrees, and padding the edges with zeros to ensure square dimensions. Inverted pixel values were corrected by subtracting the original pixel values from the maximum pixel value. Three separate models, including one using image data only, one using clinical data only, one combining image and clinical data, were trained to compare prediction performance for 1-year PROMs after TKA. The binary outcomes were defined as whether the patients achieved an MCID threshold for KOOS Pain (a delta of 7.99), whether the patients achieved a PASS threshold for KOOS Pain (1-year KOOS Pain of 77.7), and similarly for their counterparts for KOOS PS (delta of 8.04; 1-year KOOS PS of 70.3). A ResNet-18 architecture was used as an image feature extractor for raw pixel data. Deep learning models were trained as follows: for image only model, this was achieved through the replacement of the final layer from ResNet-18 with a linear layer. The clinical data only model was a two-layer neural network with one linear layer and one classification layer. Lastly, the combined model was trained by concatenating the image features obtained from ResNet-18 and clinical data features obtained from the linear layer. All model parameters were trained using the Adam optimizer with a learning rate of 0.001, weighted cross-entropy loss, and a batch size of 64. Early stopping was applied when the validation loss did not improve for five epochs. As a comparative benchmark, the traditional machine learning model XGBoost was trained for all three models. Results were reported using Area under the ROC curve (AUC). Missing data in clinical variables were singly imputed using multiple imputation using *mice* package in R software (Version 4.3.1). The models were trained using the PyTorch library (version 2.3.1), and scikit learn library (version 1.5.0) in Python (version 3.9) on high performance computing cluster equipped with NVIDIA Tesla V100 Tensor Core graphics processing unit.

RESULTS:
Among the 17,958 unique patients identified at baseline, 10,988 of them have 1-year follow up PROMs, with 9,941 of them having at least one bilateral knee radiograph at baseline. Detailed descriptive statistics can be found in Table 1. On test data, for the KOOS Pain PASS threshold models, the image only model achieved an AUC of 0.54, whereas the clinical data only model and the combined model achieved AUCs of 0.65; for the KOOS Pain MCID threshold models, the image only model achieved an AUC of 0.52, whereas the clinical data only model and the combined model achieved AUC of 0.70, and 0.65 respectively. For the KOOS PS PASS threshold models, the image only model achieved an AUC of 0.51, whereas the clinical data only model and the combined model achieved AUC of 0.66, and 0.63 respectively. Lastly, for the KOOS PS MCID models, the image only model achieved an AUC of 0.54, whereas the clinical data only model and the combined model achieved AUCs of 0.73. The XGBoost models showed similar results.

DISCUSSION AND CONCLUSION:
Multimodal deep learning models incorporating preoperative image and clinical variables can predict 1-year PROMs after TKA with moderate to good performance. However, the results showed that the value of adding images to 1-year PROMs prediction was limited when they were used as sole inputs to the models or included in the models with clinical variables. These findings suggest that extracting specific radiographic features—such as alignment angles or implant positioning—may offer greater clinical utility than raw pixel data. Future work should focus on incorporating such quantitative imaging biomarkers to enhance surgical planning, patient counseling, and risk stratification in knee arthroplasty.

Table 1. Descriptive Statistics of the analysis cohort

Variable	Level	Total (N=9941)	N
Age		67.0 [61.0;73.0]	9941
Sex	Male	3988 (40.1%)	9941
	Female	5953 (59.9%)	
BMI		30.9 [27.2;35.4]	9941
Race	White	8439 (86.8%)	9723
	Black	208 (2.14%)	
	Other	1076 (11.1%)	
Education		14.0 [12.0;16.0]	9941
Smoking	Never	5716 (57.5%)	9941
	Quit 6m+	3385 (34.1%)	
	Quit <6m	303 (3.05%)	
	Current	537 (5.40%)	
ADI		54.0 [34.0;73.0]	9598
CCI		0.00 [0.00;2.00]	9877
Insurance	Private	3160 (35.3%)	8949
	Medicaid	312 (3.49%)	
	Medicare	5364 (59.9%)	
	Other	113 (1.26%)	
Baseline KOOS Pain		41.7 [30.6;52.8]	9941
Baseline KOOS PS		51.5 [42.1;59.7]	9936
Baseline MCS		53.0 [43.4;61.1]	9941
PROMPhenotype	Pain+ PS+ MCS+	2446 (24.6%)	9936
	Pain+ PS+ MCS-	1460 (14.7%)	
	Pain+ PS- MCS+	339 (3.41%)	
	Pain+ PS- MCS-	343 (3.45%)	
	Pain- PS+ MCS+	835 (8.40%)	
	Pain- PS+ MCS-	785 (7.90%)	
	Pain- PS- MCS+	1398 (14.1%)	
	Pain- PS- MCS-	2330 (23.5%)	
Diagnosis	OA	9682 (97.4%)	9941
	non-OA	259 (2.61%)	
Anesthesia	General	2283 (23.0%)	9939
	Spinal	7364 (74.1%)	
	Other	292 (2.94%)	
Narx		110 [0.00;190]	5880
Baseline KOOS JR		44.9 [36.9;52.5]	9077
1-year KOOS Pain		86.1 [72.2;97.2]	9864
1-year KOOS PS		75.1 [66.4;85.2]	9456
KOOS Pain MCID Improvement	Treatment Failure	633 (6.42%)	9864
	MCID Improved	9231 (93.6%)	
KOOS PS MCID Improvement	Treatment Failure	1397 (14.8%)	9452
	MCID Improved	8055 (85.2%)	
KOOS Pain PASS Threshold	Treatment Failure	2986 (30.3%)	9864
	PASS Improved	6878 (69.7%)	
KOOS PS Threshold	Treatment Failure	3220 (34.1%)	9456
	PASS Improved	6236 (65.9%)	