Deep Learning-Based Total Knee Replacement Prediction from Magnetic Resonance Images and Biomarker Analysis

Aniket Tolpadi[12], Jinhee Lee[2], Valentina Pedoia[2], Sharmila Majumdar[2]
[1]*Department of Bioengineering, University of California, Berkeley*
[2]*Department of Radiology and Biomedical Imaging, University of California, San Francisco, USA*

Introduction

Knee Osteoarthritis (OA) is a common musculoskeletal disorder in the United States that frequently causes disability [1-2]. Its progression is typically assessed with the Kellgren-Lawrence (KL) scale, a 0-4 scale in which higher scores imply more advanced OA [3]. When diagnosed early (KL = 0, 1), nonsurgical treatments such as exercise, weight loss, and injections can slow OA progression, but at late stages (KL = 4), no noninvasive option exists, making total knee replacement (TKR) the only option [4-5]. The procedure is effective but imperfect: only 66% of patients report their knees feeling "normal," and 33% report pain post-implant [6]. This, as well as potential complications that necessitate revisions, makes delaying TKR preferable whenever possible [7-8]. Thus, a model predicting if patients will undergo TKR with high sensitivity and specificity would have utility, particularly at early-stage OA.

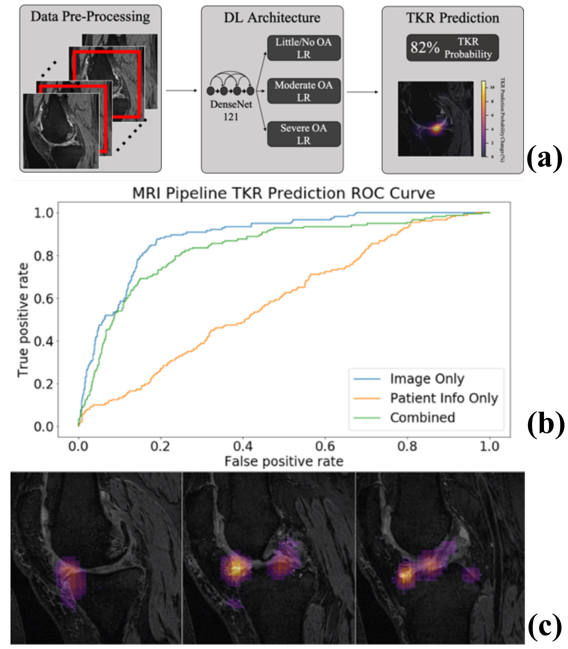

Figure 1: (a) Deep Learning (DL) pipeline integrating MRI images, non-image features to predict TKR; (b) ROC curves of pipeline; (c) Slices of occlusion map of true positive from pipeline overlaid on DESS MRI image.

Methods

3D Double Echo Steady-State MRI images from the Osteoarthritis Initiative (OAI) were center-cropped to a 120×320×320 region, normalized, and converted to 14-bit [9]. Cases were defined as patients who underwent a first TKR within 5 years of an image; controls were patients who did not undergo a first TKR within 5 years and for whom another image was taken 5 years after the given image.

A scheme of the modeling pipeline used is shown in Figure 1a. A DenseNet-121 classifier was pretrained to predict OA from MRI images and fine-tuned to predict TKR. Image-based predictions were fed to one of three logistic regression (LR) ensembles based on OA severity, yielding a final prediction. Ensembles integrated image-based predictions with 21 non-imaging variables relevant to TKR, covering pain metrics, physical performance tests, and demographics [10-12]. Performance of 2 ensemble versions are reported: one where non-imaging variables were used to predict TKR (patient only), and one where image-based predictions were added (combined). DenseNet-121 output is also reported (image only). 75% of test cases were randomly sampled 100 times to calculate mean, standard error, and confidence intervals. Occlusion maps were developed for true positives, and anatomic regions with pixels among the top 5% of TKR probability change were designated as hotspots.

OA Status	Model Type	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Non-TKR Cases	TKR Cases
None	Patient Only	77.1 +/- 0.271	26.3 +/- 1.804	78.6 +/- 0.307	2892	12
	Image Only	95.1 +/- 0.067	68.5 +/- 2.208	95.3 +/- 0.068		
	Combined	82.4 +/- 0.129	74.3 +/- 0.768	82.6 +/- 0.127		
Moderate	Patient Only	78.3 +/- 0.375	31.1 +/- 0.956	79.6 +/- 0.409	2056	83
	Image Only	68.7 +/- 0.179	78.4 +/- 0.709	68.3 +/- 0.178		
	Combined	83.4 +/- 0.136	64.7 +/- 1.051	84.0 +/- 0.132		
Severe	Patient Only	72.6 +/- 0.661	33.8 +/- 3.920	74.6 +/- 0.772	141	57
	Image Only	34.9 +/- 0.532	98.3 +/- 0.283	9.1 +/- 0.358		
	Combined	77.1 +/- 0.940	51.9 +/- 3.871	78.3 +/- 1.146		
ALL	Patient Only	77.4 +/- 0.089	27.8 +/- 0.546	78.9 +/- 0.089	5089	152
	Image Only	82.1 +/- 0.094	85.1 +/- 0.407	82.0 +/- 0.095		
	Combined	82.6 +/- 0.090	70.0 +/- 0.624	83.0 +/- 0.089		

Table 1: Improved performance of pipeline at early-stage OA with integration of non-image features justifies decreased AUC in integrated model.

Results

Pipeline performance is shown in Table 1 and Figure 1b. Furthermore, areas under the Receiver-Operating Characteristic (ROC) curve are as follows, $p < 0.01$: 0.885 ± 0.021 (image only), 0.587 ± 0.042 (patient only), 0.830 ± 0.033 (combined).

Figure 1c is an example occlusion map. The following were hotspots in at least 83% of true positives: medial patellar retinaculum (MPR), synovium, tibiofemoral joint cartilage and bone (medial and lateral), anterior and posterior meniscus (medial and lateral), and Hoffa fat pad. All but the MPR are implicated in OA progression [13].

Conclusions

We present a pipeline that predicts TKR with strong sensitivity and specificity, particularly for less advanced OA. From occlusion map analysis, we find many tissues from which the pipeline drew most consistently to predict TKR have been implicated as OA biomarkers, which is logical, as TKR is an outcome of OA progression. This finding shows promise that further studies can facilitate discovery of additional imaging biomarkers, including those preferentially affecting TKR prediction probability for patients with no or moderate OA.

Highlights: We present a deep learning pipeline that uses 3D DESS MRI images and 21 non-imaging variables to predict TKR with high sensitivity and specificity, particularly for patients with less-advanced OA. The model may have utility in identifying an at-risk population so nonsurgical treatments can be implemented that delay TKR onset.

References

[1] Kremers HM et al. J Bone Surg Am 2015;97(17):1386-97. [2] Murphy LB et al. Arthritis Care Res (Hoboken) 2017;70(6):869-76. [3] Kellgren JH et al. Ann Rheum Dis 1957;16:494-502. [4] Ringdahl E et al. Am Fam Physician 2011;83(11):1287-92. [5] Tiulpin A et al. Sci Rep 2018;8(1727):1-10. [6] Parvizi J et al. Clin Orthop Relat Res 2014;472:133-37. [7] Chang MJ et al. Knee Surg Relat Res 2014;26(2):61-7. [8] Kim KT et al. Knee Surg Relat Res 2014;26(1):13-9. [9] Peterfy CG et al. Osteoarthr Cartilage 2008;16(12):1433-41. [10] Riddle DL et al. Knee 2009;16(6):494-500. [11] Hawker GA et al. Arthritis Rheum 2006;54(10):3212-20. [12] Lewis JR et al. PLoS One 2013;8(12):1-8. [13] Collins JE et al. Arthritis Rheumatol 2016;68(10):2422-31.