Predicting bilateral knee pain from MR imaging using deep neural networks

Gary H. Chang[1], David T. Felson[2], Shangran Qiu[3], Terence D. Capellini[4,5], Vijaya B. Kolachalama[1, 6, 7, *]

[1]Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA – 02118

[2]Clinical Epidemiology Research and Training Unit, Department of Medicine, Boston University School of Medicine, Boston, MA, USA – 02118

[3]Department of Physics, College of Arts and Sciences, Boston University, Boston, MA, USA – 02215

[4]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA – 02138

[5]Broad Institute of MIT and Harvard, Cambridge, MA, USA 02142

[6]Whitaker Cardiovascular Institute, Boston University School of Medicine, Boston, MA, USA - 02118

[7]Hariri Institute for Computing and Computational Science and Engineering, Boston University, Boston, MA, USA – 02215

*Corresponding author:

Vijaya B. Kolachalama, PhD

Boston University School of Medicine

72 E. Concord Street, Evans 636, Boston, MA, USA – 02118

Email: vkola@bu.edu

Phone: 617-358-7253

ABSTRACT

Background and objective: It remains difficult to characterize pain in knee joints with osteoarthritis solely by radiographic findings. We sought to understand how advanced machine learning methods such as deep neural networks can be used to analyze MRI scans and predict bilateral knee pain.

Methods: We developed a deep learning framework to associate information from MRI slices taken from the left and right knees of subjects from the Osteoarthritis Initiative with bilateral knee pain. Model training was performed by first extracting features from two-dimensional (2D) sagittal intermediate-weighted turbo spin echo slices. The extracted features from the 2D slices were subsequently combined to directly associate using a fused deep neural network with the output of interest as a binary classification problem.

Results: The deep learning model resulted in predicting bilateral knee pain on test data with 70.1% mean accuracy, 51.3% mean sensitivity, and 81.6% mean specificity. Systematic analysis of the predictions on the test data revealed that the model performance was consistent across subjects of various Kellgren-Lawrence grades.

Conclusion: The study demonstrates a proof of principle that a machine learning approach can be applied to associate MR images with bilateral knee pain.

INTRODUCTION

Osteoarthritis (OA) is the most common musculoskeletal disease and one of the leading causes of disability globally [1]. It has become widely recognized that OA is a disease of the whole joint [2], and as OA progresses, the tissue and structures of the whole joint can be affected, including the degradation of the cartilage and lesions of the bone marrow, menisci, and synovium. For knee OA the frequency and severity of pain increase during OA progression, and severe OA-induced pain often leads to disability. Currently, there is no effective cure for advanced stage knee OA other than total joint replacement surgery. Identifying pre-clinical stages of OA, preferably when knee pain is the only reported output of interest, can hopefully facilitate better patient management.

The occurrence of pain in knee joints with OA can be correlated with a variety of structural findings [3], and neuropathic mechanisms such as central sensitization and hyperalgesia [4-6]. For example, structural findings such as bone marrow lesions (BMLs), cartilage damage, synovitis, and effusion are related to pain in knee joints with OA [3, 7-10]. Also, the frequency and severity of pain are self-reported and usually defined subjectively [4, 11]. As a consequence, the correlation between pain and radiographic findings is weak and there has been little success in correlating OA-induced pain with a specific type and location of structural damage. One review found that the proportion of subjects with knee pain who have radiographic OA ranged from 15 to 76% [12]. Identifying the source of OA-induced pain in each patient can greatly benefit the design of targeted, individualized treatments to reduce symptoms and to limit disability [11].

Magnetic resonance imaging (MRI) techniques are capable of providing more detailed structural information about the knee joint than radiographs. It was found in a systematic review of MRI measures that knee pain may arise from BMLs, effusion, and synovitis; however, the correlation between pain and MRI findings was inconsistent and moderate [13-15]. MRI readings are also limited by the interpretation of the individual radiologist. Hence, there is a need for an unbiased method to objectively and accurately associate MR imaging data with knee pain.

Machine learning (ML) is a discipline within computer science that uses computational algorithms for the analysis of various forms of data. Machine learning algorithms applied on medical images have shown remarkable success in predicting various outcomes of interest. Over the past few years, a new ML modality known as deep learning is gaining popularity because of its ability to analyze large volumes of data for pattern recognition and prediction with unprecedented levels of success [16]. Specifically, deep learning frameworks such as convolutional neural networks (CNN) are increasingly being leveraged for object recognition tasks and specifically for disease classification [16, 17]. Traditional ML algorithms require visual or statistical features to be manually identified and selected ("handcrafted"), and researchers need to decide which of these handcrafted features are related to the problem at hand. On the contrary, deep learning algorithms extract visual features automatically, and one can utilize these features simultaneously for various applications related to classification, segmentation, and detection [18]. CNN model training is associated with learning a series of image filters through numerous layers of feed-forward neural networks. The filters are then projected on the original input image, and the image features that are most correlated with the outcome are extracted through the training process. Recently, deep learning techniques were applied on MRI scans of the knee joint for automatic cartilage and meniscus segmentation [19], on X-ray imaging for automatic Kellgren-Lawrence (KL) grade classification [20], and on the localization of cartilage [21].

In this paper, we investigated the feasibility of a deep learning framework to predict bilateral knee pain directly from MRI scans. Imaging and clinical assessment data were taken from the Osteoarthritis Initiative (OAI) study. In this study, bilateral knee pain was defined when an individual had pain, aching or stiffness for more than half of the days in a month on both knees. We used sagittal intermediate-weighted turbo spin echo (SAG-IW-TSE) sequence images as the inputs and bilateral knee pain as the output to solve a binary classification problem using the deep learning framework. Initially, we built deep learning models using individual two-dimensional (2D) slices as inputs to the binary classifier. Later, for the fused deep learning model, training was performed by first extracting features from 2D SAG-IW-TSE slices, and then the extracted features were combined to directly associate with the output of interest, knee pain. Mean performance across multiple model runs was reported using standard metrics from

the literature. To the best of our knowledge, our article is the first to leverage a deep learning framework to associate MR images of the knee with bilateral knee pain.

METHODS

Subjects and MR imaging data

Our sample was drawn from OAI, an NIH funded cohort study of persons with or at risk of knee OA [22, 23]. The baseline MR image dataset (0.E.1) from the OAI database was used for training and testing our deep learning model (**Table 1**). The dataset consists of imaging and clinical data from 2059 subjects. Among them, 1383 subjects with pain, aching or stiffness for more than half the days of a month in both knees (bilateral knee pain, labeled 1) or neither knee (no knee pain, labeled 0) were selected. From these cases, 1349 subjects passed initial quality checks and used for training and testing the deep learning model. Among the selected cases, 512 subjects had bilateral knee pain (class 1), and 837 subjects had no pain in either knee (class 0). We selected SAG-IW-TSE-FS sequence images as the inputs and bilateral knee pain as the output to solve a binary classification problem using the deep learning framework. Previously, IW sagittal MRI images were found to be sufficient in finding focal cartilage damage [24], and BML [25]. The original 2D MRI slices had dimensions of 480×448 pixels. The imaging was performed with a 3.0 T magnet using imaging sequence with TR/TE of 3200/30 ms. The in-plane spatial resolution was 0.357×0.511 mm and the slice thickness was 3.0 mm [22].

Image registration and quality check

The majority of the SAG-IW-TSE sequence imaging comprised of 37 slices per knee. In order to optimize output using our deep learning framework, we first manually examined all the MRI slices oriented in the sagittal direction and selected the slice showing the most complete view of posterior cruciate ligament (PCL) and indexed it as the center slice (Red colored box in **Figure 1**). The remaining slices were indexed relative to the center slice for each knee, with a plus sign indicating the lateral direction and a minus sign indicating the medial direction. After the images were indexed, for each 2D MRI slice of the knee for a subject, we performed Euclidean transformation to align the slice with respect to a template that we previously picked from the MRI scans of a subject. A region (Red colored box in **Figure 2**) with dimensions of 294×294 pixels was subsequently selected and all the registered slices were cropped to this region of

interest, which contained femoral, tibial, and patellar components of bone, cartilage and meniscus. All the cropped slices were further resized into images of 224×224 pixels, and 12 adjacent slices on the lateral side of the center slice, 12 adjacent slices on the medial side of the center slice along, and the center slice were selected for deep learning. In total 50 MRI slices were selected for each subject (25 from each knee). After image registration, we performed a quality check operation that involved manually observing each sagittal slice of each scan for artifacts such as missing data, abnormal misalignment within a slice, presence of a foreign object, and cases with sub-regions of high contrast (**Figure 3**). All these cases were not considered for model training or testing.

Neural network architecture

A pre-trained neural network architecture known as Residual Network (or ResNet) was used on each 2D SAG-IW-TSE slice, and the features extracted from each MRI slice were then combined to directly associate with the output of interest (i.e. bilateral knee pain) (**Figure 4**). For feature extraction from each of the 2D MRI slice, we utilized the 18-layer ResNet model. The ResNet model allowed us to perform transfer learning to efficiently extract the features from each 2D MRI slice. This approach utilizes pre-trained CNN models whose features were already learned using millions of images with different object classes. Moreover, transfer learning usually results in faster training times than it would were we to construct a new CNN from scratch, since there is no need to estimate all the parameters in a pre-trained CNN.

The extracted features from each 2D MRI slice were combined using fully-connected layers with regularization methods of dropout and batch normalization to create the final binary prediction (**Figure 4**). The dropout layers improve the generalizability of the model by randomly deactivating units in the hidden layers during the feed-forward and back propagation operations [26]. A batch normalization layer normalizes the outputs from its previous layer using their mean and variance. It has been shown that both methods can effectively improve the generalizability of the model and reduce overfitting. In addition, we used L_2-norm regularization (coefficient = $1e^{-2}$) as another step to minimize overfitting. The training was performed using Adam optimizer and cross-entropy loss function for binary classification, and a batch size of 32.

Model construction and performance metrics

The OAI dataset was randomly split in the ratio 7:3 such that 70% of the data was used for model training and the remaining for testing. For completeness, the process of randomly splitting the data in 7:3 ratio and then performing model training and testing was repeated 10 times, and the model performance on the test data was averaged across all the runs. During each time, the ratio of cases with knee pain and the ones that do not have knee pain was kept the same for the training data. These steps ensured that we created a consistent and unbiased deep learning model to predict bilateral knee pain. After every run, model accuracy, sensitivity and specificity was computed based on the model performance on the test data, and mean values of these performance metrics were computed across the 10 model runs. We also computed area under curve (AUC) of the receiver operating characteristic (ROC) curves of the binary classifier.

Statistical analysis

Descriptive statistics are presented as the mean along with 95% confidence intervals. Unpaired Student's t-test was used to compare two different groups as appropriate. A p-value < 0.05 was considered statistically significant.

RESULTS

The slice number of the MRI scan containing the PCL from the original OAI dataset varied between individuals (Assigned an ID - '0' in **Figure 1**). Scan alignment allowed us to select the appropriate slice, which then led to presenting similar information from each subject to the deep learning model. Also, the Euclidean transformation to align a specific 2D slice with respect to the template, followed by cropping the region comprising the knee joint, allowed us to present consistent information from each subject to the deep neural network (**Figure 2**).

Our initial set of experiments focused on leveraging pre-trained ResNet models to train individuals 2D slices on either knee to predict knee pain. Given that we selected cases with only bilateral knee pain or no knee pain, for consistency, we simply kept the dataset unchanged even for predicting slice-specific knee pain. Two-dimensional model performance across different slices and averaged over 10 runs resulted in mean accuracy of 62.45% (95% CI: 61.96%, 62.94%), mean sensitivity of 53.08% (95% CI: 51.57%, 54.59%), and mean specificity of 68.2% (95% CI: 66.98%, 69.42%) on the left knee for the test data (**Figure 5**). Similarly, for the right knee, 2D model performance across different slices and averaged over 10 runs resulted in mean accuracy of 61.73% (95% CI: 61.4%, 62.06%), mean sensitivity of 54.41% (95% CI: 53.48%, 55.34%), and mean specificity of 66.23% (95% CI: 65.51%, 66.95%) on the test data (**Figure 5**). When the features from these 2D models were combined at the fully connected layer, the fused deep neural network resulted in an improved performance in terms of predicting the output of interest. Specifically, the fused deep neural network resulted in mean accuracy of 70.27% (95% CI: 69.44%, 71.11%), mean sensitivity of 51.90% (95% CI: 50.22%, 53.57%) and mean specificity of 81.57% (95% CI: 79.74%, 83.40%) (**Figure 6A**). Also, across the 10 model runs, the mean AUC of the ROC curve was 0.7268 (95% CI: 0.7101, 0.7435) (**Figure 6B**).

While the goal of our deep learning model was to predict bilateral knee pain from MR images, the cases that we considered for model training had persons with no radiographic features of OA and also others with varying degrees of OA as defined using the Kellgren-Lawrence (KL) grading scale [27]. After examining the performance of the deep learning model on the test subjects, we found that our model did not perform significantly differently between the group of

subjects with no reported radiographic features of OA (KL=0), and radiographic feature of OA (KL>0) in both the left (p=0.16) and the right (p=0.181) knees (**Table 2**). For completeness, we also examined our model performance across subjects defined by the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain score, another well-known standard used to evaluate pain severity. Interestingly, we found no statistical differences between the model performance on WOMAC=0 and WOMAC≥1 (p=0.158 and p=0.138 for the left and right knees, respectively) (**Table 2**).

DISCUSSION

We investigated the feasibility of a fused deep learning framework to predict bilateral knee pain directly from knee MRI scans. Since the origin of pain can vary between subjects, we selected 25 MRI slices centered on the middle knee slice from each knee of each subject. By combining information from these slices, our model was presumably able to synthesize needed structural information from multiple locations to predict pain. The ability of the fused deep neural network to predict the output was evident in the form of the performance metrics (accuracy, sensitivity and specificity) that were generated by averaging all model runs. Compared with a machine learning approach using posteroanterior and lateral knee x-rays to predict knee pain, in which the area under the ROC curve ranged from 0.64 to 0.67, the MRI-based prediction of pain generated using our model had higher accuracy in classifying persons with/without knee pain at a single visit [28].

While deep learning algorithms such as CNNs are now increasingly considered for image-based classification, most of the literature has focused on using datasets containing single 2D MRI slices or radiography images converted to 2D grayscale formats. An important reason for using such data is because developing a three-dimensional (3D) deep neural network would require higher amounts of GPU memory as more parameters need to be learned to fully train the deep neural network. To circumvent this problem, we extracted features from each 2D MRI slice and then combined them using fully-connected layers and then associated them with the output of interest, pain. This framework therefore served as the best compromise between using as much information as possible from the volumetric data available within an MRI scan and the ability of utilizing such datasets to train without causing memory-related errors. Note that model training was performed on a high-performance GPU (Tesla P100, Nvidia, Santa Clara, CA).

The series of pre-processing steps including manual quality check and image registration of each MRI slice turned out to be important to train an accurate deep learning model for predicting bilateral knee pain. Using a straightforward Euclidean transformation, we aligned all the slices with respect to a pre-defined template, and then isolated a region representing the knee joint from all the registered images. While such linear transformations allowed us to generate a model that

resulted in consistent performance across multiple runs, more sophisticated linear or even nonlinear image registration techniques may also be applied to improve the alignment of structures of interest between different subjects. Nonetheless, certain forms of nonlinear registration techniques may introduce unwanted distortion, which then may lead to invalid representations of important anatomical features.

While our deep learning model demonstrates promising results for predicting bilateral knee pain using MR images, there is room for improvement in model performance. Other neural network architectures can be explored such as deep autoencoders or 3D CNNs. An important reason for us to focus on bilateral knee pain as opposed to using single knee-based MRI scans to predict unilateral knee pain is because we wanted to utilize as much information as possible from all the 2D MRI slices per subject to generate a model with high accuracy. In the case of predicting bilateral knee pain, we had in total 50 slices per subject that were carefully selected to train the model. Despite that, we believe that significant work needs to be done to increase the model performance, especially its sensitivity. In the recent examination of x-rays and their prediction of knee pain, limiting the pain outcome to subjects who repeatedly reported knee pain increased the accuracy of x-ray prediction. Alternate definitions of pain or tenderness could facilitate development of models with higher performance. Other aspects to consider while developing deep learning models for knee pain and OA in general are to identify ways to locate the source of pain in the joints. Our fusion model framework used a series of sagittal MRI slices, and it has the capability to combine coronal and axial MRI slices as well.

In conclusion, this work demonstrates the use of a fused deep neural network that associated MR imaging data of the knees with bilateral knee pain. The fusion framework allowed us to combine data from multiple 2D MRI slices from both the knees to efficiently construct the deep learning model. Such a modeling strategy can be easily extended to predict other clinical outcomes of interest. Further validation of the deep learning model across different imaging datasets is necessary to validate this technique across the full spectrum of OA.

ACKNOWLEDGMENTS

This article was prepared using an Osteoarthritis Initiative (OAI) public-use data set, and its contents do not necessarily reflect the opinions or views of the OAI Study Investigators, the NIH, or the private funding partners of the OAI. The OAI is a public–private partnership between the NIH (contracts N01-AR-2-2258, N01-AR-22259, N01-AR-2-2260, N01-AR-2-2261, and N01-AR-2-2262) and private funding partners (Merck Research Laboratories, Novartis Pharmaceuticals, GlaxoSmithKline, and Pfizer, Inc.) and is conducted by the OAI Study Investigators. Private sector funding for the OAI is managed by the Foundation for the NIH. The authors of this article are not part of the OAI investigative team. The OAI was also funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant HHSN-268201000019C).

This work was supported in part by the National Center for Advancing Translational Sciences, National Institutes of Health, through BU-CTSI Grant (1UL1TR001430), a Scientist Development Grant (17SDG33670323) from the American Heart Association, and a Hariri Research Award from the Hariri Institute for Computing and Computational Science & Engineering at Boston University, and NIH grants to VBK, DTF, and TDC (5U01AG-018820 & 1R01AR070139).

CONFLICTS OF INTEREST

None

FIGURE CAPTIONS

Figure 1: Pipeline for slice selection. For each subject's knee joint, we manually examined all the MRI slices oriented in the sagittal view and selected the slice showing the posterior cruciate ligament (PCL) and indexed it as the center slice (red colored box). The remaining slices were indexed relative to the center slice for each knee. Two-dimensional MRI slices for three different subjects (A, B & C) are shown.

Figure 2: Slice-specific registration. For each 2D MRI slice of the knee for a subject, we performed linear registration to align the slice with respect to a template that was already selected after manual examination. Later, a region (red colored box) containing the center of the knee joint with dimensions 294x294 pixels was cropped for all registered slices and used for model training.

Figure 3: Sample cases not used for model training due to the presence of various artifacts.

Figure 4: Deep learning framework. (A) Modified ResNet model used for training a single 2D MRI slice. (B) Each output of the 2D ResNet model was fused together and trained as a binary classification problem. A total of 25 2D slices were used for each knee for model fusion.

Figure 5: Performance of the 2D ResNet models. (A) Mean accuracy on the test data for both the left and right knee are shown as a function of slice number. Similarly, (B) mean sensitivity and (C) mean specificity on the test data for the left and right knee are shown.

Figure 6: Fused deep learning model performance. (A) Box plots indicating the median, range, and interquartile ranges of the model accuracy, sensitivity and specificity are shown. (B) Mean receiver operating characteristic (ROC) curve showing the performance of the binary classifier. The ROC curve was averaged across 10 independent model runs. The blue shaded region around the mean ROC curve shows the standard error across model runs.

TABLE CAPTIONS

Table 1: Study population and characteristics.

Table 2: Performance of the fused deep learning model across different groups defined by KL-grade and WOMAC pain score.

REFERENCES

1. Cross, M., et al., *The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study.* Ann Rheum Dis, 2014. **73**(7): p. 1323-30.
2. Loeser, R.F., et al., *Osteoarthritis: a disease of the joint as an organ.* Arthritis Rheum, 2012. **64**(6): p. 1697-707.
3. Hunter, D.J., et al., *Structural correlates of pain in joints with osteoarthritis.* Osteoarthritis Cartilage, 2013. **21**(9): p. 1170-8.
4. Dimitroulas, T., et al., *Neuropathic pain in osteoarthritis: a review of pathophysiological mechanisms and implications for treatment.* Semin Arthritis Rheum, 2014. **44**(2): p. 145-54.
5. Finan, P.H., et al., *Discordance between pain and radiographic severity in knee osteoarthritis: findings from quantitative sensory testing of central sensitization.* Arthritis Rheum, 2013. **65**(2): p. 363-72.
6. Clauw, D.J. and J. Witter, *Pain and rheumatology: thinking outside the joint.* Arthritis Rheum, 2009. **60**(2): p. 321-4.
7. Felson, D.T., et al., *The association of bone marrow lesions with pain in knee osteoarthritis.* Ann Intern Med, 2001. **134**(7): p. 541-9.
8. Felson, D.T., et al., *Correlation of the development of knee pain with enlarging bone marrow lesions on magnetic resonance imaging.* Arthritis Rheum, 2007. **56**(9): p. 2986-92.
9. Hill, C.L., et al., *Synovitis detected on magnetic resonance imaging and its relation to pain and cartilage loss in knee osteoarthritis.* Ann Rheum Dis, 2007. **66**(12): p. 1599-603.
10. Hill, C.L., et al., *Knee effusions, popliteal cysts, and synovial thickening: association with knee pain in osteoarthritis.* J Rheumatol, 2001. **28**(6): p. 1330-7.
11. O'Neill, T.W. and D.T. Felson, *Mechanisms of Osteoarthritis (OA) Pain.* Curr Osteoporos Rep, 2018. **16**(5): p. 611-616.
12. Bedson, J. and P.R. Croft, *The discordance between clinical and radiographic knee osteoarthritis: a systematic search and summary of the literature.* BMC Musculoskelet Disord, 2008. **9**: p. 116.
13. Sowers, M.F., et al., *Magnetic resonance-detected subchondral bone marrow and cartilage defect characteristics associated with pain and X-ray-defined knee osteoarthritis.* Osteoarthritis Cartilage, 2003. **11**(6): p. 387-93.
14. Yusuf, E., et al., *Do knee abnormalities visualised on MRI explain knee pain in knee osteoarthritis? A systematic review.* Ann Rheum Dis, 2011. **70**(1): p. 60-7.
15. Wenham, C.Y. and P.G. Conaghan, *Imaging the painful osteoarthritic knee joint: what have we learned?* Nat Clin Pract Rheumatol, 2009. **5**(3): p. 149-58.
16. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* Nature, 2015. **521**(7553): p. 436-44.
17. Kolachalama, V.B., et al., *Association of Pathological Fibrosis With Renal Survival Using Deep Neural Networks.* Kidney Int Rep, 2018. **3**(2): p. 464-475.
18. Litjens, G., et al., *A survey on deep learning in medical image analysis.* Med Image Anal, 2017. **42**: p. 60-88.
19. Norman, B., V. Pedoia, and S. Majumdar, *Use of 2D U-Net Convolutional Neural Networks for Automated Cartilage and Meniscus Segmentation of Knee MR Imaging Data to Determine Relaxometry and Morphometry.* Radiology, 2018. **288**(1): p. 177-185.

20. Tiulpin, A., et al., *Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach.* Sci Rep, 2018. **8**(1): p. 1727.
21. Yong, C.W., et al., *Train Convolutional Neural Networks Without Well-Segmented Ground Truth Images for Cartilage Localization: Data from the Osteoarthritis Initiatives.* Advanced Science Letters, 2018. **24**(3): p. 1771-1774.
22. Peterfy, C.G., E. Schneider, and M. Nevitt, *The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee.* Osteoarthritis Cartilage, 2008. **16**(12): p. 1433-41.
23. Fawaz-Estrup, F., *The osteoarthritis initiative: an overview.* Med Health R I, 2004. **87**(6): p. 169-71.
24. Roemer, F.W., et al., *Semiquantitative assessment of focal cartilage damage at 3T MRI: a comparative study of dual echo at steady state (DESS) and intermediate-weighted (IW) fat suppressed fast spin echo sequences.* Eur J Radiol, 2011. **80**(2): p. e126-31.
25. Hayashi, D., et al., *Semiquantitative assessment of subchondral bone marrow edema-like lesions and subchondral cysts of the knee at 3T MRI: a comparison between intermediate-weighted fat-suppressed spin echo and Dual Echo Steady State sequences.* BMC Musculoskelet Disord, 2011. **12**: p. 198.
26. Srivastava, N., et al., *Dropout: A Simple Way to Prevent Neural Networks from Overfitting.* Journal of Machine Learning Research, 2014. **15**: p. 1929-1958.
27. Kellgren, J.H. and J.S. Lawrence, *Osteo-arthrosis and disk degeneration in an urban population.* Ann Rheum Dis, 1958. **17**(4): p. 388-97.
28. Minciullo, L., et al., *Comparing image analysis approaches versus expert readers: the relation of knee radiograph features to knee pain.* Ann Rheum Dis, 2018. **77**(11): p. 1606-1609.

Table 1

Description	Value
Total number of cases in the OAI 0.E.1 dataset	2059
Selected cases for deep learning	1383
Case with bilateral knee pain	528
Cases with no pain	855
MRI scan type	SAG_IW_TSE_RIGHT SAG_IW_TSE_LEFT
2D MRI slice size	448x444 pixels

Figure 1

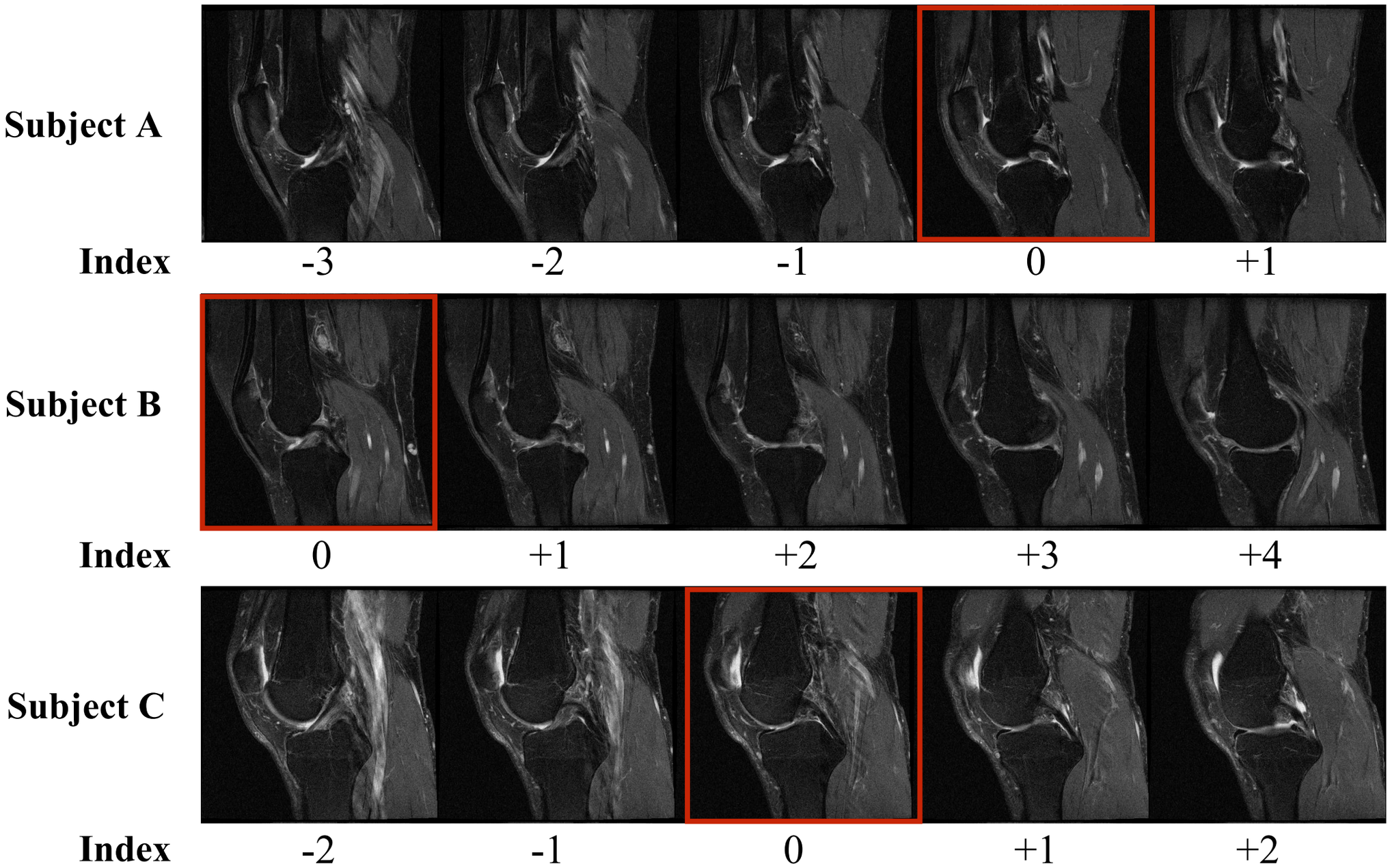
Subject A
Index
-3
-2
-1
0
+1
Subject B
Index
0
+1
+2
+3
+4
Subject C
Index
-2
-1
0
+1
+2

Figure 2

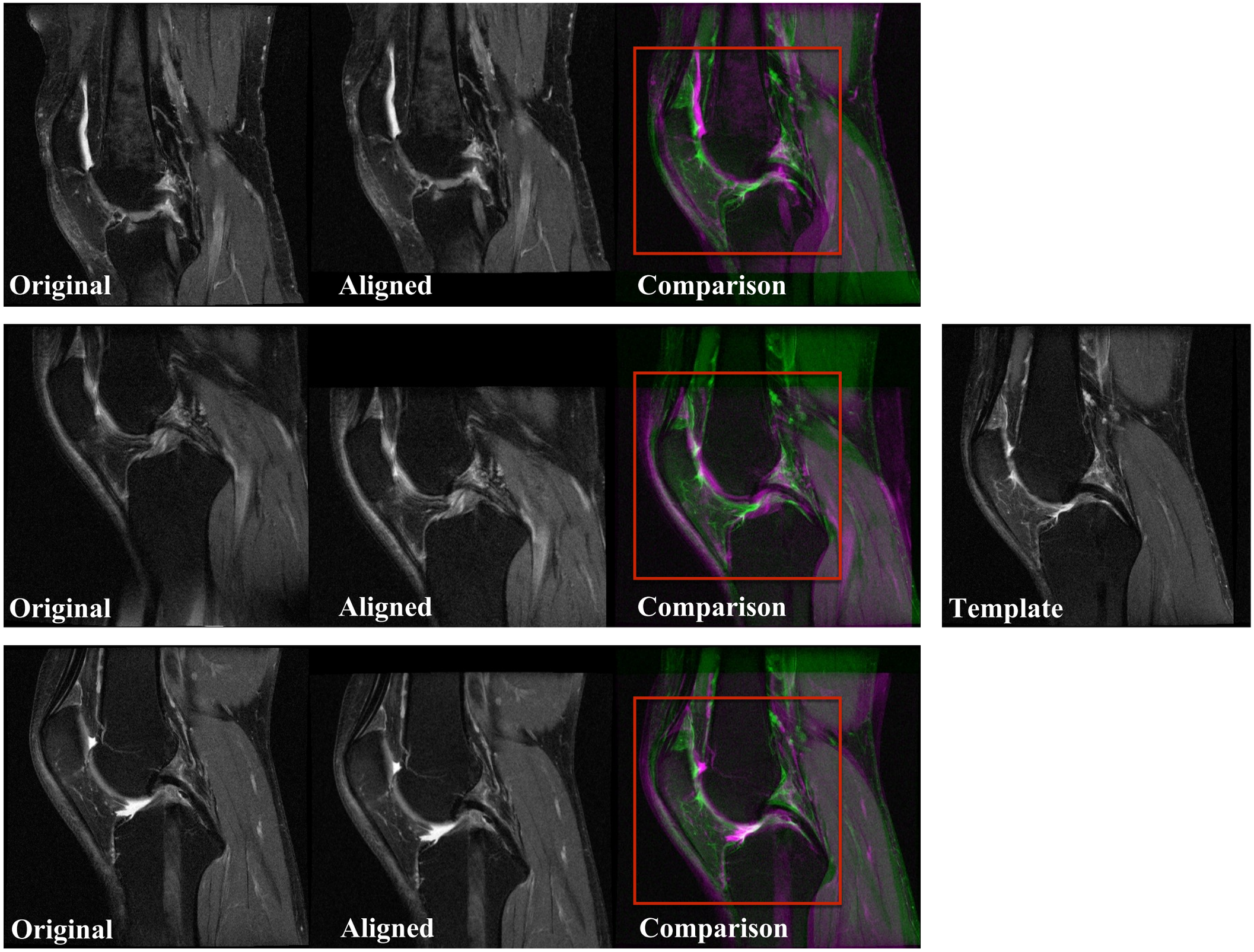
Subject A
Original
Aligned
Comparison
Subject B
Original
Aligned
Comparison
Template
Subject C
Original
Aligned
Comparison

Figure 3

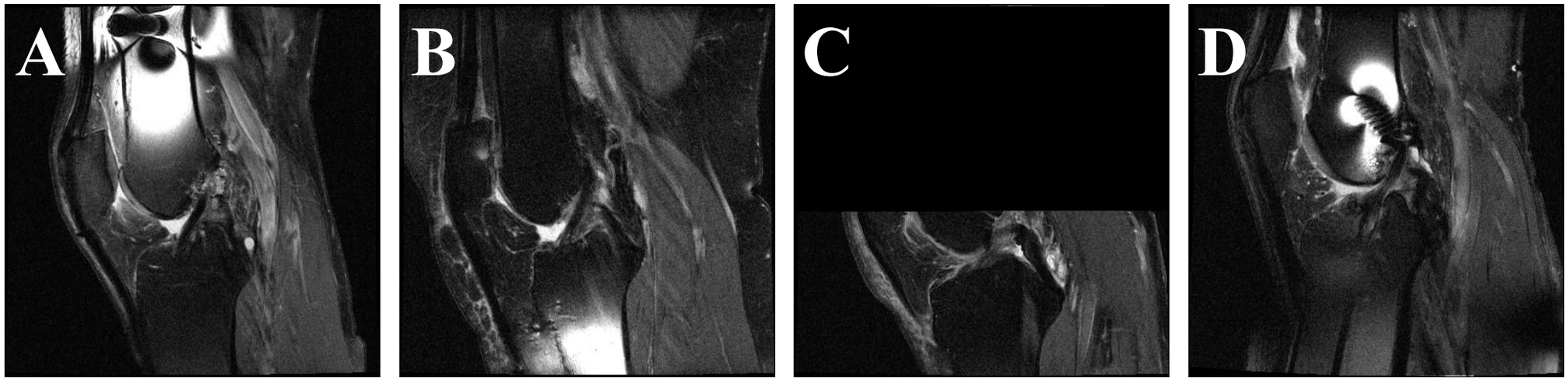

Figure 4

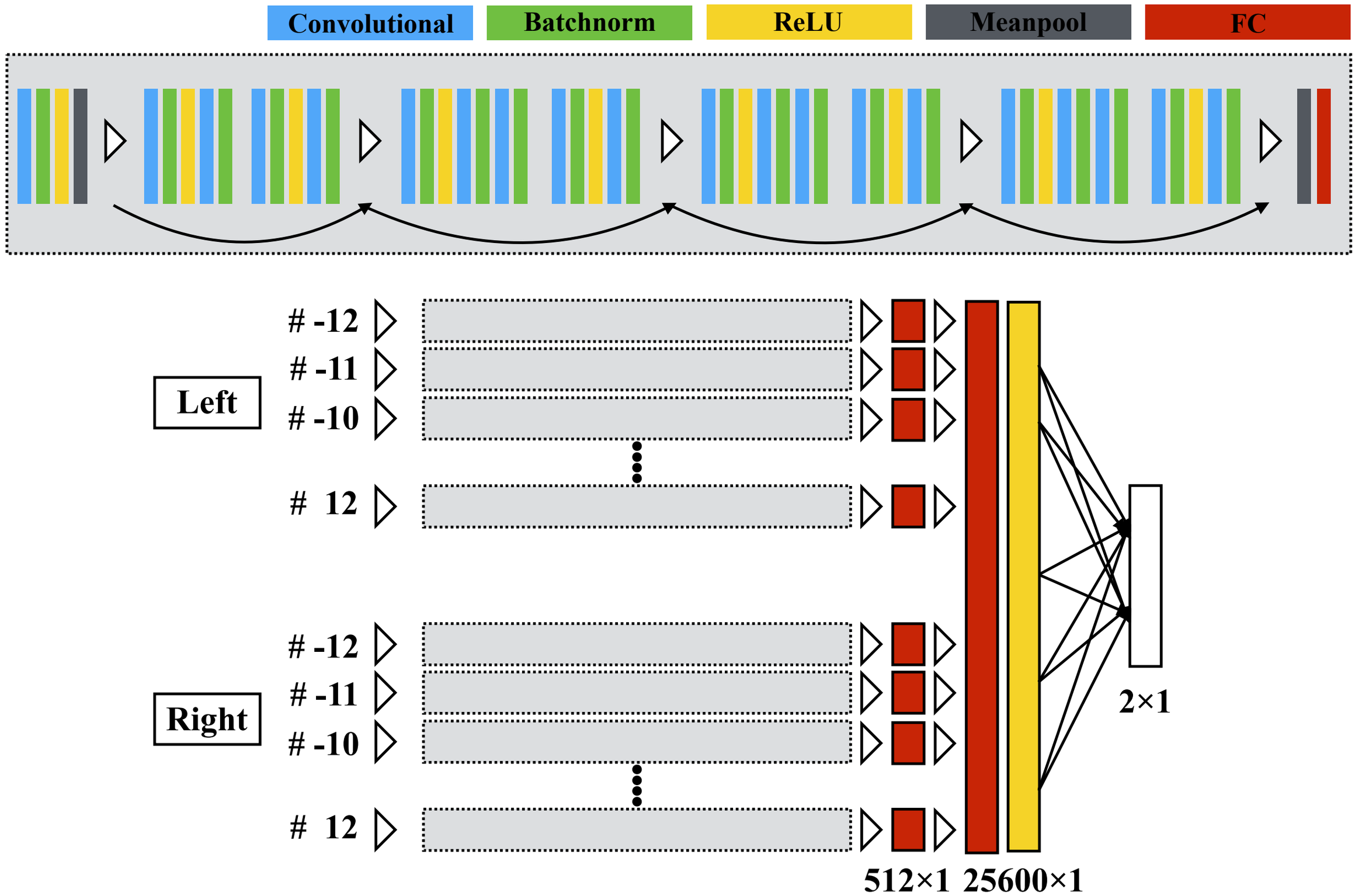
Convolutional
Batchnorm
ReLU
Meanpool
FC
Left
-12
-11
-10
12
Right
-12
-11
-10
12
512×1
25600×1
2×1

Figure 5

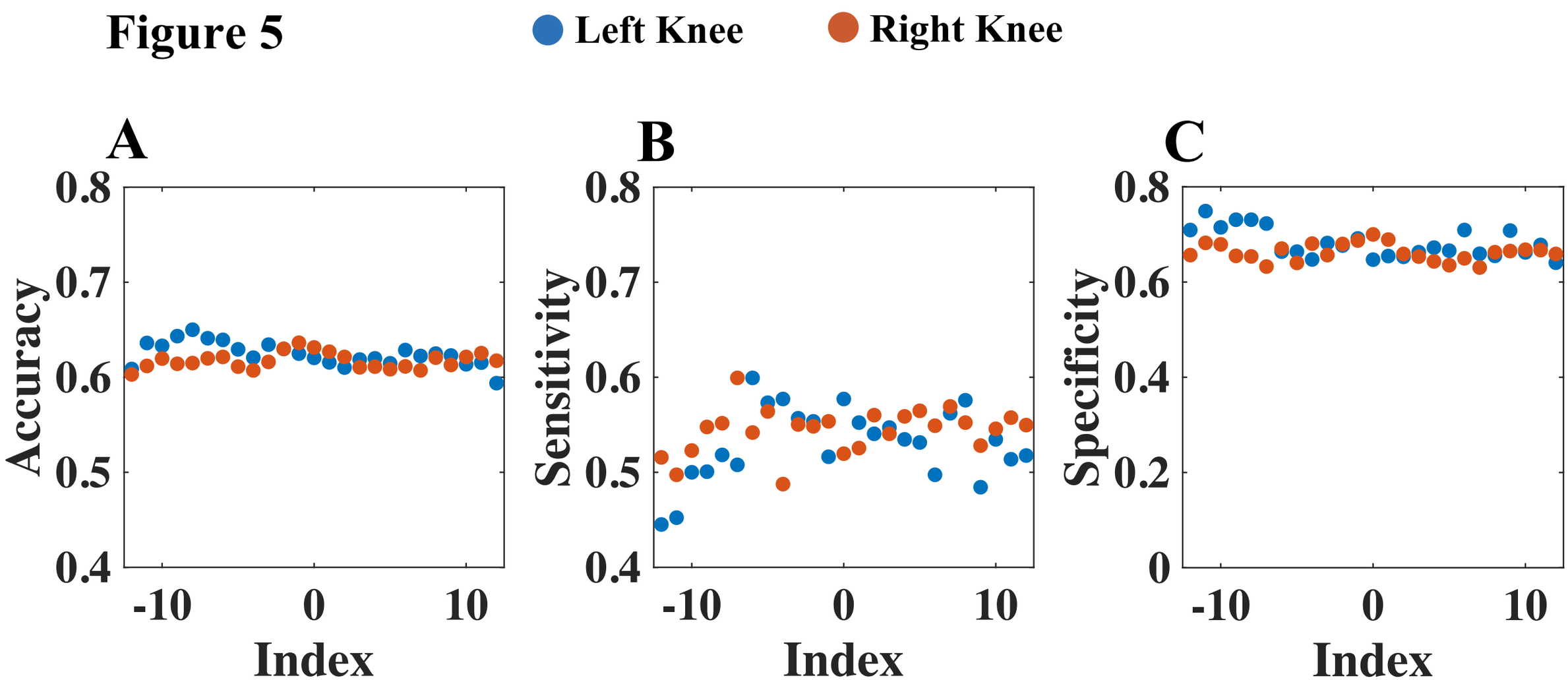
Left Knee
Right Knee
A
B
C
Accuracy
Sensitivity
Specificity
0.8
0.7
0.6
0.5
0.4
0.2
0
-10
0
10
Index

Figure 6

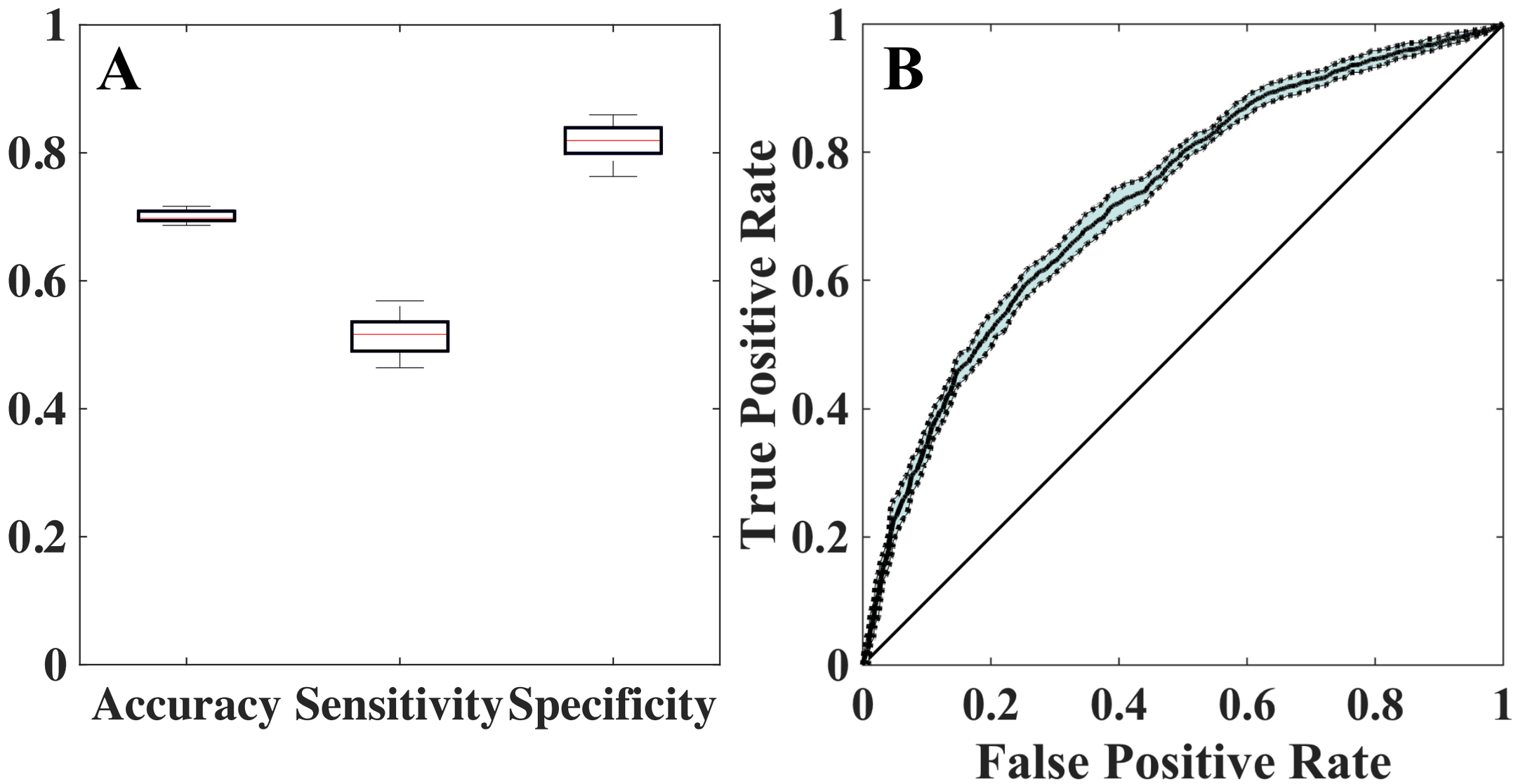
A
1
0.8
0.6
0.4
0.2
0
Accuracy
Sensitivity
Specificity
B
True Positive Rate
False Positive Rate
0
0.2
0.4
0.6
0.8
1

Table 2

KL (Left)	= 0	>= 1
Accuracy	0.715	0.695

KL (Right)	= 0	>= 1
Accuracy	0.716	0.695

WOMAC Pain (Left)	= 0	>= 1
Accuracy	0.728	0.680

WOMAC Pain (Right)	= 0	>= 1
Accuracy	0.730	0.683