Assessment of knee pain from MRI scans using a convolutional Siamese network

Gary H. Chang[1], David T. Felson[2, 3], Shangran Qiu[1], Ali Guermazi[4], Terence D. Capellini[5, 6], Vijaya B. Kolachalama[1, 7, 8, *]

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

[2]Section of Rheumatology, Department of Medicine, Boston University School of Medicine, Boston, MA, USA – 02118

[3]Centre for Epidemiology, University of Manchester and the NIHR Manchester BRC, Manchester University, NHS Trust, Manchester, UK

[4]Department of Radiology, Boston University School of Medicine, Boston, MA, USA – 02118

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

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

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

[8]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

Objectives: It remains difficult to characterize pain in knee joints with risk of osteoarthritis solely by radiographic findings. We sought to understand if advanced machine learning methods such as deep neural networks can be used to predict and identify the structural features that are associated with knee pain.

Methods: We constructed a convolutional Siamese network to associate MRI scans obtained on subjects from the Osteoarthritis Initiative (OAI) with frequent unilateral knee pain (n=1,529) comparing their knee with frequent pain to the contralateral knee without pain. The Siamese network architecture enabled pairwise learning of information from two-dimensional (2D) sagittal intermediate-weighted turbo spin echo slices obtained from similar locations on both knees. Class activation mapping (CAM) was utilized to create saliency maps, which highlighted the regions that were most associated with knee pain. The MRI scans and the CAMs of each subject were reviewed by a radiologist to identify the presence of abnormalities within the model-predicted regions of high association.

Results: Using 10-fold cross validation, our model achieved an area under curve (AUC) value of 0.808. When individuals whose knee WOMAC pain scores were not discordant were excluded, model performance increased to 0.853. The radiologist review revealed that about 86% of the cases that were predicted correctly had effusion-synovitis within the regions that were most associated with pain.

Conclusions: This study demonstrates a proof of principle that deep learning can be applied to assess knee pain from MRI scans.

Keywords: Deep learning, Osteoarthritis, Knee pain, MRI, Siamese network

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 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 using MRI scans, 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 individual can greatly benefit the design of targeted, individualized treatments to reduce symptoms and to limit disability [11].

MRI scans are capable of providing more detailed structural information about the knee joint than radiographs. A systematic review of MRI measures found 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 to develop a method to objectively and accurately associate MRI scans with knee pain.

Machine learning (ML) is a discipline within computer science that uses computational algorithms for the analysis of various forms of data. ML algorithms applied to medical images have shown remarkable success in predicting various outcomes of interest [16, 17]. 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. Specifically, deep learning frameworks such as convolutional neural networks (CNN) are increasingly being leveraged for object recognition tasks and specifically for disease classification [18, 19]. 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 them simultaneously for various applications related to classification, segmentation, and detection [16]. 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 [20], on X-ray imaging for automatic Kellgren-Lawrence (KL) grade classification [21], and for localization of cartilage [22].

The purpose of this study was first, to investigate the performance of a deep learning framework to predict knee pain and second to identify the structural regions that are most relevant to knee pain using MRI scans of both knees of individuals enrolled in the Osteoarthritis Initiative (OAI). Unilateral frequent knee pain was defined when an individual had pain, aching or stiffness for more than half of the days in a month on one knee and no pain in the contralateral knee. We used sagittal intermediate-weighted turbo spin echo (SAG-IW-TSE) sequence images that capture structural regions thought to be critical in generating knee pain, BMLs, synovitis, effusion and cartilage loss [7, 8, 23-27], to train a convolutional Siamese network. We subsequently leveraged class activation mapping (CAM) to identify the structural regions that were most relevant to frequent knee pain. An expert radiologist then independently reviewed the MRI scans and identified possible presence of knee abnormalities, which were then compared with CAM-based findings.

METHODS

Study selection

Our samples were drawn from the OAI, an NIH funded study of persons with or at risk of knee OA [28, 29]. baseline data from the OAI database was used for training and testing our deep learning model (**Table 1**). The dataset consists of MRI scans and clinical data from 4,759 subjects, of whom 1,529 subjects had pain, aching, or stiffness for more than half the days of a recent month in one knee and who reported they did not have pain on more than half the days in the contralateral knee. Out of them, 784 subjects had frequent knee pain in the left knee (label 0), and 745 subjects had frequent knee pain in the right knee (label 1). From these cases, 1,505 subject-specific MRI scans passed initial quality check (see below), and were used for construction of the deep learning model (Model A). Additionally, we excluded subjects with similar Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain scores between the left and right knee to construct another model. For this case, we selected a subset of subjects who had a WOMAC pain score difference greater than 2 between the knees. In total, 721 subjects met the criteria and 710 subjects were used for model construction after quality check (Model B). For the selected subjects, sagittal intermediate-weighted turbo spin echo (SAG-IW-TSE) sequence images were extracted as the inputs and unilateral knee pain as the output to solve a binary classification problem using the deep learning framework. The SAG-IW-TSE sequence of OAI images provided great details in the knee including bone marrow abnormality, synovial effusion, osteophytes, and cysts [28]. 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 [28].

No patient and public involvement

This research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient relevant outcomes or interpret the results.

Image registration and quality check

The majority of the SAG-IW-TSE sequence scans comprised 37 slices per knee. 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 1A**). 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 (**Figure 1B**). A region with dimensions of 294×294 pixels (105×105 mm) was subsequently selected, which contained the femoral, tibial, and patellar components of bone, cartilage and meniscus and all the registered slices were cropped to this region of interest. All the cropped slices were further resized into images of 224×224 pixels, and 11 adjacent slices on the lateral side of the center slice, 11 adjacent slices on the medial side of the center slice, and the center slice were selected for deep learning. In total, 46 MRI slices were selected for each subject (23 from each knee). After image registration, we performed a quality check 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. All these cases were not considered for model training or testing (**Figure 1C**).

Neural network architecture

We developed a deep learning architecture inspired by the Siamese network architecture [30], to associate the MRI scans from the left and right knee of the same subject with knee pain. Siamese networks were originally proposed to estimate similarities between a pair of images, and it was accomplished by feeding each of the two input images to two identical neural networks that shared the same parameters. The outputs from the two branches were subsequently joined by a contrastive loss function. In the present study, our model architecture was designed such that a pair of MRI slices from the two knees, each extracted from a relatively similar location within each knee were learned together **(Figure 2A)**. The deep learning architecture consisted of convolutional, batch normalization, nonlinear activation and two max pooling layers.

The present network was designed such that the final convolutional layer of the model had high spatial resolution, which dictates the ability to localize pain-related structural regions from the saliency maps. In the present model, only the first convolutional layer and the two max pooling layers had a stride of 2. Consequently, the final convolutional layer had a high in-plane resolution with a dimension of 28×28×512 **(Figure 2B)**.

During the training process, the two branches of the network learned the same set of shared parameters by presumably focusing on the similarities as well as the unique aspects of each knee associated with pain. This also allowed us to reduce the number of learnable parameters, considering both the left and right knee were evaluated by the same network. The training was performed using stochastic gradient descent optimizer and cross-entropy loss function for binary classification, and a batch size of 64. The neural network model, optimizer and training algorithms were implemented using an open-source deep learning library (PyTorch).

Class activation mapping and radiologist review

We used the technique of class activation mapping (CAM) to generate saliency maps that highlighted regions that were most associated with pain, at least as determined by the neural network. CAMs have the ability to localize the discriminative image regions from CNN models trained for classification without any prior locational knowledge [31]. This was achieved by first extracting the final feature maps of the left and right knee of the subject ($f_{L,n}$ and $f_{R,n}$ in **Figure 2A**, respectively). These feature maps were generated by the final convolutional layer of the network and had a dimension of 512×28×28. The maps were subsequently multiplied by weights of their respective fully-connected layer (FC_n in **Figure 2A**), which indicated the importance of each feature stored in the extracted maps. This resulted in CAMs with a dimension of 28×28 pixels for each MRI slice. For each subject, we identified a CAM with the most pain-relevant regions by selecting the CAM with highest average value from the 23 CAMs generated from all the MRI slices. This process resulted in the generation of saliency maps that highlighted the structural regions that were most associated with unilateral knee pain.

The dataset of 710 subjects (unilateral knee pain, between knee WOMAC pain difference ≥ 3) was divided into training, validation and testing sets in a 70:15:15 ratio. The sampling was

stratified based on risk factors of OA, including gender, age and body mass index (BMI). A musculoskeletal radiologist with extensive experience in knee MRI interpretation reviewed CAMs of MRI of the last 15%, consisting of 107 subjects. The radiologist reviewed each case and identified the presence of abnormalities using the MRI scan. The radiologist then reviewed the model-derived CAMs and identified the specific lesion that was co-localized with the highlighted region in CAM. Additionally, we compared the saliency maps generated by our model with the ones from a popular deep neural network architecture known as VGGNet [32]. We used the 16-layer VGGNet, and removed the final convolutional layer as in Zhou et. [31], and fine-tuned the network using the same images used for training, validation and testing the Siamese network.

To evaluate the statistical performance of the model, we first performed 10-fold cross-validation on the model using the selected dataset of 1,505 subjects (Model A, unilateral knee pain) and 710 subjects (Model B, unilateral knee pain, WOMAC pain difference ≥ 3). In each fold, the selected dataset was divided into the training and testing sets in a 9:1 ratio. In this scenario, every subject in the entire selected dataset appeared in the testing set exactly once. We subsequently computed the 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 the 95% confidence intervals. Unpaired Student's t-test was used to compare the mean value of two different groups, and Fisher's exact test was used to examine the non-random association between two groups of categorical variables. A p-value < 0.01 was considered statistically significant.

RESULTS

Among the 1,505 subjects from the baseline OAI (n=4,759) (Model A in Tables 1a & 1b), the mean age was 60.7±9.1 years and the mean BMI was 28.7±4.7. About 56.9% of the selected subjects were women. When cases with WOMAC score difference <3 were excluded (Model B in Tables 1a & 1c), the sub-group demographic characteristics remained relatively similar with respect to the overall group (Mean age: 60.9±9.2; Mean BMI: 29.2±4.8 and Percentage women: 60.1). For the cases that were reviewed by the radiologist (Table 1d), stratified sampling based on age, BMI, and gender helped us to generate a sub-group with demographic characteristics that were similar to the ones considered for models A & B.

Our methodology of generating CAMs provided a means by which to identify and examine the regions that were highly associated with knee pain. These CAMs are generated by extracting the features learned from the final convolutional layer in the neural network. Therefore, the ability of the CAMs to precisely identify a region of interest is directly dependent on the spatial resolution of this convolutional layer. Previously, researchers who proposed CAMs have used model architectures that had an in-plane resolution of 14×14 pixels (**Figure 3A**). In the present study, we generated CAMs with higher resolution containing 28×28 pixels, which resulted in qualitatively improved identification of the regions of association with pain (**Figure 3B**).

After the radiologist review, the location and the type of lesions that were co-localized with the highlighted region in the CAMs of subjects were identified (**Figures 3C-H**). The identified lesions included effusion, synovitis, BML, Hoffa fat pad lesion, cartilage loss, and meniscal damage. The effusion on the selected intermediate-weighted MRI scans included effusion and synovitis, and therefore, we combined effusion/synovitis into a single category, as used in MRI Osteoarthritis Knee Score (MOAKS). Out of the 107 cases reviewed by the radiologist (**Figure 3I**), effusion/synovitis was found to be the most relevant structural abnormality related to frequent knee pain in 95 (88.8%) subjects. BML was found to be the most relevant abnormality for 5 (5.6%) subjects. Hoffa fat pad abnormalities were found for 4 (3.7%) subjects, cartilage loss was found for 2 (1.9%) subjects, and meniscal damage was found for 1 subject.

The overall performance of the model was evaluated using 10-fold cross validation. For the model trained with 1,505 subjects with unilateral knee pain (Model A), we observed an AUC of 0.808 on the test data (**Figure 4A**). For the case with 710 subjects with unilateral knee pain and with a difference in WOMAC pain score larger than 2, the model achieved an AUC of 0.853 on the test data (**Figure 4B**). This observation indicates that pain-associated information can be present at several locations within the knee.

Sub-group analysis further revealed that model performance varied as a function of KL grade, BMI, age and gender (**Figure 4C**). Regardless of the difference in WOMAC pain scores between the knees (Model A vs Model B), model performance on the test data was better on individuals who had radiographic osteoarthritis (KL grade ≥ 1) in the knee. This could also imply that the model is able to better extract needed information on individuals who are likely to have detectable changes in morphology, at least as determined by MRI. The models also had higher values of AUC for subjects who were older, male and had higher BMI.

DISCUSSION

We developed a deep learning-based approach to distinguish MRI features of painful knees from nonpainful contralateral ones. We were ultimately able to do so with a high level of accuracy producing an area under the ROC curve of 0.853. Further, a radiologist reviewing the areas that were identified as painful knees suggested that these were primarily regions of synovitis or effusion, which are known sources of knee pain in OA.

Several studies have examined the association between radiographic features in the knee and knee pain within individual subjects and across multiple cohorts [14, 33-35]. While some studies relied on x-ray imaging, other relied on more sophisticated modalities such as MRI scans [36]. For some cases, associations between imaging features and unilateral pain were observed by comparing the knee with pain in the individual with the contralateral knee without pain in the same individual [37]. This strategy of using both knees for the same individual to study unilateral pain effectively allowed one knee to serve as a control to the other knee with pain. Moreover, this dual-knee paradigm to assess unilateral knee pain becomes attractive because the effect of common confounding factors such as age, gender, and BMI, on the outcome of interest no longer applies in such situations. Also, unlike the manual extraction of image-based or radiographic features which were then associated with knee pain, we investigated the feasibility of using deep learning to correlate structural regions from MRI scans of both knees with unilateral frequent knee pain. By combining information from a series of 2D slices, our model synthesized needed information from multiple locations to predict knee pain.

The Siamese neural network provided us with a viable strategy to simultaneously learn needed information from both the knees. This strategy resulted in a model that achieved high AUC (0.808), as evaluated using 10-fold cross validation. The performance of the model was improved further by excluding subjects with similar WOMAC pain score between the two knees (AUC=0.853). This improvement in the AUC value (5.6%) suggests that the Siamese network was able to identify those image features that were associated with a strong pain signature arising from one knee. This also implies that a set of image features that are common between the knees was also learned, but were not considered by the model to play a role in predicting unilateral

pain. When compared to a machine learning approach using postero-anterior and lateral knee x-rays to predict knee pain, our model generated a significantly higher AUC in predicting unilateral knee pain [38].

While deep learning algorithms such as CNNs are now increasingly considered for image-based classification, most of the previous work 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. 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.

The series of pre-processing steps, including manual quality check and image registration of each MRI slice, helped our model focus on a region representing the knee joint. With the network architecture designed to preserve high in-plane resolution in the CAM, we were able to extract slice-specific CAMs from each subject. Using a straightforward Euclidean transform, 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, it is important to note that certain forms of nonlinear registration techniques may introduce unwanted distortion, which then may lead to invalid representations of important anatomical features [39].

For the present study, we constructed *de novo* neural network models, instead of using pre-trained models. While pre-trained deep neural networks have the advantage of possessing a rich set of features learned from various image categories, these categories are different from the ones present in MRI scans. Moreover, the learned features in pre-trained models do not necessarily

resemble the features that one could extract from MRI scans [40]. As such, constructing neural network models from scratch also allowed us to limit the number of parameters while achieving a high level of performance. Importantly, the present network was designed to have high spatial resolution in the CAMs to allow us to associate the results of CAMs with specific anatomical regions, and consequently to identify lesions that were highly correlated with knee pain.

While our deep learning model demonstrates promising results for predicting unilateral knee pain using MRI scans, there is room for improvement in model performance. Other neural network architectures can be explored such as deep autoencoders or 3D CNNs [41]. In the case of predicting knee pain, we had in total 23 slices per subject carefully selected to train the model. Despite that, we believe that work needs to be done to increase the model performance. In a 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. 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 convolutional Siamese network that associated MR imaging data with unilateral knee pain. This 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. Our results provide a means by which to understand and evaluate early imaging markers of OA and other joint disorders. 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 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 and supported by the NIHR Manchester Biomedical Research Centre).

This article was prepared using the 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).

CONFLICTS OF INTEREST

Ali Guermazi is shareholder of BICL and consultant to Pfizer, AstraZeneca, TissueGene, Roche, Galapagos and MerckSerono.

FIGURE CAPTIONS

Figure 1: Image pre-processing pipeline. **(A)** 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 are shown. **(B)** 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 294×294 pixels was cropped for all registered slices and used for model training. Three cases from the baseline OAI dataset are shown. **(C)** Sample cases not used for model training due to the presence of various artifacts.

Figure 2: Architecture of the convolutional Siamese network. **(A)** After imaging pre-processing, the indexed MRI slices were fed into the corresponding Siamese network model. The outputs from all the 2D models were concatenated for the binary prediction task. **(B)** The 2D MR slices for the left and right knee were fed into the network simultaneously. A series of convolutional operations, batch normalization, nonlinear activation, max pooling and average pooling were applied to predict knee pain, by solving a binary classification problem.

Figure 3: Generated CAMs on selected subjects within the test data. **(A)** Examples of CAMs generated from the fine-tuned VGGNet model, resulting in CAMs with an in-plane resolution of 14×14 pixels. **(B)** Examples of CAMs generated from the present model from the same MRI images with an in-plane resolution of 28x28 pixels. Both the heat maps and the overlap of the MR image with the heat map are shown. In some cases of the test data **(C, D)**, effusion/synovitis was identified as the lesion present within the hot spots. Also, in few other cases, **(E)** bone marrow lesions, **(F)** Hoffa fat pad abnormality, **(G)** cartilage loss, and **(H)** meniscus damage, were identified as the lesions present within the hot spots. The red arrows indicate the locations of the identified structural regions. **(I)** Radiologist's assessment on the test cases (n=107). For each case, the MR scan and the model-derived heat map of the knee with confirmed pain were

reviewed by the radiologist, who then identified the presence of any lesions within the regions highlighted by the heat map.

Figure 4: Performance of the convolutional Siamese network model. (A) Receiver operating characteristic (ROC) curve of Model A trained with 1,505 subjects with unilateral frequent knee pain. (B) ROC curve of Model B trained with 710 subjects with unilateral frequent knee pain and contralateral knee difference in WOMAC pain ≥ 3. (C) Sub-group analysis. Performance of the models as a function of KL grade, BMI, age, and gender.

TABLE CAPTIONS

Table 1: Schematic of study selection. (Table 1a): In total, 1,529 subjects with unilateral frequent knee pain were selected from the OAI's baseline study, consisting of 4,759 subjects. After image quality check, 1,505 subjects were selected to train the neural network (Model A). Subjects who had the contralateral difference in WOMAC pain score (≥ 3) were further selected, leading to 710 subjects which were then used to train another neural network (Model B). Stratified sampling was then used to divide the dataset in the ratio of 70:15:15, where 70% of the dataset was used for training, 15% for validation, and the remaining cases (n=107) for independent testing. The stratified sampling was performed to match for age, gender, and body mass index. These cases were then reviewed by the expert radiologist. (Tables 1b-1d): Baseline characteristics of the sub-group of individuals selected for model training (Models A & B), and for radiologist review.

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 & Rheumatism, 2007. **56**(9): p. 2986-2992.
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 and Cartilage, 2003. **11**(6): p. 387-393.
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. Litjens, G., et al., *A survey on deep learning in medical image analysis.* Med Image Anal, 2017. **42**: p. 60-88.
17. Ker, J., et al., *Deep Learning Applications in Medical Image Analysis.* IEEE Access, 2018. **6**: p. 9375-9389.
18. 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.

19. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* Nature, 2015. **521**(7553): p. 436-44.
20. 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.
21. Tiulpin, A., et al., *Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach.* Sci Rep, 2018. **8**(1): p. 1727.
22. 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.
23. Englund, M., et al., *Effect of meniscal damage on the development of frequent knee pain, aching, or stiffness.* Arthritis & Rheumatism, 2007. **56**(12): p. 4048-4054.
24. Hill, C.L., et al., *Synovitis detected on magnetic resonance imaging and its relation to pain and cartilage loss in knee osteoarthritis.* Annals of the Rheumatic Diseases, 2007. **66**(12): p. 1599-1603.
25. Eckstein, F., et al., *Magnetic resonance imaging-based cartilage loss in painful contralateral knees with and without radiographic joint space narrowing: Data from the osteoarthritis initiative.* Arthritis & Rheumatism, 2009. **61**(9): p. 1218-1225.
26. Cibere, J., et al., *Natural history of cartilage damage and osteoarthritis progression on magnetic resonance imaging in a population-based cohort with knee pain.* YJOCA, 2011. **19**(6): p. 683-688.
27. Kim, H.A., et al., *The association between meniscal and cruciate ligament damage and knee pain in community residents.* YJOCA, 2011. **19**(12): p. 1422-1428.
28. 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.
29. Fawaz-Estrup, F., *The osteoarthritis initiative: an overview.* Med Health R I, 2004. **87**(6): p. 169-71.
30. Chopra, S., et al., *Learning a similarity metric discriminatively, with application to face verification.* cs.utoronto.ca.
31. Zhou, B., et al., *Learning Deep Features for Discriminative Localization*, in *IEEE Conference on Computer Vision and Pattern Recognition*. 2016. p. 2921-2929.
32. Simonyan, K. and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition.* 2014.
33. Bedson, J. and P.R. Croft, *The discordance between clinical and radiographic knee osteoarthritis: A systematic search and summary of the literature.* BMC Musculoskeletal Disorders, 2008. **9**(1): p. 467-11.
34. Wenham, C.Y. and P.G. Conaghan, *Imaging the painful osteoarthritic knee joint: what have we learned?* Nature Reviews Rheumatology, 2009. **5**(3): p. 149-158.
35. Sayre, E.C., et al., *Associations between MRI features versus knee pain severity and progression: Data from the Vancouver Longitudinal Study of Early Knee Osteoarthritis.* PLOS ONE, 2017. **12**(5): p. e0176833-12.
36. Neogi, T., et al., *Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies.* BMJ, 2009. **339**(aug21 1): p. b2844-b2844.

37. Javaid, M.K., et al., *Individual magnetic resonance imaging and radiographic features of knee osteoarthritis in subjects with unilateral knee pain: The Health, Aging, and Body Composition Study.* Arthritis & Rheumatism, 2012. **64**(10): p. 3246-3255.
38. 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.
39. Klein, A., et al., *Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration.* NeuroImage, 2009. **46**(3): p. 786-802.
40. Tajbakhsh, N., et al., *Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?* IEEE Transactions on Medical Imaging, 2016. **35**(5): p. 1299-1312.
41. Dolz, J., C. Desrosiers, and I. Ben Ayed, *3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study.* NeuroImage, 2018. **170**: p. 456-470.

Table 1

A

Description	Value
Number of subjects in OAI baseline dataset	4796
Cases with unilateral knee pain	1606
Unilateral knee pain cases selected after quality check (Model A)	1505
Unilateral knee pain cases with contralateral WOMAC pain score difference (>2) (Model B)	710
Cases selected for radiologist review	107
MRI scan type	SAG_IW_TSE_RIGHT SAG_IW_TSE_LEFT
2D MRI slice size	448x444 pixels

B

Model A (n = 1505)			
Parameter	**Left Knee Pain**	**Right Knee Pain**	**p-value**
Subjects	n = 734	n = 771	
Percent female	57.9	55.9	0.700
Age (years)	60.13 ± 9.18	61.24 ± 8.98	0.018
BMI	28.97 ± 4.88	28.38 ± 4.53	0.015

C

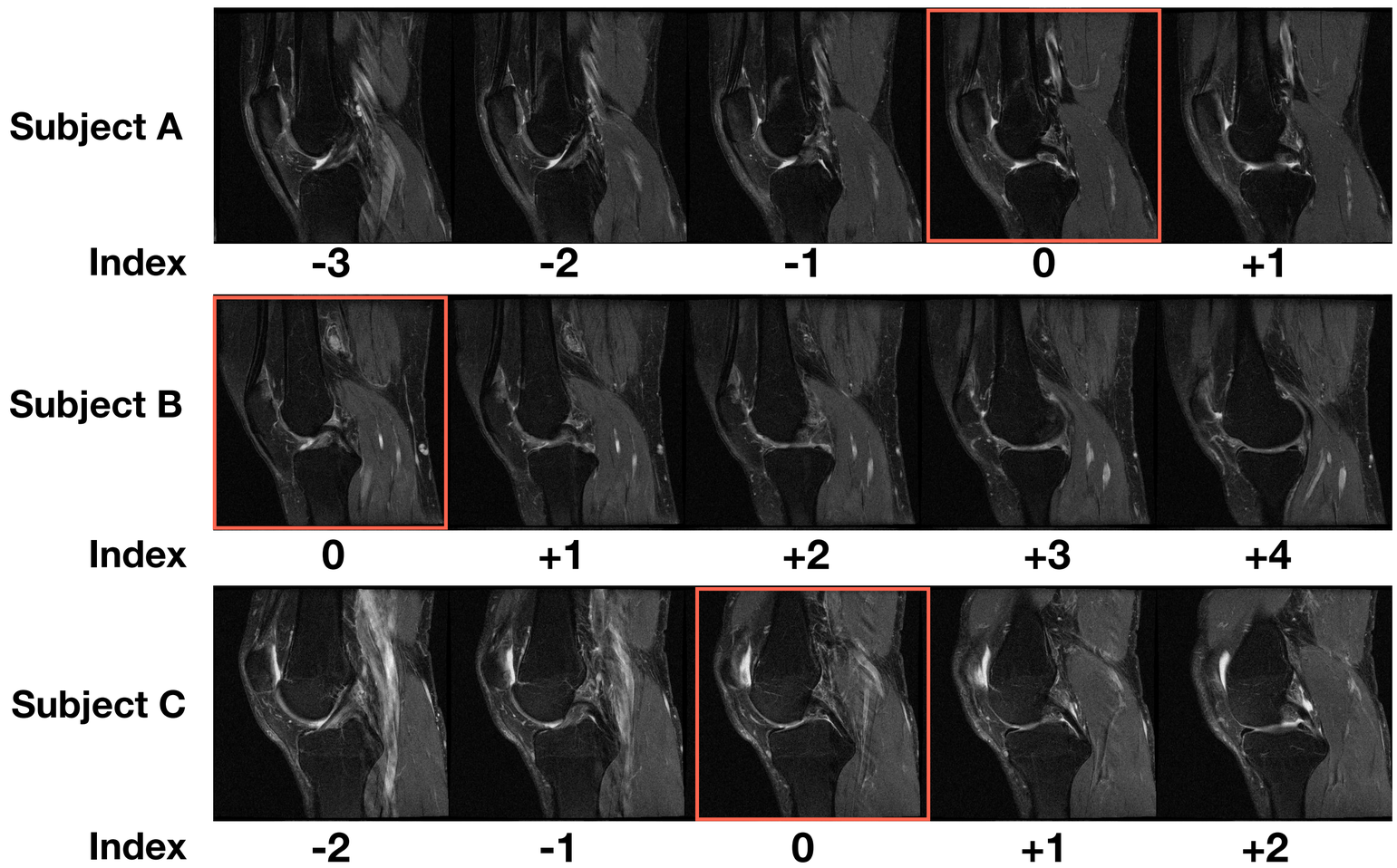
Model B (n = 710)			
Parameter	**Left Knee Pain**	**Right Knee Pain**	**p-value**
Subjects	n = 350	n = 360	
Percent female	59.4	60.8	0.854
Age (years)	60.62 ± 9.39	61.21 ± 9.01	0.395
BMI	29.50 ± 4.96	28.94 ± 4.63	0.120

D

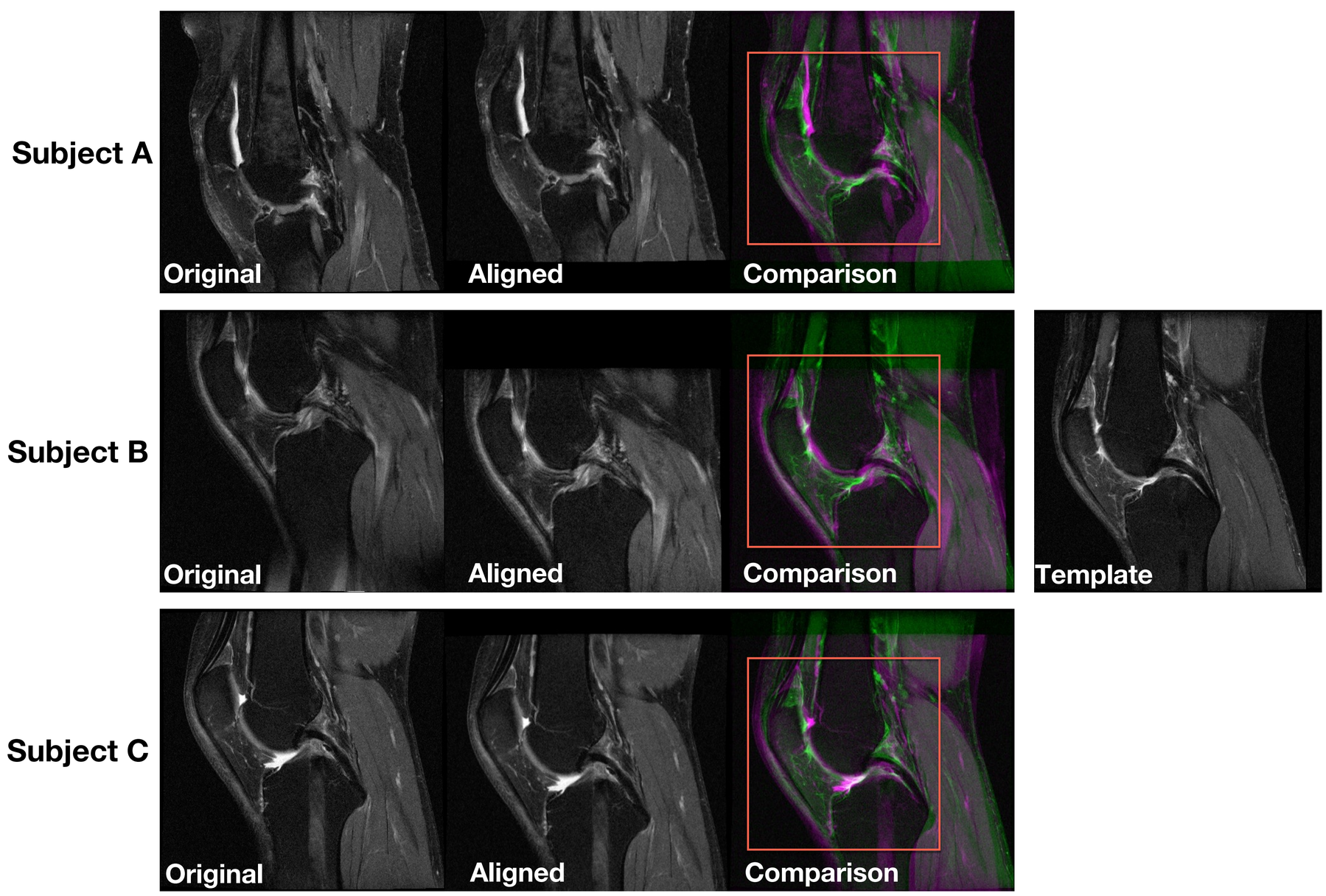
Cases for radiologist review (n = 107)			
Parameter	**Left Knee Pain**	**Right Knee Pain**	**p-value**
Subjects	n = 46	n = 61	
Percent female	60.9	59.0	1.000
Age (years)	60.83 ± 9.34	60.67 ± 8.90	0.932
BMI	29.38 ± 5.30	29.01 ± 4.89	0.711

Figure 1

A


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

B


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

C

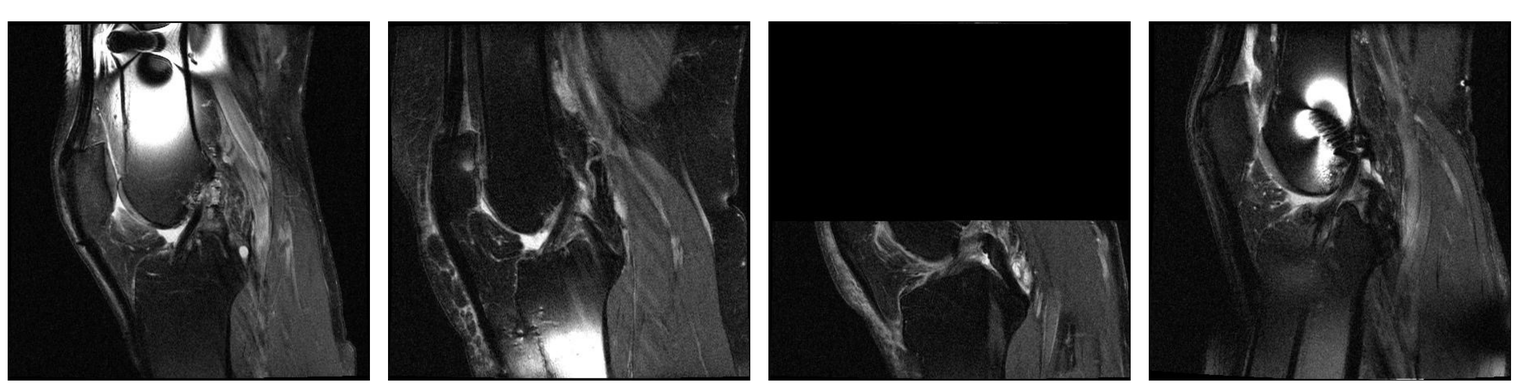
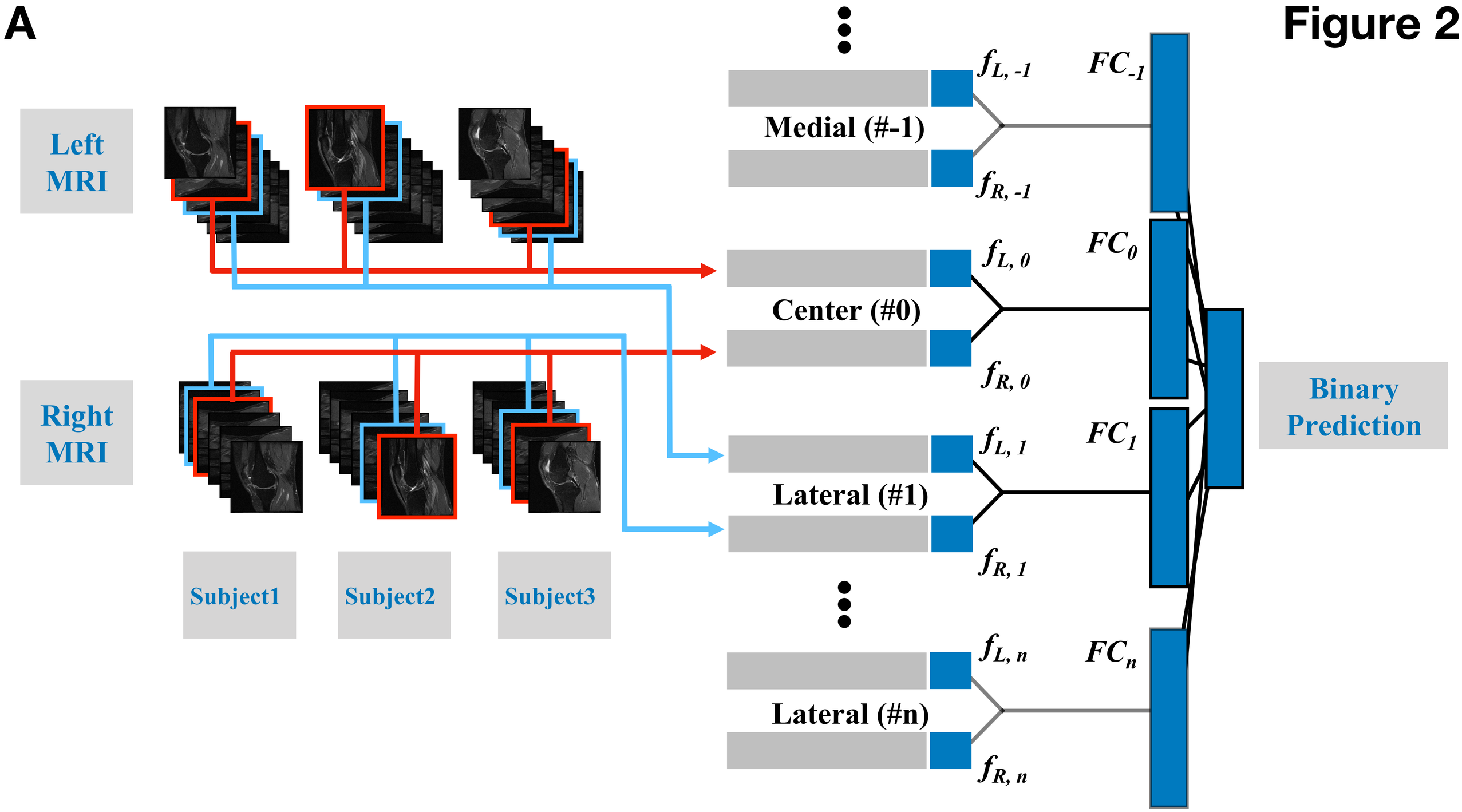


Figure 2

A


Left MRI
Right MRI
Subject1
Subject2
Subject3
Medial (#-1)
Center (#0)
Lateral (#1)
Lateral (#n)
$f_{L,-1}$
$f_{R,-1}$
$f_{L,0}$
$f_{R,0}$
$f_{L,1}$
$f_{R,1}$
$f_{L,n}$
$f_{R,n}$
FC_{-1}
FC_0
FC_1
FC_n
Binary Prediction

B

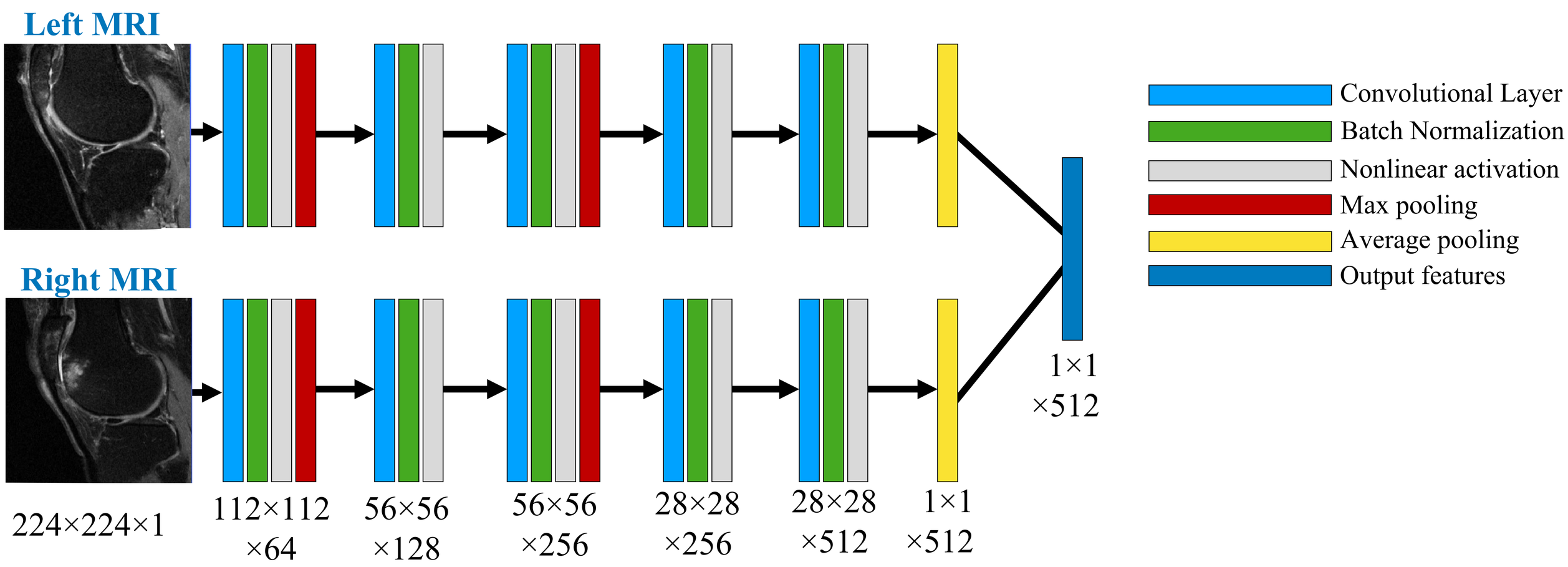
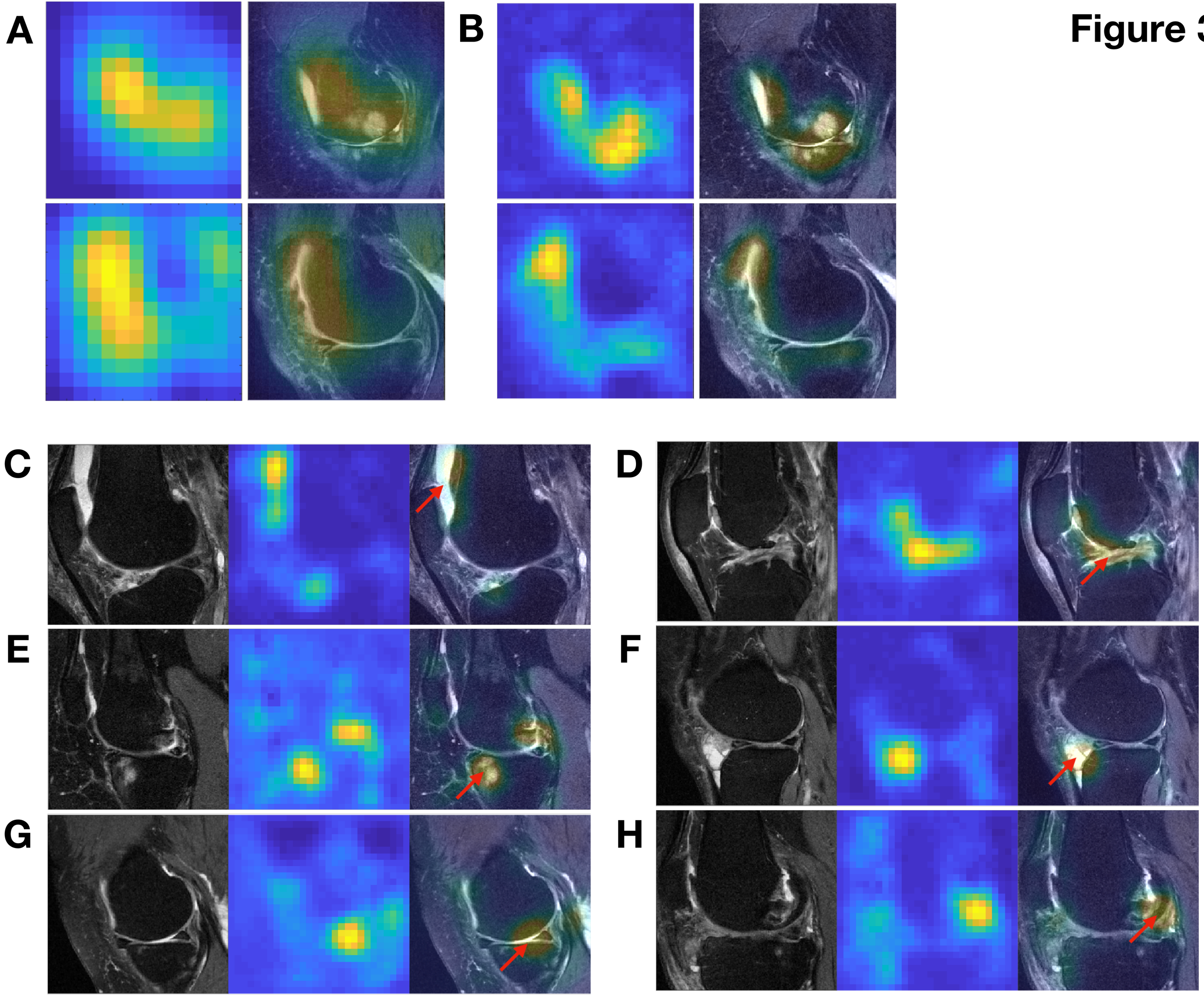

Left MRI
Right MRI
224×224×1
112×112 ×64
56×56 ×128
56×56 ×256
28×28 ×256
28×28 ×512
1×1 ×512
1×1 ×512
Convolutional Layer
Batch Normalization
Nonlinear activation
Max pooling
Average pooling
Output features

Figure 3

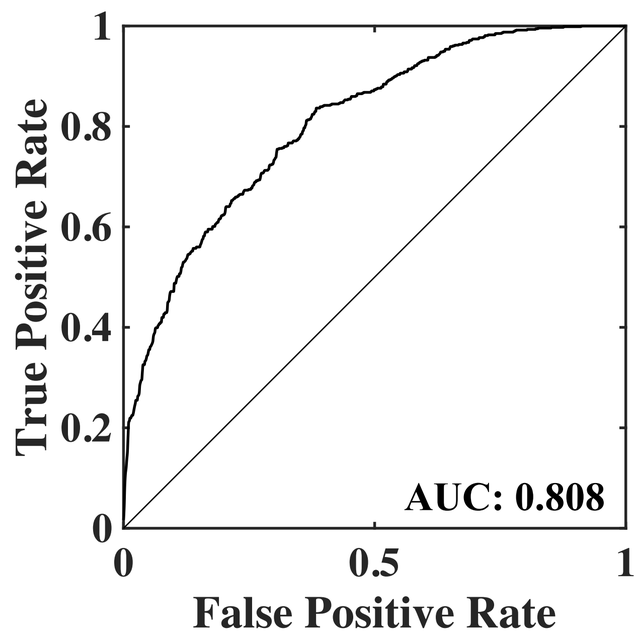


I

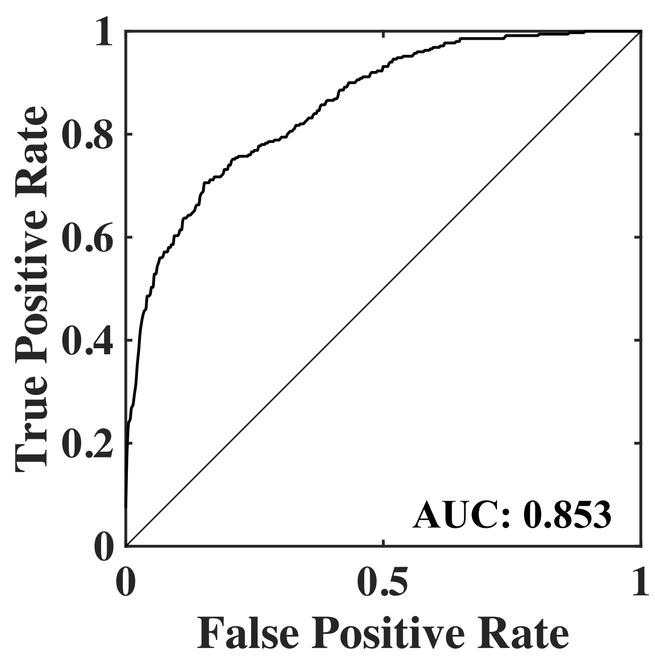
	Value	Pred. Correct	Pred. Wrong
Total	107	81	26
Effusion/Synovitis	95 (88.8%)	70	25
Bone Marrow Lesion	5 (4.7%)	5	0
Hoffa Fat Pad	4 (3.7%)	3	1
Cartilage Loss	2 (1.9%)	2	0
Meniscus Damage	1 (0.9%)	1	0

Figure 4

A



B



C

Model A	
Description	**AUC**
Overall	0.808
KL grade of the knee with frequent pain = 0	0.593
KL grade of the knee with frequent pain ≥ 1	0.866
BMI ≥ 30	0.853
BMI < 30	0.775
Age ≥ 60	0.819
Age < 60	0.791
Male	0.854
Female	0.766

Model B	
Description	**AUC**
Overall	0.853
KL grade of the knee with frequent pain = 0	0.582
KL grade of the knee with frequent pain ≥ 1	0.907
BMI ≥ 30	0.879
BMI < 30	0.832
Age ≥ 60	0.863
Age < 60	0.839
Male	0.891
Female	0.822