Automatic Knee Meniscus Tear Detection and Orientation Classification with Mask-RCNN

Vincent Couteaux, S. Si-Mohamed, O. Nempont, T. Lefevre, A. Popoff, Guillaume Pizaine, N. Villain, Isabelle Bloch, A. Cotten, L. Boussel

Automatic knee meniscus tear detection and orientation classification with Mask-RCNN

Vincent Couteaux[a,b*]

Salim Si-Mohamed[c,d]

Olivier Nempont[a]

Thierry Lefevre[a]

Alexandre Popoff[a]

Guillaume Pizaine[a]

Nicolas Villain[a]

Isabelle Bloch[b],

Anne Cotten[e]

Loïc Boussel[c,d]

[a] Philips Research France, 33 rue de Verdun, 92150 Suresnes, France.

[b] LTCI, Télécom ParisTech, Université Paris-Saclay, 46 rue Barrault, 75013 Paris, France.

[c] Claude Bernard-Lyon 1 University, CREATIS, CNRS UMR 5220, INSERM U1206, INSA-Lyon, 69100 Villeurbanne,

[d] Department of Radiology, Hospices Civils de Lyon, 69002 Lyon, France.

[e] Department of Musculoskeletal Radiology, CHRU de Lille, 59000 Lille, France.

*** Corresponding author** vincent.couteaux@telecom-paristech.fr

Philips Research France, 33 rue de Verdun, 92150 Suresnes, France

Abstract

Purpose: This work presents our contribution to a data challenge organized by the French Radiology Society during the *Journées Francophones de Radiologie* in October 2018. This challenge consisted in classifying MR images of the knee with respect to the presence of tears in the knee menisci, on meniscal tear location, and meniscal tear orientation.

Materials and methods: We trained a mask region-based convolutional neural network (R-CNN) to explicitly localize normal and torn menisci, made it more robust with ensemble aggregation, and cascaded it into a shallow ConvNet to classify the orientation of the tear.

Results: Our approach predicted accurately tears in the database provided for the challenge. This strategy yielded a weighted AUC score of 0.906 for all three tasks, ranking first in this challenge.

Conclusion: The extension of the database or the use of 3D data could contribute to further improve the performances especially for non-typical cases of extensively damaged menisci or multiple tears.

Keywords: Knee meniscus; Artificial intelligence; Mask region-based convolutional neural network (R-CNN); Meniscal tear detection; Orientation classification

Introduction

Meniscal lesions are a frequent and common cause of knee pain, responsible for approximately 700,000 arthroscopic partial meniscectomies per year in the United States [1]. They are defined as a tear within the meniscus, and can lead to articular cartilage degeneration over time, further necessitating surgical treatment. Magnetic resonance imaging (MRI) plays a central role in the diagnosis of meniscus lesions, the preoperative planning and the postoperative rehabilitation of the patient [2,3]. As meniscal lesions are very frequent, their diagnosis could certainly benefit from a quantitative and automated solution giving more accurate results in a faster way. Computer-aided detection systems for meniscal tears were thus proposed whereby regions of interest in the image are extracted and classified based on handcrafted image features [4-8].

The *Journées Francophones de Radiologie* was held in Paris in October 2018. For the first time, the French Society of Radiology organized an artificial intelligence (AI) competition involving teams of industrial researchers, students and radiologists.

This paper presents our contribution to the knee meniscus tear challenge, where participants had to classify sagittal MRI slices cropped around the knee depending on the presence of tears in anterior and posterior menisci and on their orientation (horizontal or vertical). We proposed a method that takes advantage of recent advances in deep learning [9, 10]. More precisely, we propose to localize, segment, and classify healthy and torn menisci using a mask region-based convolutional neural network (R-CNN) approach that is cascaded into a shallow ConvNet to classify tear orientation.

Method

Knee meniscus tear challenge

Sagittal MR images centered around the knee were provided with the following annotations: (*i*), position of the image (medial or lateral); (*ii*), presence of a tear in the posterior meniscus; (*iii*), presence of a tear in the anterior meniscus; (*iiii*), orientation of the tear in the posterior meniscus (if any); (*v*), and orientation of the tear in the anterior meniscus (if any).

Two training batches were provided; the first made of 257 images was shared one month before the conference and the other, made of 871 images, 2 days before the end of the challenge. The first batch contained 55/257 (21.4 %) images with horizontal posterior tears, 46/257 (17.9 %) with vertical posterior tears, 13/257 (5.1 %) with horizontal anterior tears and 8/257 (3.1 %) with vertical anterior tears. The second batch contained 107/871 (12.3 %) images with horizontal posterior tears, 60/871 (6.9 %) with vertical posterior tears, 8/871 (0.9 %) with horizontal anterior tears and 3/871 (0.3 %) with vertical anterior tears. The classes were imbalanced, with horizontal tears and posterior meniscus tears being more frequent, and a low number of anterior tears were available for training. We reviewed the database and removed any ambiguous annotations from the training set.

Images of size 256 × 256, either of the medial or of the lateral plane of the knee, were provided, as illustrated in Figure 1a,b. The femur was always on the left and the tibia on the right, with the anterior meniscus at the top and the posterior meniscus at the bottom of the image. Horizontal tears appeared vertical and vice versa. The grey level scale was in an

arbitrary unit scaled between 0 and 1, and the images did not have consistent brightness and contrast.

On MRI, meniscal tears appear as thin, hyperintense lines that cut across the menisci, but we can observe various hyper-intense signals in the menisci (Figure 1). For instance, hyper-intense lines that only partially cut across the menisci should not be classified as tears (Figure 2a). Moreover, the menisci may be barely visible (Figure 2b) or have multiple tears (Figure 2c). In the latter case, the provided orientation may be ill-defined.

Menisci are small structures and tears present as thin abnormalities within menisci on MRI. To facilitate tear detection, we first localized the menisci. However, the localization of the menisci was not part of the provided annotations. To efficiently perform this annotation, we chose to approximate menisci by triangles resulting in a coarse segmentation of menisci (Figure 3).

To detect tears in both menisci and identify their orientation, we opted for a cascaded approach. First, menisci were localized and tears were identified. Then the orientation of torn menisci was classified. For both tasks, we applied a morphological pre-processing, as described below, to enhance the relevant structures in the image.

Morphological pre-processing of images

In a first attempt to better understand the dataset and the classification task at hand, we trained simple neural network classifiers on the training dataset in order to classify the posterior (resp. anterior) meniscus into healthy and torn cases. Both networks were based on a simplified VGG-like architecture: the whole image is taken as the input, on which four 2D-convolution/ ReLu/Maxpool layers are applied followed by two dense layers and a final softmax-activated output layer [11]. Note that these neural networks were only meant to explore the given dataset.

An accuracy of 83% could readily be obtained (precision, 0.76; recall, 0.8). However, when analyzing the interpretation of the ConvNet classification, it appeared that the network used non-relevant features in the image to provide the result (Figure 4). This phenomenon was consistently observed on other images in the dataset and is probably attributed to the variability of the images. Given that the dataset is small, the network is not able to properly generalize on so few samples. This prompted us to consider pre-processing the images in

order to bring robustness to the classification. The approach we propose is close to that for detecting white matter hyper-intensities in brain MRI [12].

Since meniscus tears are primarily characterized by distinct morphological features, it seems reasonable to assume that any image pre-processing step that would retain these characteristics while limiting the influence of other structures may be beneficial. Figure 4c shows the result of applying a black top-hat morphological filter on the image of Figure 4a. A black top-hat filter outputs an image wherein the bright regions correspond to regions in the original image which are smaller than the structuring element and darker than their surroundings. As can be clearly observed, the torn posterior meniscus can still be identified. A ConvNet classifier trained on black top-hat filtered images shows performance similar to the initial one (accuracy, 83%; precision, 86%; recall, 67%) but is able to focus precisely on the meniscus region (Figure 4d). In this case, the saliency map clearly indicates that only the meniscus region is relevant for the classification.

Meniscus localization and tear detection

To localize both menisci and identify tears in each meniscus, we used the Mask R-CNN framework, a state-of-the-art approach for Instance Segmentation. It performs object detection, segmentation and classification in a single forward pass [13]. We trained the model to detect and segment four objects: (*i*) healthy anterior meniscus; (ii), torn anterior meniscus; (*iii*) healthy posterior meniscus, and (*iiii*), torn posterior meniscus. In this way, we obtained the localization of each meniscus, the classification of healthy vs. torn, and a classification score. We chose to perform the classification of tear orientation independently on the segmented meniscus region only, as explained below because the classes would have been too imbalanced otherwise (only 11 vertical tears in the anterior meniscus for instance).

We used a Mask R-CNN model pre-trained on the common object in context (COCO) dataset [14] whose input is a three channel image. We applied three white top-hat filters (the dual of the black top-hat filters described above) on original MRI slices with square structuring elements of size 5×5, 11×11 and 21×21 (Figure 5) to generate network inputs. Note that we did not constrain the model to return exactly one result for each meniscus because the two menisci were correctly detected in almost all cases. We illustrate in Figure 6 the output of the Mask R-CNN. In Figure 6a, the two healthy menisci are properly detected. In Figure 6b, the posterior meniscus is appropriately identified as torn. However, the posterior meniscus is too widely segmented and incorrectly labeled as torn (Figure 6).

Training

We fine-tuned a Mask R-CNN with a ResNet-101 backbone, pretrained on COCO dataset) [13-15]. The training was done using an Adam optimizer, 1.10 −3 learning rate and batches of 8 images, during 1000 epochs of 100 batches.

Ensemble aggregation

To improve the robustness of our model, we applied ensemble aggregation. We trained five models on random folds of the full training data set (1128 images) and retained five additional models trained on random folds of the first training batch only (the first 257 images). We aggregated the results differently for anterior and posterior menisci. We classified the anterior meniscus as torn when at least one network had detected a torn anterior meniscus, with a probability $P_{ant}(F)$ equal to the mean classification score of all detected torn anterior menisci by the ensemble. We classified the posterior meniscus as torn when the strict majority of the networks had detected a torn posterior meniscus. The probability $P_{post}(F)$ is equal to the mean classification score of all detected torn posterior menisci by the ensemble. We used different aggregation methods as a large majority of anterior menisci are healthy. Some networks may not have seen enough torn anterior menisci in order to recognize them.

Tear orientation classification

To classify the orientation of torn menisci as horizontal or vertical, we trained a neural network on images cropped to the bounding boxes of detected torn menisci, resized to 47 × 47 pixels. This network was fed with pre-processed patches, each input having five channels illustrated in Figure 7: 1) unprocessed patch, 2) local orientation map, computed with $\sigma = 3$ (see below), 3) local orientation map, computed with $\sigma = 1$, 4) black top-hat transform, with a disk structuring element of radius 4 pixels, and 5) black top-hat transform, with a disk structuring element of radius 8 pixels. The local orientation map represents the angle of the smallest eigenvector of the Hessian matrix at each pixel. The Hessian matrix was computed with the second derivative of a Gaussian kernel, whose standard deviation σ is a parameter.

Only 300 torn menisci were provided for training. Therefore, we trained a very shallow CNN based on a VGG-like architecture:

• Convolution, 3 × 3 kernel, 8 filters, ReLU activation,

• Max-pooling, 2 × 2,

• Convolution, 3 × 3 kernel, 16 filters, ReLU activation,

• Max-pooling, 2 × 2,

• Convolution, 3 × 3 kernel, 32 filters, ReLU activation,

• Max-pooling, 2 × 2,

• Dense Layer with 1024 units, ReLU activation, P = 0.5 dropout,

• Dense Layer with 1024 units, ReLU activation, P = 0.5 dropout,

• Dense Layer with 2 units and a softmax activation.

We trained this network on 246 torn menisci of the training database with a Stochastic Gradient Descent, 1.10 −3 learning rate and batches of 32 images, during 800 epochs (approximately 5 min). We validated the method on the remaining 54 cases and selected the model with the highest validation accuracy.

Results

Score and ranking

Teams were ranked according to a weighted average of the area under the ROC curves (AUC) of the tear detection task *Det* (tear in any meniscus), the tear localization task *Loc* (anterior or posterior) and the orientation classification task *Or* (horizontal or vertical), according to Equation 1 (E1):

E1 $$Score = 0.4 \times \mathrm{AUC}(Det) + 0.3 \times \mathrm{AUC}(Loc) + 0.3 \times \mathrm{AUC}(Or)$$

The organizers therefore removed from the database cases where both menisci had tears and the following values were submitted for each image:

• Probability of a tear in any meniscus *P(F)*,

• Probability that the tear (if any) is in the anterior meniscus *P(Ant)*,

• Probability that the tear (if any) is horizontal *P(H)*.

The Mask R-CNN ensemble outputs a probability $P_{ant}(F)$ that the anterior meniscus is torn, and a probability $P_{post}(F)$ that the posterior meniscus is torn, both being independent a priori. This results in Equation 2 (E2)

E2 $$P(F) = P_{post}(F) + P_{ant}(F) - P_{post}(F)P_{ant}(F)$$

where *P(Ant)* is defined by Equation 3 (E3)

E3 $P(Ant) = P_{ant}(F)/(P_{ant}(F)+P_{post}(F))$.

To obtain *P(H)*, we applied the orientation classifier on the anterior meniscus when *P(Ant)* > 0.5 and on the posterior meniscus otherwise.

A test set of 700 images was used for ranking. We obtained a score of 0.906 and shared the first place with another team (score 0.903).

Visual inspection

In most cases, the prediction was in line with our interpretation as illustrated in Figure 8, but a few cases seemed suspicious. The resulting classification scores were almost binary, either very close to 1 or very close to 0, especially *P(F)*. However, for some images, the predictor returned classification scores close to 0.5 (Figure 9).

Discussion

The knee meniscus tear challenge posed an image classification problem. Image classification tasks in computer vision aim to discriminate many classes from the prominent object in the image [16]. However, in this problem, the classification result should be based on thin details at a specific location in the image. Moreover, only a small database was available. Training a standard classifier from the image may therefore result in sub-optimal performances as observed in our initial experiments.

We chose to localize and segment the menisci and perform the classification within the anterior and posterior menisci. We opted for a Mask R-CNN approach as it can perform both tasks jointly [13]. Due to the class imbalance, we did not classify the tear orientation using this approach, but cascaded the Mask R-CNN into a shallow ConvNet to classify tear orientation [17]. Moreover, we provided pre-processed images to the networks to focus on relevant parts of the images by enhancing the tears.

In conclusion, our approach ranked first in the challenge by predicting accurately tears in the database provided for the challenge. The extension of the database or the use of 3D data could contribute to further improve the performances especially on untypical cases such as very damaged menisci or multiple tears.

Conflict of interests

The authors declare that they have no conflicts of interest concerning this article.

References

[1] Cullen KA, Hall MJ, Golosinskiy A. Ambulatory surgery in the United States, 2006. Natl Health Stat Rep 2009;11:1-25

[2] S. Pache, Z.S. Aman, M. Kennedy, G.Y. Nakama, G. Moatshe, C. Ziegler, et al., Meniscal root tears: current concepts review, Bone Joint Surg. 6 (2018) 250-9.

[3] Lecouvet F, Van Haver T, Acid S, Perlepe V, Kirchgesner T, Vande Berg B, et al. Magnetic resonance imaging (MRI) of the knee: Identification of difficult-to-diagnose meniscal lesions. Diagn Interv Imaging 2018;99(2):55-64.

[4] I. Boniatis, G.S. Panayiotakis, E. Panagiotopoulos. A computer-based system for the discrimination between normal and degenerated menisci from magnetic resonance images. International Workshop on Imaging Systems and Techniques. IEEE 2008:335-9.

[5] C. Köse, O. Gençalioğlu, U. U. Şevik, An automatic diagnosis method for the knee meniscus tears in MR images, Expert Systems with Applications. 36 (2009) 1208-16.

[6] B. Ramakrishna, W. Liu, G. Saiprasad, N. Safdar, C.I. Chang, K. Siddiqui, et al., An automatic computer-aided detection system for meniscal tears on magnetic resonance images, IEEE Trans Med Imaging. 28 (2009) 1308-16.

[7] A. Saygili, S. Albayrak. Meniscus segmentation and tear detection in the knee MR images by fuzzy c-means method. Signal Processing and Communications Applications Conference (SIU). 2017:1-4.

[8] A. Saygili, S. Albayrak. Meniscus tear classification using histogram of oriented gradients in knee MR images. Signal Processing and Communications Applications Conference (SIU). 2018:1-4.

[9] A. Garcia-Garcia, S. Orts, S. Oprea, V. Villena-Martinez, J.G. Rodríguez. A review on deep learning techniques applied to semantic segmentation. Computer Vision and Pattern Recognition. Cornell University; 2017.

[10] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature. 521 (2015) 436-44.

[11] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[12] Y. Xu, T. Géraud, E. Puybareau, I. Bloch, J. Chazalon, White matter hyperintensities segmentation in a few seconds using fully convolutional network and transfer learning, Brainlesion: glioma, multiple sclerosis, stroke and Traumatic Brain Injuries. Lecture Notes in Computer Science (LNCS) 1067 0 (2017).

[13] K. He, G. Gkioxari, P. Dollár, R. Girshick. Mask R-CNN. International Conference on Computer Vision (ICCV). IEEE; 2017:2980-8.

[14] TY. Lin, M. Maire, S.J. Belongie, L.D. Bourdev, R.B. Girshick, J. Hays, et al. Microsoft COCO: Common objects in context. European Conference on Computer Vision (ECCV). 2014:740-55.

[15] K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning for image recognition. Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2016:770-8.

[16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., Imagenet large scale visual recognition challenge, Int Comput Vision 115 (2015) 211-52.

[17] A. Krizhevsky, I. Sutskever, G.E. G. E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems. 25 (2012).

Figure legends

Figure 1. Database contains either medial e.g. (a) or lateral e.g. (b) MR images of the knee. (a-b) MR images shows healthy menisci. (c-f) MR images shows examples of tears as present in the database. (c) Horizontal tear in posterior meniscus, (d) Horizontal tear in anterior meniscus, (e) Vertical tear in posterior meniscus and (f), Vertical tear in anterior meniscus. Arrows point out tears.

Figure 2. MR images illustrate challenging cases. (a) Potentially misleading lesion. (b) Barely visible meniscus. (c) Multiple tears in the same meniscus.

Figure 3. Figure shows two MR images from the training database illustrating the menisci annotation process. Each colored dot is a vertex of a triangle that approximates the segmentation of a meniscus.

Figure 4. (a) Image from the training database with a clearly visible torn posterior meniscus that was correctly classified by the ConvNet. (b) Same image with a superimposed saliency mask indicating that the network focuses on non-relevant regions and barely considers the posterior meniscus itself. (c) Image (a) after applying a black top-hat filter with a disk structuring element of radius 5 pixels. (d) Saliency map for the processed image.

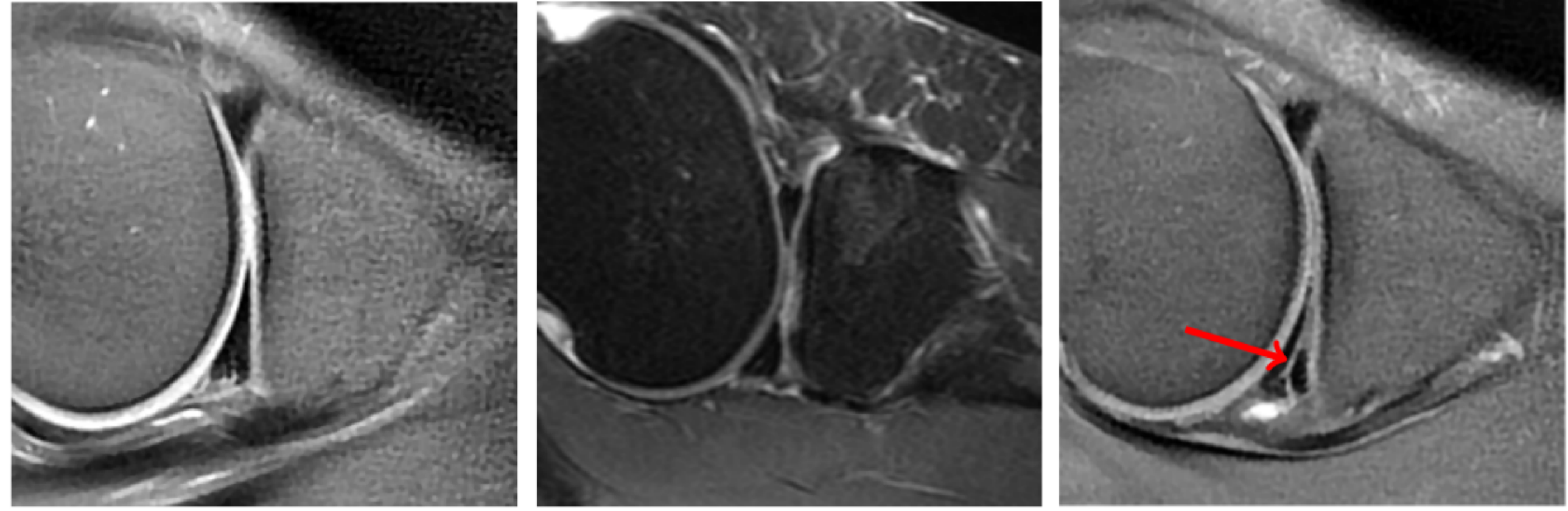
Figure 5. Pre-processing of the data used as input of the Mask R-CNN. (a) Original image. (b) 5 × 5 white top-hat. (c) 11 × 11 white top-hat. (d) 21 × 21 white top-hat.

Figure 6. Output of Mask R-CNN. (a-b) Correct results. (c) Posterior meniscus incorrectly segmented and labeled as torn.

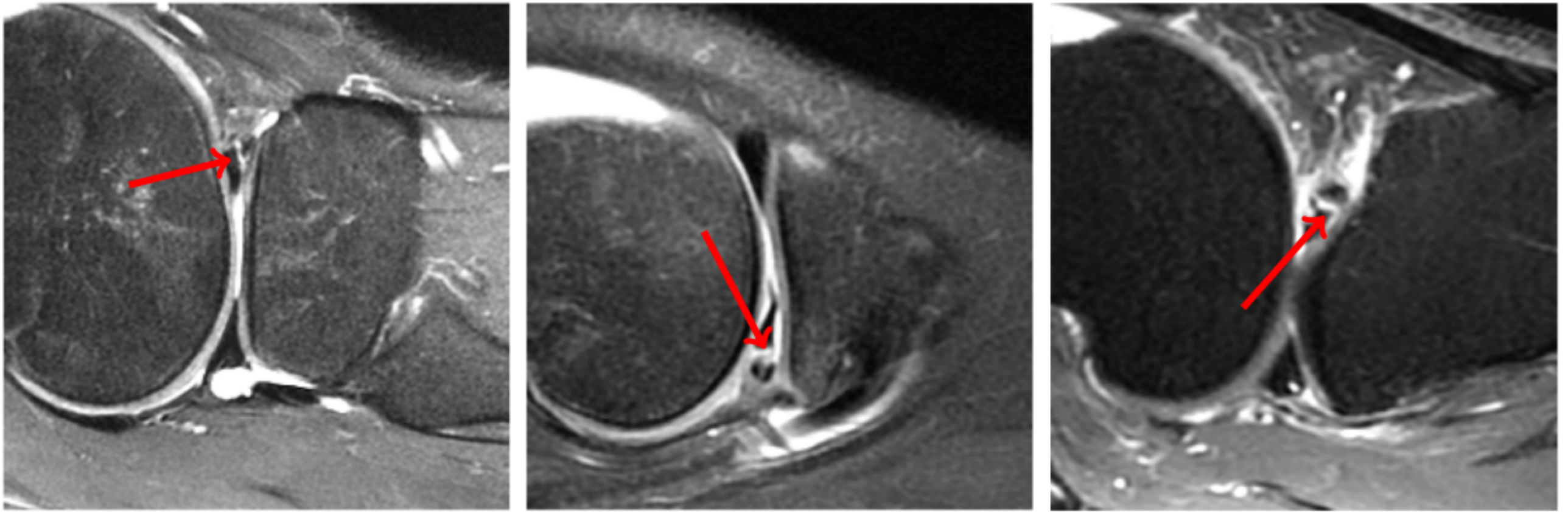
Figure 7. Patch extraction for orientation classification. (a) Extracted patch, resized to 47 × 47. (b) Local orientation map, $\sigma = 3$. (c) Local orientation map, $\sigma = 1$. (d) Black top-hat, $r = 4$. (e) Black top-hat, $r = 8$.

Figure 8. Prediction results on the testing batch. Most results seem correct, *e.g.* (a-b). However, some predictions are suspicious, e.g. (d-e). (a) No tear. (b) Horizontal tear on the posterior meniscus. (d) $P(Ant) \sim 0.45$ but the anterior meniscus looks torn. (e) $P(F) \sim 0$ but a tear is visible in the anterior meniscus. (c) Distribution of $P(F)$. (f) Distribution of P (H) for cases satisfying $P(F) > 0.5$.

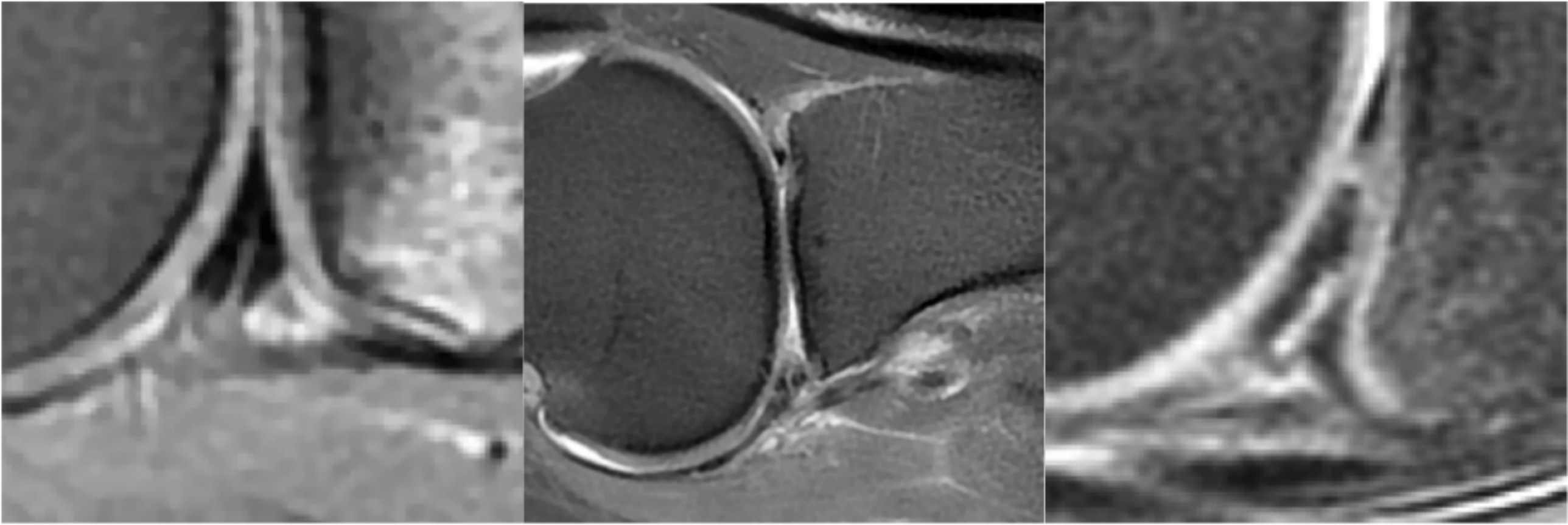
Figure 9. Cases for which P (F) (a-c) or P (H) (d-f) were close to 0.5. (a) Tear on the anterior meniscus but a slice where the menisci are connected was selected which does not meet the inclusion criteria. (b) Damaged anterior meniscus, but the presence of a tear is unclear. Yet the algorithm focused on the anterior meniscus: P (Ant) > 0.99. (c) Untypical lesion on the anterior meniscus. (d-e) Extensively damaged meniscus. (f) Several tears in one meniscus.



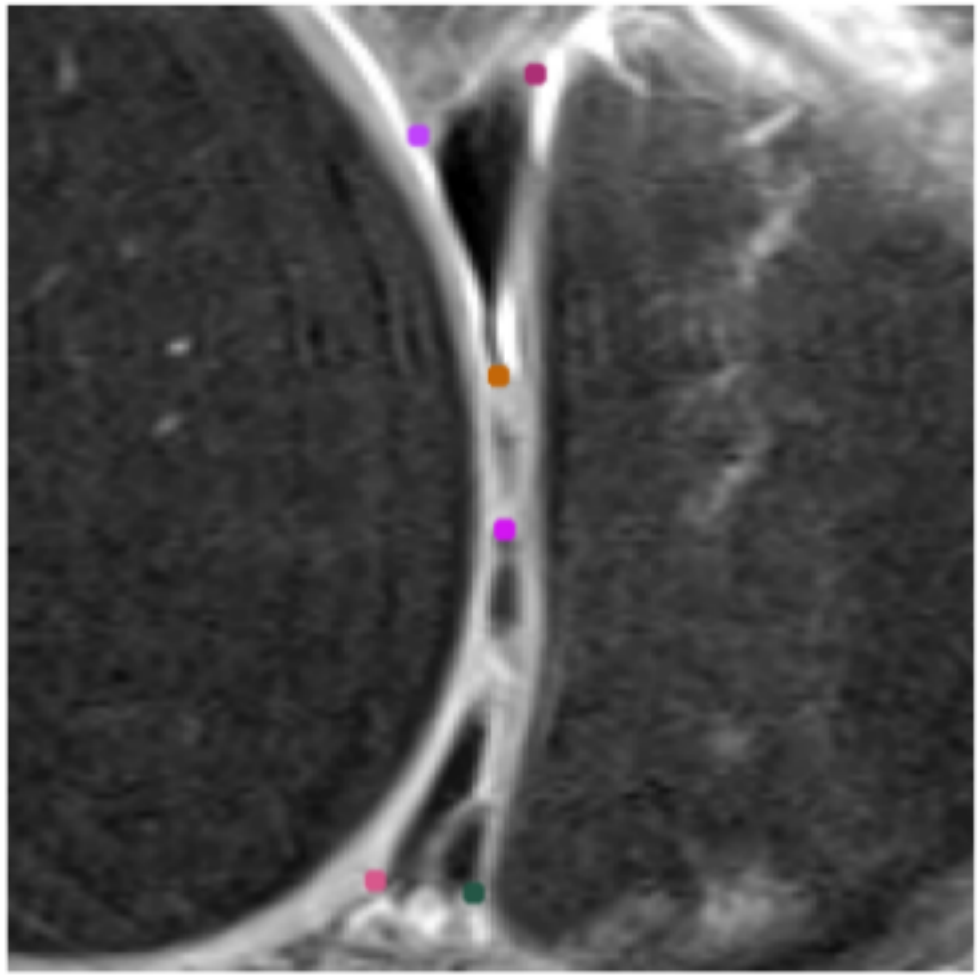
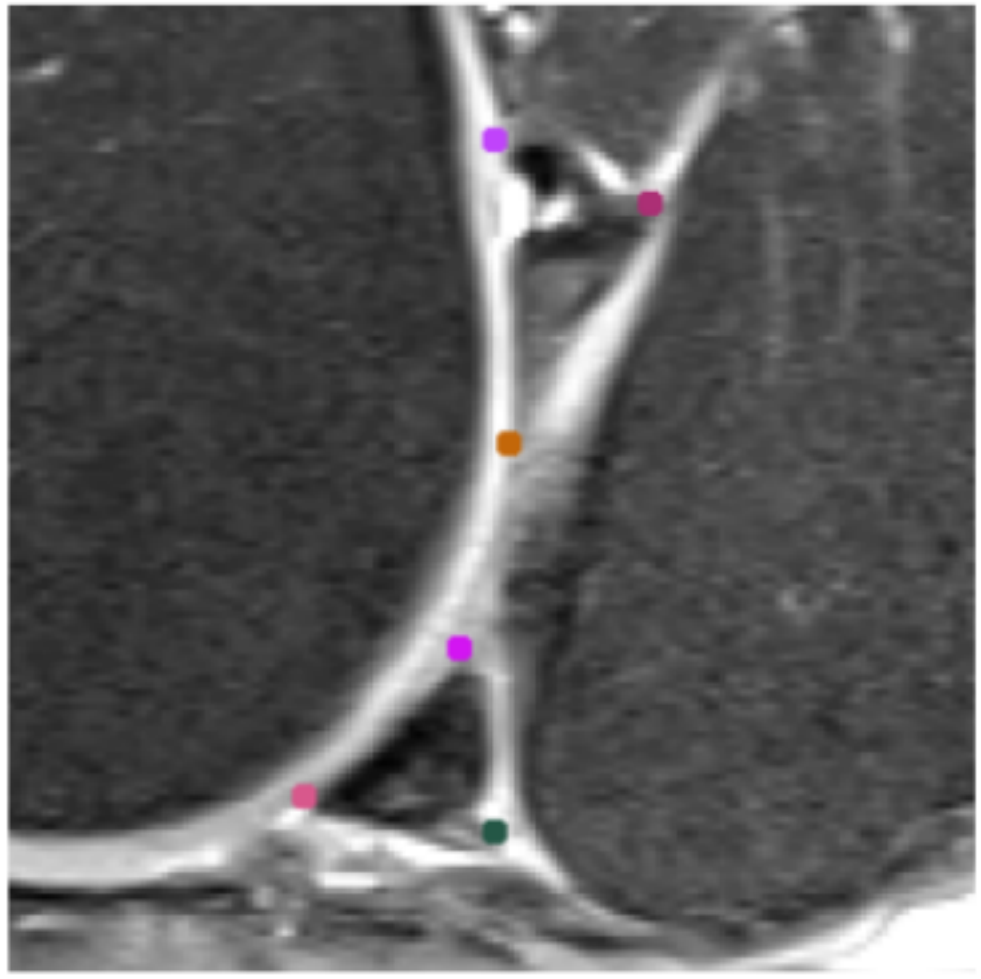
(a) (b) (c)

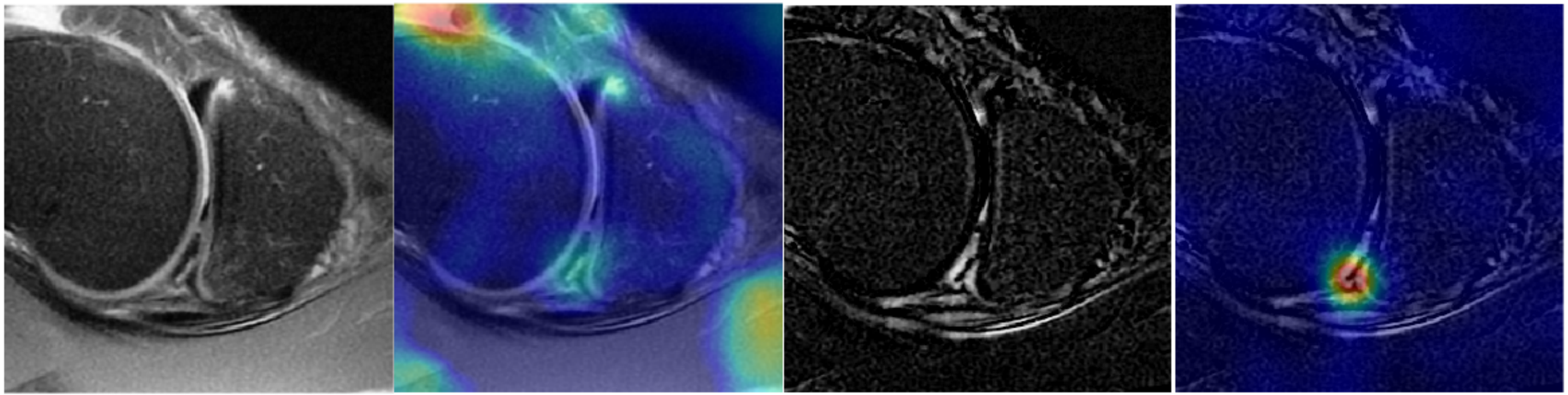


(d) (e) (f)

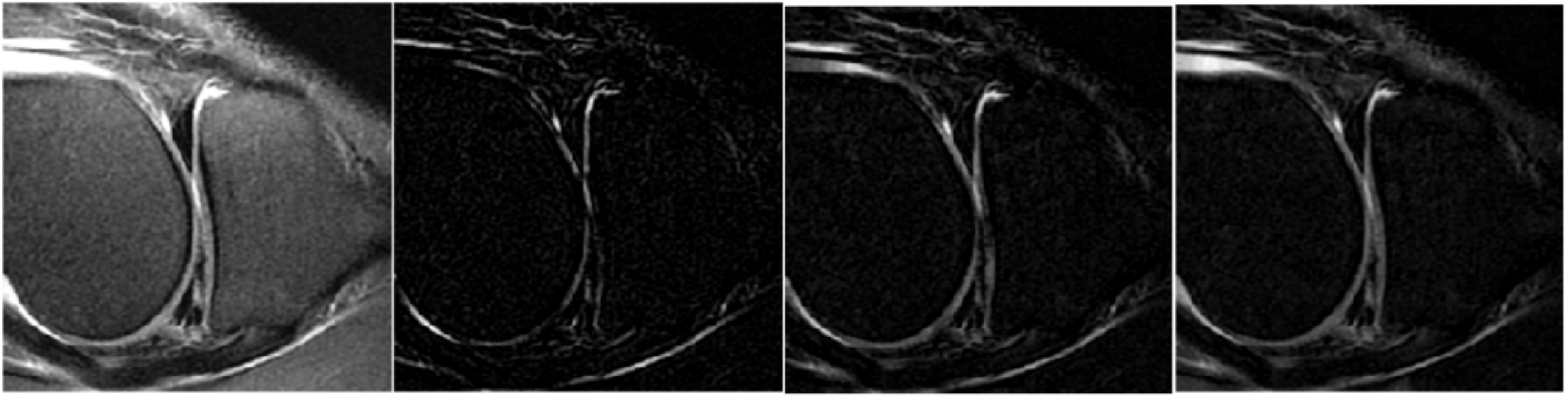


(a) (b) (c)

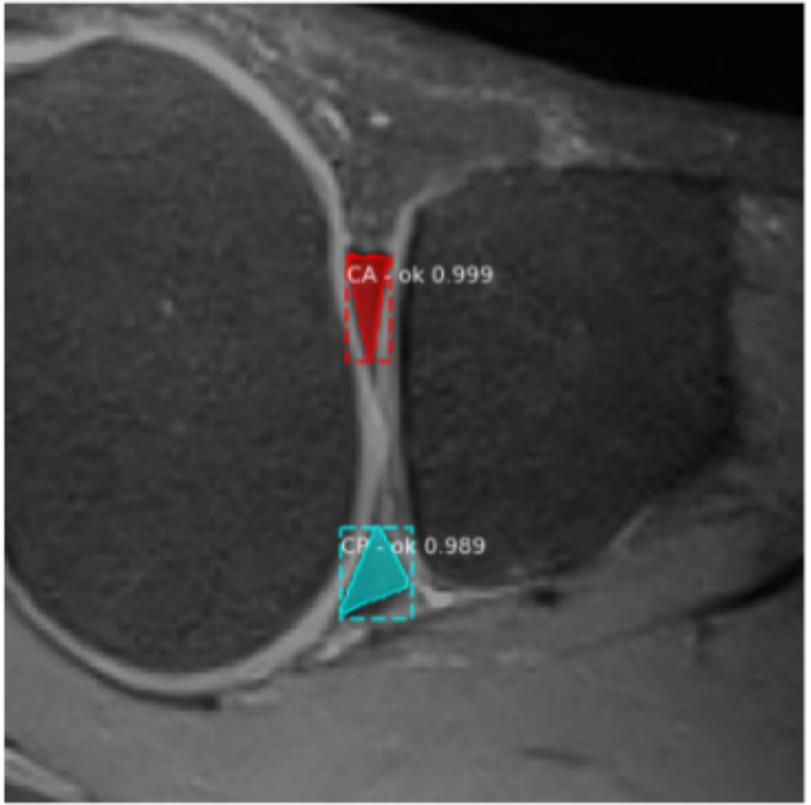





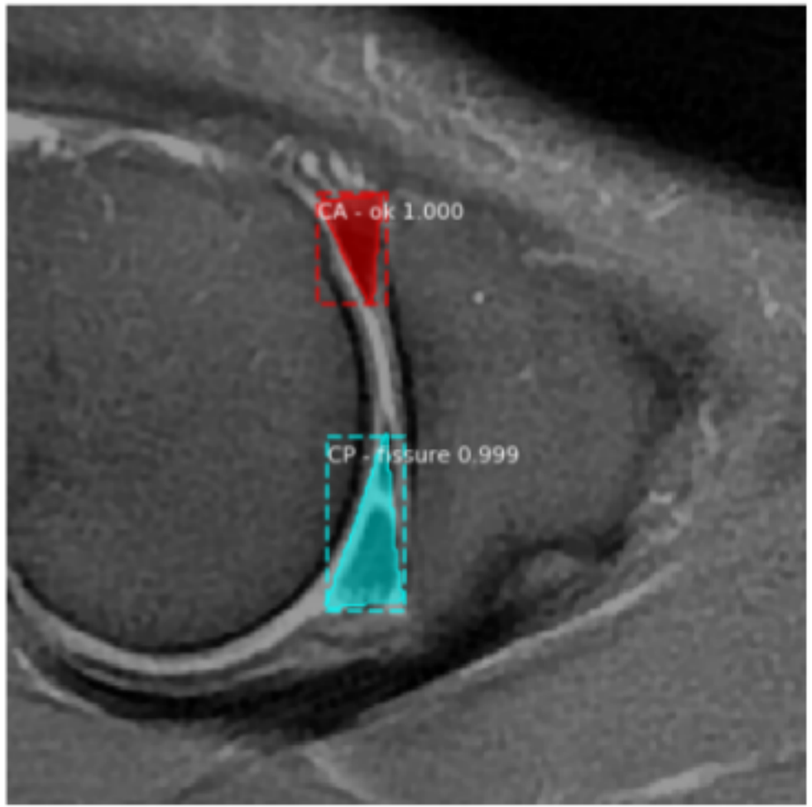
(a) (b) (c) (d)



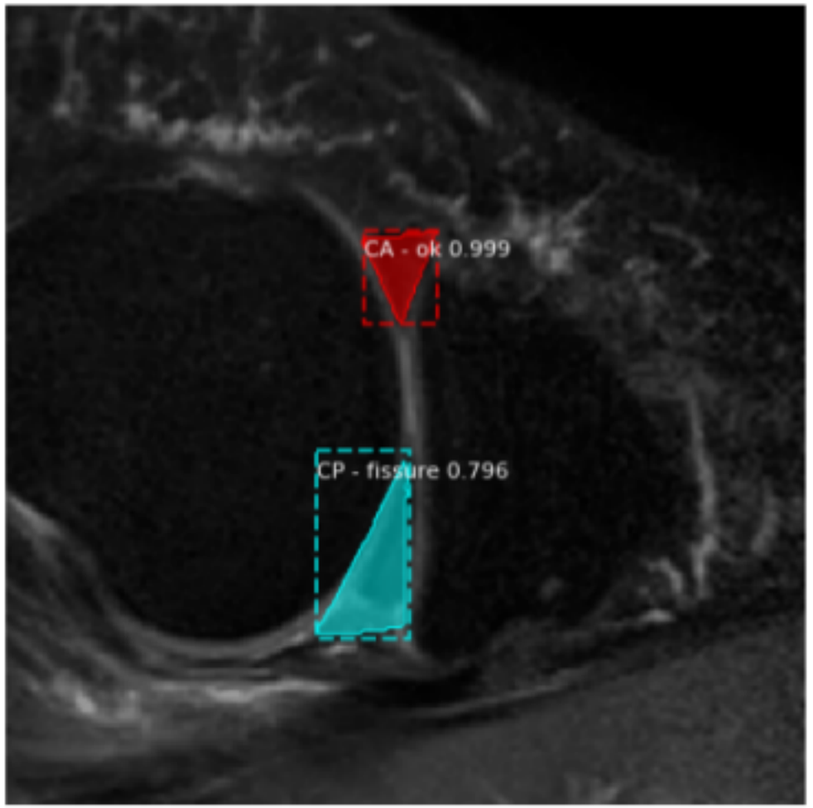
(a) (b) (c) (d)


CA - ok 0.999
CP - ok 0.989

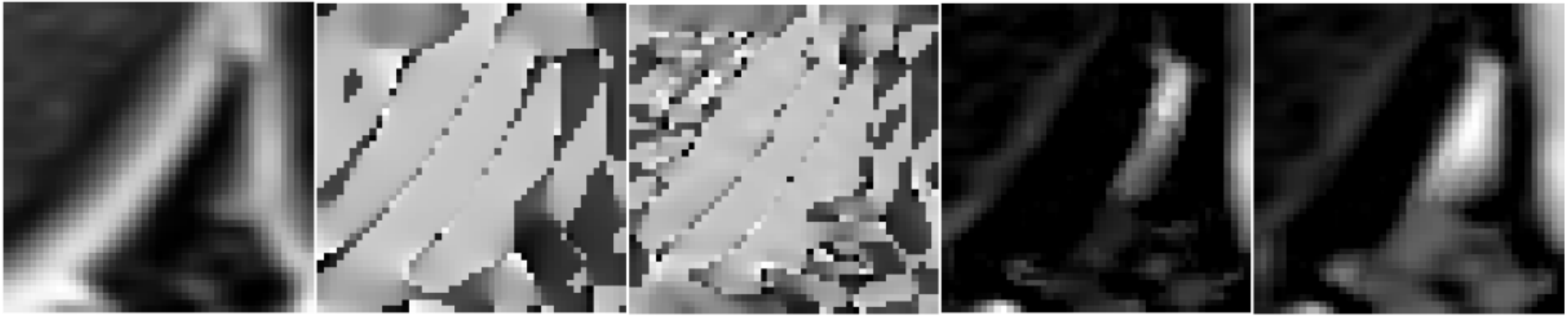
(a)


CA - ok 1.000
CP - fissure 0.999

(b)

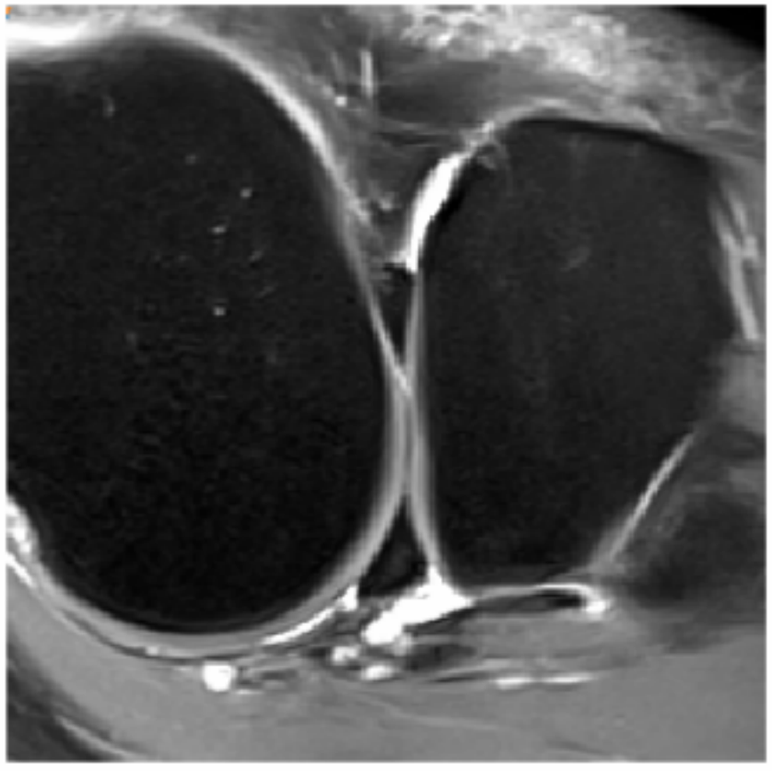

CA - ok 0.999
CP - fissure 0.796

(c)



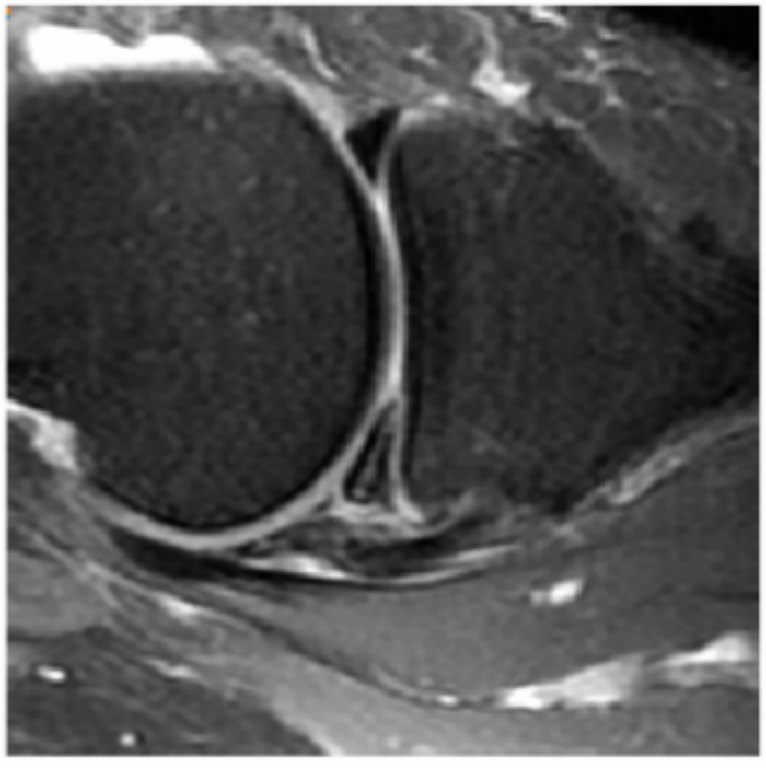
(a) (b) (c) (d) (e)

P(F)=0.0006, P(H)=0.9999, P(Ant)=0.8627

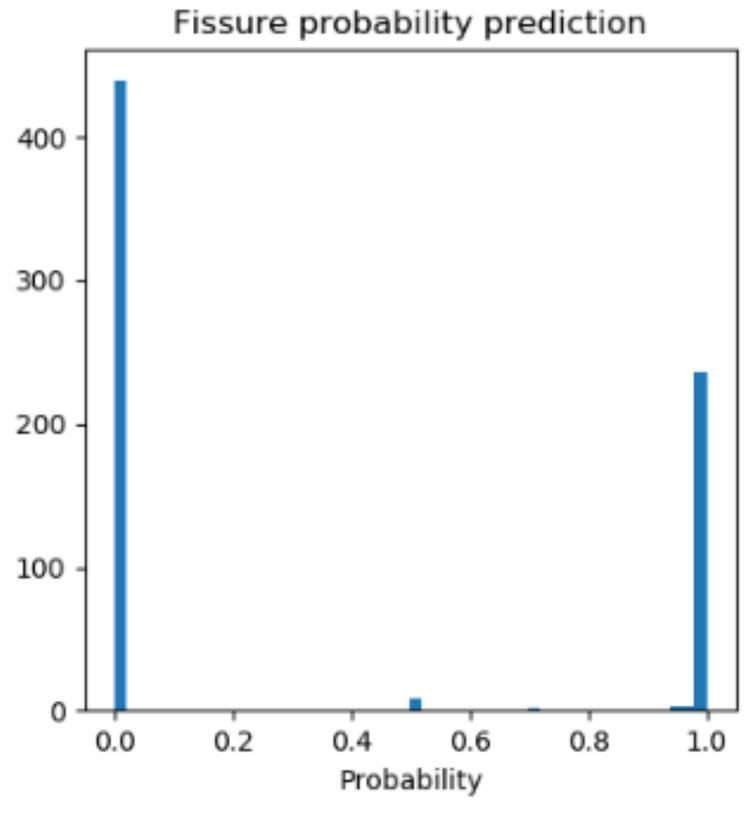


(a)

P(F)=0.9999, P(H)=1.0000, P(Ant)=0.0001

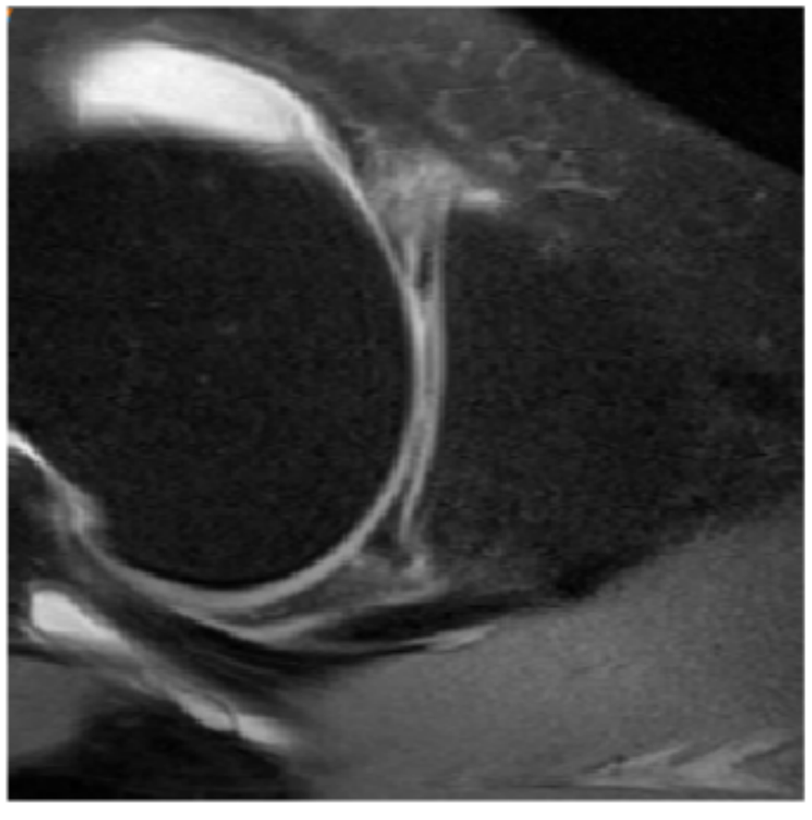


(b)



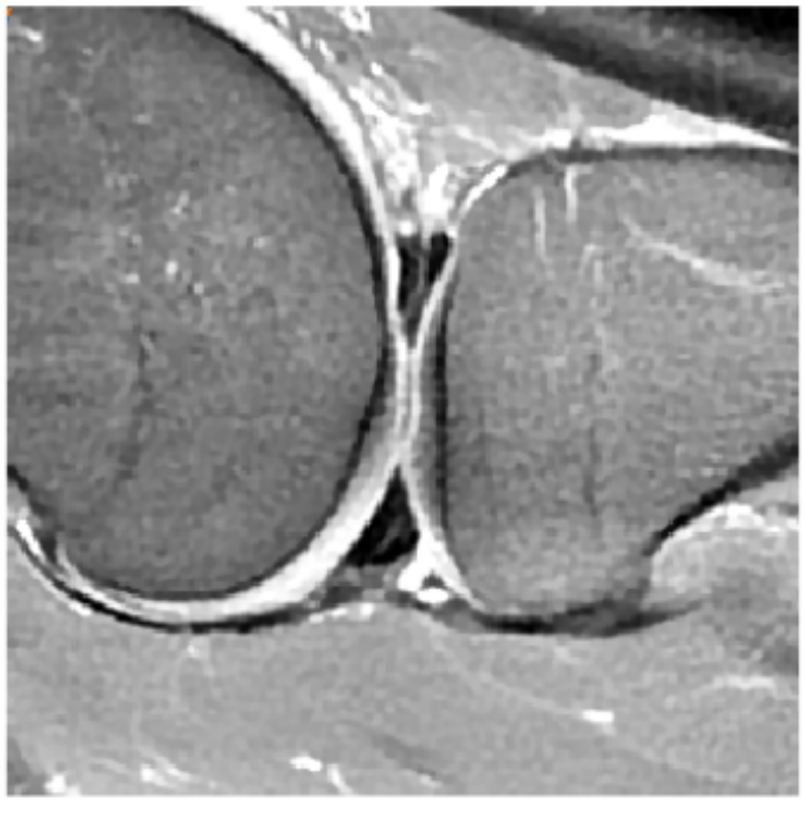
(c)

P(F)=0.9999, P(H)=0.9999, P(Ant)=0.4516

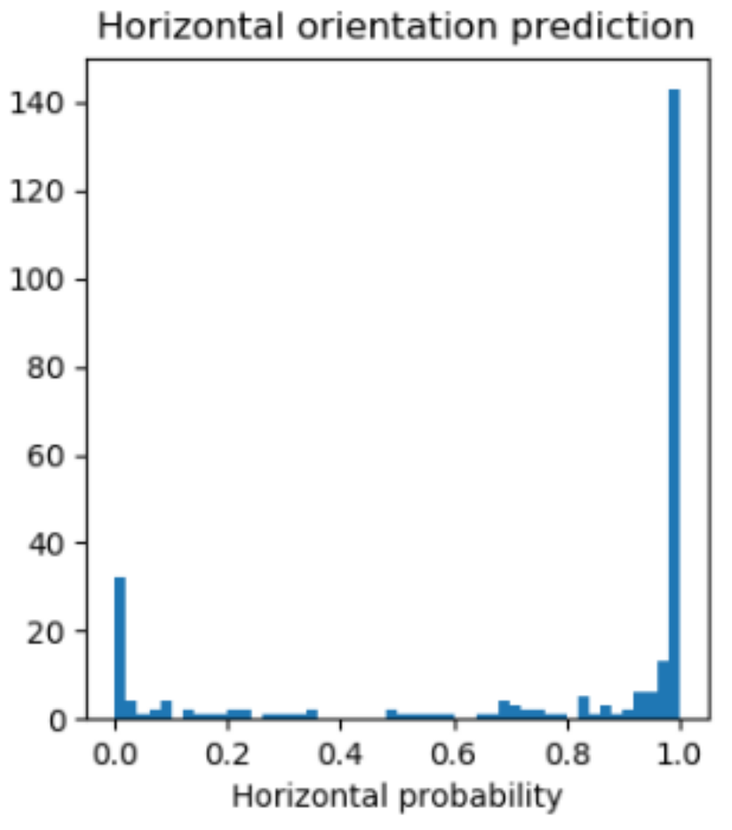


(d)

P(F)=0.0004, P(H)=0.2190, P(Ant)=0.4590

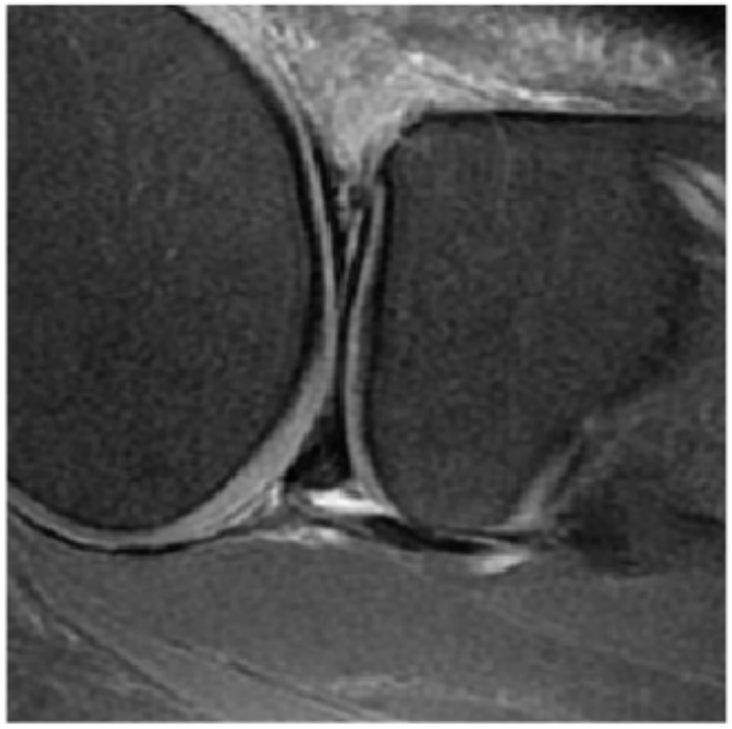


(e)



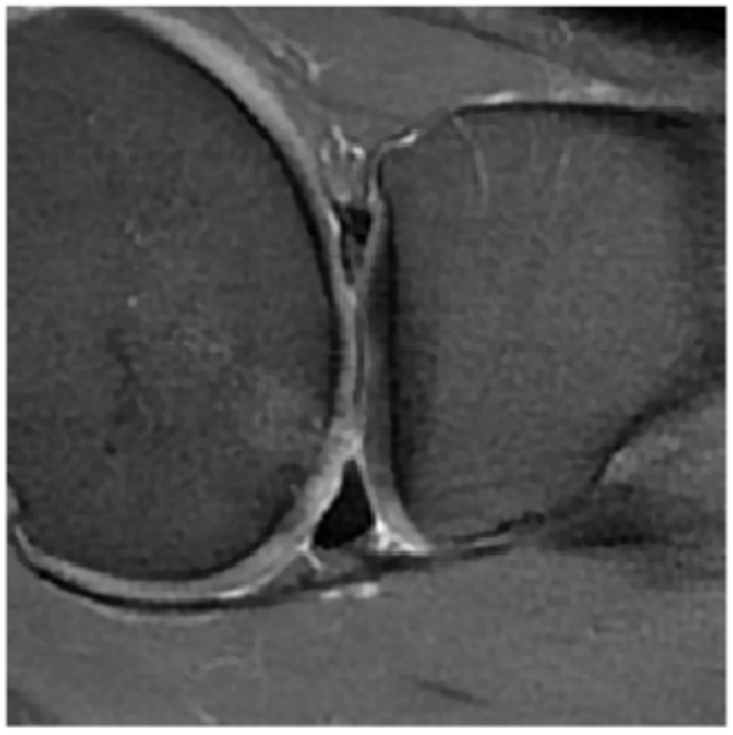
(f)

P(F)=0.5000, P(H)=0.9967, P(Ant)=0.9998



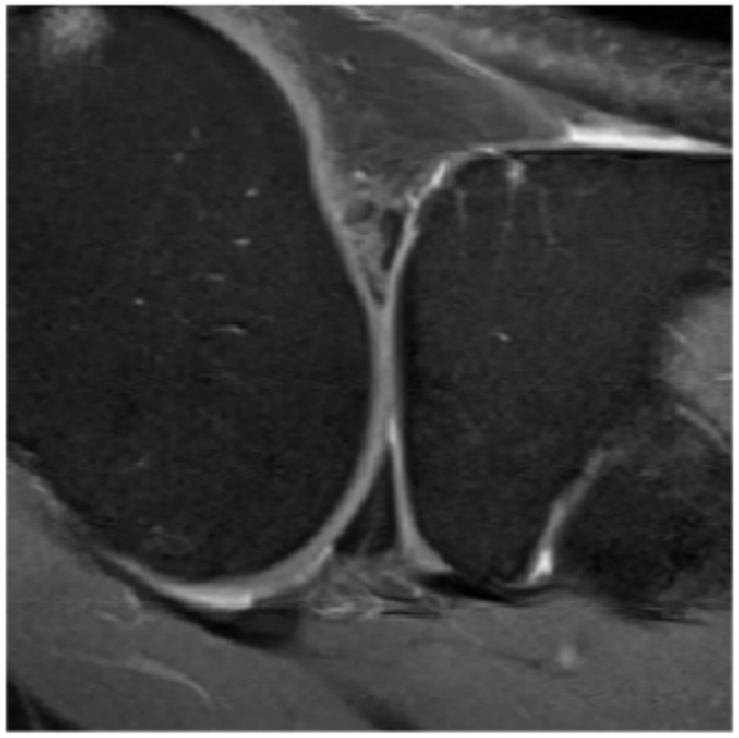
(a)

P(F)=0.5000, P(H)=0.0832, P(Ant)=0.9997



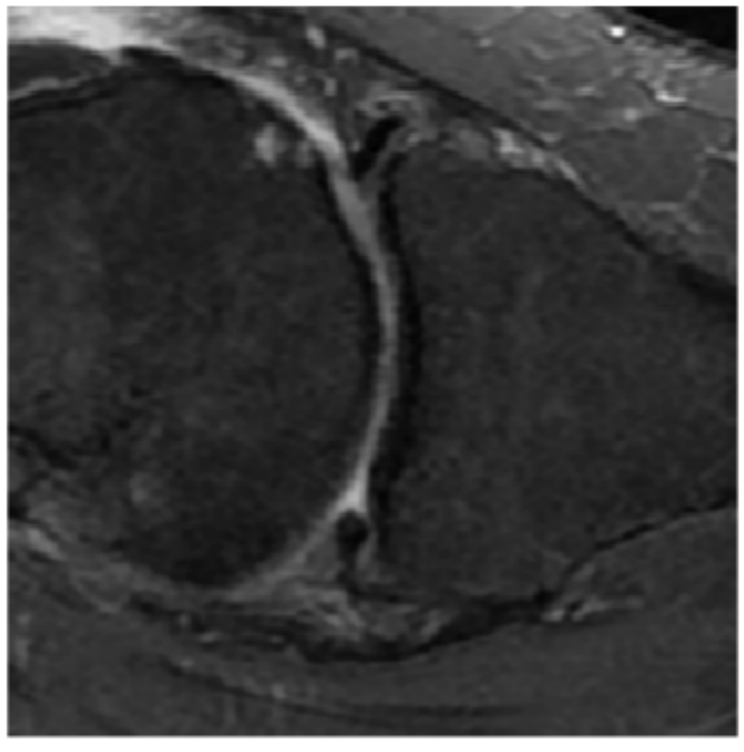
(b)

P(F)=0.5000, P(H)=0.7088, P(Ant)=0.9996



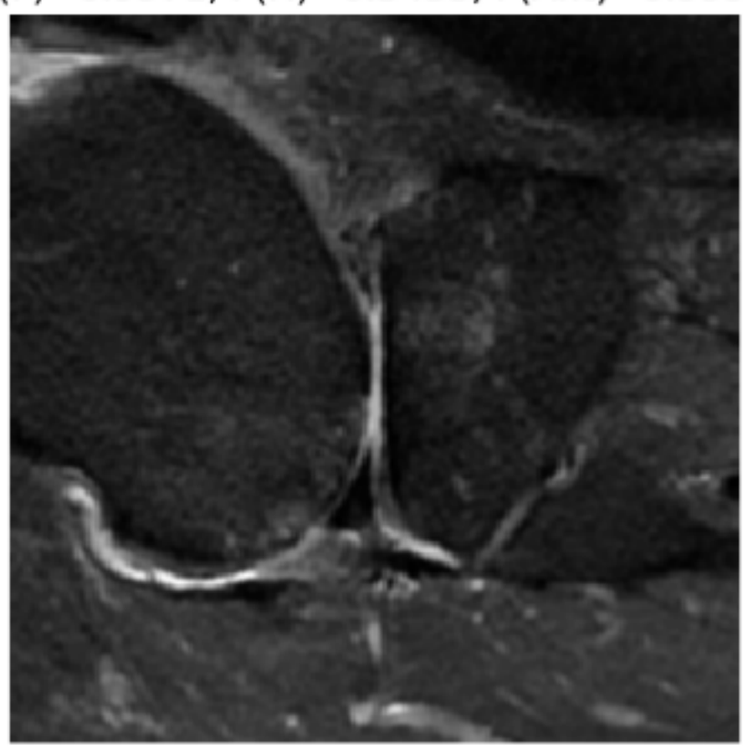
(c)

P(F)=0.9998, P(H)=0.6882, P(Ant)=0.0002



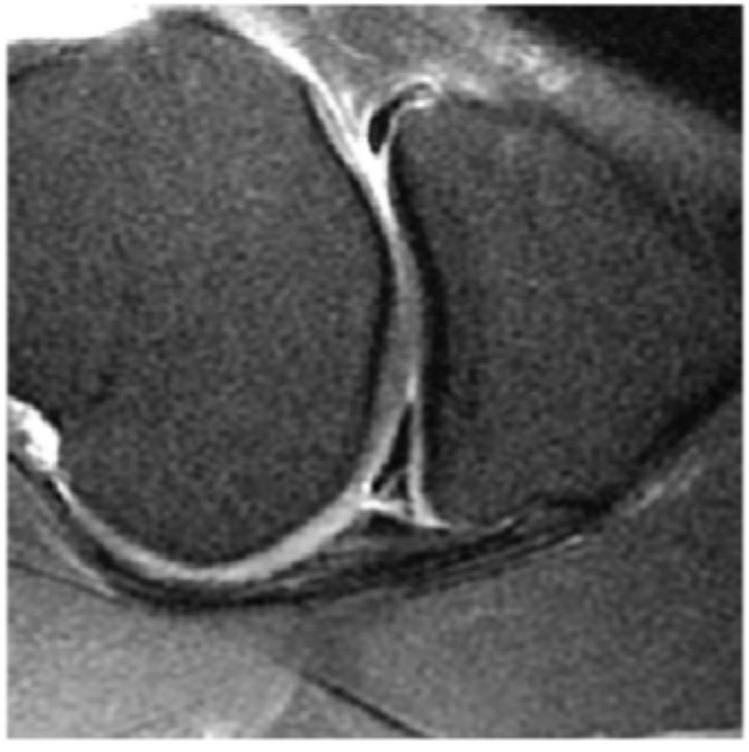
(d)

P(F)=0.9971, P(H)=0.5433, P(Ant)=0.9997



(e)

P(F)=0.9994, P(H)=0.6937, P(Ant)=0.0002



(f)