Zuse Institute Berlin

Takustr. 7
14195 Berlin
Germany

ALEXANDER TACK[1], ALEXEY SHESTAKOV, DAVID LÜDKE, STEFAN ZACHOW[2]

A deep multi-task learning method for detection of meniscal tears in MRI data from the Osteoarthritis Initiative database[3]

[1] 0000-0002-2418-7629, corresponding author
[2] 0000-0001-7964-3049
[3] to appear in: Frontiers in Bioengineering and Biotechnology, section Biomechanics

tack@zib.de

Zuse Institute Berlin
Takustr. 7
14195 Berlin
Germany

Telephone: +49 30 84185-0
Telefax: +49 30 84185-125

E-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Abstract

We present a novel and computationally efficient method for the detection of meniscal tears in Magnetic Resonance Imaging (MRI) data. Our method is based on a Convolutional Neural Network (CNN) that operates on a complete 3D MRI scan. Our approach detects the presence of meniscal tears in three anatomical sub-regions (anterior horn, meniscal body, posterior horn) for both the Medial Meniscus (MM) and the Lateral Meniscus (LM) individually. For optimal performance of our method, we investigate how to preprocess the MRI data or how to train the CNN such that only relevant information within a Region of Interest (RoI) of the data volume is taken into account for meniscal tear detection. We propose meniscal tear detection combined with a bounding box regressor in a multi-task deep learning framework to let the CNN implicitly consider the corresponding RoIs of the menisci. We evaluate the accuracy of our CNN-based meniscal tear detection approach on 2,399 Double Echo Steady-State (DESS) MRI scans from the Osteoarthritis Initiative database. In addition, to show that our method is capable of generalizing to other MRI sequences, we also adapt our model to Intermediate-Weighted Turbo Spin-Echo (IW TSE) MRI scans. To judge the quality of our approaches, Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values are evaluated for both MRI sequences. For the detection of tears in DESS MRI, our method reaches AUC values of 0.94, 0.93, 0.93 (anterior horn, body, posterior horn) in MM and 0.96, 0.94, 0.91 in LM. For the detection of tears in IW TSE MRI data, our method yields AUC values of 0.84, 0.88, 0.86 in MM and 0.95, 0.91, 0.90 in LM. In conclusion, the presented method achieves high accuracy for detecting meniscal tears in both DESS and IW TSE MRI data. Furthermore, our method can be easily trained and applied to other MRI sequences.

Keywords: knee joint, meniscal lesions, convolutional neural networks, residual learning, explainable AI, multi-task deep-learning, bounding box regression, object detection

1 Introduction

Menisci are hydrated fibrocartilaginous soft tissues within the knee joint that absorb shocks, provide lubrication, and allow for joint stability during movement [1]. In patients with symptomatic osteoarthritis, meniscal damage is also found very frequently with a prevalence of up to 91% [2]. Meniscal tears are usually caused by trauma and degeneration [3] and might lead to a loss of function, early osteoarthritis, tibiofemoral osteophytes, and cartilage loss [4, 5]. Magnetic Resonance Imaging (MRI) is commonly used for the noninvasive assessment of meniscal morphology since MRI provides a three-dimensional view of the knee joint with high contrast between soft tissues. Hence, MRI is the recognized screening tool for diagnostic assessment before performing therapeutic arthroscopy or any other treatment [6]. Among other factors, a proper treatment concept of meniscal damage depends highly on the type of tear and its location [3, 7]. An appropriate medical intervention can delay further development of arthritic changes, improve quality of life, and reduce healthcare expenditures. However, in practice, it is not always clear

what the optimal treatment actually is [8, 9], while an improper procedure might even lead to an acceleration of osteoarthritis progression [10]. For this reason, an accurate and reliable diagnosis of meniscal tears in view of their location, type, and orientation is important.

The diagnosis of meniscal tears in MRI is a time consuming and tedious procedure. These defects are often difficult to detect due to their small sizes and arbitrary orientations. It is frequently necessary to go back and forth in the MRI slices and to switch view directions for a thorough assessment of occurrences and spatial extents of pathological changes. In addition, the meniscal representation in the image data depends on the chosen MRI sequence. What appears clearly visible in one sequence may be barely noticeable in another due to insufficient contrast. Computer-Aided Diagnosis (CAD) attempts to overcome some of these limitations. CAD tools can be employed to increase the sensitivity and specificity of physicians in detecting and classifying meniscal tears [11, 12, 13]. Moreover, CAD could speed up the diagnosis, reduce the number of unintentionally missed defects, avoid unnecessary interventions (e.g. arthroscopic interventions), and lead to fewer treatment delays. Several CAD approaches for an automated detection of meniscal tears in MRI data have been proposed in recent years. A distinction can be made between methods that evaluate the 2D contents of cross-sectional images often coming from a set of curated slices (2D approaches) and those that evaluate 3D image information in the MRI data volume (3D approaches). In the context of image analysis by means of Convolutional Neural Networks (CNNs), we distinguish between 2D CNNs and 3D CNNs. In the case of the 2D approaches, there exists a pseudo-3D variant in which sets of (neighboring) sectional images are included in the evaluation. In these pseudo-3D variants, 2D CNNs are employed to encode 2D slices of a 3D MRI dataset. Afterwards, the respective 2D encodings are condensed (e.g. by global max- or average-pooling), concatenated, and passed to a classifier.

[14] proposed a method to detect meniscal tears from a curated set of sagittal 2D MRI slices. Their approach is based on the 2D "faster R-CNN" [15] and comprises three steps: Firstly, the positions of both meniscal horns are detected; secondly, the presence of a tear is classified; and thirdly, the respective tear orientation is determined. The method yields an Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) of 0.92 for the detection of the meniscal horns' positions, an AUC of 0.94 for detecting the presence of meniscal tears, and an AUC of 0.83 for the determination of the tear orientations. [16] presented a similar method, also detecting meniscal tears from a curated set of sagittal 2D MRI slices. They employed a masked region-based 2D CNN [17] to locate the anterior and the posterior horns of the Medial Meniscus (MM) as well as the Lateral Meniscus (LM). Their method yields on average an AUC of 0.906 for all three tasks, i.e. the location of the respective region, the detection of meniscal tears, and the classification of the tear orientation.

Processing of all MRI slices instead of individually selected ones was performed by [12] who proposed a 2D CNN for the detection of meniscal tears. Their method achieves an AUC of 0.847. [11] adopted a method that combined a 2D CNN for meniscus segmentation with a 3D CNN for detection and severity staging of meniscal tears. This approach was able to differentiate between

tears and no tears with an AUC of 0.89. [18] proposed a so-called "Efficiently-Layered Network" for detection of meniscal tears, reaching an AUC of 0.904 and 0.913 for two different datasets. [19] demonstrated the use of a 2D CNN as a pseudo-3D variant for detection of torn menisci. Their method relies on transfer learning while using data augmentation and reaches an AUC of 0.934. [20] presented a deep 3D CNN to detect tears in MRI data for MM and LM, respectively. Their method reaches AUC values of 0.882, 0.781, and 0.961 for the detection of medial, lateral, and overall meniscal tears. [21] also proposed a 3D CNN for meniscal tear detection in MRI data for MM and LM individually. Their approach yields an AUC of 0.93 for MM and 0.84 for LM.

A common limitation among many of the methods listed above is their strong reliance on segmentations of the menisci (or at least of bounding boxes), which can be challenging to obtain due to the inhomogeneous appearance of pathological menisci in MRI data as well as an insufficient contrast to adjacent tissues [22]. Furthermore, some approaches merely operate on 2D slices. A major limitation of such methods is that the trained 2D CNNs cannot take whole MRI volumes into account, thus possibly missing important feature correlations in 3D space. Besides, the selection of curated slices requires expert knowledge. Therefore, the applicability of these methods to 3D volumes is unclear since they were not trained on 3D data. Finally, none of the presented methods is able to detect meniscal tears for all anatomical sub-regions of the menisci individually, i.e., the anterior horn, the meniscal body, and the posterior horn.

Our motivation is to detect meniscal tears in MRI data more accurately than previous methods in terms of correctness and localization. For this purpose, we present a method that detects tears in anatomical sub-regions of both the MM and the LM. We design our study in a manner that allows for a comparison of different possible approaches. Moreover, the study shows our progression in addressing the task of meniscal tear detection in 3D MR images. We investigate how to handle best the input data such that the least pre-processing is required for inference and the best accuracy is achieved. Furthermore, we show that our proposed method generalizes well to different MRI sequences. We employ two ResNet architectures [23, 24] to classify meniscal tears in each sub-region of the MM and the LM, respectively, utilizing three different approaches. In a first approach (i), we train a 3D CNN on the complete 3D MRI dataset as input. We call it ***Full-scale*** **approach** within the remainder of this article.

Since large input data requires a lot of GPU memory, longer time for training and inference, and contains image information not necessarily needed for an assessment of meniscal tears, we decided to crop the data to the Regions of Interest (RoI) of both menisci in an automated pre-processing step that requires segmentations of sufficient quality for training and testing [25]. Hence, in a second approach (ii), a 3D CNN is trained on these cropped MRIs detecting meniscal tears more accurately than in our first approach. We refer to the second approach as ***BB-crop*** **approach**. We enhanced the performance of our first approach by adding a bounding box regression task. Thus, our final approach (iii) trains a CNN to detect meniscal tears in complete 3D MRI, combined with an additional bounding box regression task leading to an auxiliary loss (the ***BB-loss*** **approach**). Framing the problem of meniscal tear detection in this multi-task learning

setting - simultaneously solving meniscal tear detection *and* meniscal bounding box regression - allows our model to implicitly learn to focus on the meniscal regions. Furthermore, segmentation masks are only required during training. Hence, our final approach requires the least data pre-processing at inference time and achieves the best results.

This study presents a method that detects meniscal tears in 3D MRI data on a sub-region level, i.e., the anterior horn, the meniscal body, and the posterior horn for both MM and LM. Formulating the problem in a multi-task learning setting, by adding the information of the location of the menisci as an auxiliary loss to our 3D CNN, state-of-the-art results are achieved. In order to provide an explanation to our CNN's decision, SmoothGrad saliency maps [26] are computed and visualized. That way a visual guidance can be given to the clinical domain experts for confirming the results of our approach.

2 Materials and Methods

In section 2.1 of this chapter, the data to our method is presented. Thereafter, we introduce our data pre-processing and bounding box generation in 2.2. Section 2.3 is a description of the model architectures utilized in our approach and of their respective components. The particular configuration of our three approaches is illustrated in detail in sections 2.4, 2.5, and 2.6, followed by an explanation of our experimental set-up and training in 2.7. Finally, the statistical evaluation is summarized in 2.8 and a method for saliency maps is proposed in 2.9.

2.1 Data from the OAI database

The publicly available database of the Osteoarthritis Initiative (OAI)[1] was established to provide researchers with resources to promote the prevention and treatment of knee osteoarthritis. We use 2,399 sagittal Double Echo Steady-State (DESS) 3D MRI scans from the OAI database acquired using Siemens Trio 3.0 Tesla scanners [27]. Additionally, 2,396 sagittal Intermediate-Weighted Turbo Spin-Echo (IW TSE) MRI scans are investigated for the same patients. The demographics of our study are shown in Table 1. The OAI database includes multiple reading studies of respective osteoarthritis characteristics, which can be assessed in medical image data. As a gold standard, we utilize labels from MOAKS [28] image reading studies performed by clinical experts. In the MOAKS scoring system, the menisci are divided into three anatomical sub-regions: anterior horn, body, posterior horn. We consider regions as not containing a tear if the MOAKS score was "normal" or if the sub-region contains a signal abnormality (which is not extending through the meniscal surface and, hence, is no tear). We considered any other type of abnormality (radial, horizontal, vertical, etc.) as a meniscal tear (c.f. Table S1). Examples of the MRI sequences, signal abnormalities, and meniscal tears are shown in Fig. 1.

[1] `https://nda.nih.gov/oai/`

Figure 1: Examples of normal menisci, signal abnormalities, and subjects with meniscal tears shown for DESS as well as IW TSE MRI data. For a summary of different types of meniscal tears per sub-region the reader is referred to Supplementary Table S1.

	DESS	IW TSE
Number of MR images	2,399	2,396
In-plane resolution	0.36 mm × 0.36 mm	0.36 mm × 0.36 mm
Usual slice dimension	384 × 384	442 × 448
Slice thickness	0.7 mm	3 mm
Number of slices	160	35 to 43
Side (left; right)	1104; 1295	1104; 1292
Sex (female; male)	1489; 910	1487; 909
Age [years]	61.88 ± 8.87	61.89 ± 8.86
BMI [kg/m^2]	29.01 ± 4.79	29.08 ± 4.79
MM (% normal)	60.0%	59.9%
LM (% normal)	80.0%	79.9%

Table 1: Demographics: In this study, 2,399 DESS and 2,396 IW TSE MRI scans from the OAI database are analyzed. In these data, slightly more normal than diseased medial menisci (MM) and lateral menisci (LM) are contained. Here, normal is defined as no conspicuous features with respect to the MOAKS scoring system in any sub-region.

2.2 Data pre-processing and localization of menisci

In a first step of our pre-processing, the intensities of all MR images are scaled to a range of $[0, 1]$ using min-max normalization. Following that, a standardization is applied to each MR image $\mathcal{I}_i$ according to the equation below:

$$\widetilde{\mathcal{I}}_i = \frac{\mathcal{I}_i - \mu}{\sigma} \ , \tag{1}$$

where μ is the mean intensity and σ is the standard deviation of the training population of normalized scans. Leveraging meniscal segmentations generated by the method of [25] RoIs spanning the MM and LM are created for DESS MRI data (see Fig. 2). RoIs are computed by querying the minimum and maximum position of the menisci along each dimension of the binary segmentation masks: x_{min}, x_{max}, y_{min}, y_{max}, z_{min}, z_{max}. The bounding boxes are uniquely defined as the 3D center coordinate

$$BB_{center} = \begin{bmatrix} (x_{max} - x_{min})/2 + x_{min} \\ (y_{max} - y_{min})/2 + y_{min} \\ (z_{max} - z_{min})/2 + z_{min} \end{bmatrix} , \tag{2}$$

and with the respective height $(x_{max} - x_{min})$, width $(y_{max} - y_{min})$, and depth $(z_{max} - z_{min})$. These values are represented as relative image coordinates. Hence, a bounding box is defined by 6 floating values: $[BB^x_{center}, BB^y_{center}, BB^z_{center}, height, width, depth]$.

For the IW TSE data 600 segmentations are generated in a semi-automated fashion using Amira ZIB Edition[2] [29]. These masks are defined as voxel-wise annotations of the tissue belonging to the respective meniscus. The method of [25] was originally developed and evaluated on DESS

[2] `https://amira.zib.de`

MRI data. Since the DESS and IW TSE MRI sequences differ significantly in the image resolution (number of slices), that could pose an issue, we have decided to train the self-adapting nnU-net framework [30] on these 600 training datasets. The nnU-net offers 2D and 3D architectures and 3D architectures usually yields better results [30]. For this reason, we have used a 3D variant of the nnU-net that employs 3D convolutions in an encoder-decoder framework with skip-connections. For the IW TSE data, the nnU-net has been automatically configured to have an input size of $24 \times 256 \times 256$ pixels and seven layers of 3D convolutions [30]. We train the nnU-net with data augmentation such as random rotations and random cropping using a dice similarity coefficient loss [30] until convergence is reached. Hereby, the dice similarity coefficient is computed between the output of the nnU-net and the respective hand-labelled target segmentation masks. Afterwards, the nnU-net is employed to segment all 2,396 IW TSE MRI scans to yield the respective meniscal RoIs. In order to achieve this, multiple patches of the MRI with a size of $24 \times 256 \times 256$ pixels are being processed by the nnU-net. These patches overlap by half of the patch size in each dimension. Afterwards, the nnU-net framework merges all patches to a final 3D segmentation mask employing a majority voting for every pixel.

2.3 Model architecture

Two distinct models, which are based on 3D counterparts of ResNet architectures [23, 24] are introduced. ResNets have been widely applied to the medical domain and provide good properties due to the employed skip connections. In theory, the residual connections allow the design of very deep ResNets without exhibiting problems of vanishing gradients [31]. We have chosen 3D counterparts of 2D ResNets since 3D convolutions are able to comprehend three-dimensional context inherently. It has previously been shown in the context of musculoskeletal MRI analysis that 3D convolutions are more powerful than concatenation of 2D slices as well as a provision of multiple 2D slices as input to a CNN that employs 2D convolutions [32, 33]. We adapt these 3D ResNet architectures to the three different approaches and their associated input volume sizes. Each model consists of a ResNet encoder followed by one or two Multi-Layer Perceptron (MLP) heads. The *BB-crop* approach has a dilation ResNet-C-26 architecture with an MLP head for the multi-label classification. The *Full-scale* approach has a ResNet50 encoder with a classifier MLP head, and the *BB-loss* approach consists of a ResNet50 encoder with two MLP heads. The performance of the classification task is improved in the *BB-loss* approach by solving additionally a second task, which is to learn a bounding box regression simultaneously. Again, the first MLP head is employed for multi-label classification. The second MLP head is responsible for the bounding box regression task. All ResNets comprise of a series of convolutional layers, each followed by batch normalization [34] and a Rectified Linear Unit (ReLU) activation function [35]. Our approaches that will be presented in the following sections are designed based on (a selection of) encoders and MLP heads:

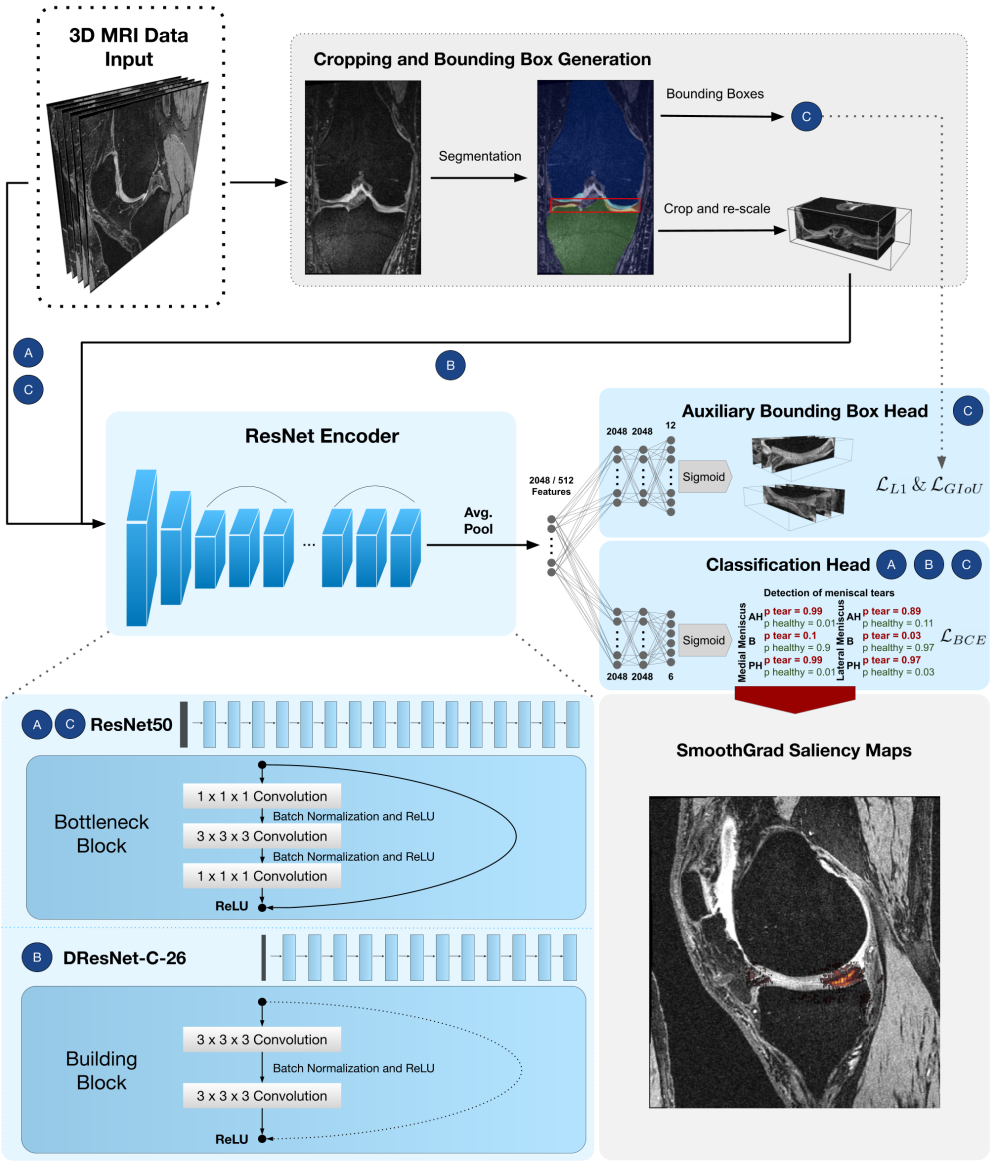

Figure 2: CNN pipeline for detection of meniscal tears in six sub-regions. Approach *Full-scale* uses a ResNet50 encoder followed by a classifier head with $\mathcal{L}_{BCE}$ for classification of meniscal tears in 3D MRI data (A). Approach *BB-crop* reduces the 3D MRI input to the meniscal RoI and uses a DRN-C-26 encoder followed by a classifier head with $\mathcal{L}_{BCE}$ to detect meniscal tears (B). Approach *BB-loss* uses a ResNet50 encoder followed by a classifier head with $\mathcal{L}_{BCE}$ as well as another bounding box regression head with $\mathcal{L}_{L1}$ and $\mathcal{L}_{GIoU}$ in order to predict bounding boxes of the menisci in the 3D MRI data (C). The ResNet50 is made up of an initial convolutional layer followed by max-pooling before 16 ResNet bottleneck blocks with residual connections are stacked. The DRN-C-26 starts with the same convolutional layer but is immediately followed by ten residual building blocks and, lastly, two building blocks without a residual connection. After average pooling, the encoders generate 2048 and 512 features, respectively. Finally, SmoothGrad saliency maps are presented as overlaid heatmaps on top of the respective MR image to highlight these regions that mostly influenced the detection of tears (button right corner).

ResNet50 encoder

[23] proposed a residual layer connection as a way to train deep neural networks without suffering from vanishing gradients. One of their proposed architectures is the ResNet50, with a total of 50 convolutional layers (see Fig. 2). The network comprises an initial convolutional layer with kernel size $7 \times 7 \times 7$ followed by a max-pooling layer with kernel size $3 \times 3 \times 3$ and stride 2. The following residual layers are grouped in so-called "bottleneck blocks" (see Fig. 2), which are constructed of three convolutional layers. The first and the last are convolutional layers, with kernel size $1 \times 1 \times 1$, where the first one downsamples the number of volume features, and the last one applies feature upsampling. Between these layers, there is a convolutional layer with kernel size $3 \times 3 \times 3$. The bottleneck blocks are arranged in four groups of sizes 3, 4, 6 and 3, where each group starts with a stride of 2 in the first convolutional layer to downsample the feature volumes' spatial dimensions. Finally, the residual blocks are terminated with a global average pooling [36] over the 2048 individual 3D feature volumes coming from the last layer of the ResNet encoder. Computing the average value of each feature map via global average pooling results in a 1D tensor with 2048 features.

Dilation ResNet-C-26 (DRN-C-26) encoder

The DRN-C-26 is a dilated residual CNN architecture with 26 layers introduced by [24]. The original ResNet downsamples the input images by a factor of 32. Downsampling our cropped and uneven sized image volumes by such an amount would result in a loss of information about small and salient parts caused by less expressive feature maps. However, simply reducing the convolutional stride restricts the receptive field of subsequent layers. For this reason, [24] presented an approach with which downsampling could be reduced while sustaining a sufficiently large receptive field and improving classification results. To construct the DRN-C-26 [24] applied the following changes to the ResNet18 [23] made of so-called ResNet "building blocks" with two convolutional layers with kernel size $3 \times 3 \times 3$ (see Figure 2). First, the convolutional stride in the last two groups is replaced by dilation. Second, the initial max-pooling layer is replaced by two residual building blocks. Lastly, to reduce aliasing artefacts, a decrease in dilation is added with two final building blocks without residual connections. Again, the residual blocks of the DRN-C-26 are followed by a global average pooling over the 512 feature maps of the last ResNet layer, resulting in a 1D tensor with 512 features.

MLP heads

The features obtained by the respective ResNet encoders are passed through a simple three-layered feed-forward network, also known as MLP, to achieve the respective classifications and regressions. As shown in Fig. 2, the MLP input dimension matches the feature dimensions of the CNN (i.e., 2048 neurons in case of ResNet50 and 512 neurons for a DRN-C-26). The hidden layers of all MLP's consist of 2048 neurons. The classifier head has six output nodes. In the *BB-loss* setting, an additional three-layered MLP with twelve output nodes was added to perform a bounding box

regression.

2.4 *Full-scale* approach: Detection of meniscal tears in complete MRI scans

In our first and most straightforward approach, the complete 3D MRI is provided as input to the CNN. The CNN consists of a ResNet50 encoder followed by an MLP head. The outputs of the MLP after a sigmoid activation represent probabilities for the six meniscal sub-regions to contain a tear or not.
The CNN is trained by minimizing the binary cross-entropy loss $\mathcal{L}_{BCE}$ for a given batch of N samples. With a target matrix $\mathbf{Y} \in \mathbb{Z}_2^{N \times C}$ and an output matrix $\hat{\mathbf{Y}} \in \mathbb{R}^{N \times C}$ for all C meniscal sub-region labels the definition of $\mathcal{L}_{BCE}$ is:

$$\mathcal{L}_{BCE} = \frac{1}{N} \sum_{c=1}^{C} \sum_{i=1}^{N} w_c [y_{i,c} log(\sigma(\hat{y}_{i,c})) + (1 - y_{i,c}) log(1 - \sigma(\hat{y}_{i,c}))], \tag{3}$$

where w_c is an inverse weighting of label frequencies and $\sigma(\cdot)$ is a sigmoid activation function. The *Full-scale* approach is visualized under A) in Fig. 2.

2.5 *BB-crop* approach: Detection of meniscal tears in cropped MRI datasets

Cropping 3D MRI data to the meniscal RoI is expected to provide two desirable properties. First, it provides smaller volumes reducing the required GPU memory as well as the run time. Second, the full-scale 3D MR images can be considered noisy as they provide additional and unnecessary information about surrounding anatomical structures. By cropping the data to the RoI of the menisci, this unnecessary information is suppressed. Leveraging the RoI generated as described in 2.2 the 3D MR images are cropped with a 5% margin around the menisci. Each cropped image is then resampled with trilinear interpolation to the closest multiples of 16, given the biggest bounding box in the training set. Figure 2 visualizes the cropping and resampling process. Consequently, the cropped and resampled images have a size of $(64, 64, 176)$ for the DESS data and $(16, 64, 176)$ for the IW TSE data. *BB-crop* utilizes a Dilation Resnet-C-26 encoder followed by an MLP classifier head. The CNN is trained by minimizing the $\mathcal{L}_{BCE}$ as given in (3). The framework is visualized under B) in Figure 2.

2.6 *BB-loss* approach: Detection of meniscal tears in complete MRI scans enhanced by regression of meniscal bounding boxes

The *BB-crop* approach requires segmentation of both menisci (or at least the determination of a meniscal region) in training and testing. Since generating segmentations is time-consuming (the method of [25] requires approximately 5 minutes of run time), it is beneficial to avoid this step. Moreover, this approach heavily relies on high-quality bounding boxes in training and inference, which are difficult to obtain and strongly influence the performance quality. Thus, the motivation for our final *BB-loss* approach is to detect meniscal tears in 3D MRI data without extensive pre-processing requirements such as segmenting the menisci or computing bounding boxes for meniscal regions. Instead, the location of the menisci is added as an additional loss term for the training. The encoder is kept identical to the *Full-scale* approach, namely a ResNet-50 encoder. Furthermore, an identical MLP head is utilized for the meniscal tear detection. Additionally, we show that the meniscal position information helps the CNN to focus on these regions in the image yielding better results. A second MLP head is employed in the *BB-loss* approach to regress the coordinates of the meniscal RoI. By incorporating this knowledge as a loss in the training process, the locations of the menisci must not be explicitly provided at test time. The total loss in the *BB-loss* setting is computed considering the multi-label classification and the bounding box regression task. For detection of meniscal tears $\mathcal{L}_{BCE}$ is employed (3). In the bounding box regression the outputs of the MLP head are 6 coordinates d for the MM and LM, respectively. Utilizing a sigmoid activation function, these values are given as relative positions within the image in a range of $[0, 1]$ of the respective dimension. For a detailed description of the bounding box generation procedure, we refer the reader to Section 2.2. The first component of the bounding box loss is an L1-term $\mathcal{L}_{L1}$ defined as

$$\mathcal{L}_{L1} = ||B - \hat{B}|| \,, \tag{4}$$

with a predicted bounding box $\hat{B}$ and a target bounding box B that is derived from the automated segmentation masks. These $N \times 2d$ matrices contain N rows with medial and lateral bounding box values. Where $b_{n,i}$ and $\hat{b}_{n,i}$ describe the nth element of the batch and the ith value of the concatenated bounding boxes. With this formulation the loss is given as

$$\mathcal{L}_{L1} = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{2d} |b_{n,i} - \hat{b}_{n,i}| \,. \tag{5}$$

The second component of the bounding box loss is a modified Intersection over Union (IoU) term, more specifically the Generalized-IoU (GIoU) $\mathcal{L}_{GIoU}$ [37] defined as:

$$\mathcal{L}_{GIoU} = 1 - IoU + \frac{|C \setminus (B \cup \hat{B})|}{|C|} \,, \tag{6}$$

where C is a convex hull enclosing the predicted and the target box. The operator $|\cdot|$ computes the box volume. The convex hull is the smallest possible region that encloses both the output and the target bounding boxes. It can be defined as a bounding box, fully characterised by the 6 coordinates elaborated above. It is computed by taking the minimum and maximum extent of both the target bounding box and the predicted bounding box coordinates along the x-, y- and z-axis. The numerator of the third term of the $\mathcal{L}_{GIoU}$ is the convex hull volume subtracted by the volume of B and $\hat{B}$, and the denominator is the volume of the convex hull. Hence, the third term of the $\mathcal{L}_{GIoU}$ can be considered as the relative volume of the convex hull not covered by the union of predicted and target bounding box. The IoU is defined as $\frac{|B \cap \hat{B}|}{|B \cup \hat{B}|}$, that is, the ratio of the intersecting voxels of B and $\hat{B}$ to their union. The $\mathcal{L}_{GIoU}$ is computed for each meniscal RoI and averaged for the given batch.The overall loss $\mathcal{L}$ for the *BB-loss* approach is given as

$$\mathcal{L} = \mathcal{L}_{BCE} + \mathcal{L}_{L1} + \mathcal{L}_{GIoU} \,. \tag{7}$$

The *BB-loss* approach is visualized under C) in Fig. 2.

2.7 Experimental setup and training of CNNs

The given MRI data of the OAI are randomly split into 50% training data, 15% validation data and 35% testing data. Hence, our two experiments have 1200/359/840 and 1197/359/840 training/validation/testing scans for the DESS data and the IW TSE data, respectively. We implemented the CNNs of all approaches in PyTorch 1.9. Convolutional weights are initialized using a normal distribution as in [38] tailored towards our deep neural networks with asymmetric ReLU activation functions. While, batch normalization weights and biases are initialized constant with 1 and 0. We train our CNNs on an Nvidia A100 GPU with 40 GB memory. Training our three ResNets, separate learning rates and dropout probabilities for the ResNet-encoders and the MLP-heads are introduced. Suitable learning rate, dropout and batch size hyper-parameters are found using the validation data of the DESS scans. The learning rate values for all parts (ResNet encoder, classifier head and bounding box head) are evaluated in an interval of $[1\text{e}{-5}, 0.01]$. Dropout percentages are varied in an interval of $[0.1, 0.9]$. Further, the training batch size limited by the input size is varied from 2 to 64 for the *BB-crop* approach. Due to a larger input volume in approach *Full-scale* and *BB-loss* batch size was kept constant at a value of 4. For a complete summary of our hyper-parameter values, please refer to Supplementary Table S2. Training is performed using the ADAM optimizer [39] with $\beta_1 = 0.9, \beta_2 = 0.999$ and $\epsilon = 1e-08$ with a learning rate decay of 0.5 every 50 epochs. Training on the IW TSE sequence is not performed from scratch, instead, both ResNet encoder and MLP weights are fine-tuned. In both DRN-C-26 and ResNet50 cases, we use the CNNs that achieve the lowest validation loss on the DESS sequence.

On-the-fly data augmentation is performed during training. Specifically, this means, random

cropping around the RoI, horizontal flips, rotations, Gaussian noise, and intensity scaling are applied with 50% probability. For the *Full-scale* approach, we perform random cropping of up to 10% along coronal, 20% sagittal and 20% axial direction. In the *BB-crop* approach, random crops are performed by uniformly cropping within a 20% margin around the menisci. The *BB-loss* approach uniformly samples possible crops around the menisci. All cropped images are resampled with trilinear interpolation to attain consistent sizes per approach and dataset. Input images for the *Full-scale* and *BB-loss* approach are sampled for the DESS sequence data to $(160, 384, 384)$ and for IW TSE images to $(44, 448, 448)$. The *BB-crop* approach resamples to $(64, 64, 176)$ and $(16, 64, 176)$, respectively. The added Gaussian noise is pixel-wise sampled as $\epsilon \in \mathcal{N}(0.1, 0.5)$. The random rotation is uniformly sampled from $\mathcal{U}(-5°, +5°)$ and image intensity is scaled by a uniformly sampled multiplication factor $b \in \mathcal{U}(0.9, 1.1)$.

2.8 Statistical assessment of detection quality

For all experiments, we plot the true positive rate (TPR = sensitivity) against the false positive rate (FPR = 1 - specificity) at various decision thresholds to create ROC curves [40]. Additionally, we compute the ROC AUC to assess the quality of our classifiers. The quality of our predicted bounding boxes is assessed by computing the IoU with the target bounding boxes. We consider IoU values over 0.5 as successful localization of the menisci since this is a common value in object detection tasks [41].

2.9 SmoothGrad saliency map visualizations for areas addressed by the CNN

Gradient saliency maps [42] (otherwise called pixel attribution maps or sensitivity maps) highlight pixel regions in the input image that mostly influenced a neural network's decision. To attain such pixel attributions, one computes the derivative of the final linear layer in a neural network with respect to the input via back-propagation. More formally, a gradient saliency map S_c for a sub-region c for which our neural network f yields a detection of meniscal tears is calculated as:

$$S_c(\widetilde{\mathcal{I}}_i) = \frac{\partial f_c(\widetilde{\mathcal{I}}_i)}{\partial \widetilde{\mathcal{I}}_i}. \tag{8}$$

For our two most promising approaches *BB-crop* and *BB-loss*, these maps are computed by applying a slight enhancement to the original mechanism - the SmoothGrad method [26]. Similar to the SmoothGrad approach of [26] we augmented the input image slightly, introducing noise, such that through averaging, the saliency maps of different noise levels are smoothed out. We apply Gaussian distributed noise $\epsilon \in \mathcal{N}(0.1, 0.5)$, random horizontal flips, uniformly sampled rotations $r \in \mathcal{U}(-5°, +5°)$ and uniformly sampled pixel intensity shift with a multiplication factor $b \in \mathcal{U}(0.9, 1.1)$. Each image is augmented 20 times with a probability of 50% per augmentation, and the resulting maps are averaged.

3 Results

We applied all approaches to DESS as well as IW TSE data from the OAI database. Each of our approaches detects meniscal tears for the MM and the LM. In particular, tears are detected in the three anatomical sub-regions anterior horn, meniscal body, and posterior horn. All results are presented in this section.

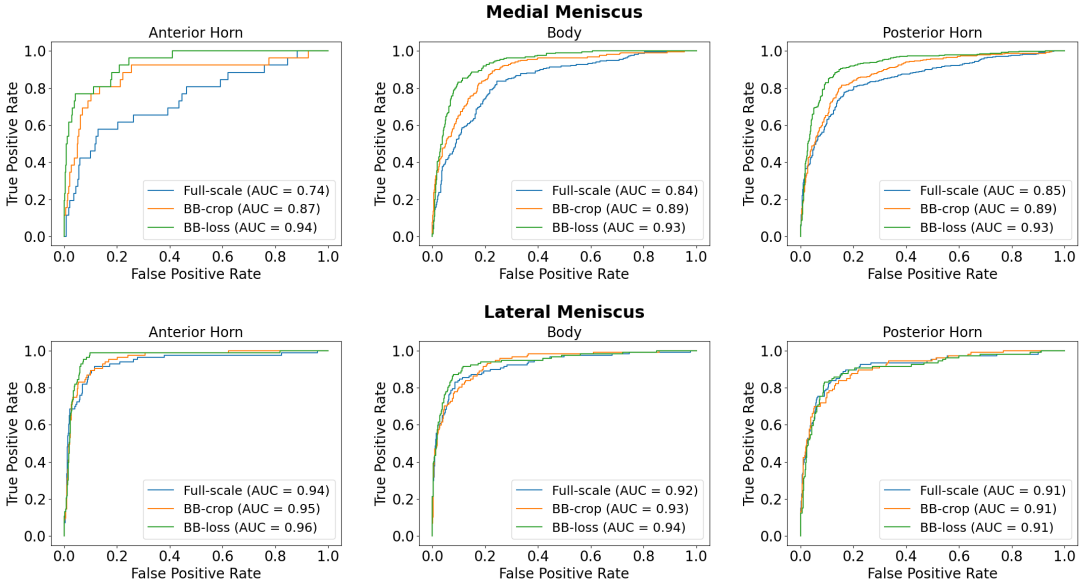

Figure 3: ROC curves for detection of meniscal tears in DESS MRI data.

3.1 Detection of meniscal tears in DESS MRI data

Employing the *Full-scale* approach, the AUC values are 0.74, 0.84, 0.85 for the anterior horn, body, and posterior horn of the MM. For the LM, the AUC values are 0.94, 0.92, 0.91. The *BB-crop* approach usually yields higher AUC values, being 0.87, 0.89, 0.89 and 0.95, 0.93, 0.91. The *BB-loss* gives the highest AUC values, being 0.94, 0.93, 0.93 and 0.96, 0.94, 0.91. The ROC curves employing all three approaches are shown in Fig. 3. In addition, all ROC AUC results are summarized in Table 2.

	MM			LM		
Method	Anterior	Body	Posterior	Anterior	Body	Posterior
Full-scale	0.74	0.84	0.85	0.94	0.92	0.91
BB-crop	0.87	0.89	0.89	0.95	0.93	**0.91**
BB-loss	**0.94**	**0.93**	**0.93**	**0.96**	**0.94**	**0.91**

Table 2: ROC AUC results for medial menisci (MM) and lateral menisci (LM) in DESS MRI data. The best results for each anatomical sub-region are highlighted.

3.2 Detection of meniscal tears in IW TSE MRI data

Employing the *Full-scale* approach, the AUC values are 0.82, 0.87, 0.82 for the anterior horn, body, and posterior horn of the MM. For the LM, the AUC values are 0.88, 0.85, 0.85. The *BB-crop* approach usually yields higher AUC values, being 0.84, 0.89, 0.86 and 0.92, 0.90, 0.90.

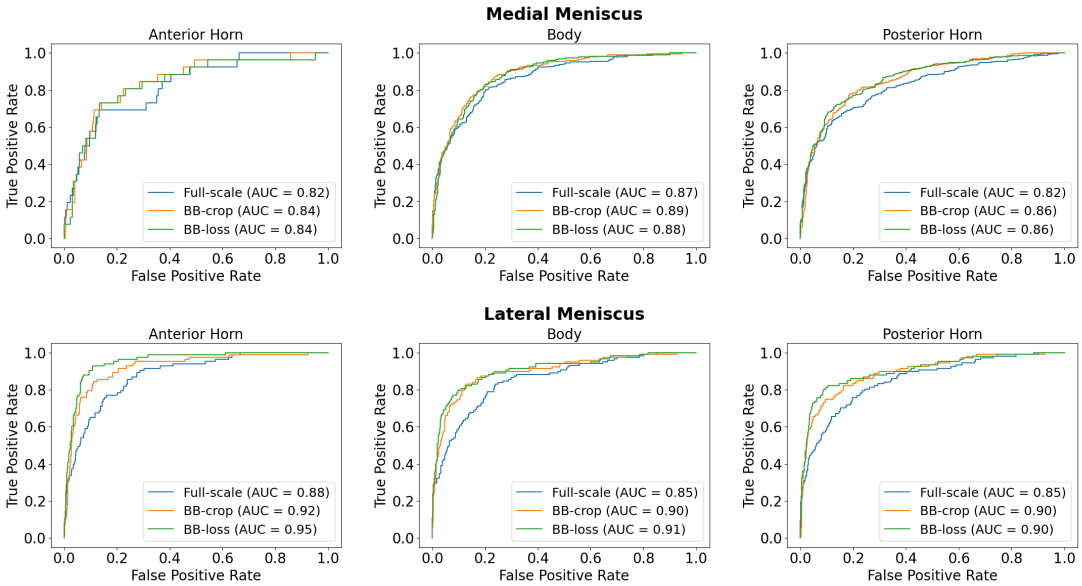

Figure 4: ROC curves for detection of meniscal tears in IW TSE MRI data.

The *BB-loss* gives similar AUC values, being 0.84, 0.88, 0.86 and 0.95, 0.91, 0.90. The ROC curves of all approaches are shown in Fig. 4. Further, all AUC values are summarized in Table 3.

	MM			LM		
Method	Anterior	Body	Posterior	Anterior	Body	Posterior
Full-scale	0.82	0.87	0.82	0.88	0.85	0.85
BB-crop	**0.84**	**0.89**	**0.86**	0.92	0.90	**0.90**
BB-loss	**0.84**	0.88	**0.86**	**0.95**	**0.91**	**0.90**

Table 3: ROC AUC results for medial menisci (MM) and lateral menisci (LM) in IW TSE data. The best results for each anatomical sub-region are highlighted.

3.3 Localization of menisci via the *BB-loss* approach

To investigate the bounding box regression quality of the proposed method we evaluate the distribution of the IoU values for the predicted bounding boxes (Fig. 5). For the DESS dataset (our primary benchmark), we observed a very high quality of MM and LM bounding box predictions. With the values being close to normally distributed around a mean value of 0.71 (95% confidence interval (CI): 0.71 - 0.72) and standard deviation of 0.13. With the IoU threshold of 0.5, we conclude that 95% of the resulted bounding boxes are identified correctly. Unfortunately, we observed a clear decrease in the object detection performance in the IW TSE dataset. With a mean value of 0.58 (95% CI: 0.57 - 0.59) and a standard deviation of 0.14. Applying the same detection threshold as above we testify, that only around 76% of menisci were detected correctly,

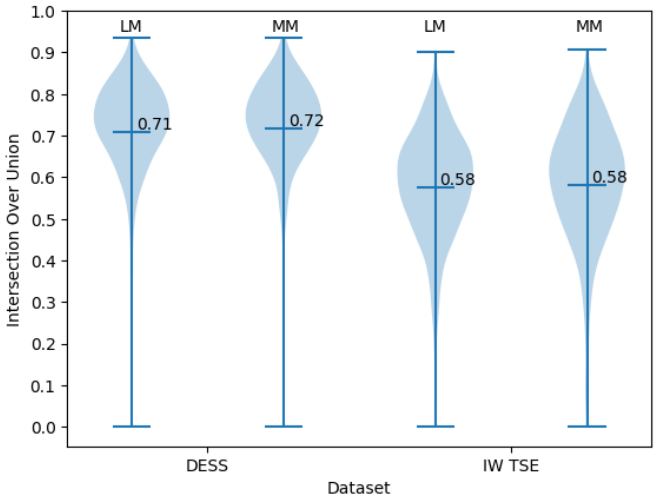

Figure 5: The distribution of the IoU values for the bounding boxes of MM and LM in DESS and IW TSE MRI data.

with the overall quality of the bounding boxes being more widely spread.

3.4 Visualization of areas addressed by the CNN

Figure 6 shows SmoothGrad saliency maps for the *BB-crop* and *BB-loss* approach overlaid to MR images. Examples are shown for randomly selected test cases, displaying different kinds of meniscal tears for DESS and IW TSE data. The RoIs for the *BB-loss* approach were extracted using predicted bounding boxes and the respective close-ups are shown. Red arrows point at the location of meniscal defects. Most saliency maps obtained this way display a plausible localization of the meniscal tears. The plausibility of these maps was qualitatively evaluated by their correspondence to the target labels of the regions in which the tears could also be confirmed with the help of visual inspection of the image data. SmoothGrad saliency maps are capable of highlighting more than just one affected sub-region, i.e. in the presence of defects in multiple sub-regions of one meniscus, one similarly observes these being correctly highlighted. With the Dilation ResNet-C-26 employed in the *BB-crop* approach, we observed that this CNN yields smoother and less noisy SmoothGrad saliency maps. However, in many cases, ResNet-50 saliency maps targeted the affected region better, but did not outline this region sharply.

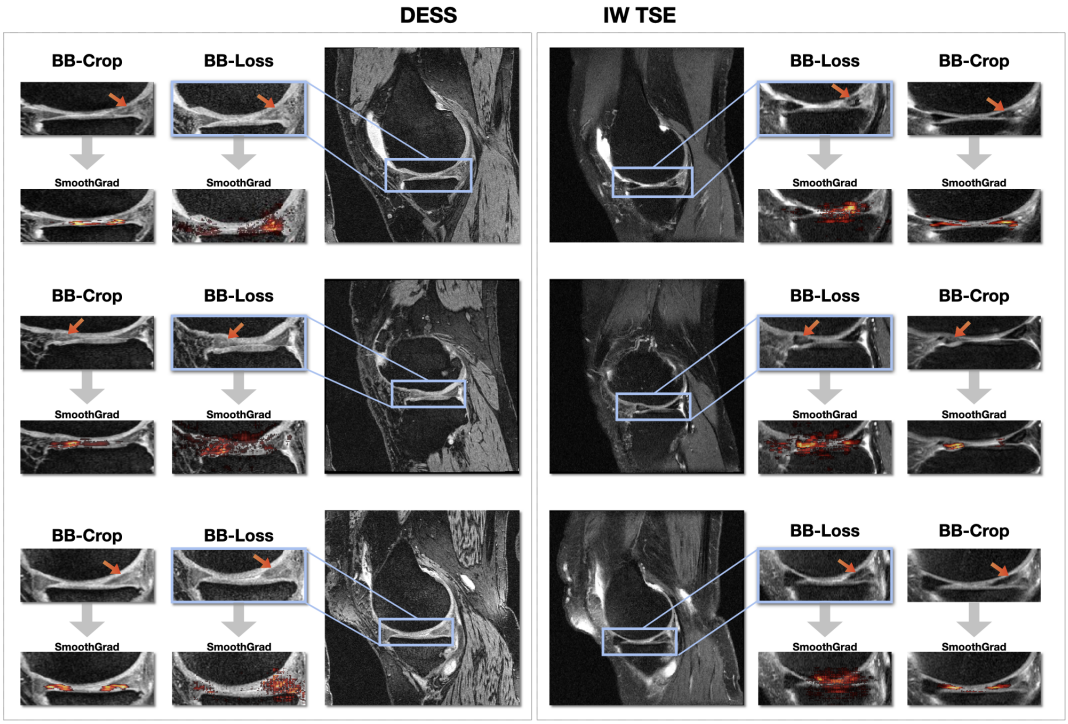

Figure 6: SmoothGrad saliency maps overlaid over DESS MRI data (left) and IW TSE MRI data (right).

	[14]*	[16]*	[12]	[11]	[18]	[19]	[20]	[21]	Ours: *Full-scale*	Ours: *BB-crop*	Ours: *BB-loss*
3D data	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
XAI	✗	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓
Anywhere	0.94	0.906	0.847	0.89	0.904 and 0.913	0.934	**0.961**	—	0.81	0.89	0.94
Any MM	—	—	—	—	—	—	0.882	0.93	0.79	0.89	**0.94**
Any LM	—	—	—	—	—	—	0.781	0.84	0.87	0.92	**0.93**
MM-AH	✓	✓	—	—	—	—	—	—	0.85	0.84	**0.94**
MM-B	—	—	—	—	—	—	—	—	0.82	0.89	**0.93**
MM-PH	✓	✓	—	—	—	—	—	—	0.78	0.89	**0.93**
LM-AH	✓	✓	—	—	—	—	—	—	0.90	0.95	**0.96**
LM-B	—	—	—	—	—	—	—	—	0.86	0.92	**0.94**
LM-PH	✓	✓	—	—	—	—	—	—	0.88	**0.91**	**0.91**

*[14] and [16] detected meniscal tears in 2D slices for AH and PH, but reported overall results only.

Table 4: Comparison of our results on DESS MRI data to the related work. The "3D data" column indicates whether the method is trained on and applied to complete 3D MR images. The explainable AI "XAI" column indicates if concepts of saliency maps are employed in order to highlight the areas responsible for the CNNs' decisions.

3.5 Detection performance – different sub-regions and defect types

Even though the occurrence of defects varies between meniscal sub-regions (see Supplementary Figure S2), we observe only minimal differences between AUC values of sub-regions in DESS MRI data (c.f. Table 2). However, we analyzed the false positive classifications and found that for all sub-regions, signal abnormalities were more often misclassified than normal menisci were (see Supplementary Figure S1). The misclassification rate of signal abnormalities is highest for the posterior horn of the lateral meniscus, the region with the least AUC for the DESS data. Conversely, the lowest signal abnormality misclassification rate is prevalent in the posterior horn of the medial meniscus, the sub-region with the highest number of signal abnormalities (supplementary Table S1).

The least common types of tears occurring in the data are radial and vertical tears, amounting to 72 and 69, respectively. Vertical tears were most challenging for our method to detect in DESS data and led to the most false negative results (see Supplementary Figure S2). Radial meniscal tears were the ones yielding the second highest rate of misclassifications.

4 Discussion

The primary goal of our work was to develop a method that provides an efficient, robust and automated way to detect and better locate meniscal tears in MRI data, that is, the detection of tears with respect to the anatomical regions in which they occur. We devised a procedure that utilizes a 3D CNN to process arbitrary 3D MRI data without the need for any extensive pre-processing.

Many previously proposed methods already yield a high accuracy in the detection of meniscal tears. To compare our results to the related work, we focus our assessment on the results of our *BB-loss* approach on the DESS MRI data. Our method detects meniscal tears in anatomical sub-regions of MM and LM. However, it has not been explicitly trained for menisci tear detection in the entire knee as well as the two menisci. Therefore, to obtain the respected values, we performed max operations on our CNNs' outputs. A comparison of the different approaches with their respective detection AUC is summarized in Table 4. Our *BB-loss* approach achieved state-of-the-art results in detecting meniscal tears in the medial and lateral meniscus with an AUC of 0.94 and 0.93. For the task of meniscal tear detection in the entire knee *BB-loss* approach had an AUC of 0.94 is second to the approach of [20]. However, the proposed methods from the related work still leave a desire for a more precise spatial assignment of the findings. For instance, localizing tears per meniscus or in anatomical sub-regions thereof. For tear detection per meniscus, our method performs better than related work [20, 21]. However, the novelty of our method is the detection of tears for each anatomical sub-region of the menisci in 3D MRI data, providing an anatomically more detailed localization.

With AUC values being consistently higher than 0.90 for DESS MRI data, our approach achieves excellent detection quality for all meniscal sub-regions using uncropped 3D MRI volumes. We also show that our method generalizes well to other MRI sequences, that is, from DESS to IW TSE data. IW TSE data provides a more challenging setting with a higher slice thickness in the mediolateral direction. Moreover, for certain meniscal defects, such as horizontal tears in the meniscal body, a lower resolution in the acquired MR image direction significantly reduces the visibility of the features required for an accurate classification. The result could be improved by using an input image with an isotropic resolution. Such an image can be obtained by either upsampling an existing image or, even better – acquiring a new image, at a higher resolution.

Signal abnormalities are still a challenge. In cases where menisci with tears are to be distinguished from menisci without tears, signal abnormalities are currently regarded as the latter. A fine-grained differentiation between tears and signal abnormalities is likewise a challenge to our method, primarily through the ambiguous image appearance. Potentially, more training data, as present for the region with the most signal abnormalities – the MM posterior horn, would allow our CNN to better learn to distinguish signal abnormalities from tears.

We expected our model to generalize to all meniscal pathologies but observed problems detecting vertical and radial tears. However, these tears were less common in the available training data, and we believe that more data on such cases would enable our method to detect vertical and radial tears with higher accuracy. Furthermore, coronal and axial imaging sequence orientation could provide additional insights [12], possibly improving the detection of otherwise barely visible tears.

One major limitation that we see is that our method still requires a localization of the menisci in training. However, other segmentation approaches or (non-automatic) approaches could be applied to attain bounding boxes, possibly improving results by providing more accurate bounding

boxes for training.

5 Conclusions and future work

We present a method in an efficient and fully automated multi-task learning setting that accurately detects meniscal tears on a sub-region level in MM and LM. Our method yields the best results on sagittal DESS MRI data and generalizes well to sagittal IW TSE data. Further, visual support for clinical detection of meniscal tears is provided by SmoothGrad saliency maps highlighting regions that mainly contributed to the decision.
Future work could comprise an analysis of anomaly detection (normal vs. signal abnormality vs. torn menisci) or a classification of different types of tears (horizontal, radial, complex, etc.). Since some of these types occur only rarely for specific sub-regions, deep learning-based methods probably require a lot more image data or data generated with generative models. Also, new issues of class imbalances will arise for the classification of tear types.
From the method perspective, the choice of an encoder provides opportunities for improvement. For instance, recent self-attention mechanisms, so-called "transformer" architectures [43, 44] are worth an investigation. Since transformers typically require a vast amount of training data, they might not necessarily lead to better accuracy, but the self-attention maps [45] may result in a more meaningful explanatory power than classical methods of saliency mapping. Also, generative adversarial networks have been recently employed for explaining the decision of CNN's [46, 47]. As deep learning methods become more precise in localizing meniscal tears coupled with further sophisticated concepts on explainability, CAD tools will become practical for clinical decision support. In future work, we plan to investigate whether our method better assists physicians in their diagnostic tasks.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

AT, AS, DL and SZ designed the study. AT, AS and DL implemented the proposed methods. AT, AS and DL collected the data, performed the statistical evaluation and executed the experiments. SZ obtained the funding resources for this project. AT, AS, DL and SZ drafted and wrote the manuscript.

Funding

The authors gratefully acknowledge the financial support by the German federal ministry of education and research (BMBF) within the research network on musculoskeletal diseases, grant no. 01EC1408B (Overload/PrevOP). Furthermore, the authors are funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within research project ZA 592/4-1.

Acknowledgments

The Osteoarthritis Initiative is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health. This manuscript was prepared using an OAI public use data set and does not necessarily reflect the opinions or views of the OAI investigators, the NIH, or the private funding partners.

Supplemental Data

Data Availability Statement

The datasets analyzed for this study as well as the medical image annotations can be found in the NIH OAI archive `https://nda.nih.gov/oai/`. The employed segmentation masks of all MM and LM will be made publicly available at `https://pubdata.zib.de` upon publication of this paper.

References

[1] Alexander R Markes, Jonathan D Hodax, and Chunbong Benjamin Ma. Meniscus form and function. *Clinics in sports medicine*, 39(1):1–12, 2020.

[2] Timothy Bhattacharyya, Daniel Gale, Peter Dewire, Saara Totterman, M Elon Gale, Sara McLaughlin, Thomas A Einhorn, and David T Felson. The clinical importance of meniscal tears demonstrated by magnetic resonance imaging in osteoarthritis of the knee. *JBJS*, 85(1):4–9, 2003.

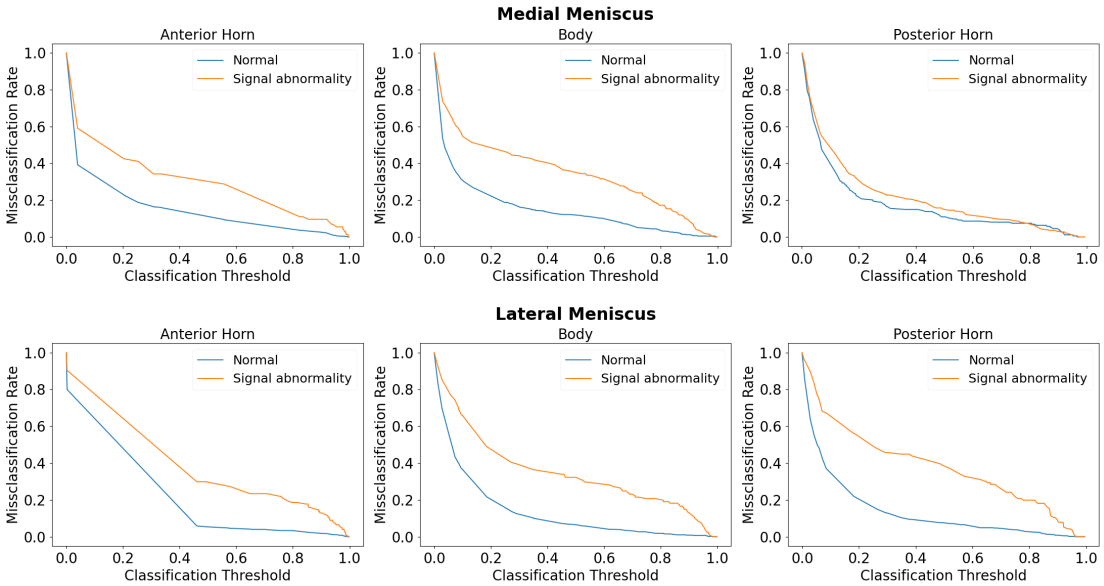

Figure S1: Visualization of relative misclassification rates for signal abnormalities and normal menisci with the *BB-loss* model. For a given classification threshold, signal abnormalities are considerably more likely to be classified as false positives than normal menisci.

[3] P Beaufils and N Pujol. Management of traumatic meniscal tear and degenerative meniscal lesions. save the meniscus. *Orthopaedics & Traumatology: Surgery & Research*, 103(8):S237–S244, 2017.

[4] Changhai Ding, Johanne Martel-Pelletier, Jean-Pierre Pelletier, François Abram, Jean-Pierre Raynauld, Flavia Cicuttini, and Graeme Jones. Meniscal tear as an osteoarthritis risk factor in a largely non-osteoarthritic cohort: a cross-sectional study. *The Journal of rheumatology*, 34(4):776–784, 2007.

[5] BAM Snoeker, M Ishijima, J Kumm, F Zhang, AT Turkiewicz, and M Englund. Are structural abnormalities on knee mri associated with osteophyte development? data from the osteoarthritis initiative. *Osteoarthritis and Cartilage*, 2021.

[6] Ruth Crawford, Gayle Walley, Stephen Bridgman, and Nicola Maffulli. Magnetic resonance imaging versus arthroscopy in the diagnosis of knee pathology, concentrating on meniscal lesions and acl tears: a systematic review. *British medical bulletin*, 84(1):5–23, 2007.

[7] M Englund, Ewa M Roos, HP Roos, and LS Lohmander. Patient-relevant outcomes fourteen years after meniscectomy: influence of type of meniscal tear and size of resection. *Rheumatology*, 40(6):631–639, 2001.

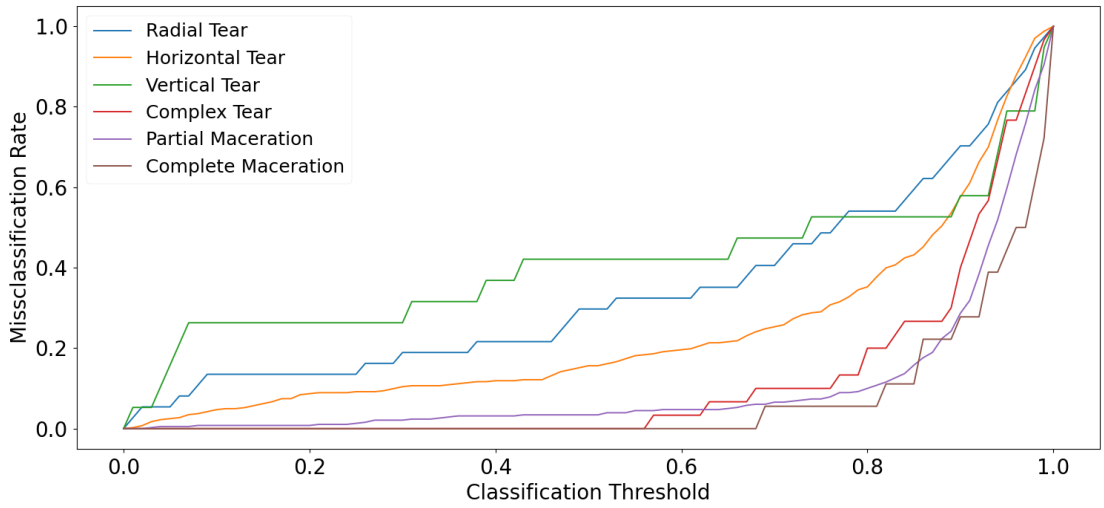

Figure S2: Visualization of relative misclassification rates for meniscal tears with the *BB-loss* model. For a given classification threshold, radial and vertical tears result in the most false negatives, while complex tears and partial and complete maceration show little false negatives. The most prominent tear type of horizontal tears lies in between the two groups.

[8] Moin Khan, Nathan Evaniew, Asheesh Bedi, Olufemi R Ayeni, and Mohit Bhandari. Arthroscopic surgery for degenerative tears of the meniscus: a systematic review and meta-analysis. *Cmaj*, 186(14):1057–1064, 2014.

[9] Nina Jullum Kise, May Arna Risberg, Silje Stensrud, Jonas Ranstam, Lars Engebretsen, and Ewa M Roos. Exercise therapy versus arthroscopic partial meniscectomy for degenerative meniscal tear in middle aged patients: randomised controlled trial with two year follow-up. *bmj*, 354, 2016.

[10] Frank W Roemer, C Kent Kwoh, Michael J Hannon, David J Hunter, Felix Eckstein, Jason Grago, Robert M Boudreau, Martin Englund, and Ali Guermazi. Partial meniscectomy is associated with increased risk of incident radiographic osteoarthritis and worsening cartilage damage in the following year. *European radiology*, 27(1):404–413, 2017.

[11] Valentina Pedoia, Berk Norman, Sarah N Mehany, Matthew D Bucknor, Thomas M Link, and Sharmila Majumdar. 3d convolutional neural networks for detection and severity staging of meniscus and pfj cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects. *Journal of Magnetic Resonance Imaging*, 49(2):400–410, 2019.

[12] Nicholas Bien, Pranav Rajpurkar, Robyn L Ball, Jeremy Irvin, Allison Park, Erik Jones, Michael Bereket, Bhavik N Patel, Kristen W Yeom, Katie Shpanskaya, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of mrnet. *PLoS medicine*, 15(11):e1002699, 2018.

Meniscal morphology per sub-region (DESS)

Type	MM-AH	MM-B	MM-PH	LM-AH	LM-B	LM-PH
Normal	2,109	1,241	545	1,819	1,721	1,780
Signal Abnormality	209	576	971	301	317	327
Radial Tear	2	5	33	6	8	18
Horizontal Tear	17	206	476	121	207	143
Vertical Tear	1	11	30	6	11	10
Complex Tear	0	13	44	4	9	11
Partial Maceration	54	345	294	105	123	104
Complete Maceration	5	2	4	34	3	2
NaN	2	0	2	3	0	4

Meniscal morphology per sub-region (IW TSE)

Type	MM-AH	MM-B	MM-PH	LM-AH	LM-B	LM-PH
Normal	2,107	1,238	544	1,818	1,718	1,778
Signal Abnormality	209	576	971	300	317	327
Radial Tear	2	5	33	6	8	18
Horizontal Tear	17	206	475	121	207	143
Vertical Tear	1	11	30	6	11	10
Complex Tear	0	13	44	4	9	11
Partial Maceration	54	345	294	105	123	104
Complete Maceration	5	2	4	34	3	2
NaN	1	0	1	2	0	3

Table S1: Meniscal tear type frequencies across meniscal sub-regions for the DESS and IW TSE sequence.

[13] Kyle N Kunze, David M Rossi, Gregory M White, Aditya V Karhade, Jie Deng, Brady T Williams, and Jorge Chahla. Diagnostic performance of artificial intelligence for detection of anterior cruciate ligament and meniscus tears: A systematic review. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, 2020.

[14] V Roblot, Y Giret, M Bou Antoun, C Morillot, X Chassin, A Cotten, J Zerbib, and L Fournier. Artificial intelligence to diagnose meniscus tears on mri. *Diagnostic and interventional imaging*, 100(4):243–249, 2019.

[15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99, 2015.

[16] Vincent Couteaux, Salim Si-Mohamed, Olivier Nempont, Thierry Lefevre, Alexandre Popoff, Guillaume Pizaine, Nicolas Villain, Isabelle Bloch, Anne Cotten, and Loïc Boussel. Automatic knee meniscus tear detection and orientation classification with mask-rcnn. *Diagnostic and interventional imaging*, 100(4):235–242, 2019.

Method	LR-ENC	LR-MLP	DO-ENC	DO-MLP	BS
Full-scale	5.0e-4	5.0e-5	50%	10%	4
BB-crop	4.4e-4	6.1e-5	76%	19%	13
BB-loss	5.0e-4	5.0e-5	50%	10%	4

Table S2: Hyper-parameters of our methods. This table shows the learning rate of the respective encoder (LR-ENC) and MLP head (LR-MLP). Also, the dropout probabilities of the respective encoder (DO-ENC) and MLP head (DO-MLP) are provided. Finally, the batch size (BS) used for training the method is given. The same set of hyper-parameters was used for the DESS as well as IW TSE dataset.

[17] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[18] Chen-Han Tsai, Nahum Kiryati, Eli Konen, Iris Eshed, and Arnaldo Mayer. Knee injury detection using mri with efficiently-layered network (elnet). In *Medical Imaging with Deep Learning*, pages 784–794. PMLR, 2020.

[19] David Azcona, Kevin McGuinness, and Alan F Smeaton. A comparative study of existing and new deep learning methods for detecting knee injuries using the mrnet dataset. In *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, pages 149–155. IEEE, 2020.

[20] Benjamin Fritz, Giuseppe Marbach, Francesco Civardi, Sandro F Fucentese, and Christian WA Pfirrmann. Deep convolutional neural network-based detection of meniscus tears: comparison with radiologists and surgery as standard of reference. *Skeletal radiology*, 49(8):1207–1217, 2020.

[21] B Rizk, H Brat, P Zille, R Guillin, C Pouchy, C Adam, R Ardon, and G d'Assignies. Meniscal lesion detection and characterization in adult knee mri: A deep learning model approach with external validation. *Physica Medica*, 83:64–71, 2021.

[22] Muhammed Masudur Rahman, Lutz Dürselen, and Andreas Martin Seitz. Automatic segmentation of knee menisci–a systematic review. *Artificial Intelligence in Medicine*, 105:101849, 2020.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[24] Fisher Yu, Vladlen Koltun, and Thomas Funkhouser. Dilated residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 472–480, 2017.

[25] Alexander Tack, Anirban Mukhopadhyay, and Stefan Zachow. Knee menisci segmentation using convolutional neural networks: data from the osteoarthritis initiative. *Osteoarthritis and cartilage*, 26(5):680–688, 2018.

[26] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825*, 2017.

[27] CG Peterfy, G Gold, F Eckstein, Flavia Cicuttini, B Dardzinski, and R Stevens. Mri protocols for whole-organ assessment of the knee in osteoarthritis. *Osteoarthritis and cartilage*, 14:95–111, 2006.

[28] David J Hunter, Ali Guermazi, Grace H Lo, Andrew J Grainger, Philip G Conaghan, Robert M Boudreau, and Frank W Roemer. Evolution of semi-quantitative whole joint assessment of knee oa: Moaks (mri osteoarthritis knee score). *Osteoarthritis and cartilage*, 19(8):990–1002, 2011.

[29] Gutha Vaishnavi Reddy. Automatic Classification of 3D MRI data using Deep Convolutional Neural Networks. Master's thesis, Otto-von-Guericke-Universität Magdeburg, Germany, 2017.

[30] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

[31] Hidenori Ide and Takio Kurita. Improvement of learning for cnn with relu activation by sparse regularization. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 2684–2691. IEEE, 2017.

[32] Felix Ambellan, Alexander Tack, Moritz Ehlke, and Stefan Zachow. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative. *Medical image analysis*, 52:109–118, 2019.

[33] Alexander Tack and Stefan Zachow. Accurate automated volumetry of cartilage of the knee using convolutional neural networks: data from the osteoarthritis initiative. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 40–43. IEEE, 2019.

[34] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[35] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

[36] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

[37] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 658–666, 2019.

[38] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[39] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[40] Christopher D Brown and Herbert T Davis. Receiver operating characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, 80(1):24–38, 2006.

[41] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[42] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.

[43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[44] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[45] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. *arXiv preprint arXiv:2104.14294*, 2021.

[46] Alexander Katzmann, Oliver Taubmann, Stephen Ahmad, Alexander Mühlberg, Michael Sühling, and Horst-Michael Groß. Explaining clinical decision support systems in medical imaging using cycle-consistent activation maximization. *Neurocomputing*, 458:141–156, 2021.

[47] Sheng-Min Shih, Pin-Ju Tien, and Zohar Karnin. Ganmex: One-vs-one attributions using gan-based model explainability. In *International Conference on Machine Learning*, pages 9592–9602. PMLR, 2021.