

Automated analysis of rabbit knee calcified cartilage morphology using micro-computed tomography and deep learning segmentation

Santeri J. O. Rytty^{1,*}, Lingwei Huang², Petri Tanska², Aleksei Tiulpin^{1,3,4}, Egor Panfilov¹, Walter Herzog⁵, Rami K. Korhonen², Simo Saarakkala^{1,3}, Mikko A. J. Finnilä^{1,2,3}

¹Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Oulu, Finland;

²Department of Applied Physics, University of Eastern Finland, Kuopio, Finland;

³Department of Diagnostic Radiology, Oulu University Hospital, Oulu, Finland

⁴Ailean Technologies Oy, Oulu, Finland

⁵Human performance laboratory, Faculty of Kinesiology, University of Calgary, AB, Calgary, Canada

*Corresponding author:

Santeri J. O. Rytty,

Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Oulu, Finland,

POB 5000, FI-90014 Oulu, Finland

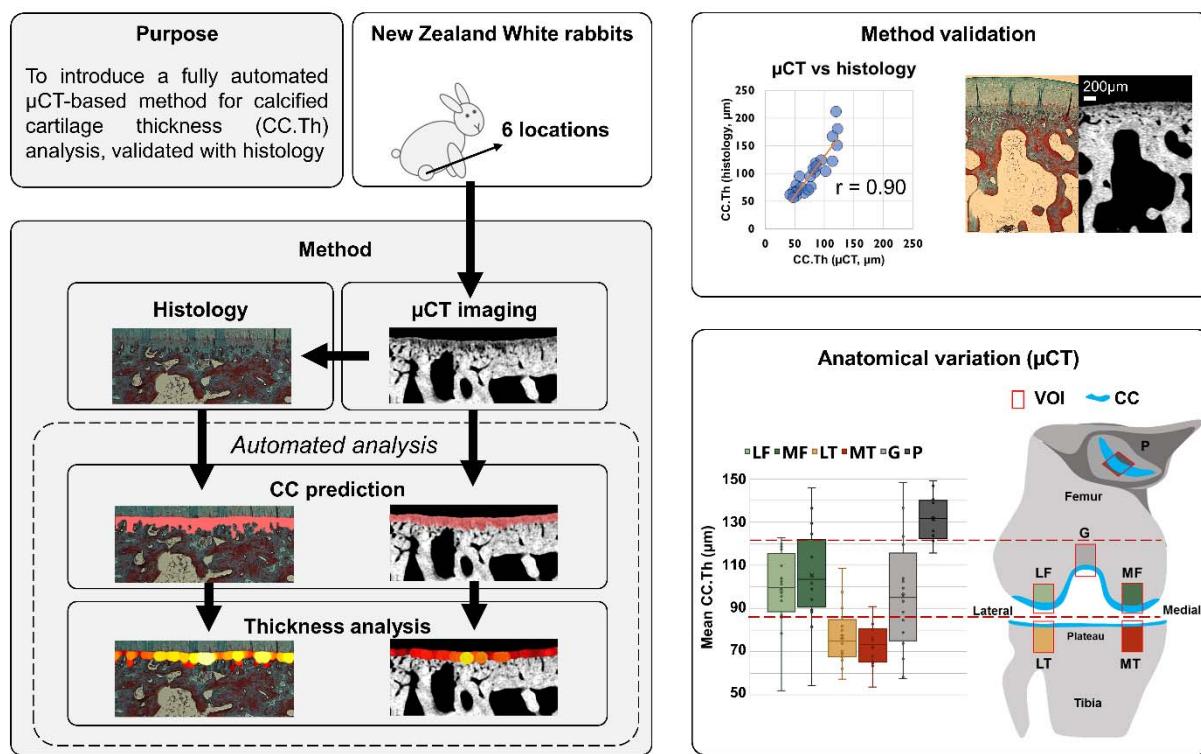
Email: santeri.rytty@oulu.fi

Tel: +358 45 672 6367

Running title: μ CT of rabbit calcified cartilage

Disclosures

PT received funding from the Finnish Cultural Foundation (Central Fund No. 191044) and Maire Lisko Foundation. AT is a CTO and a shareholder of Ailean Technologies Oy. WH was supported by The Canadian Institutes of Health Research, The Canada Research Chair Programme, and Killam Foundation. RK was supported by the Academy of Finland (no. 324529) and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 713645. SS was supported by the Academy of Finland (no. 303786) and the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement No. 336267. MAJF received funding from the Finnish Cultural Foundation (North Ostrobothnia Regional Fund No. 60172246).



Graphical abstract: We present a μCT-based method with deep learning segmentation for analyzing calcified cartilage thickness (CC.Th). The method is compared throughout the study against conventional histology. The comparison against co-registered regions yielded a strong Pearson correlation ($r = 0.90$). Both methods were able to separate the CC.Th properties between tibia, femur, and patella.

Abstract

Purpose: Only little is known how calcified cartilage (CC) structure changes during exercise, aging and disease. CC thickness (CC.Th) can be analyzed using conventional histological sections. Micro-computed tomography (μ CT) allows for three-dimensional (3D) imaging of mineralized tissues, however, the segmentation between bone and CC is challenging. Here, we present state-of-the-art deep learning segmentation for μ CT images to enable assessment of CC morphology.

Methods: Sixteen knees from twelve New Zealand White rabbits were dissected into osteochondral samples from six anatomical regions: lateral and medial femoral condyles, lateral and medial tibial plateaus, femoral groove and patella ($n = 96$). Samples were imaged with μ CT and processed for conventional histology. Manually segmented CC from the histology and reconstructed μ CT images was used as the gold standard to train segmentation models with different encoder-decoder architectures. The models with the greatest out-of-fold evaluation Dice score were used for automated CC.Th analysis. Subsequently, the automated CC.Th analysis was compared across a total of 24 regions, co-registered between the imaging modalities, using Pearson correlation and Bland-Altman analyses. Finally, the anatomical variation in CC.Th was assessed via a Linear Mixed Model analysis.

Results: The best segmentation models yielded average Dice scores of 0.891 and 0.807 for histology and μ CT segmentation, respectively. The correlation between the co-registered regions across the modalities was strong ($r = 0.897$). The Bland-Altman analysis yielded a bias of 21.9 μ m and a standard deviation of 21.5 μ m between the methods. Finally, both methods could separate the CC morphology between the patella, femoral, and tibial regions ($p < 0.001$).

Conclusion: The presented method allows for *ex vivo* 3D assessment of CC.Th in an automated and non-destructive manner. We demonstrated its utility by quantifying CC.Th in different anatomical regions. CC.Th was the thickest in the patella and the thinnest in the tibial plateau.

Keywords: Bone μ CT; Bone histomorphometry; Animal models, Osteoarthritis

Introduction

Calcified cartilage (CC) is a mineralized tissue delineated from the non-calcified articular cartilage by the tidemark, and from the subchondral bone by the cement line⁽¹⁾. The CC has an important role in anchoring the articular cartilage to the subchondral bone via individual collagen fibrils⁽²⁾. For healthy conditions, the relative CC thickness (CC.Th) to the total cartilage is nearly constant, but the CC volume relative to the total cartilage volume varies and has been shown to range from 3.23% to 8.8%⁽³⁾. Blood vessels from the subchondral bone extend into the CC layer, providing nutrients to the local chondrocytes⁽¹⁾. Furthermore, based on the current literature, CC is a dynamic tissue undergoing changes with mechanical loading, aging and joint pathology, *e.g.* osteoarthritis⁽⁴⁾.

The thickness of articular cartilage^(5,6) and subchondral bone⁽⁷⁾ varies greatly in different areas of the knee joint with a high thickness in heavily loaded areas. It can be hypothesized that similar changes are present in the CC as well. Early study on CC.Th revealed regional differences within the human femoral head⁽³⁾. Furthermore, clear regional differences in equine CC have been reported^(8,9). In contrast, in canine knees only minor regional differences have been found⁽⁵⁾. These differences related to anatomical location could be linked to the local loading environment.

In general, exercise and loading are thought to affect the CC structure. The intensity of exercise on heavily loaded joint regions is associated with thicker CC in equine tarsi⁽¹⁰⁾ and carpus, even without changes in the overlying non-calcified cartilage⁽¹¹⁾. An increase in the canine CC.Th was observed with high-intensity exercise⁽¹²⁾. In contrast, unloading of knees with immobilization resulted in thinner CC in canine knees⁽¹³⁾. In the human knee joint, similar findings have been reported; both articular and

calcified cartilage are thick in load-bearing areas and thin under the menisci of the knee⁽¹⁴⁾.

Two competing events occur in aging CC: calcification of the deep articular cartilage via advancement of the tidemark⁽¹⁵⁾ and endochondral ossification (bone replacing CC at the cement line)⁽¹⁶⁾. The latter is likely dominant since aging accelerates the thinning of CC and increases the number of tidemarks^(16,17). Although CC.Th varies across humans and different animal species⁽¹⁸⁾, similar changes in aging CC have been found in animal models. Thinning of CC, increases in vessel invasion⁽¹⁹⁾, as well as chondrocyte apoptosis⁽²⁰⁾ have been reported in murine CC with aging. On the other hand, *Murray et al.* reported an age-related increase in CC.Th in the equine tarsometatarsal joint⁽²¹⁾. Joint pathology can also induce tissue responses in the CC. Remodeling of CC^(16,17) occurs during OA progression, contributing to a decrease in articular cartilage thickness⁽²²⁾. Microfractures in the CC, subchondral bone plate, and the trabeculae, lead to the formation of cysts and channels, thereby affecting the cross-talk between articular cartilage and subchondral bone⁽¹⁾.

Traditionally, CC imaging has been performed on images obtained from histological sections⁽³⁾ as well as backscattered scanning electron microscopy (SEM) in equine⁽¹⁶⁾ and human joints^(23,24). Both histology and SEM require extensive and time-consuming sample processing protocols, and allow for two-dimensional (2D) imaging only. Nowadays, three-dimensional (3D) volumetric reconstruction of histological⁽²⁵⁾ and SEM images⁽²⁶⁾ is possible with serial sectioning and imaging, but the associated processing is laborious and has the potential to introduce errors.

Micro-computed tomography (μ CT) has been widely used to characterize 3D morphology in micron

level, including CC^(27,28). In contrast to histology and SEM, only minimal sample processing is required in μ CT. We showed previously that μ CT images of the human subchondral plate contain both the mineralized CC and the subchondral bone⁽²⁹⁾. Indeed, CC cannot be separated from bone with low-resolution μ CT imaging but becomes visible only in high-resolution μ CT images⁽³⁰⁾. However, because of the very minor difference in mineralization between the subchondral bone and CC, it is challenging to delineate the interface between CC and subchondral bone also in high-resolution μ CT imaging.

The identification of the tidemark and cement line from μ CT images is often conducted manually by researchers. This is a subjective and highly time-consuming endeavor, especially for tissues with complex shapes. Deep convolutional neural networks (CNNs) have recently shown great promise for automating various segmentation problems. U-Net⁽³¹⁾ has been the most popular segmentation architecture for biomedical images in recent years, and it has also been applied to μ CT data⁽³²⁾. However, the newly introduced Feature Pyramid Networks (FPN) allow for capturing both low-resolution global features as well as high-resolution local features at a low computational cost⁽³³⁾. Conventional training of CNNs is conducted by initializing the coefficients from a random distribution. An alternative training approach is transfer learning, in which the network is initialized from an existing model, often pre-trained on ImageNet dataset^(34,35). Notably, such approach works efficiently across domains beyond natural images^(36,37). For example, transfer learning from deep residual networks⁽³⁸⁾ has been used to classify pulmonary nodules from CT images⁽³⁹⁾, or segment the lungs in chest X-rays⁽⁴⁰⁾.

In this study, we propose an accurate framework for automated μ CT-based evaluation of the CC.Th in 3D. This requires introducing state-of-the-art deep learning architectures for CC segmentation. To

demonstrate the validity of the method, we perform direct comparison of CC.Th between μ CT and conventional histology. We utilized osteochondral samples of New Zealand White rabbits, a frequently used animal model for various musculoskeletal diseases. Furthermore, we hypothesize that the CC.Th varies in different anatomical locations of the knee. We demonstrate the capability of our automatic framework by assessing differences in CC.Th between the different anatomical locations.

Materials and Methods

Sample collection

Sixteen knees were collected from twelve healthy, skeletally mature female New Zealand White rabbits (strain 052 CR). Eight knees were collected from four rabbits (age: 14 months) and eight knees from eight rabbits (age: 12.5 months). Each knee was dissected and divided into six anatomical regions: lateral and medial femoral condyle, lateral and medial tibial plateau, femoral groove and patella (n = 96, Table 1). Details on animal housing, husbandry conditions and diet are detailed in a previous study⁽⁴¹⁾. All experiments were carried out under the guidelines of the Canadian Council on Animal Care and were approved by the committee on Animal Ethics at the University of Calgary (Renewal 3 for ACC Study #AC11-0035).

Table 1. Descriptive statistics of the rabbits used in the study. On the right, the number of images and samples (separated by / mark) segmented manually is described. These segmentations are used as training data for the deep learning models.

# animals	# knees	# Samples	# Histology slices	Manual segmentations	
				Histology	μ CT
12	16	96	3 / sample	253 / 87	1050 / 60

Imaging

The dissected osteochondral samples were formalin-fixed. Prior to imaging, samples were wrapped in moist paper, and placed in plastic vials (Cryo.sTM) for positional stability. The samples were subsequently imaged using a desktop μ CT scanner (Skyscan 1272, Bruker microCT, Kontich, Belgium) with a tube voltage of 50 kV, current of 200 μ A, and a 0.5 mm aluminum filter. The scanning was conducted in a step of 0.2° over 360° and finally, 1800 projection images with an isotropic pixel size of 3.2 μ m were obtained.

The images were reconstructed using the manufacturer's software (NRecon, version 1.7.0.4). A narrow window with attenuation coefficients 0.085-0.141 was used to provide high contrast between the bone and CC. The volumes-of-interest (VOI) of all samples were selected from the central load-bearing area (VOI size = 2 mm × 2 mm × sample height). This selection reduced the μ CT image stacks to a reasonable size (from ~12 GB to ~700 MB per sample) for the subsequent analysis. See Supplementary Figure 1 for examples of the preprocessing steps.

After the μ CT imaging, samples were prepared for histological analysis. Samples were decalcified using a standard protocol (with EDTA solution), paraffin-embedded, and cut into 5- μ m-thick sections using a microtome (three sections from each region). The sections were stained with Masson-Goldner's trichrome for identification of the CC layer and imaged with a light microscope (Axioimager 2; Carl Zeiss MicroImaging GmbH, Jena, Germany; control software = AxioVision; resolution = 2.56 μ m). A total of 281 sections were used in this study.

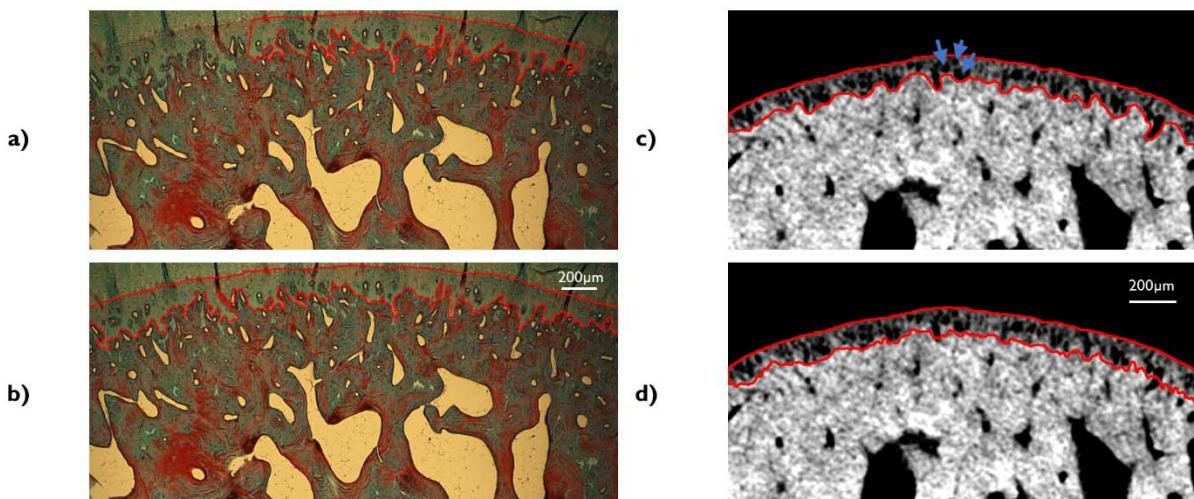


Figure 1. A histological section from the rabbit femoral condyle segmented manually (a) and automatically with the neural network (b). μ CT image from a similar anatomical region (different animal) with manual (c) and automatic (d) segmentation. Scale bar for 200 μ m is shown in the corresponding images. The blue arrows refer to chondrocytes inside the CC layer.

Training CC segmentation models

For the histology images, 253 continuous CC structures were segmented manually from 87 samples based on the distinct collagen staining in CC against articular cartilage and subchondral bone (Figure 1a). For the μ CT, manual annotation was conducted for 60 samples from 10 knees according to two criteria: 1) CC region with a distinct grayscale gradient and 2) presence of chondrocytes inside the CC layer (Figure 1c, blue arrows). Annotations were done for 10-30 slices per sample, evenly spaced within each volume. The manual annotations were used as the gold standard for the automated segmentation algorithms and for conducting a reference analysis for the CC morphology.

The fully automatic CC segmentation was conducted using a deep learning pipeline inspired by *Solovyev et al.*⁽⁴⁰⁾ on Python 3.7. The pipeline was built using in-house developed *Collagen*-framework (<https://github.com/MIPT-Oulu/Collagen>). For the histology segmentation, we used ResNet-34⁽³⁸⁾ pre-trained on ImageNet⁽³⁴⁾. We used a U-Net decoder with batch normalization in this model. The network was trained for 100 epochs under 4-fold cross-validation, splitting the training and validation folds with respect to rabbit ID. For the μ CT segmentation, we used ResNet-18 as our base model, and also an FPN decoder, which had instance normalization as well as the spatial dropout. Briefly, the normalization reduces bias for individual features with large values, while dropout reduces model overfitting by zeroing random nodes of the network. This model was also trained in a 4-fold cross-validation but for 60 epochs due to faster convergence.

We used a combination of binary cross-entropy and soft Jaccard index as the optimization loss function. Binary cross-entropy is one of the most popular segmentation metrics and can result in stable convergence. However, Jaccard index can account for class imbalance, such as imbalance between the

CC and the surrounding tissue. To facilitate a robust segmentation model, we used several image augmentation techniques (Supplementary Table 1) from the SOLT library⁽⁴²⁾ to diversify the training data. To assess the final segmentation performance, we calculated the loss and Dice score coefficient as an average from the evaluation folds. The selection of the encoder and decoder was done based on an ablation study (Figure 2, Supplementary Figure 2).

Model application on new images (inference)

During inference, CC was predicted for the full histology images, by combining smaller tiles with a sliding window (512×1024 -pixel window with 256×512 -pixel steps), averaging the overlapping predictions. The tiling was used to avoid memory issues on the graphical processing unit while segmenting larger areas of CC. The tiles were combined, averaging the overlapping areas and predictions from every fold. Subsequently, a threshold was applied to the prediction map by using a probability of 0.8 (a high threshold was used for the exclusion of ambiguous areas from the maps, especially for the μ CT images). In the case of the μ CT stacks, the inference was conducted slice-by-slice with similar tiling. The predictions were averaged from every fold as well as the coronal and sagittal planes for obtaining the final probability map.

The histology masks were post-processed by removing small isolated areas (< 500 pixels). This ensured the removal of small artifacts while retaining large CC regions that could be disconnected due to a fold in the histology section (Supplementary Figure 3). In the μ CT post-processing, masks were subjected to a sweep operation to keep only the largest object. This ensured the removal of possible false positives occurring on the tiles far from the actual CC layer. Finally, all CC masks were median filtered with a radius of 12 pixels (3D filtering in case of μ CT).

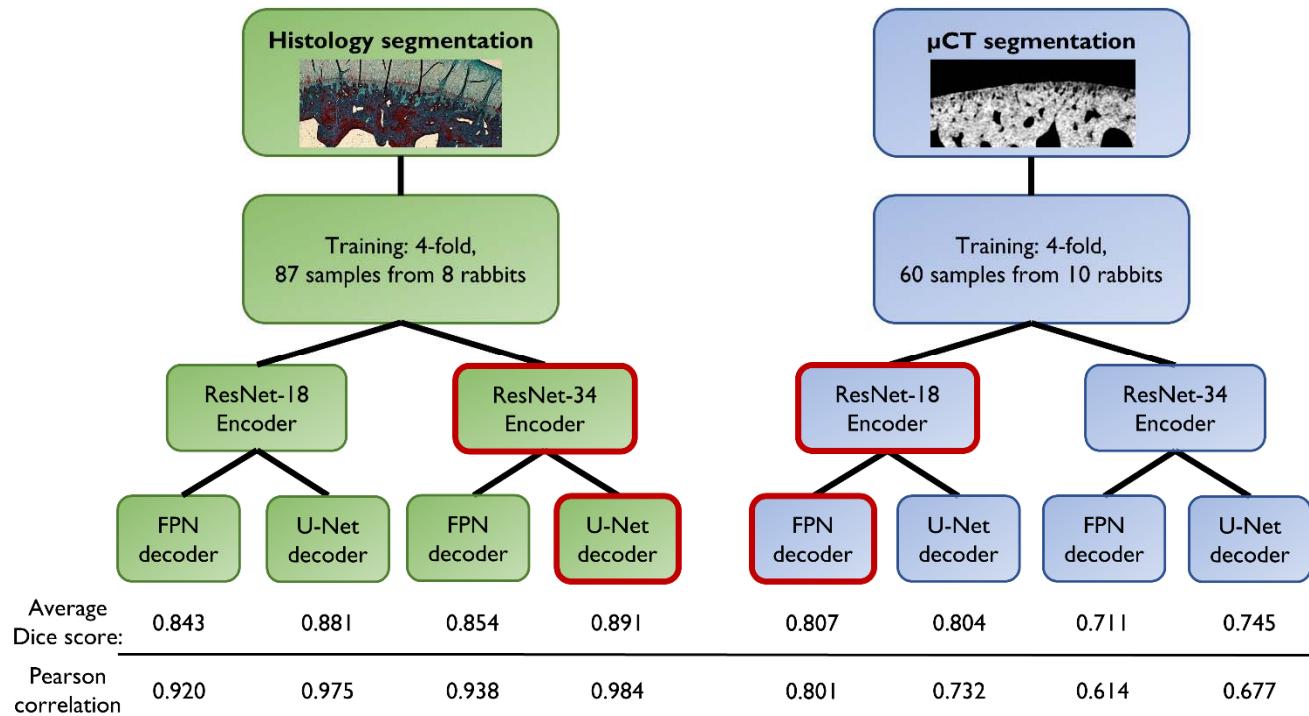


Figure 2. Illustration of the model training process. For both histology- and μ CT segmentation, a total of four models were trained with two different encoder and decoder designs. Based on the experiments, ResNet-34 and U-Net were more suitable for the complex histology masks (Dice score = 0.891), while ResNet-18 and FPN yielded higher performance for the smoother μ CT masks (Dice score = 0.807). Pearson correlation of the subsequent CC.Th analysis (bottom row) supported the choice of the segmentation models.

Morphological analysis

The full analysis procedure of CC.Th is summarized in Figure 3. The thickness estimation of the CC layer was performed automatically using a Python-based implementation of the local thickness algorithm. In the 2D case, the thickness assessment relies on mask skeletonization, a Euclidean distance transformation, and finally a simple circle-fitting algorithm⁽⁴³⁾. The 3D CC.Th analysis of the μ CT volumes was conducted with a similar sphere-fitting algorithm. From the estimated thickness maps, quantitative parameters such as mean-, median-, maximum CC.Th or standard deviation of

CC.Th can be calculated. In this study, we used the mean CC.Th as the quantitative parameter. The source code for the full segmentation and analysis procedure is published on our research unit's GitHub page (<https://github.com/MIPT-Oulu/RabbitCCS>). For the μ CT volumes, the thickness analysis took 2-3 hours per sample (on a high-end 12-core CPU), whereas the analysis for the histology slices took roughly three seconds per image. For this study, the 3D thickness analysis was calculated with parallel processing on the Puhti supercomputer (<https://research.csc.fi/csc-s-servers>). This effectively reduced the computation time for the μ CT volumes to roughly six minutes per sample.

To further investigate the applicability of the automatic segmentation on CC.Th analysis, a 2D analysis was performed between the manual segmentations and the out-of-fold predictions of the selected models. The thickness values were averaged for each sample with multiple histology sections or μ CT slices.

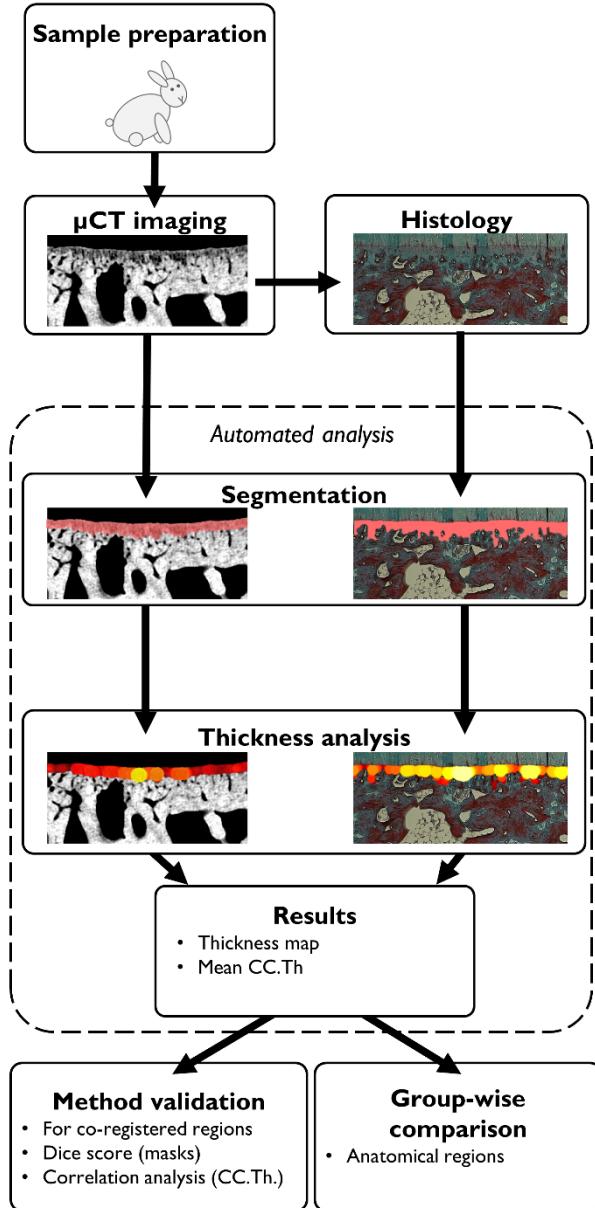


Figure 3. A flowchart summarizing the present study. After sample preparation, the tissue samples were imaged with μCT. Subsequently, the samples underwent histology processing, sectioning, and imaging with a light microscope. The preprocessing steps for the μCT data are illustrated in Supplementary Figure 1. During the automated analysis process, the CC layer is predicted using the deep learning models, thickness analysis is conducted, and finally, quantitative parameters are estimated from the estimated thickness maps. The obtained values were used in the validation of the methods as well as for comparison between the anatomical regions of the knee.

Validation with histology

To compare the CC analysis between histology and μ CT in 2D, matched μ CT slices (Figure 4) were estimated using co-registration based on rigid transformations with DataViewer (Bruker, Kontich, Belgium; version 1.5.2.4). A total of 24 samples (from four animals) were co-registered with the corresponding histology sections to find the matching subchondral structures. Since the search space is large when aligning the few μ m thick histology sections with the full sample, the remaining samples in paraffin blocks were imaged again using the μ CT scanner. The co-registration of two μ CT-imaged samples is straightforward and allows for locating the cutting orientation and approximating the location of the histological sample. Final co-registration was fine-tuned by performing a second co-registration between the original μ CT datasets and the histology images. Five serial μ CT images closest to the co-registered histology image were selected. Finally, we calculated the CC.Th from the co-registered histology image, while the CC.Th for μ CT-imaged samples was averaged from the five selected images.

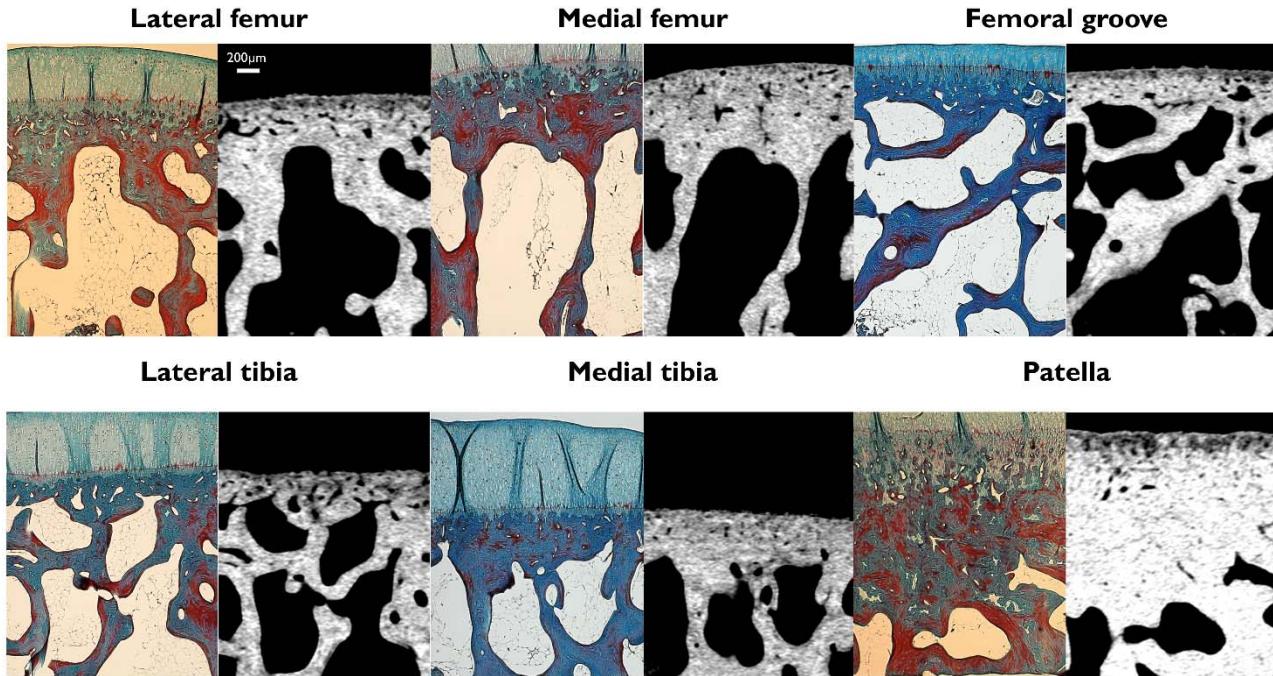


Figure 4. Examples from the co-registered histology slices and μ CT images. Scalebar for 200 μ m is shown in the top left. The CC can be assessed using both imaging modalities, although the thinnest CC areas are not visible in the μ CT images. Likely, these areas have a similar level of mineralization as the subchondral bone.

Statistical analysis and performance evaluation

For the co-registration experiment, a two-tailed Pearson correlation and Bland-Altman analyses were conducted to compare CC.Th between the μ CT and histology. The deep learning segmentation models were validated against the manual CC segmentations from μ CT and histology using the Dice score. The thickness analyses using out-of-fold predictions and manual segmentations were compared using Pearson correlations. The anatomical differences of CC.Th were assessed using mean comparisons with Linear Mixed Effect Models, accounting for the rabbit ID as the random effect, and the anatomical location as the fixed effect. The significance was assessed with Least Significant Difference without Bonferroni correction.

Results

Deep learning-based segmentation

For both imaging modalities, the quality of the deep learning model predictions against the manual annotations (out-of-fold validation) is summarized in Figure 2 and Supplementary Figure 2. By comparing the four different model architectures, ResNet-34 with the U-Net decoder yielded the highest mean Dice score for histology (Dice score = 0.891), while ResNet-18 with FPN yielded the best performance for μ CT segmentation (Dice score = 0.807). The quality of the segmentation on the full dataset was visually confirmed from virtual sections on orthogonal planes (Supplementary Figure 4).

In addition, we compared the 2D CC.Th analysis for the manual and predicted CC segmentations for both modalities (Figure 2 bottom, Supplementary Figure 5). With the selected model architecture, a high Pearson correlation was achieved between the manual and automatic CC.Th quantification from histology ($r = 0.984$). The correlation between predicted CC.Th and manually segmented CC.Th in μ CT images was also strong, although considerably smaller ($r = 0.801$). This correlation analysis further supported the choice for model architecture (Figure 2, bottom)

Validation with histology

Examples of μ CT images co-registered with histology are shown in Figure 4. The results of the quantitative comparisons are shown in Figure 5 (predicted CC) and Supplementary Figure 6 (manual segmentation). The automated μ CT-based measurements of CC.Th had a strong correlation ($r = 0.897$)

with a similar analysis on the co-registered histology images. Furthermore, the μ CT analysis had a good agreement (bias = 21.9 μ m, standard deviation = 21.5 μ m) with histology, based on the Bland-Altman analysis. Manual segmentation yielded a smaller correlation ($r = 0.852$) as well as greater bias (36.9 μ m) and standard deviation (30.9 μ m) than the comparison using predicted masks.

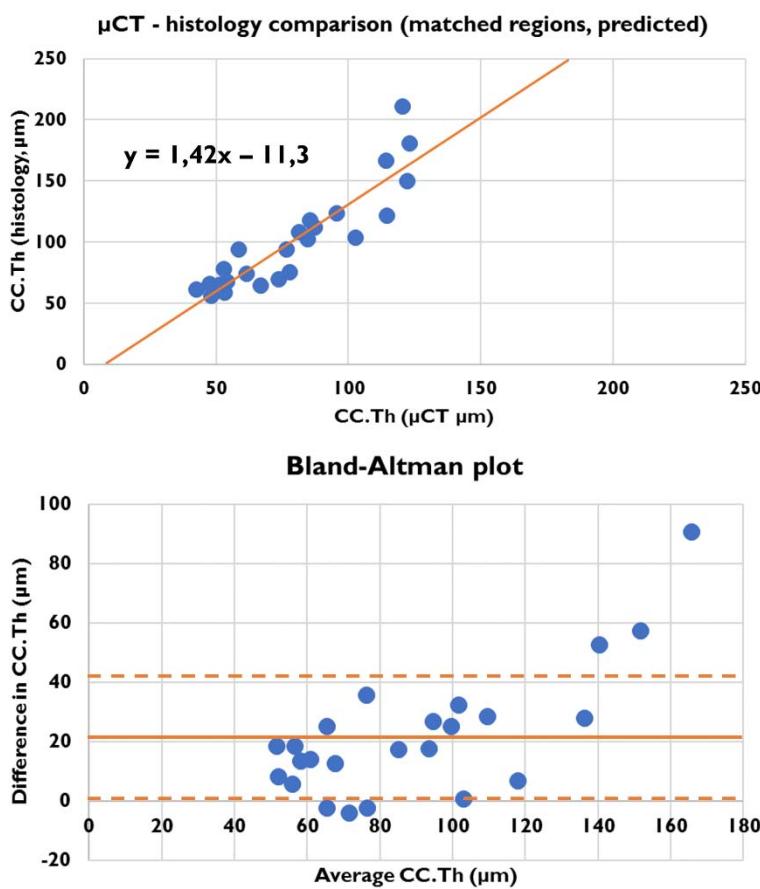


Figure 5. Quantitative CC.Th comparison of the matched histology and μ CT regions based on automated segmentation. The equation for the linear fit is shown in the top image. For the Bland-Altman plot, the bias is indicated with a horizontal line, and the distance of one standard deviation with a dashed line. The estimated values are highly correlated ($r = 0.897$) and the Bland-Altman analysis reveals that the μ CT method yields 21.9 μ m thinner CC.Th on average. The areas with a high CC.Th (mainly the patellar region) have the highest absolute differences between methods.

Anatomical locations

An example of a thickness map and VOI inside a lateral plateau sample is shown in the Supplementary Video. The differences in CC.Th based on anatomical variability are illustrated in Figure 6. According to the Linear Mixed Effects Model analysis on the histology and μ CT results (Table 2), the mean CC.Th varies greatly between the studied anatomical regions ($p < 0.001$). The thickest CC was in the patellar region, while the thinnest CC was in the tibial regions (lateral and medial plateau). The histology analysis allowed for further separation of the lateral and medial femoral condyles ($p = 0.026$). Although the absolute differences in CC.Th were larger using histology analysis than with the μ CT approach, the μ CT results had a smaller variance for individual regions than that observed with histology, allowing for separation of the anatomical locations.

Anatomical regions

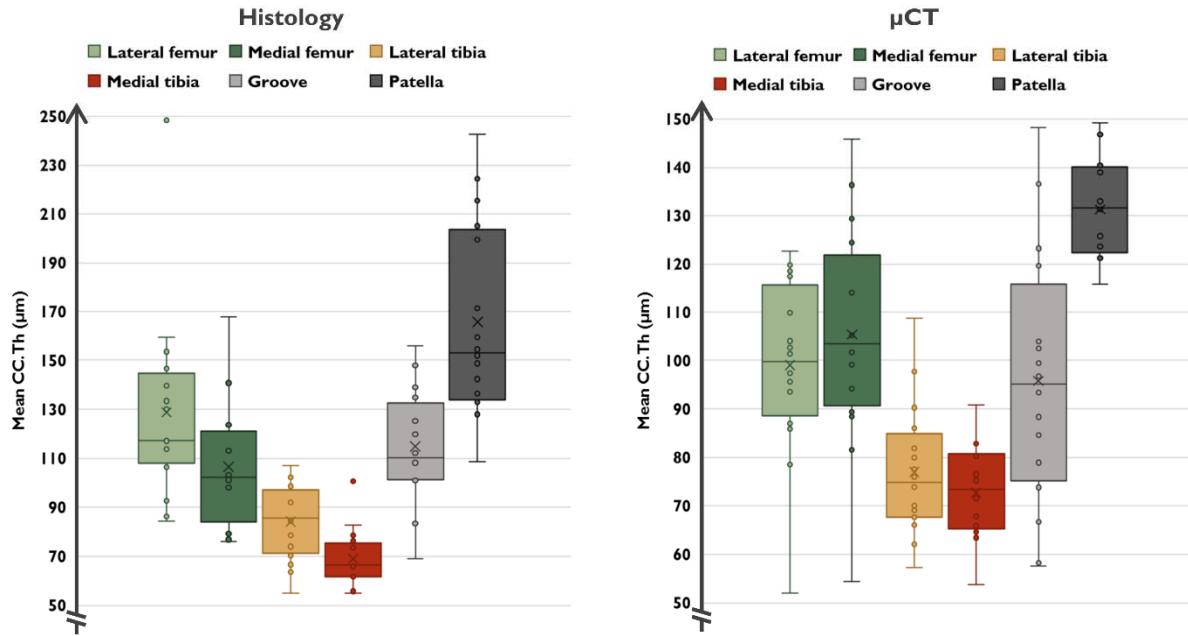


Figure 6. Boxplots illustrating the group-wise CC.Th values obtained from the histology and μCT modalities. The median value for each group is shown with the horizontal line and mean value with the cross. From the graph, the anatomical regions can be divided into three categories: thin CC (lateral and medial tibia), intermediate CC (lateral and medial femoral condyles, femoral groove) and thick CC (patella).

Table 2. Mean differences of mean CC thickness between the six regions (in μm): Lateral (LF) and medial (MF) femoral condyle, lateral (LP) and medial (MP) tibial plateau, femoral groove (G) and patella (P). The differences were assessed using a Linear Mixed Effects Model analysis, with Least Significant Difference. Detailed p-values are shown for $p \geq 0.001$.

Histology

	MF	LP	MP	G	P
LF	22.2 p = 0.026	44.5*	59.7*	13.9 $p = 0.162$	-36.9*
MF		22.2 p = 0.026	37.4*	-8.4 $p = 0.397$	-59.1*
LP			15.2 $p = 0.126$	-30.6 p = 0.002	-81.4*
MP				-45.8*	-96.6*
G					-50.8*

μCT

LF	-6.3 $p = 0.317$	22.1 p = 0.001	26.3*	3.3 $p = 0.604$	-32.3*
MF		28.4*	32.6*	9.6 $p = 0.131$	-26.0*
LP			4.2 $p = 0.503$	-18.8 p = 0.004	-54.4*
MP				-23.0*	-58.6*
G					-35.6*

* $p < 0.001$

Discussion

Morphological analysis of CC may reveal novel understanding of musculoskeletal physiology and pathology. A suitable tool for structural analysis of CC would be μ CT, however, the separation between bone and CC is extremely challenging. In this study, we developed a μ CT-based framework for 3D analysis of CC morphology. The framework utilizes state-of-the-art deep learning segmentation and automated analyses of CC.Th. Finally, we compared CC morphology on different joint surfaces within the healthy rabbit knees. Our results demonstrate that CC.Th can be quantified not only from histology but also from μ CT, which is feasible and efficient due to an automatic segmentation approach. The proposed method enables studying the 3D morphology of the mineralized CC without the time-consuming and destructive histological processing and with minimal user-induced bias.

Our results revealed that different CNN architectures were best suited for CC segmentation from histology and μ CT images. The FPN decoder is computationally more efficient, but it introduces an up-sampling layer for the model output. As a result, U-Net provides more detailed predictions since the CC is predicted without a subsequent interpolation. The results show that the U-Net decoder provided a slight advantage for segmenting the more complex CC structures in histology images. In the μ CT images, such details are not visible, and FPN decoder yielded better results than the U-Net one. Encoder-wise, the deeper ResNet-34 might yield even better performance than the ResNet-18 encoder⁽³⁸⁾. However, the ResNet-18 encoder with fewer layers than ResNet-34 performed better on the μ CT data than ResNet-34. Thus, we suspect that the more complex ResNet-34 may overfit when images become ambiguous, as in the case of the μ CT images.

The automated CC segmentation performed particularly well for the histology samples. A relatively high Dice score coefficient (0.891) and similar CC.Th results compared to the manual annotations ($r = 0.984$) suggest that the automated and manual methods give virtually identical results. For the μ CT data, the performance was weaker than for the histology data (Dice = 0.807, $r = 0.801$). However, the segmentation of CC from the μ CT images is much more difficult than segmentation from histology slides. Therefore, this result was expected. Based on our experience, there is also a significant variation in manual CC segmentation between human annotators. However, when comparing the estimated 2D CC.Th between histology and μ CT for co-registered regions, there was strong agreement ($r = 0.897$).

We have previously shown that the subchondral bone plate imaged with μ CT contains also the CC layer⁽²⁹⁾. Consequently, automated labeling of the CC layer could identify the true subchondral bone tissue accurately. The proposed method requires high-resolution for resolving the mineralized cartilage. We believe that this is of high interest for studies that focus on the subtle changes in the bone plate, such as thinning due to increased remodeling. Such thinning of the bone plate has been suggested to occur already in the early stages of OA⁽⁴⁴⁾.

The greatest differences in CC.Th between imaging modalities were seen in the samples with the highest CC.Th (Figure 5). We hypothesize that the less mineralized CC measured with μ CT accounts for “young” tissue, which has distinct attenuation properties compared to the bone layer. Lower mineralization (hydroxyapatite content) of CC compared to bone has previously been reported using X-ray diffraction⁽⁴⁵⁾. However, many studies have reported a higher mineralization of CC in backscattered electron imaging studies^(23,24,46), at least for human tissue. Thus, we considered that there might be a possible contribution of partial volume effects related to cellularity. Most of the cellular structures are

visible, however, the observed changes are likely related to tissue mineralization. On the other hand, the deep CC appears more mineralized with similar attenuation properties as the subchondral bone, making it impossible to identify it solely based on X-ray methods. The samples with high CC.Th likely contain large, partly ossified areas of deep CC (such as the patellar region in Figure 4), leading to differences in CC.Th between the imaging methods. Therefore, we propose that our method could provide novel 3D information on tidemark advancement and other dynamic processes in calcified cartilage.

Interestingly, CC.Th depends greatly on anatomical location, as identified with both imaging methods. This is also consistent with our hypothesis. In the patellar region, CC.Th was the thickest among all locations of the rabbit knee. Femoral regions had intermediate CC.Th, while the thinnest regions were found in the medial tibial plateau region. We hypothesize that these variations in CC.Th are due to the distinct biomechanical environment in the different regions. First, the tibial plateau predominantly experiences compressive load due to body weight, while the patella experiences mainly shear forces that arise from the sliding joint articulation. Second, in the femur, the environment is a mix of these phenomena, i.e. the femoral condyles experience more compressive stress compared to higher shear forces on the groove. However, we did not find statistically significant differences in CC.Th between the condyles and groove. Finally, the higher shear stress experienced by the patella and femoral groove likely requires a stronger connection between the articular cartilage and the underlying subchondral bone plate, thus, resulting in higher CC.Th. Other studies have shown that the CC.Th of rabbit knees increases when subjected to chronic compression and that the CC is thicker in the lateral compared to the medial knee compartment⁽⁴⁷⁾.

This study has several limitations: First, the decalcification process required for preparation of the histology slides may cause structural alterations in the tissue. Second, the intensity gradient between CC and subchondral bone can be ambiguous. This is especially the case for ultra-thin or non-existent CC. An ambiguous interface may appear because of endochondral remodeling resulting in bony protrusions into CC. Third, although an acceptable performance was achieved, the amount of training data used for the deep learning segmentation was relatively low. Examples from a greater number of animals may give a better performance, especially in the case of the challenging μ CT segmentation. Fourth, our automated thickness analysis method is computationally expensive and does not scale well for large volumes. For routine use, more advanced scalable algorithms should be implemented, for example utilizing a distance ridge calculation⁽⁴⁸⁾. Finally, the segmentation models might require fine-tuning to data acquired from a different microscope or μ CT scanner to ensure sufficient performance on new samples.

In conclusion, we have presented a promising method for the morphological analysis of CC with μ CT. To the best of our knowledge, this is the first automated method for quantitative 3D analysis of CC. That has been sufficiently validated against the histological gold standard. As a proof of concept, we could detect anatomical variation in the rabbit knee; the patellar region has the thickest CC and the tibial plateau region the thinnest. This structural difference between regions is presumably related to the diverse biomechanical environments, and thus the different requirements of the joint surfaces in different regions of the knee.

Acknowledgements

The full source code of the project is openly available on our research unit's GitHub page (<https://github.com/MIPT-Oulu/RabbitCCS>). This project received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 713645; European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement No. 336267; Academy of Finland [grant numbers 286526, 303786, 324529]; Saastamoinen Foundation, Päivikki ja Sakari Sohlberg Foundation; Finnish Cultural Foundation (Central Fund No. 191044, North Ostrobothnia Regional Fund No. 60172246); Maire Lisko Foundation; The Canadian Institutes of Health Research, The Canada Research Chair Programme; The Killam Foundation. The strategic funding of the University of Oulu and the University of Eastern Finland are acknowledged. CSC – IT Center for Science, Espoo, Finland is acknowledged for generous computational resources.

Author's roles

Study conception and design: SJOR, LH, PT, AT, RKK, SS, WH, MAJF

Data collection: LH, PT, RKK, WH, MAJF

Method development: SJOR, AT, EP, SS, MAJF

Data analysis and interpretation: SJOR, LH, AT, SS, MAJF

Drafting the manuscript: SJOR, LH, MAJF

Critical revision and approving the final version of the manuscript: All authors

SJOR takes responsibility for the integrity of the work.

References

1. Madry H, van Dijk CN, Mueller-Gerbl M. The basic science of the subchondral bone. *Knee Surg Sports Traumatol Arthroscopy* 2010;18(4):419-433 doi.org/10.1007/s00167-010-1054-z.
2. Sophia Fox AJ, Bedi A, Rodeo SA. The basic science of articular cartilage: structure, composition, and function. *Sports Health* 2009;1(6):461-468 doi.org/10.1177/1941738109350438.
3. Müller-Gerbl M, Schulte E, Putz R. The thickness of the calcified layer of articular cartilage: a function of the load supported? *J Anat* 1987;154:103-111.
4. Hoemann CD, Lafantaisie-Favreau CH, Lascau-Coman V, Chen G, Guzmán-Morales J. The cartilage-bone interface. *J Knee Surg* 2012;25(2):85-97 doi.org/10.1055/s-0032-1319782.
5. Kiviranta I, Tammi M, Jurvelin J, Helminen HJ. Topographical variation of glycosaminoglycan content and cartilage thickness in canine knee (stifle) joint cartilage. Application of the microspectrophotometric method. *J Anat* 1987;150:265-276.
6. Cohen ZA, McCarthy DM, Kwak SD, Legrand P, Fogarasi F, Ciaccio EJ, et al. Knee cartilage topography, thickness, and contact areas from MRI: in-vitro calibration and in-vivo measurements. *Osteoarthritis Cartilage* 1999 Jan;7(1):95-109 doi.org/10.1053/joca.1998.0165.
7. Milz S, Putz R. Quantitative morphology of the subchondral plate of the tibial plateau. *J Anat* 1994;185(Pt 1):103-110.
8. Martinelli MJ, Eurell J, Les CM, Fyhrie D, Bennett D. Age-related morphometry of equine calcified cartilage. *Equine Vet J* 2002;34(3):274-278 doi.org/10.2746/042516402776186100.
9. Kim W, McArdle BH, Kawcak CE, McIlwraith CW, Firth EC, Broom ND. Histomorphometric evaluation of the effect of early exercise on subchondral vascularity in the third carpal bone of horses. *Am J Vet Res* 2013;74(4):542-549 doi.org/10.2460/ajvr.74.4.542.
10. Tranquille CA, Blunden AS, Dyson SJ, Parkin TDH, Goodship AE, Murray RC. Effect of exercise on thicknesses of mature hyaline cartilage, calcified cartilage, and subchondral bone of equine tarsi. *Am J Vet Res* 2009;70(12):1477-1483 doi.org/10.2460/ajvr.70.12.1477.
11. Murray RC, Whitton RC, Vedi S, Goodship AE, Lekeux P. The effect of training on the calcified zone of equine middle carpal articular cartilage. *Equine Vet J Suppl* 1999;30:274-278 doi.org/10.1111/j.2042-3306.1999.tb05234.x.
12. Oettmeier R, Arokoski J, Roth AJ, Helminen HJ, Tammi M, Abendroth K. Quantitative study of articular cartilage and subchondral bone remodeling in the knee joint of dogs after strenuous running training. *J Bone Miner Res* 1992;7(S2):S419-S424 doi.org/10.1002/jbmr.5650071410.
13. Kiviranta I, Jurvelin J, Tammi M, SääMänen A, Helminen HJ. Weight bearing controls

glycosaminoglycan concentration and articular cartilage thickness in the knee joints of young beagle dogs. *Arthritis Rheum* 1987;30(7):801-809 doi.org/10.1002/art.1780300710.

14. Thambyah A, Nather A, Goh J. Mechanical properties of articular cartilage covered by the meniscus. *Osteoarthritis Cartilage* 2006;14(6):580-588 doi.org/10.1016/j.joca.2006.01.015.

15. Havelka S, Horn V, Spohrova D, Valouch P. The calcified-noncalcified cartilage interface: The tidemark. *Acta Biol Hung* 1984;35(2-4):271-279.

16. Doube M, Firth EC, Boyde A. Variations in articular calcified cartilage by site and exercise in the 18-month-old equine distal metacarpal condyle. *Osteoarthritis Cartilage* 2007;15(11):1283-1292 doi.org/10.1016/j.joca.2007.04.003.

17. Lane LB, Bullough PG. Age-related changes in the thickness of the calcified zone and the number of tidemarks in adult human articular cartilage. *J Bone Jt Surg* 1980;62(3):372-375 doi.org/10.1302/0301-620X.62B3.7410471.

18. Frisbie DD, Cross MW, McIlwraith CW. A comparative study of articular cartilage thickness in the stifle of animal species used in human pre-clinical studies compared to articular cartilage thickness in the human knee. *Vet Comp Orthop Traumatol* 2006;19(3):142-146 doi.org/10.1055/s-0038-1632990.

19. Pan J, Wang B, Li W, Zhou X, Scherr T, Yang Y, et al. Elevated cross-talk between subchondral bone and cartilage in osteoarthritic joints. *Bone* 2012;51(2):212-217 doi.org/10.1016/j.bone.2011.11.030.

20. Adams CS, Horton Jr. WE. Chondrocyte apoptosis increases with age in the articular cartilage of adult animals. *Anat Rec* 1998;250(4):418-425 doi.org/10.1002/(SICI)1097-0185(199804)250:43.0.CO;2-T.

21. Murray RC, Blunden TS, Branch MV, Tranquille CA, Dyson SJ, Parkin TDH, et al. Evaluation of age-related changes in the structure of the equine tarsometatarsal osteochondral unit. *Am J Vet Res* 2009;70(1):30-36 doi.org/10.2460/ajvr.70.1.30.

22. Goldring MB, Goldring SR. Osteoarthritis. *J Cell Physiol* 2007;213(3):626-634 doi.org/10.1002/jcp.21258.

23. Gupta HS, Schratter S, Tesch W, Roschger P, Berzlanovich A, Schoeberl T, et al. Two different correlations between nanoindentation modulus and mineral content in the bone-cartilage interface. *J Struct Biol* 2005;149(2):138-148 doi.org/10.1016/j.jsb.2004.10.010.

24. Ferguson VL, Bushby AJ, Boyde A. Nonmechanical properties and mineral concentration in articular calcified cartilage and subchondral bone. *J Anat* 2003;203(2):191-202 doi.org/10.1046/j.1469-7580.2003.00193.x.

25. Gerstenfeld LC, Alkhiary YM, Krall EA, Nicholls FH, Stapleton SN, Fitch JL, et al. Three-dimensional reconstruction of fracture callus morphogenesis. *J Histochem Cytochem*

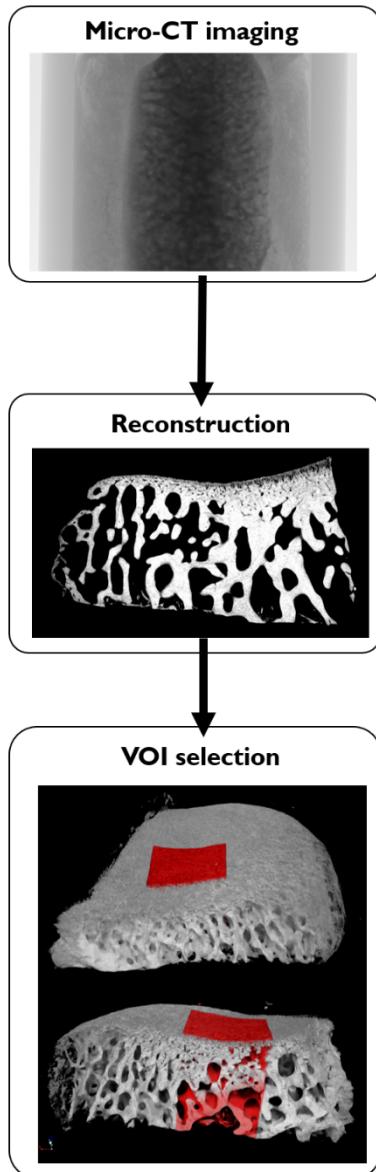
2006;54(11):1215-1228 doi.org/10.1369/jhc.6A6959.2006.

26. Guo H, Xu Z, He B, Hao D, Bian W. Three dimensional reconstruction of bone-cartilage transitional structures based on semi-automatic registration and automatic segmentation of serial sections. *J Tissue Eng Regen Med* 2014;11(5):387-396 doi.org/10.1007/s13770-014-0027-6.
27. Kerckhofs G, Sainz J, Wevers M, Van De Putte T, Schrooten J. Contrast-enhanced nanofocus computed tomography images the cartilage subtissue architecture in three dimensions. *Eur Cells and Mater* 2012;25:179-189 doi.org/10.22203/ecm.v025a13.
28. Mehadji B, Ahmed Y, Berteau J. A novel approach for computing 3D mice distal femur properties using high-resolution micro-computed tomography scanning. *Micron* 2019;121:1-7 doi.org/10.1016/j.micron.2019.02.011.
29. Finnilä MAJ, Thevenot J, Aho O-, Tiitu V, Rautiainen J, Kauppinen S, et al. Association between subchondral bone structure and osteoarthritis histopathological grade. *J Orthop Res* 2017;35(4):785-792 doi.org/10.1002/jor.23312.
30. Rytty SJO, Tiulpin A, Frondelius T, Finnilä MAJ, Karhula SS, Leino J, et al. Automating three-dimensional osteoarthritis histopathological grading of human osteochondral tissue using machine learning on contrast-enhanced micro-computed tomography. *Osteoarthritis Cartilage* 2020.
31. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *Lect Notes Comput Sci* 2015;9351 doi.org/10.1007/978-3-319-24574-4_28.
32. Tiulpin A, Finnilä M, Lehenkari P, Nieminen HJ and Saarakkala S. Deep-Learning for Tidemark Segmentation in Human Osteochondral Tissues Imaged with Micro-computed Tomography. *Advanced Concepts for Intelligent Vision systems*; 2020 doi.org/10.1007/978-3-030-40605-9_12.
33. Lin T, Dollar P, Girshick R, He K, Hariharan B and Belongie S. Feature Pyramid Networks for Object Detection. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017 doi.org/10.1109/CVPR.2017.106.
34. J. Deng, W. Dong, R. Socher, L. Li, Kai Li, Li Fei-Fei. ImageNet: A large-scale hierarchical image database. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2009:248-255 doi.org/10.1109/CVPR.2009.5206848.
35. Ng H-, Nguyen VD, Vonikakis V and Winkler S. Deep learning for emotion recognition on small datasets using transfer learning. *ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction*; 2015 doi.org/10.1145/2818346.2830593.
36. Tiulpin A, Saarakkala S. Automatic Grading of Individual Knee Osteoarthritis Features in Plain Radiographs using Deep Convolutional Neural Networks. *arXiv* 2019 doi.org/1907.08020.
37. Shin H-, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE*

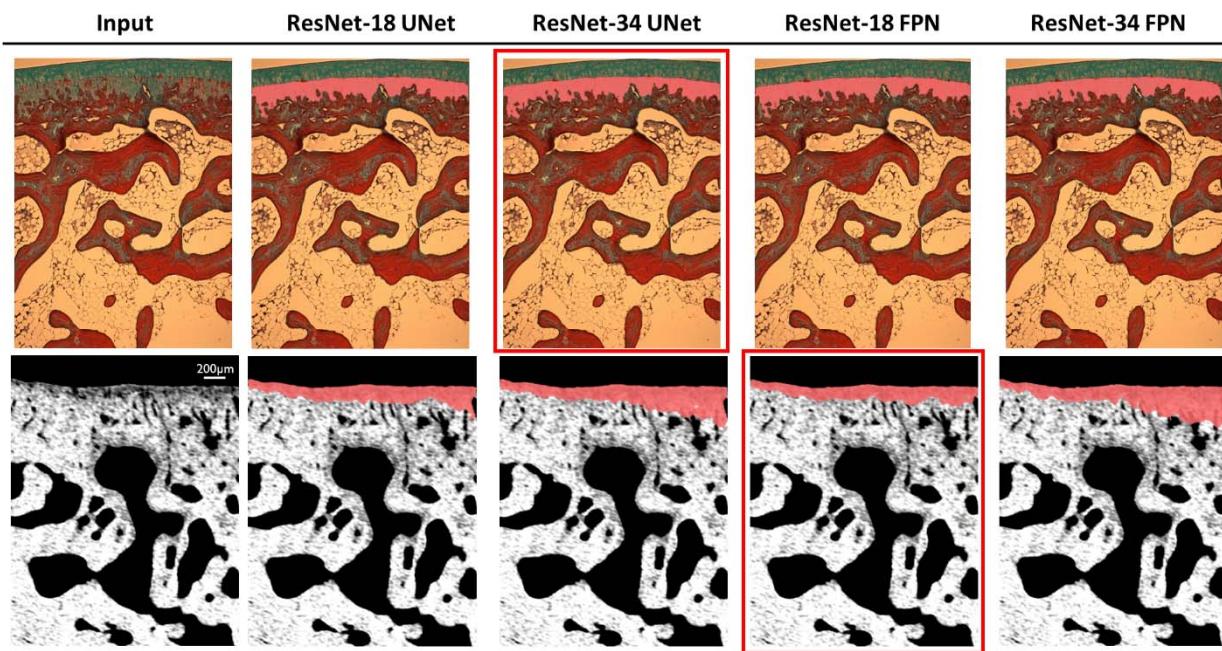
Trans Med Imaging 2016;35(5):1285-1298 doi.org/10.1109/TMI.2016.2528162.

38. He K, Zhang X, Ren S and Sun J. Deep residual learning for image recognition. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 doi.org/10.1109/CVPR.2016.90.
39. Nibali A, He Z, Wollersheim D. Pulmonary nodule classification with deep residual networks. International Journal of Computer Assisted Radiology and Surgery 2017;12(10):1799-1808 doi.org/10.1007/s11548-017-1605-6.
40. Solovyev R, Melekhov I, Lesonen T, Vaattovaara E, Tervonen O and Tiulpin A. Bayesian Feature Pyramid Networks for Automatic Multi-Label Segmentation of Chest X-rays and Assessment of Cardio-Thoracic Ratio. Advanced Concepts for Intelligent Vision Systems (ACIVS); 2020 doi.org/10.1007/978-3-030-40605-9_11.
41. Mustonen A, Käkelä R, Finnilä MAJ, Sawatsky A, Korhonen RK, Saarakkala S, et al. Anterior cruciate ligament transection alters the n-3/n-6 fatty acid balance in the lapine infrapatellar fat pad. Lipids Health Dis 2019;18(1):67 doi.org/10.1186/s12944-019-1008-5.
42. Tiulpin A. SOLT: Streaming over Lightweight Transformations. 2019 doi.org/10.5281/zenodo.3351977.
43. Hildebrand T, Rüegsegger P. A new method for the model-independent assessment of thickness in three-dimensional images. J Microsc 1997;185(1):67-75 doi.org/10.1046/j.1365-2818.1997.1340694.x.
44. Burr DB, Gallant MA. Bone remodelling in osteoarthritis. Nature Reviews Rheumatology 2012;8(11):665-673 doi.org/10.1038/nrrheum.2012.130.
45. Zhang Y, Wang F, Tan H, Chen G, Guo L, Yang L. Analysis of the mineral composition of the human calcified cartilage zone. International journal of medical sciences 2012;9(5):353-360 doi.org/10.7150/ijms.4276.
46. Burr DB. Anatomy and physiology of the mineralized tissues: Role in the pathogenesis of osteoarthritis. Osteoarthritis Cartilage 2004;12:20-30 doi.org/10.1016/j.joca.2003.09.016.
47. Roemhildt ML, Beynnon BD, Gardner-Morse M, Badger G, Grant C. Changes induced by chronic in vivo load alteration in the tibiofemoral joint of mature rabbits. J Orthop Res 2012;30(9):1413-1422 doi.org/10.1002/jor.22087.
48. Dougherty R, Kunzelmann K. Computing Local Thickness of 3D Structures with ImageJ. Microscopy and Microanalysis 2007;13(S02):1678-1679 doi.org/10.1017/S1431927607074430.

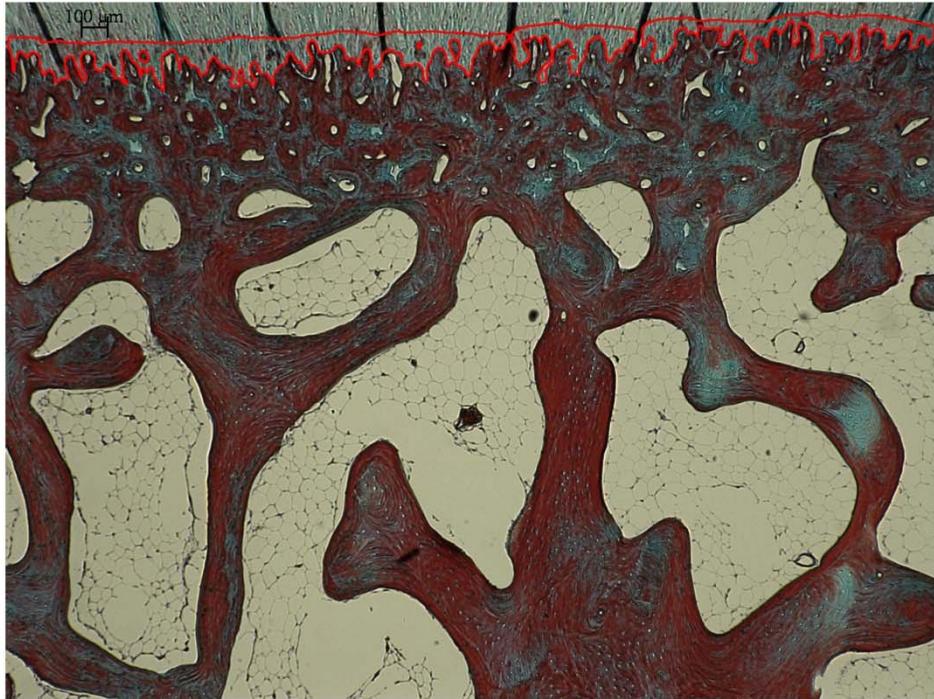
Supplementary material



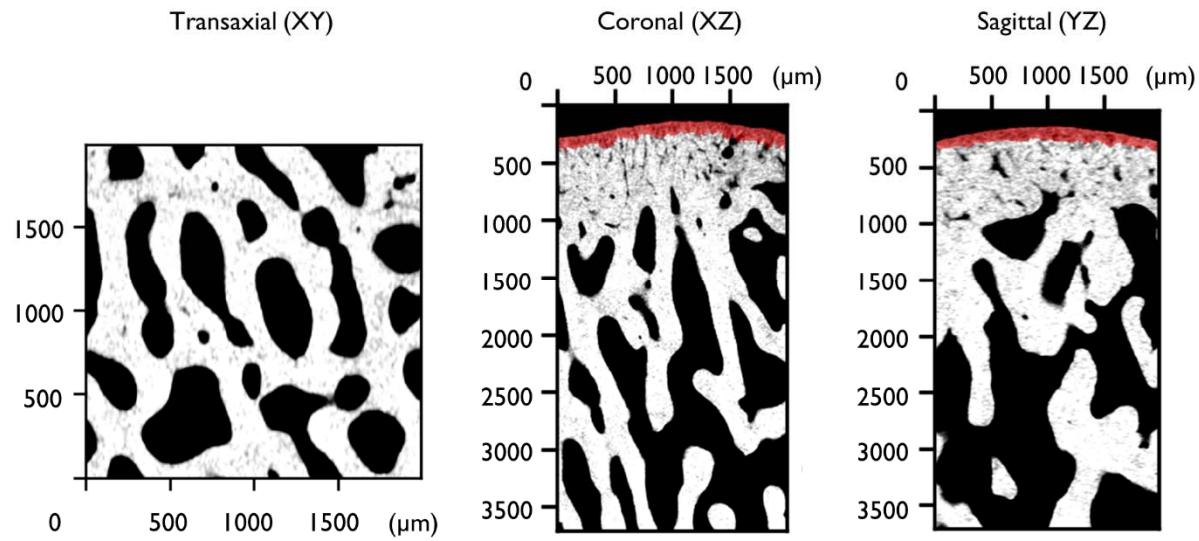
Supplementary Figure 1. The main preprocessing steps for the μ CT data. In the top image, an example projection image is shown from a lateral tibial plateau sample. In the middle, a coronal section from the reconstruction result is displayed. The bottom image shows the 3D rendered sample with an example of the VOI selection. The top part shows an overview, while the bottom part includes a virtual section inside the sample.



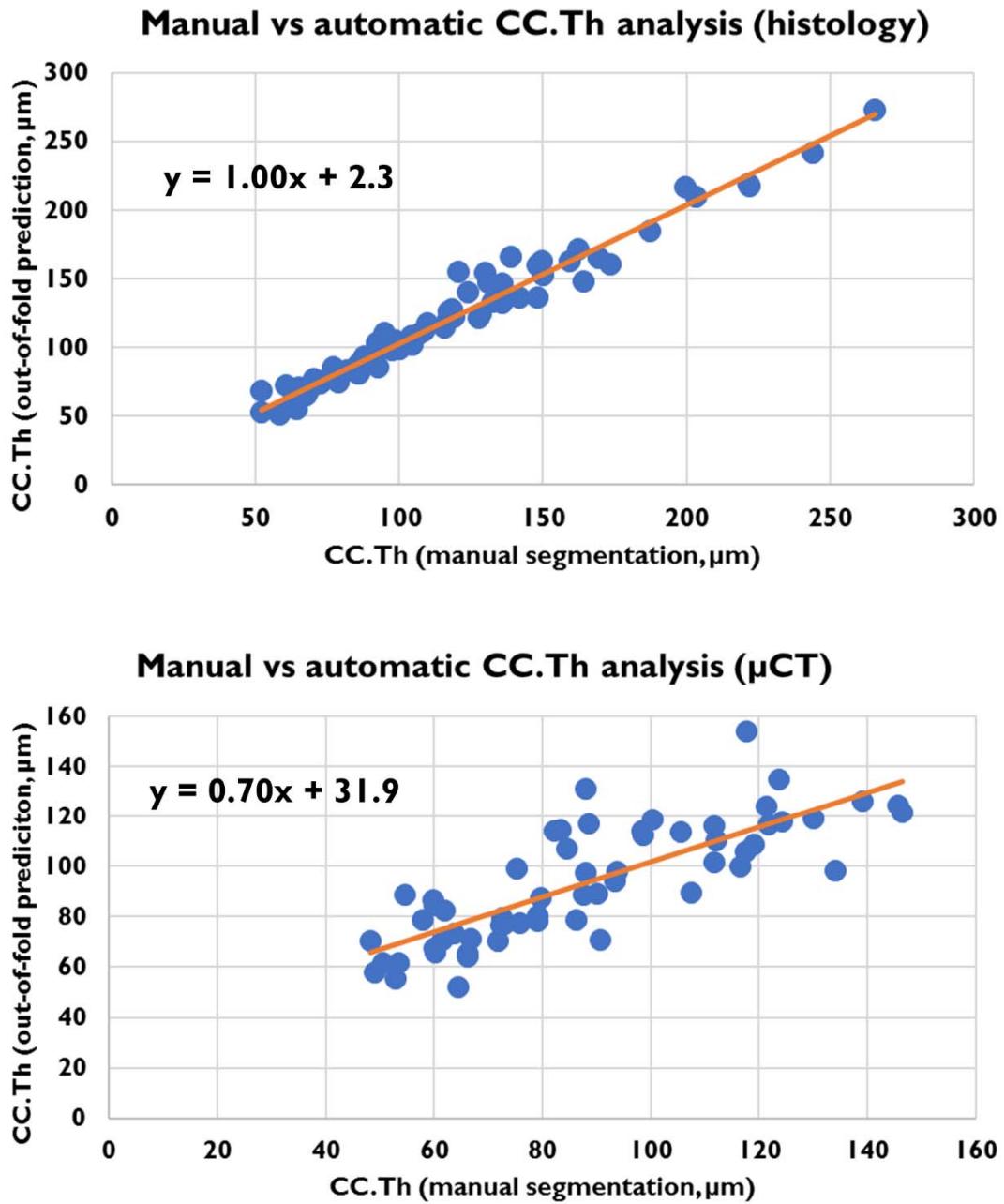
Supplementary Figure 2. Qualitative comparison of the predicted CC masks. The selected models are highlighted with a red rectangle. Since the FPN decoder requires upsampling of the segmentation result, small details of the histology masks are easily overlooked. This makes it better suited for the smooth µCT masks. U-Net preserves the complexity of the histology mask slightly better than FPN. For the µCT models, the complex ResNet-34 model may overfit the training data, thus yielding higher errors on the validation images.



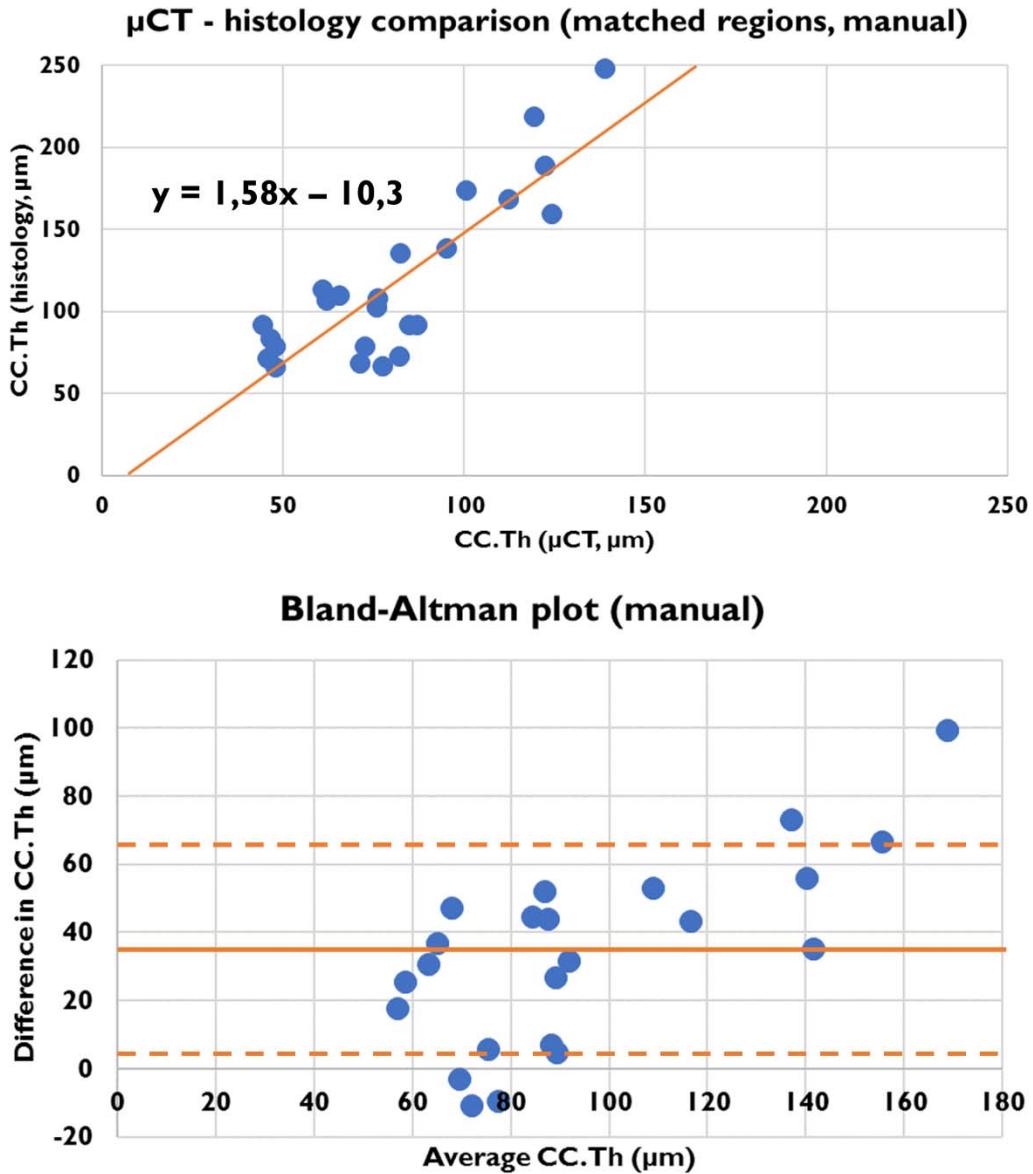
Supplementary Figure 3. Example histology slice with a disconnected prediction. The red outline illustrates the mask post-processed with the despeckle operation, removing areas smaller than 500 pixels. If only the largest mask would be retained, much of the real CC layer would be lost.



Supplementary Figure 4. Illustration for visual output of the predicted CC mask. A figure with three orthogonal planes on the sample is drawn after the inference, displaying the model output in red.



Supplementary Figure 5. Scatterplots for the CC.Th analysis based on manual and automatic segmentation. The equation for the linear fit is shown. With the histology images, the correspondence between the methods is extremely high. The more challenging μCT segmentation results in discrepancies between prediction and the gold standard.



Supplementary Figure 6. Quantitative CC.Th comparison of the matched histology and µCT regions based on manual segmentation (compared to the predicted CC in Figure 5). The equation for the linear fit is shown in the top image. For the Bland-Altman plot, the bias is indicated with a horizontal line, and the distance of one standard deviation with a dashed line. The analysis yielded a Pearson correlation coefficient of 0.852, a bias of 36.9 μm , and a standard deviation of 30.9 μm .