

Automated Segmentation of Knee Menisci Using U-Net Deep Learning Model: Preliminary Results

Alexei BOTNARI^a, Manuella KADAR^b, Daniela Rodica PUIA^b,
Jenel Marian PATRASCU^{a, c}, Jenel Marian PATRASCU Jr^{a, c}

^aDepartment of Orthopedics and Traumatology, Faculty of Medicine,
“Victor Babes” University of Medicine and Pharmacy, 300041 Timisoara, Romania

^bDepartment of Computer Science, Faculty of Informatics and Engineering,
“1 Decembrie 1918” University of Alba Iulia, 510009 Alba Iulia, Romania

^cDepartment of Orthopedics and Traumatology, “Professor Teodor Șora” Research Center,
“Victor Babes” University of Medicine and Pharmacy, Timișoara, Romania



ABSTRACT

Objectives: The present study describes the initial findings of the detection and segmentation of the knee meniscus in magnetic resonance imaging (MRI) scans using the U-Net deep learning model. The primary goal was to develop a model that automatically identified and segmented the meniscus from the region of interest (ROI) in knee MRI scans.

Material and methods: The current study was conducted in two phases. Initially, a U-Net deep learning model was developed to automatically detect the meniscus using a training dataset comprising 104 knee MRI images. In the second phase, the model was fine-tuned with an additional 50 MRI scans featuring manually segmented images to segment the meniscus from the ROI accurately.

Results: After performing 14 training tests, the U-Net model achieved a detection accuracy of 0.91. The average Dice score for ROIs after training at 100 epochs was 0.7259. With training extended to 300 epochs, the Dice score improved to 0.7525. Finally, the model reached a Dice score of 0.7609 after 500 epochs.

Conclusions: The present study introduces a practical deep learning-based approach for segmenting the knee meniscus, which is validated against ground truth annotations from orthopedic surgeons. Despite challenges such as data scarcity and the need for sequence-specific optimization, our method demonstrates significant potential for advancing automated meniscus segmentation in clinical settings.

Keywords: meniscus, knee, MRI, segmentation, U-Net, deep learning.

Address for correspondence:

Alexie Botnari, Manuella Kadar

Email: alexei.botnari@umft.ro; mkadar@uab.ro

Article received on the 9th of November and accepted for publication on the 13th of December 2024

BACKGROUND AND AIMS

The meniscus between the knee joint's tibial plateau and femoral condyles is critical for stabilizing articular cartilage (1, 2). It withstands shear, tension and compression forces and has an essential role in load distribution, shock absorption and cartilage lubrication (3, 4-7). During weight-bearing, it redistributes 30–55% of body weight across the joint, using the alignment of collagen fibers to convert axial forces into circumferential tensile forces (8, 9). Magnetic resonance imaging (MRI) is instrumental in detecting meniscal injuries, aiding in preoperative planning and postoperative recovery (11); it often surpasses arthroscopy in identifying meniscal tears, with 93% sensitivity and 88% specificity, especially in sagittal and coronal planes, though it can sometimes produce artifacts and false negatives (12). Despite volumetric and technical limitations, MRI remains an invaluable tool for guiding clinical management and surgical decisions by providing detailed pathology evaluations (13). Segmenting the meniscus from MRI images is fundamental for investigating tissue changes and their impact on knee conditions (14). This segmentation supports quantitative analysis, including volume measurements, mean intensity, intensity distribution, texture and thickness (15). Accurately segmenting musculoskeletal tissues is a critical first step for extracting quantitative metrics from MRIs to assess joint degeneration. Traditionally, this process has been time-consuming as it was requiring manual boundary delineation on each MRI slice and was strongly influenced by the user's expertise (16). Thus, there is a pressing demand for a fully automated knee segmentation method that is swift, accurate and capable of extracting detailed features and relaxation times from MRI data (17). Deep learning methods, like U-Net convolutional neural networks (CNNs) and their variations, have gained recognition in addressing these imaging challenges, especially in radiology and osteoarthritis (OA) diagnostics (18). Researchers have recently used 2D U-Net CNNs to segment the meniscus in double-echo steady-state MRIs of healthy individuals and those with OA. Automated segmentation has demonstrated its utility in assessing meniscal morphology and extracting meaningful biomarkers for OA diagnosis (19, 20).

This research showcases our initial results in identifying and segmenting the knee meniscus on MRI scans utilizing the U-Net deep learning architecture.

METHODS

The present study was carried out in two stages. Initially, we created a U-Net-based deep learning model to automate meniscus detection. In the second stage, we segmented the meniscus within the defined region of interest (ROI).

Data collection

With authorization from the hospital Ethical Committee, knee MRI scans were retrieved from the PACS system of the County Emergency Hospital in Alba Iulia, Romania. A total of 188 knee MRI scans were retrieved from the archive database for the present study, with each being accompanied by a radiology report containing detailed descriptions and diagnoses. All MRI examinations were conducted using a Siemens Magnetom Essenza scanner with a magnetic field strength of 1.5 T. We utilized sagittal proton density turbo spin echo (TSE) fat saturation images to target the meniscus. The imaging protocol utilized a repetition time of 2800 ms, an echo time of 26 ms, a slice thickness of 3 mm, a fat saturation level of 1.5 and a matrix resolution of 256 × 256 pixels, covering a field of view of 180 mm × 180 mm.

Model development

We employed a deep learning method built on the U-Net architecture (Figure 1). Previous re-

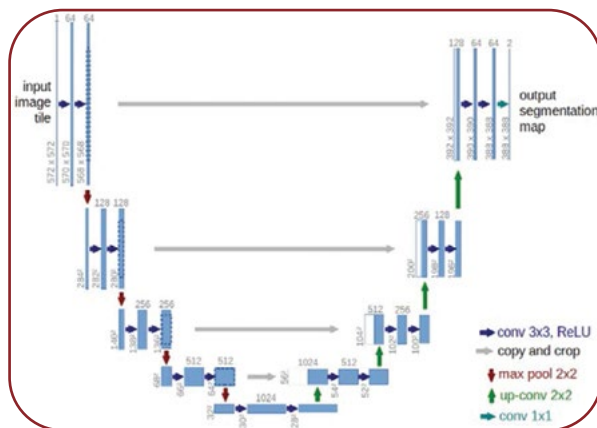


FIGURE 1. U-net architecture (32x32 pixels at the lowest resolution)

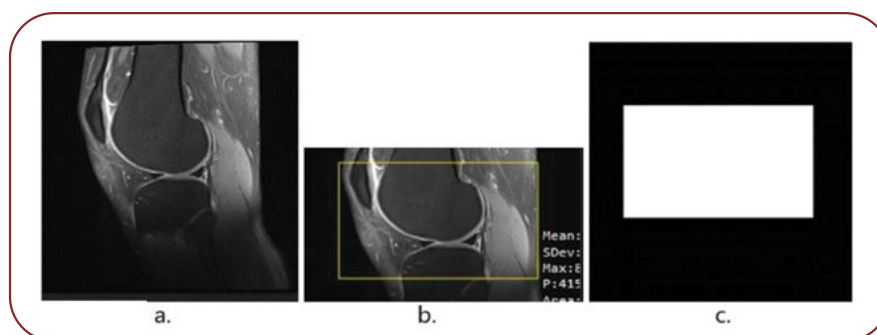


FIGURE 2. Magnetic resonance imaging of the knee: a) original image (sagittal proton density); b) image with the region of interest marked; and c) the mask created for the region of interest

search (21, 22) has demonstrated that models derived from this architecture achieved promising outcomes in meniscus segmentation. Designed explicitly for biomedical image segmentation, U-Net is a convolutional neural network capable of efficiently and accurately segmenting images.

Each blue box represents a multi-layer feature map, with the number of layers being indicated at the top and the xy dimensions specified at the lower left corner. White boxes denote copied feature maps, while arrows illustrate the various operations that were performed.

At baseline, we used GIMP, a free software tool, to annotate and create masks for each image (Figure 2). The original dataset was fully anonymized and sourced from the Alba Iulia County Emergency Hospital archives.

Image processing method

The training dataset consisted of 104 knee MRI scans with a 256 x 256 pixels resolution. This number of scans was insufficient for effectively training a deep-learning neural network. Thus, we performed data augmentation to expand the dataset. This process generated additional samples, enabling the model to generalize more effectively by simulating various scenarios and variations it might encounter in real-world data. Data augmentation increased the adequate size of the training dataset. It enhanced the robustness and accuracy of the model by reducing the risk of overfitting and ensuring it learned meaningful features rather than memorizing the training data. Following data preprocessing, the dataset was ready for input into the U-Net network.

U-Net model training

The deep neural network was constructed using the Keras functional API, which allowed for fle-

xibility in exploring various architectural designs. The network produces a 256 x 256 mask as output, which undergoes training to optimize its performance. A sigmoid activation function ensures that the mask's pixel values remain within the range (0, 1).

The model was trained in over 50 epochs and achieved an accuracy of approximately 0.9948 by the final epoch. Binary cross-entropy was used as the loss function during training, which optimized the ability of the model to distinguish between classes. The implementation required dependencies, including TensorFlow and Keras (version 1.0 or above).

In the second phase of the study, the meniscus was segmented from the detected region of interest. We included 50 MRI scans with manually segmented images to enhance the robustness of the model. The dataset was divided into training, validation and test sets in a 30:10:10 ratio. Each knee MRI scan, which was acquired in Proton Density (PD) and Fast Spin Echo (FSE) sequences in the sagittal plane, underwent preprocessing. Images were converted from DICOM to NIFTI format and subsequently imported into ITK-SNAP software for annotation purposes. We utilized ITK-SNAP version 4.0.1, an open-source application widely used for manual and automated anatomical structure segmentation (23).

In this study, the first author performed manual segmentation on 50 MRI scans for each region of interest (ROI), encompassing both intact menisci and those with tears in the medial and lateral regions. The segmentation process involved meticulously outlining the contours of the structures manually to ensure accurate delineation of the anatomical features.

After training, the U-Net model was applied to compute ROIs using a test set comprising 150 2D images from 15 menisci. Once the model

was fine-tuned, we proceeded with the automated meniscus segmentation process.

The "ground truth" was defined as the precise and accurate annotation or labeling of meniscus structures in medical images, typically performed by expert radiologists or experienced orthopedic surgeons. These annotations served as a benchmark for assessing the performance of the automated segmentation algorithm. Ground truth data was created by manually delineating the meniscus on knee MRI scans. During the training phase, the ground truth data was used as labeled input for supervised learning, which enabled the model to predict segmentations that closely align with the expert-annotated ground truth.

Evaluation of segmentation performance

To evaluate the performance of the U-Net model comprehensively, the following metrics were employed:

- **accuracy** – it is a fundamental performance metric that quantifies the proportion of correct predictions relative to the total number of predictions; it provides an overall measure of the model's effectiveness in classification tasks with the following formula (24):

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

- **confusion matrix** – it is a detailed tool for assessing classification model performance, also referred to as an error matrix (Table 1), which categorizes predictions into four key outcomes as follows: a) true positives (TP), including the number of instances where the model correctly identified the presence of the condition; b) true negatives (TN), including the number of instances where the model correctly predicted the absence of the condition; c) false positives (FP), including the number of instances where the model incorrectly predicted the presence of the condition; and d) false negatives (FN), including the number of instances where the model failed to identify the condition, despite its presence.

These outcomes allow for the calculation of advanced metrics, including precision (P), recall

(R), F1-score and accuracy, which collectively provide a nuanced understanding of the model's performance (24), as detailed below.

- **precision** – it measures the proportion of true positives among all positive predictions, offering insight into the accuracy of the model's positive classifications.

- **recall** – it is also known as sensitivity or the true positive rate (TPR) and evaluates the model's ability to correctly identify positive instances by calculating the proportion of true positives out of all actual positives (24) using the following formulas:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1-score** – it is the harmonic mean of precision and recall, with particular usefulness for imbalanced datasets, where one class significantly outweighs the other (Figure 6); it balances the trade-off between precision and recall, ensuring neither metric is overlooked; the formula is as follows:

$$F1 = \frac{2 * TP}{2 * TP + FP + FN}$$

- **Dice similarity coefficient (DSC)** – this statistical metric, also referred to as the Sørensen–Dice index or Dice coefficient, measures the similarity between two datasets and is widely used in medical imaging to quantify overlap between predicted and actual segmentations, providing a robust evaluation of segmentation accuracy (25); the formula for the Dice score is as follows:

$$\text{Dice score} = \frac{2 |M \cap A|}{|M| + |A|}$$

By employing the above-described metrics, the evaluation framework ensures a thorough analysis of the U-Net model's performance, emphasizing its strengths and identifying areas for improvement, particularly in the context of medical image segmentation.

RESULTS

In this section, we discuss the results regarding the accuracy of detection and segmentation, along with the computational performance of

TABLE 1. Confusion matrix

		Predicted	
		Yes	No
Actual	Yes	TP	FN
	No	FP	TN

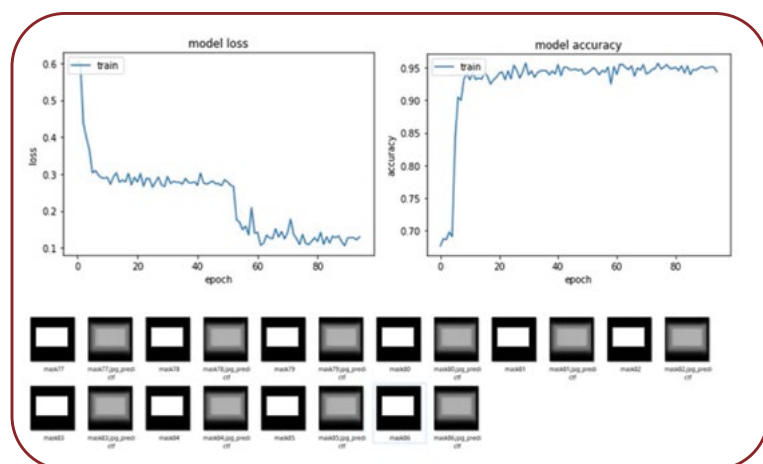


FIGURE 3. Graphical representation of the accuracy of the U-Net model and the output

TABLE 1. The performance results of the U-Net model

Number of tests	Precision	Recall	F1	Accuracy
Test 1	0.36	1.0	0.53	0.45
Test 2	0.41	1.0	0.58	0.55
Test 3	0.37	1.0	0.54	0.47
Test 4	0.36	1.0	0.53	0.44
Test 5	0.41	1.0	0.58	0.55
Test 6	0.0007	0.01	0.001	0.99
Test 7	N/A	N/A	N/A	0.75
Test 8	0.0	0.0	0.0	0.95
Test 9	N/A	N/A	N/A	0.99
Test 10	0.74	0.99	0.81	0.78
Test 11	0.68	0.99	0.76	0.73
Test 12	0.84	0.99	0.90	0.92
Test 13	N/A	N/A	N/A	N/A
Test 14	0.81	0.99	0.89	0.91

FIGURE 4. Diagram of the Dice score and cross entropy loss at 500 epochs

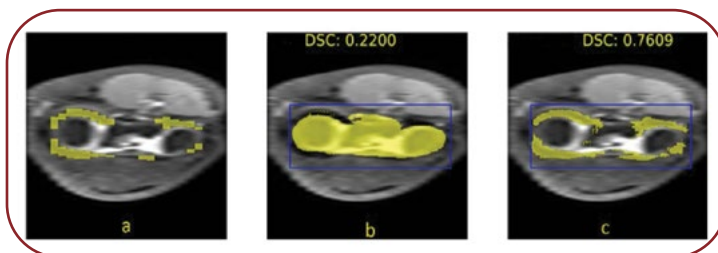
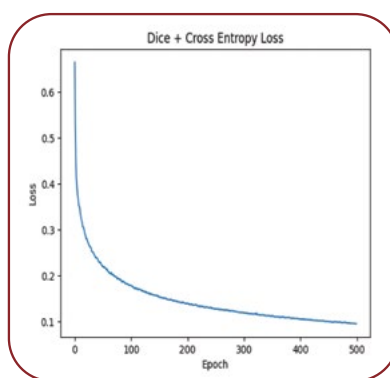


FIGURE 5. Segmentation of menisci: a) ground truth; b, c) automatically segmented menisci

our fully automated approach. After conducting 14 training tests of the U-Net model, we achieved a detection accuracy of 0.91 (Figure 3).

Detection of the meniscus after training the U-net model for 100 epochs provided the following results: - precision 0.8199845638709029; - recall 0.9988682663966786; - IoU 0.818865617977352; - accuracy 0.9125228881835937; and - F1 0.8903130266804892.

In the initial phase of the study, we conducted 14 training sessions. Table 2 summarizes the performance results of the U-Net model.

Segmentation

The average Dice score for ROIs was 0.7259 in the DL model trained at 100 epochs and 0.7525 in that trained at 300 epochs, while in the U-Net model it was 0.7609 for 500 epochs (Figures 4 and 5).

DISCUSSION

In this study, we introduced an efficient deep learning-based approach for knee meniscus segmentation, utilizing ROIs annotated by orthopedic surgeons as ground truth for evaluation. Our findings demonstrate that the proposed method achieves segmentation accuracy comparable to that of orthopedic surgeons. Specifically, the agreement between the U-Net model and the surgeons, quantified by the Dice score, was 0.769. However, a fundamental limitation of our study is the limited data availability. This issue could be mitigated through data augmentation techniques or transfer learning approaches, as demonstrated by previous research (26, 27). For instance, Olmez et al successfully fine-tuned a convolutional neural network (CNN) with only 100 meniscus images pre-trained for a different task, which has led to significant performance improvements. Similarly, Byra et al reported a Dice score of 0.818, surpassing our study's results. At the same time, Paprochi et al achieved Dice similarity index (DSI) values of 77.1% and 83.5% for medial and lateral menisci in individuals with osteoarthritis (OA), which highlighted the potential of alternative semi-automated methods (28). Additionally, Tack et al employed a 2D U-Net model and obtained a Dice score of 0.808 for medial and lateral meniscus segmentation, which reflected the ongoing advancements

in automated segmentation techniques (20). Despite these advancements, our approach of detecting and subsequently segmenting the meniscus from ROIs demonstrates potential for clinical applications. The automated segmentations consistently showed substantial similarity to manual segmentations, indicating the suitability of our method for quantitative meniscus analysis. Collaboration with musculoskeletal radiologists will ensure accurate segmentation as the algorithm evolves to address more complex meniscal structures and degenerated knees. Another limitation of our study is its focus on segmentation methods optimized for sagittal proton density (PD) sequences. Extending the proposed U-Net framework to other MRI sequences would require robust sequence-specific training datasets. Future research should investigate the applicability of our approach to coronal PD images and other MRI sequences, aiming to enhance its versatility and accuracy. In conclusion, the present study describes a practical deep learning-based method for knee meniscus segmentation, which has been validated against expert annotations from orthopedic surgeons. While challenges such as data scarcity and sequence-specific optimization persist, our approach holds promise for advancing automated meniscus segmentation and analysis in clinical practice.

CONCLUSIONS AND FUTURE WORK

The primary objective of our study was to develop a deep-learning model capable of automatically detecting and segmenting the meniscus from ROIs in knee MRI scans. We fine-tuned a U-Net model using meniscus images from only 50 knee MRIs. Although convolutional neural networks (CNNs) typically require large datasets for practical training, this challenge can be mitigated through transfer learning, allowing CNNs

to adapt quickly to similar tasks with smaller datasets (26). Recent advancements in deep learning have enabled the analysis of large volumes of MRI scans for knee patients. These algorithms learn patterns from individual scans and generalize this knowledge to new cases. The continued evolution of such automated techniques has significant potential to enhance the diagnosis and management of meniscal injuries and disorders. Future efforts should prioritize expanding the dataset to include a broader range of cases, improving the model's accuracy and robustness. Moreover, integrating these algorithms into clinical workflows could provide real-time diagnostic support, streamlining the diagnostic process and improving patient outcomes. By advancing automated methods for knee MRI analysis, this technology could play a transformative role in orthopedic care. 

Author's contribution: A.B. conceptualized the study and oversaw data collection. J.M.P. Jr. was involved in the manual segmentation of the menisci, while J.M.P. provided expertise in knee MRI diagnostics, reviewing and resolving issues during data analysis. A.B. prepared the initial manuscript draft, which was subsequently reviewed and refined by M.K. and J.M.P. M.K. and D.R.P. offered computational support for the development of the U-Net model and significantly contributed to the interpretation and presentation of the results. All authors reviewed, approved and contributed to the final version of the manuscript.

Acknowledgments: We extend our gratitude to the Emergency County Hospital of Alba Iulia, Romania, for granting access to the knee MRI database (2016–2019) following approval from the Ethical Board.

Conflicts of interest: none declared.

Financial support: none declared.

REFERENCES

1. Trunz LM, Morrison WB. MRI of the Knee Meniscus. *Magn Reson Imaging Clin N Am* 2022;30:307-324. doi: 10.1016/j.mric.2021.11.009.
2. Yang YP, Ma X, An H, et al. Meniscus repair with simultaneous anterior cruciate ligament reconstruction: Clinical outcomes, failure rates and subsequent processing. *Chin J Traumatol* 2022;25:37-44. doi: 10.1016/j.cjtee.2021.09.005.
3. Makris EA, Hadidi P, Athanasiou KA. The knee meniscus: structure-function, pathophysiology, current repair

- techniques, and prospects for regeneration.
Biomaterials 2011;32:7411-7431.
doi: 10.1016/j.biomaterials.2011.06.037.
4. **Cameron HU, Macnab I.** The structure of the meniscus of the human knee joint.
Clin Orthop Relat Res 1972;89:215-219.
 5. **Newman AP, Anderson DR, Daniels AU, Dales MC.** Mechanics of the healed meniscus in a canine model.
Am J Sports Med 1989;17:164-175.
doi: 10.1177/036354658901700205.
 6. **Zhu W, Chern KY, Mow VC.** Anisotropic viscoelastic shear properties of bovine meniscus.
Clin Orthop Relat Res 1994;306:34-45.
 7. **Tissakht M, Ahmed AM, Chan KC.** Calculated stress-shielding in the distal femur after total knee replacement corresponds to the reported location of bone loss.
J Orthop Res 1996;14:778-785.
doi: 10.1002/jor.1100140515.
 8. **Voloshin AS, Wosk J.** Shock absorption of meniscectomized and painful knees: a comparative in vivo study.
J Biomed Eng 1983;5:157-161.
doi: 10.1016/0141-5425(83)90036-5.
 9. **Walker PS, Erkman MJ.** The role of the menisci in force transmission across the knee.
Clin Orthop Relat Res 1975;109:184-192.
doi: 10.1097/00003086-197506000-00027.
 10. **Pandya S, Melville DM.** Evaluation of the knee joint with ultrasound and magnetic resonance imaging.
J Ultrason 2023;23:e239-e250.
doi: 10.15557/jou.2023.0032.
 11. **Lecouvet F, Van Haver T, Acid S, et al.** Magnetic resonance imaging (MRI) of the knee: Identification of difficult-to-diagnose meniscal lesions.
Diagn Interv Imaging 2018;99:55-64.
doi: 10.1016/j.diii.2017.12.005.
 12. **Young BL, Ruder JA, Trofa DP, Fleischli JE.** Visualization of Concurrent Anterolateral and Anterior Cruciate Ligament Injury on Magnetic Resonance Imaging.
Arthroscopy 2020;36:1086-1091.
doi: 10.1016/j.arthro.2019.09.039.
 13. **Zebis MK, Aagaard P, Andersen LL, et al.** First-time anterior cruciate ligament injury in adolescent female elite athletes: a prospective cohort study to identify modifiable risk factors.
Knee Surg Sports Traumatol Arthrosc 2022;30:1341-1351.
doi: 10.1007/s00167-021-06595-8.
 14. **Rahman MM, Dürselen L, Seitz AM.** Automatic segmentation of knee menisci - A systematic review.
Artif Intell Med 2020;105:101849.
doi: 10.1016/j.artmed.2020.101849.
 15. **Swanson MS, Prescott JW, Best TM, et al.** Semi-automated segmentation to assess the lateral meniscus in normal and osteoarthritic knees.
Osteoarthr Cartil 2010;18:344-353.
doi: 10.1016/j.joca.2009.10.004.
 16. **Kessler DA, MacKay JW, Crowe VA, et al.** The optimisation of deep neural networks for segmenting multiple knee joint tissues from MRIs.
Comput Med Imaging Graph 2020;86:101793. doi: 10.1016/j.compmedimag.2020.101793.
 17. **Zhou Z, Zhao G, Kijowski R, Liu F.** Deep convolutional neural network for segmentation of knee joint anatomy.
Magn Reson Med 2018;80:2759-2770.
 18. **Ronneberger O, Fischer P, Brox T.** U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention* 2015, pp 234-241.
 19. **Norman B, Padoia V, Majumdar S.** Use of 2D U-Net Convolutional Neural Networks for Automated Cartilage and Meniscus Segmentation of Knee MR Imaging Data to Determine Relaxometry and Morphometry.
Radiology 2018;288:177-185.
 20. **Tack A, Mukhopadhyay A, Zachow S.** Knee menisci segmentation using convolutional neural networks: data from the Osteoarthritis Initiative.
Osteoarthr Cartil 2018;26:680-688.
 21. Artificial intelligence in radiology (Internet). SlideShare. 2020. Available from: <https://www.slideshare.net/devlakhera/artificial-intelligence-in-radiology>
 22. Artificial Intelligence in Medical Imaging (Internet). SlideShare. 2020. Available from: <https://www.slideshare.net/ZahidulIslamJewel2/artificial-intelligence-in-medical-imaging>
 23. ITK-SNAP Version 4.x Downloads (Internet). Available from: <http://www.itksnap.org/pmwiki/pmwiki.php?n=Downloads.SNAP4>
 24. **Tigges S.** Diagnostic Test Metrics and Receiver Operating Characteristic Curves: The Graphic Novel.
J Am Coll Radiol 2022;19:1034-1041.
doi: 10.1016/j.jacr.2022.06.011.
 25. **Moore C, Bell D.** Dice similarity coefficient. Reference article, Radiopaedia.org (Accessed on 01 Jul 2024). <https://doi.org/10.5334/rld-75056>
 26. **Ölmez E, Akdoğan V, Korkmaz M, Er O.** Automatic Segmentation of Meniscus in Multispectral MRI Using Regions with Convolutional Neural Network (R-CNN).
J Digit Imaging 2020;33:916-929.
doi: 10.1007/s10278-020-00329-x.
 27. **Byra M, Wu M, Zhang X, et al.** Knee menisci segmentation and relaxometry of 3D ultrashort echo time cones MR imaging using attention U-Net with transfer learning.
Magn Reson Med 2020;83:1109-1122.
doi: 10.1002/mrm.27969.
 28. **Paproki A, Engstrom C, Chandra SS, et al.** Automated segmentation and analysis of normal and osteoarthritic knee menisci from magnetic resonance images--data from the Osteoarthritis Initiative.
Osteoarthr Cartil 2014;22:1259-1270.
doi: 10.1016/j.joca.2014.06.029.

