ORIGINAL ARTICLE Musculoskeletal imaging

Automated Segmentation of Knee Joint Cartilage: A Clinical Comprehensive Study Evaluating the Feasibility of pyKNEEr

Prizov Aleksey Petrovitch[1,2], Lutsenko Artyom Mikhailovich[1,3], Abdelrazzaq Khaled A. Jaafreh Alhabashneh[1], Brashich Stefan Aleksandrovich[1,3], Karpenko Alik Viktorovich[3], Lazko Fedor Leonidovich[1,2].

[1]*Department of Traumatology and Orthopaedics, Peoples' Friendship University of Russia named after Patrice Lumumba, Moscow, Russian Federation,*

[2]*Moscow City Clinical Hospital, named after V.M. Buyanova, Moscow, Russian Federation,*

[3]*Zhukovsky Regional Clinical Hospital, Zhukovsky, Moscow region, Russian Federation*

SUBMISSION: 14/04/24 - ACCEPTANCE: 05/09/2024

ABSTRACT

Objective: This study aims to evaluate the feasibility and accuracy of the Python Knee Cartilage Image Analysis Workflow (pyKNEEr), an open-source tool for automated segmentation of standard sagittal magnetic resonance imaging (MRI) in assessing femoral knee joint cartilage tissue changes, in comparison with the established Whole-Organ Magnetic Resonance Imaging Score (WORMS) and actual arthroscopic findings.

Materials and methods: This study, conducted between January and October 2022, involved a cohort of 10 patients with varying degrees of femoral bone cartilage changes. The patients underwent knee arthroscopy for internal meniscal damage. Sagittal MRI tomograms were analyzed using pyKNEEr v0.0.5 for cartilage tissue measurements, and manual assessment was performed using the WORMS scale. Statistical data processing was performed using Pingouin 0.5.3 and Numerical Python (NumPy) 1.24.2 for Python 3.9 (Python Software Foundation, Delaware, USA).

Results: The pyKNEEr analysis revealed an average total cartilage thickness of 2.26 ± 0.21 mm, (2.33 ± 0.26 mm for men, 2.22 ± 0.19 mm for women), and an average total cartilage volume of 10242.2 ± 1860.75 mm³, (10,380.25 ± 2,654.41 mm³ for men, 10,150.17 ± 1,406.89 mm³ for women).

A statistically significant strong inverse correlation

CORRESPONDING AUTHOR, GUARANTOR

Abdelrazzaq Khaled A. Jaafreh Alhabashneh MBBS, Traumatology and Orthopaedics Resident at RUDN University
Address: Mikluho-Maklaya st., Moscow, Russian Federation, 117198.
Telephone: +79601241195 Email address: Abdelrazzaq133@gmail.com

was found between cartilage thickness and WORMS score (r=-0.813, 95% CI -0.95 to -0.38, p=0.025). Additionally, a moderate inverse correlation was observed between cartilage volume and WORMS score (r=-0.777, 95% CI -0.94 to -0.29, p=0.049). No statistically significant correlations were identified by using the ICRS scale.

Furthermore, there was no significant association between cartilage thickness and volume as determined using pyKNEEr.

Conclusion: pyKNEEr for automated segmentation of standard sagittal MRI images, demonstrates alignment with the WORMS scale, but neither pyKNEEr's automated segmentation nor the WORMS scale showed a statistically significant correlation with the arthroscopic depiction of cartilage defects.

KEY WORDS

cartilage segmentation, Hyaline cartilage, Arthroscopic findings, Pykneer, WORMS scale

Introduction

The preoperative determination of hyaline cartilage thickness and the severity of its damage plays a pivotal role in decision-making regarding surgical treatment strategies. Standard assessment involves magnetic resonance imaging (MRI) of the knee joint, facilitating the evaluation of cartilage and subchondral bone status. Radiologists and orthopedic surgeons typically assess cartilage conditions based on their expertise [1,2]. For a more precise assessment of changes, it is possible to utilize specialized scales, such as the Whole-Organ Magnetic Resonance Imaging Score (WORMS) and Magnetic Resonance Observation of Cartilage Repair Tissue (MOCART) [3,4]. Obtaining specific data on the condition of cartilaginous tissue is achievable through the segmentation of MR images, accompanied by the calculation of the thickness of the cartilage at each slice. The manual analysis method has proven to be effective, but its application is limited by the considerable time and effort required by experts, up to 20 min per MRI slice, and a lot of effort for analyzing every individual MRI slice [5-7]. Automated medical image segmentation techniques provide a solution to this problem.

To date, approximately 30 algorithms have been designed to segment the femoral part of the knee joint [8]. These algorithms are based on different methods such as the active contour method, atlas-based (previously labeled images) method, graph-based method, machine and deep learning method, and hybrid combinations of methods, which have been developed by different research groups worldwide. Several algorithms have open-source codes that are publicly available in repositories [8-10]. It should be noted that segmentation algorithms developed by Wang and Shan have some limitations; for example, documentation on both code usage is limited, whereas, in another situation, the code itself is written using C++, requiring advanced programming skills for both compilation and execution. As a result of these limitations, we opted to use Bonaretti's algorithm which has good documentation available along with Python source code.

One of the free tools with open-source code is Python Knee Cartilage Image Analysis Workflow (pyKNEEr), developed by a team from Stanford and Palo Alto (United States) and led by Serena Bonaretti. Segmentation is carried out in three stages: 1) search for a reference image, 2) highlight the femoral cartilage, and 3) assess the segmentation quality. In the first stage, based on the reference image, convergence analysis is performed to search for a reference image on sagittal T2-mode tomograms by iterations until the convergence of intensity vectors, considering the femur bone and cartilage masks. The second stage, based on the atlas method, involves highlighting the femoral cartilage. The third stage involves evaluating the quality of the obtained data. The accuracy of the algorithm was studied on 3 MRI databases of the knee joint «OAI1», «OAI2» and «inHouse», where the algorithm showed a strong correlation with test data (Pearson's correlation coefficient 0.958) and correspondence of the obtained images to the test data (Dice coefficient 0.86) for cartilage segmentation and morphology assessment.

Our study aimed to determine the Feasibility and accuracy of applying pyKNEEr in clinical practice. Specifically, to evaluate the changes in thickness and volume within the femoral knee cartilage, comparing our findings with established scales such as WORMS, and assessing the relevance of our results in relation to data obtained during knee arthroscopy.

Materials and Methods

Between January and October 2022, a cohort of 34 patients with varying degrees of femoral bone cartilage changes were candidates for study, 10 were included in the study. All the patients provided informed consent for the use of their data. Planned surgical interventions were performed on all patients using knee arthroscopy to address internal meniscal damage. The inclusion criteria for this study encompass the following parameters: age ranging from 18 to 65 years, presence of knee joint MRI with a magnetic field strength of 1.5 Tesla, and the requirement for operative treatment using arthroscopy. Exclusion Criteria: MRI artifacts: patients with artifacts on MRI images will be excluded from the study. Prior Knee Interventions: Patients who had undergone previous surgical or interventional procedures on the knee joint were not included. Time interval: Patients with a time interval exceeding 6 months between the MRI assessment and planned operative intervention will be excluded. Chronic disease: patients known to have endocrine disorders will be excluded to ensure that no other conditions will interfere with the controlled MRI assessment.

In all patients, photo and video documentation of the arthroscopic view of the femoral condyle cartilage was performed, followed by defect assessment according to the International Cartilage Repair Society (ICRS) criteria. Sagittal MRI tomograms were analyzed using pyKNEEr version 0.0.5., to obtain measurements of the average thickness and volume of cartilage tissue. Additionally, a manual assessment of the femoral bone cartilage was performed using the WORMS scale. These two methods were independently performed by two orthopedic surgeons. Any discrepancies were resolved collaboratively by a third more experienced orthopedic surgeon. Data for each femoral bone cartilage zone was recorded in a table and subsequently aggregated. Statistical data processing was performed using the statistical libraries Pingouin 0.5.3 and NumPy 1.24.2 for Python 3.9 (Python Software Foundation, Delaware, USA). The Kolmogorov–Smirnov test (a non-parametric method for comparing distributions, essential for various applications in diverse fields) was used to assess the normality of the distribution. For normally distributed quantitative data, results are presented as mean ± standard error, whereas for non-normally distributed data, quantitative results are presented as median and interquartile range (25th and 75th percentiles). Spearman correlation (a statistical measure that assesses the monotonic relationship between two ranked variables) was conducted between the thickness and volume results of cartilage tissue obtained from pyKNEEr and the WORMS scale, as well as the ICRS, considering the non-parametric nature of the data.

The study incorporated six females and four males. The median age of the patients was 42 ± 18 years (range, 24–60 years). The right knee joint was affected in five patients and the left knee joint was affected in five patients. The duration between MRI examination and surgical treatment was 1.7 months (ranging, from 0.25 to 1.75 months), with a minimum of 1 day and a maximum of 6 months. Four examinations were conducted using a Toshiba scanner (Japan), two on Siemens (Germany), two on Philips (Japan), and two on GE scanners (USA). All scanners had a magnetic field strength of 1.5 Tesla and adhered to the standard protocol for knee joint examination using a phased array knee coil. A positioning device was used to ensure uniform placement of the knee among patients, T2-weighted images in the sagittal planes were acquired, using the following pulse sequence parameters: time to recovery (TR) of 3100 ms, time to echo (TE) of 75 ms, slice thickness of 2 mm, and field of view of 14 cm.

The associated knee joint injuries included 10 medial meniscal tears, 8 patellofemoral chondromalacia (average ICRS grade 3 B, ranging from ICRS grade 2 to ICRS grade 3C), 10 tibial plateau chondromalacia (1 patient with ICRS grade 1, 3 patients with ICRS grade 2, 3 patients with ICRS grade 3A, 1 patient with ICRS grade 3 B, and 2 patients with ICRS grade 3C), 1 partial anterior cruciate ligament tear, 2 lateral meniscus tears, and synovitis of the knee joint.

Table 1 Results of assessing cartilage tissue according to WORMS scale.

Patient №	MCA, mm	MCC, mm	MCP, mm	LCA, mm	LCC, mm	LCP, mm
1	4	4	3	1	5	3
2	1	5	3	4	3	3
3	1	5	3	1	4	4
4	1	3	3	3	4	3
5	3	5	4	1	1	1
6	1	5	5	1	3	3
7	1	5	1	1	3	3
8	3	5	3	1	3	3
9	5	5	3	1	1	3
10	1	5	1	1	1	3

Results

The assessment according to the WORMS scale for each of the six zones of the femoral trochlea (MCA: medial anterior condyle; MCC: medial central condyle; MCP: medial posterior condyle; LCA: lateral anterior condyle; LCC: lateral central condyle; and LCP: lateral posterior condyle) is presented in Table No. 1 the average score for the femoral condyle compartments was as follows: MCA 2.4 ± 1.6, MCC 4.6 ± 0.7, MCP 2.9 ± 1.2, LCA 1.5 ± 1, LCC 2.6 ± 1.5, LCP 2.7 ± 2.8.

The results from the pyKNEEr analysis are as follows: The average total cartilage thickness was 2.26 ± 0.21 mm. Among males, it was 2.33 ± 0.26 mm, while among females, it was 2.22 ± 0.19 mm. The total average cartilage volume was 10,242.2 ± 1,860.75 mm^3. Among males, it was 10,380.25 ± 2,654.41 mm^3, while among females, it was 10,150.17 ± 1,406.89 mm^3.

The clinical assessment of the femoral condyle was conducted through arthroscopy with photo and video documentation, followed by an evaluation of the damage according to the standard ICRS scale. Defects were identified in the cartilage of the femoral condyle, specifically in the medial compartment in 8 patients and the lateral compartment in 2 patients. All defects were classified as stage 3 according to the ICRS scale. Among these, four patients were at stage 3A, one patient at stage 3 B, and five patients at stage 3C. Figure 1 shows the arthroscopic view and the corresponding cartilage map of the femoral condyle, generated using PyKNEEr, where there is visual correspondence between the localization and depth of the defect in the medial femoral condyle.

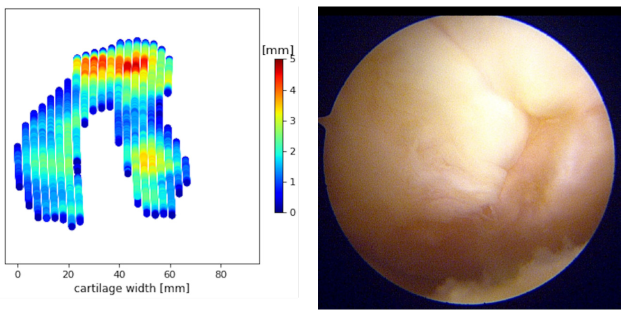

Figure 1. *Results of the colour map by pyKNEEr and the arthroscopic image of the medial compartment of the femoral knee of patient № 5.*

The average score for femoral knee cartilage on the WORMS scale was 16.7 ± 2.8. The results for each patient on the WORMS scale, as well as the pyKNEEr results and the ICRS assessment, are presented in Table 2.

Correlation analysis was performed using Spearman's rank correlation coefficient, adjusted for multiple comparisons using Bonferroni correction. The resulting correlation map is shown in Figure 2.

The summary data on the results of Spearman's correlation calculation (ρ), along with the 95% confidence interval (CI), are presented in Table 3.

A statistically significant strong inverse correlation was found between cartilage thickness and WORMS score (r=-0.813, 95% CI -0.95 to -0.38, p=0.025). Additionally, a moderate inverse correlation was observed

Table 2 WORMS Results, pyKNEEr Results, and Arthroscopic ICRS Evaluations

Patient №	Age, years	WORMS Femoral scores	pyKNEEr Thickness, mm	PyKNEEr Volume, mm³	ICRS
1	60	20	1.91	8713	3C
2	60	19	2.15	8454	3A
3	24	20	2.23	9572	3A
4	58	17	2.2	10463	3C
5	26	15	2.29	12284	3C
6	54	14	2.32	10310	3B
7	46	14	2.47	10677	3C
8	52	18	2.08	8511	3C
9	52	18	2.31	9130	3A
10	52	12	2.7	14308	3A

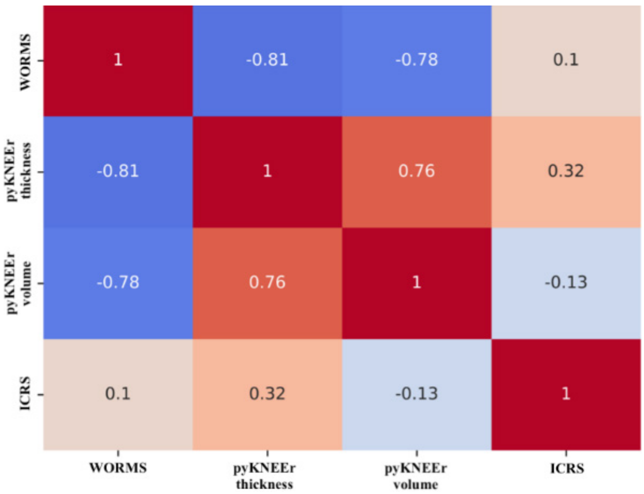

Figure 2. *Correlation map based on Spearman's rank correlation coefficients.*

between cartilage volume and WORMS score (r=-0.777, 95% CI -0.94 to -0.29, p=0.049). No statistically significant correlations were observed using the ICRS scale. Furthermore, there was no significant correlation between cartilage thickness and volume as determined using pyKNEEr.

Discussion

The statistically significant correlation between the pyKNEEr results and WORMS scores indicates that both MRI-based diagnostic techniques align with the MRI images of the patients. The existence of an inverse correlation raises a question. Considering that the WORMS scale assigns lower scores to more severe cartilage damage, the observed inverse correlation suggests an overestimation of the pyKNEEr results. We attribute this to defects in MRI segmentation using pyKNEEr and incorrect delineation of cartilage defects, resulting in an increase in cartilage tissue volume during calculations rather than a decrease. The moderate strength of the correlation between cartilage volume and WORMS scores is likely related to the resolution capabilities of stranded MRI scans and the greater slice thickness.

The lack of correlation with the ICRS arthroscopic classification is particularly noteworthy, not only for pyKNEEr, but also for WORMS. Both MRI-based methods do not exhibit a statistically significant correlation with the actual arthroscopic findings.

Importantly, there are alternative scoring systems that demonstrate good results when compared to WORMS. One such system is the Boston Leeds Osteoarthritis Knee Score (BLOKS) [11]. In various studies, BLOKS has been compared to WORMS, revealing that both systems yield equivalent information and exhibit high inter-reader reliability. However, it's important to acknowledge that BLOKS' scoring of cartilage lesions is more intricate and sophisticated than WORMS, which does require substantial effort during assessment [12–14].

Additionally, the Cartilage Lesion Score (CALS) de-

Table 3 Results of the correlation analysis

Variable 1	Variable 2	ρ	CI95%	p-value
WORMS	pyKNEEr thickness	-0.813	[-095 -0.38]	0.025
WORMS	pyKNEEr volume	-0.777	[-0.94 -0.29]	0.049
WORMS	ICRS	0.101	[-0.56 069]	1.0
pyKNEEr thickness	pyKNEEr volume	0.758	[0.24 0.94]	0.067
pyKNEEr thickness	ICRS	0.322	[-0.39 0.79]	1.0
pyKNEEr volume	ICRS	-0.127	[0.7 0.55]	1.0

ρ: the strength and direction of association between two ranked variables.
CI95%: 95% confidence interval from range of upper and lower number calculated from a sample.
p- value: the probability of obtaining test results at least as extreme as the result actually observed.

serves mention. In a comprehensive study, CALS demonstrated promising results alongside WORMS and BLOKS. This scale is considered a reproducible and valid scoring system for cartilage lesions, with improved sensitivity in detecting longitudinal changes compared to the established WORMS and BLOKS systems [15].

The primary advantage of employing pyKNEEr lies in its ability to enhance efficiency [8], while being correlated with established scales like WORMS as demonstrated by the results of our study.

To our knowledge no studies have directly compared existing automated-segmentation methods with actual arthroscopic images; these algorithms have been developed and tested on pre-annotated ideal data sets. Our study revealed an accuracy gap when comparing these results to the actual arthroscopic findings. Consequently, there exists a critical need to address this gap in the development of future scales or enhance existing ones to ensure greater relevance to actual cartilage defects.

A potential limitation of the study was the small sample size of patients, which could have influenced the absence of statistical correlation between the established WORMS assessment method and the arthroscopic images. Additionally, there may have been insufficient resolution in MRI studies that are commonly employed in routine clinical practice.

Conclusions

Automated segmentation of standard sagittal MRI images using pyKNEEr enables the evaluation of changes in femoral knee joint cartilage tissue. This method aligns with widely recognized assessment methods, such as WORMS. However, it has a calculation flaw, in addition to an increase in cartilage volume and thickness in the presence of defects. Notably, neither pyKNEErs automated segmentation nor the WORMS scale, based on MRI diagnostics, demonstrated a statistically significant correlation with the arthroscopic depiction of cartilage defects. **R**

References

1. Pedoia V, Majumdar S, Link TM. Segmentation of joint and musculoskeletal tissue in the study of arthritis. Magma N Y N. 2016 Apr; 29(2): 207–21.
2. Wirth W, Eckstein F. A Technique for Regional Analysis of Femorotibial Cartilage Thickness Based on Quantitative Magnetic Resonance Imaging. IEEE Trans Med Imaging. 2008 Jun; 27(6): 737–44.
3. Peterfy CG, Guermazi A, Zaim S, et al. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. Osteoarthritis Cartilage. 2004 Mar 1; 12(3): 177–90.
4. Schreiner MM, Raudner M, Marlovits S, et al. The MOCART (Magnetic Resonance Observation of Cartilage Repair Tissue) 2.0 Knee Score and Atlas. Cartilage. 2021 Dec; 13(1 Suppl): 571S-587S.
5. More S, Singla J, Abugabah A, et al. Machine Learning Techniques for Quantification of Knee Segmentation from MRI. Sarfraz DrS, editor. Complexity. 2020 Dec 7; 2020: 1–13.
6. Rahman MM, Dürselen L, Seitz AM. Automatic segmentation of knee menisci – A systematic review. Artif Intell Med. 2020 May 1; 105: 101849.
7. G.A. Airapetov, A.A. Vorotnikov, E.A. Konovalov, et al. Making a thickness map of the human knee hyaline cartilage. ulyanovsk Medico-Biological J. Available via: https://cyberleninka.ru/article/n/sozdanie-karty-tolschiny-gialinovogo-hryascha-kolennogo-sustava-cheloveka/viewer published 2024 Apr 13
8. Bonaretti S, Gold GE, Beaupre GS. pyKNEEr: An image analysis workflow for open and reproducible research on femoral knee cartilage. Olier I, editor. PLOS ONE. 2020 Jan 24; 15(1): e0226501.
9. Wang Q, Wu D, Lu L, et al. Semantic Context Forests for Learning-Based Knee Cartilage Segmentation in 3D MR Images. In: Menze B, Langs G, Montillo A, Kelm M, Müller H, Tu Z, editors. Medical Computer Vision Large Data in Medical Imaging. Cham: Springer International Publishing; 2014. p. 105–15.
10. Shan L, Charles C, Niethammer M. Automatic Atlas-based Three-label Cartilage Segmentation from MR Knee Images. Proc Workshop Math Methods Biomed Image Analysis. 2012; 241–6.
11. Hunter DJ, Lo GH, Gale D, Grainger AJ, Guermazi A, Conaghan PG. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston–Leeds Osteoarthritis Knee Score). Ann Rheum Dis. 2008 Feb;67(2):206–11.
12. Lynch JA, Roemer FW, Nevitt MC, Felson DT, Niu J, Eaton CB, et al. Comparison of BLOKS and WORMS scoring systems part I. Cross sectional comparison of methods to assess cartilage morphology, meniscal damage and bone marrow lesions on knee MRI: data from the osteoarthritis initiative. Osteoarthritis Cartilage. 2010 Nov;18(11):1393–401.
13. Felson DT, Lynch J, Guermazi A, Roemer FW, Niu J, McAlindon T, et al. Comparison of BLOKS and WORMS scoring systems part II. Longitudinal assessment of knee MRIs for osteoarthritis and suggested approach based on their performance: data from the Osteoarthritis Initiative. Osteoarthritis Cartilage. 2010 Nov;18(11):1402–7.
14. Hafezi-Nejad N, Zikria B, Eng J, Carrino JA, Demehri S. Predictive value of semi-quantitative MRI-based scoring systems for future knee replacement: data from the osteoarthritis initiative. Skeletal Radiol. 2015 Nov;44(11):1655–62.
15. Alizai H, Virayavanich W, Joseph GB, Nardo L, Liu F, Liebl H, et al. Cartilage Lesion Score: Comparison of a Quantitative Assessment Score with Established Semiquantitative MR Scoring Systems. Radiology. 2014 May 1;271(2):479–87.

READY - MADE CITATION

Prizov Aleksey Petrovitch, Lutsenko Artyom Mikhailovich, Abdelrazzaq Khaled A. Jaafreh Alhabashneh, Brashich Stefan Aleksandrovich, Karpenko Alik Viktorovich, Lazko Fedor Leonidovich. Automated Segmentation of Knee Joint Cartilage: A Clinical Comprehensive Study Evaluating the Feasibility of pyKNEEr. Hell J Radiol 2024; 9(4): 12-18.