Seminars in Arthritis and Rheumatism 79 (2026) 153020

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

AI automated grid placement in the OMERACT knee inflammation MRI scoring system (KIMRISS) for bone marrow lesion assessment: A multi-reader exercise

Steel M. McDonald [a,*], Stephanie Wichuk [a], Rory Gilliland [a], Soolmaz Abassi [a], Debra M. Baranec [b], Constance M. Lebrun [c], Joel Paschke [d], Abhilash Hareendranathan [a], Dan Durham [a], Dorian Nobbee [e], Maria S. Stoenoiu [f], Ulrich Weber [g], Matthew Budak [h], Omar Azmat [a], Paul Benvenuto [a,i], Chris Rothe [h], Robert G.W. Lambert [a,d,i], Walter P. Maksymowych [d,j], Jacob Jaremko [a,d,i]

[a] Department of Radiology & Diagnostic Imaging, University of Alberta, Edmonton, AB, Canada
[b] Canadian Institutes of Health Research, Institute of Musculoskeletal Health and Arthritis, Toronto, ON, Canada
[c] Department of Family Medicine, University of Alberta, Edmonton, AB, Canada
[d] CARE Arthritis Ltd., Edmonton, AB, Canada
[e] Department of Radiology, University of Calgary, Calgary, AB, Canada
[f] Department of Internal Medicine, Rheumatology Institut de Recherche Expérimentale et Clinique, Cliniques Universitaires Saint-Luc, Université Catholique de Louvain, Rheumatology, Brussels, Belgium
[g] Medical Centre Zenit, Department of Rheumatology, Bleicheplatz 3, Schaffhausen, Switzerland
[h] Qscan Radiology Clinics, Gold Coast, QLD, Australia
[i] Medical Imaging Consultants, Edmonton, AB, Canada
[j] Department of Medicine, University of Alberta, Edmonton, AB, Canada

ARTICLE INFO

Keywords:
OMERACT
Osteoarthritis
Semiquantitative scoring systems
Bone marrow lesions
Magnetic resonance imaging
Knee joint
Artificial intelligence

ABSTRACT

Objective: To validate the reliability and feasibility of AI-automated grid placement for bone marrow lesion (BML) scoring in the Knee Inflammation MRI Scoring System (KIMRISS) using the OMERACT Filter.
Methods: Eleven experts evaluated 40 MRI cases using manual and automated grid placement. Grids were compared both directly using spatial similarity metrics and indirectly using agreement metrics calculated on resulting KIMRISS BML scores. Feasibility was assessed using the System Usability Scale (SUS).
Results: In most regions, automatically- and manually-placed grids demonstrated strong spatial similarity (e.g., mean femur Dice Coefficient = 0.78) and KIMRISS BML score agreement (mean intraclass correlation coefficients of 0.86 and 0.89 for baseline and change scores, respectively). SUS scores for automated grid placement were moderate (mean = 66.1).
Conclusion: Automated grid placement is a reliable and feasible improvement to KIMRISS that could improve the ease and reproducibility of quantifying osteoarthritis in clinical trials.

Introduction

Knee osteoarthritis (OA) significantly impacts mobility and quality of life [1]. Bone marrow lesions (BML) have been associated with cartilage degradation, especially if large and persistent, and have therefore been considered targets for therapeutic intervention [2–4]. The histopathology of this lesion varies in different reports and includes both structural changes, such as trabecular fractures and infiltration with fibrous tissue, as well as inflammatory features, such as vascularization and infiltration with inflammatory cells [5,6]. Consequently, this lesion has not been formally matched with either the inflammation or structural damage domains defined for imaging outcomes but instruments

* Corresponding author.
E-mail address: steel1@ualberta.ca (S.M. McDonald).

https://doi.org/10.1016/j.semarthrit.2026.153020

Available online 12 June 2026

quantifying its extent have been developed [7]. Current BML scoring methods, such as the Magnetic Resonance Imaging (MRI) Osteoarthritis Knee Scoring System (MOAKS) [8] or the Knee Inflammation MRI Scoring System (KIMRISS) [9], are time intensive, require expert manual effort, and are subject to high inter-rater variability [10,11]. However, extensive data for MOAKS scores assessed by expert readers is provided in the Osteoarthritis Initiative archive and our research group, which developed the KIMRISS, has extensive data on this score [12].

Automating the scoring process for BML in knee OA with artificial intelligence (AI)-based methods could standardize BML scoring, reduce manual workload, and increase scoring efficiency, thereby supporting the need for accurate, accessible, and reproducible metrics in OA care [13,14]. A particularly tedious portion of manual scoring is placement of a grid template overlay to localize BML relative to other knee structures, which often takes longer than the actual BML scoring. Retaining a grid structure is helpful because there may be regional variation in the prognostic significance of BML. But after review of the literature and OMERACT Filter 2.2 requirements for instrument selection, the group decided that a new AI-based method incorporating a grid was necessary [15].

The primary objective of this study was to evaluate the validity of iKIMRISS—an AI-automated iteration of KIMRISS—in assessing OA-related inflammation according to the OMERACT Filter 2.2 [15]. Specifically, this study evaluated the automatic grid template placement feature of iKIMRISS compared to manual KIMRISS. Our evaluation aimed to assess (1) the reliability of automated grid placement and the KIMRISS BML scores it results in and (2) the feasibility of the feature, including its usability and efficiency.

Methods

Study design

This study was a multi-reader validation study on prospective longitudinal data with retrospective selection of subjects for one of the cohorts (OAI) [16] (publicly accessible at https://nda.nih.gov/oai). A two-part reading exercise was conducted remotely via REDCap (https://project-redcap.org/) with KIMRISS scoring hosted on the CARE Arthritis platform (https://www.carearthritis.com).

Reader demographics

Eleven experts (eight musculoskeletal radiologists and three rheumatologists) were tasked with applying KIMRISS scoring to fluid-sensitive sagittal knee MRI scans. Of the eleven readers, three were developers of the original KIMRISS system, three had extensive experience using KIMRISS prior to this exercise, and five were novice users of KIMRISS. The mean number of years of experience reading MRIs was 26 $\pm$13 for experienced users (n = 6) and 5 $\pm$ 5 for novice users (n = 5).

Image dataset

MRI data (Table 1) consisted of 40 cases with 2 timepoints each (baseline and 1-year follow-up). Images were sourced from two multicenter cohorts representing a diverse burden of OA: the Osteoarthritis Initiative (OAI) dataset [16] (publicly accessible at https://nda.nih.gov/oai) and the Osteoarthritis of the Knee, Inflammation, and the effect of Adalimumab (OKINADA) dataset [17]. In total, the dataset included 20 paired 2D turbo spin echo fat-suppressed scans from the OAI cohort and 20 paired proton density fat-saturated scans from the OKINADA trial.

To optimize evaluation of iKIMRISS discrimination, cases were deliberately selected to include substantial interval changes between baseline and one-year follow-up. As a result, the dataset is enriched for longitudinal change rather than representing an unselected natural history cohort. This is reflected in Table 1, where the absolute KIMRISS score changes are generally of similar magnitude to the baseline scores, indicating the presence of meaningful progression within the included cases.

Automated grid placement system

The automatic grid placement system is described in greater detail in the Supplementary Methods. In brief, the system uses an AI segmentation algorithm (Mask R-CNN) [18] to generate bone masks, to which a set of rules are applied to define grid template parameters. The templates are then propagated through the slices.

Study procedure

The reading exercise consisted of two parts. In part one, readers manually placed grid templates as instructed by Maksymowych et al. [19] but did not score BML. In part two, readers rated the quality of automatically placed grid locations and then assessed BML by marking the presence or absence of lesions within each grid square; these markings were used to calculate the iKIMRISS score. Readers completed both parts of the exercise on each case before moving onto the next one. Two readers that were system developers also completed manual grid placement and BML scoring of all cases.

Feasibility study

Following the exercise, participants were asked to complete the System Usability Scale [20,21] to evaluate the perceived utility of automated grid placement.

Reliability analysis

All quantitative analyses were conducted using Python 3.7 [22] and Medcalc (version 23.2.6).

Binary masks of the manually- and automatically-placed grids were compared spatially using Dice coefficient, Jaccard index, and Hausdorff

Table 1
Summary of the MRI datasets used in the study, including demographics, baseline BML score, and change characteristics.

Dataset				OAI			OKINADA			All
Mean (SD) Age at Baseline (years)				63.7 (9.3)			65.7 (7.9)			64.7 (8.6)
No. Males (%)				6 (30)			6 (30)			12 (30)
Mean (SD) Kellgren-Lawrence Grade at Baseline				0.8 (0.9)			2.6 (0.5)			1.7 (1.1)
Scoring Region	Femur	Tibia	Patella	Femur	Tibia	Patella	Femur	Tibia	Patella	
Mean (SD) Baseline KIMRISS Score	10.72 (19.16)	15.83 (23.81)	6.34 (8.31)	14.68 (19.35)	10.78 (10.43)	3.11 (4.01)	13.92 (15.87)	13.31 (18.32)	4.73 (6.65)	
Mean (SD) One-Year Absolute KIMRISS Score Change[†]	7.59 (6.91)	10.67 (21.48)	1.75 (2.52)	8.29 (15.91)	6.28 (8.52)	1.17 (1.06)	7.94 (12.11)	8.48 (16.28)	1.46 (1.93)	

[†] The dataset was intentionally enriched for cases demonstrating substantial one-year change to facilitate evaluation of longitudinal discrimination.

distance. For the two readers who completed full KIMRISS scoring for all cases using both manual and automated grid placement, intraclass correlation coefficients (ICCs; two-way random effects model, absolute agreement, single measures) were calculated to estimate agreement of both baseline and one-year-change KIMRISS scores. The inter-reader reliability of KIMRISS scoring with automated grid placement was assessed using pairwise ICCs.

Results

According to system developer evaluation, the mean (SD) baseline KIMRISS scores in the femur, tibia, and patella were 13.92 (15.87), 13.31 (18.32), 4.73 (6.65), respectively, with mean (SD) one year changes of 7.94 (12.11), 8.48 (16.28), and 1.46 (1.93) (Table 1).

Direct grid comparison

Spatial similarity metrics calculated between the manual and automatic grids are provided in Table 2. The mean Dice coefficient and Jaccard index across the whole knee was 0.76 and 0.63, respectively, indicating that automatically placed grids were generally similar to manually placed grids. These findings are consistent with survey ratings of automated grid placement quality, with readers reporting that automatically placed grids were displaced from their optimal position by at most a grid square 89% of the time (Supplementary Table 1). The anchor slices selected by the algorithm to define the grids tended to vary slightly more widely than those selected by human readers (especially at the medial and lateral edges of the femur and tibia), which may explain some of the discrepancy in the spatial similarity metrics (Supplementary Table 2).

A hand-selected sample of six representative slices comparing manual and automatic grid placement is presented in Fig. 1. In most cases, the automatic grids closely match a consensus of the 11 readers' manual grids, with occasional errors in the patella grid, the intercondylar femur grid, and the femur grid near the medial and lateral edges. Visualization also revealed that automatic grids sometimes failed to extend as far mediolaterally as manual grids, a discrepancy that may have affected the spatial similarity metrics (Table 2).

KIMRISS score comparison

Table 3 presents ICCs comparing KIMRISS scores from manually- and automatically-placed grids for the two readers who completed KIMRISS scoring with both methods. Agreement between methods was generally strong, with all ICCs exceeding 0.82 except patellar change (0.45 and 0.58). These findings were consistent with the results of a direct statistical comparison (paired t-tests) between the two readers' KIMRISS scores (Supplementary Table 3).

Bland-Altman plots comparing KIMRISS scores between the two grid placement methods reveal that automatic grid placement introduces moderate variability (limits of agreement = −29.6–23.5 and −19.9–31.7) but minimal bias (mean difference [95% CI] = −3.0 [−6.0–0.0] and 5.9 [2.9–8.8]) to KIMRISS scoring compared to the two manual readers (Supplementary Figure 1). Inter-reader reliability analyses (Supplementary Tables 4–11) demonstrate that KIMRISS scoring with automated grid placement is overall stable between readers, with mean pairwise ICCs exceeding 0.73 for all analyses except for one-year-change scores in the patella (mean ICC = 0.34).

Table 2
Spatial similarity metrics comparing eleven readers' manually and automatically placed KIMRISS grids across the 40 MRI cases. Hausdorff distance was calculated in pixels.

Scoring Region	Mean (SD) Dice Coefficient	Mean (SD) Jaccard Index	Mean (SD) Hausdorff Distance
Femur	0.78 (0.01)	0.67 (0.01)	19.3 (2.0)
Tibia	0.76 (0.03)	0.62 (0.04)	11.4 (4.0)
Patella	0.61 (0.10)	0.47 (0.10)	11.0 (3.4)
Whole Knee	0.76 (0.03)	0.63 (0.04)	22.7 (5.5)

Usability

System Usability Scale (SUS) scores of automatic grid placement indicated moderate overall usability, with mean scores of 66.1 ± 14.8. Experienced users (developers: 73.3 ± 12.3; previous KIMRISS users: 73.8 ± 11.2) reported higher usability than novice users (57.0 ± 14.1).

Time efficiency

Runtime analysis (detailed in Supplementary Results) revealed that automatic grid placement executed in ~7 s; meanwhile, manual grid placement took 413±82 s. Given that BML scoring took 400±136 s, automated grid placement reduced total KIMRISS scoring time by approximately half.

Discussion

In this multi-reader validation study, we evaluated the automatic grid placement feature of iKIMRISS, an AI-assisted tool designed to automate BML scoring based on the established KIMRISS system. Our results demonstrate promising evidence of reliability and feasibility, aligning with the OMERACT Filter and supporting the tool's potential for future use in clinical research, pending further refinement. Although 40 cases were included, each was independently evaluated by 11 readers, yielding a large number of paired assessments; however, this does not fully mitigate limitations related to the relatively small number of unique cases, particularly for region-specific and longitudinal analyses.

With respect to reliability, automatically placed KIMRISS grids demonstrated moderate-to-good spatial agreement with manual grids (whole knee Dice coefficient and Jaccard index of 0.76 and 0.63, respectively), with occasional discrepancies observed in the patella and femur grids. In the context of spatial overlap metrics, Dice coefficients in the mid-0.7 range and Jaccard indices around 0.6 generally reflect moderate agreement rather than near-perfect concordance, particularly for tasks involving anatomical partitioning where small boundary shifts can substantially affect overlap measures. This is further supported by lower agreement in subregions such as the patella, where smaller compartment size, complex geometry, and known anatomical variability may amplify minor misalignments [23]. Cohort-related anatomical variability, including differences across age and sex distributions, may also contribute to these observations.

Users generally endorsed the quality of the automatically placed grids, with readers reporting that automatically placed grids were displaced from their optimal position by at most a grid square 89% of the time. This high level of agreement in visual assessment underscores that the proposed framework achieves its primary objective of consistent regional localization for semi-quantitative scoring. Although spatial overlap metrics such as the Dice coefficient (0.76) and Jaccard index (0.63) may appear modest relative to fully automated segmentation benchmarks, these values should be interpreted in light of the intended application. The method is not designed for precise pixel-wise boundary delineation, but rather for reliable assignment of anatomical regions, where small positional deviations are unlikely to meaningfully affect scoring outcomes. In this context, qualitative expert evaluation demonstrating largely minimal grid displacement provides a more relevant measure of performance than strict pixel-level agreement metrics.

Manual and automatic grid placement methods were also similar regarding the KIMRISS BML scores they resulted in ICCs for both baseline and one-year-change KIMRISS scores show agreement (ICC>0.82)

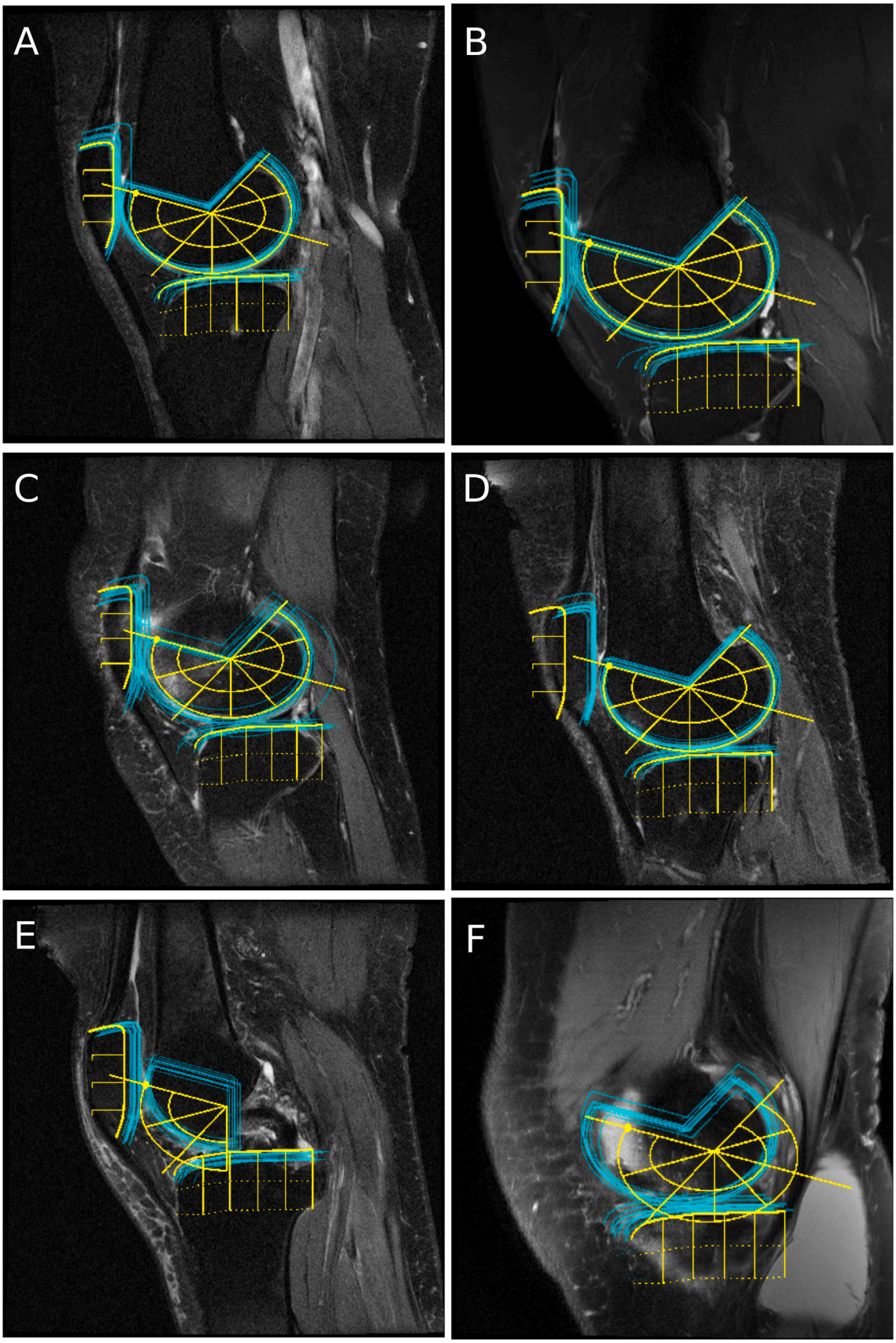

Fig. 1. A hand-selected sample of slices from the 40 MRI cases used in the study with overlays depicting the 11 readers' manual grid placements (cyan; internal grid lines have been removed for clarity) and iKIMRISS' automatic grid placement (yellow). Panes **A–C** depict examples of good automated grid positioning relative to manual consensus, while panes **D–F** exemplify three common grid mispositioning errors observed throughout the dataset: patella misidentification (**D**), intercondylar femur mispositioning (**E**), and grid misplacement at the medial femoral epicondyle (**F**).

Table 3
Intraclass correlation coefficients (ICCs; two-way random effects model, absolute agreement, single measures) comparing baseline and one-year-change KIMRISS scores for two readers using manual and automated grid placement.

Score Type	Reader	Scoring Region	ICC [95% CI]	No.(%) Cases with >|5| KIMRISS unit difference
Baseline	Expert 1	Femur	0.85 [0.74–0.92]	16 (40)
		Tibia	0.84 [0.72–0.91]	16 (40)
		Patella	0.93 [0.85–0.96]	5 (12.5)
		Whole Knee	0.83 [0.70–0.91]	29 (72.5)
	Expert 2	Femur	0.93 [0.87–0.96]	10 (25)
		Tibia	0.87 [0.76–0.93]	13 (32.5)
		Patella	0.95 [0.91–0.97]	1 (2.5)
		Whole Knee	0.88 [0.78–0.93]	25 (62.5)
Change	Expert 1	Femur	0.83 [0.70–0.91]	17 (42.5)
		Tibia	0.87 [0.76–0.93]	17 (42.5)
		Patella	0.45 [0.17–0.67]	2 (5)
		Whole Knee	0.88 [0.79–0.94]	32 (80)
	Expert 2	Femur	0.89 [0.81–0.94]	10 (25)
		Tibia	0.91 [0.84–0.95]	11 (27.5)
		Patella	0.58 [0.34–0.75]	4 (10)
		Whole Knee	0.90 [0.82–0.95]	24 (60)

between manual and automatic grid placement methods for almost all regions, with the exception of patellar one-year-change scores (ICCs of 0.45 and 0.58). This discrepancy is likely multifactorial, reflecting both the relatively low burden of BML in the patella (where small numbers of positive cases can disproportionately affect ICC estimates) and inherent anatomical and structural variability of the patella. Because the ICC is a ratio of between-case variance to total variance, it is highly sensitive to the magnitude of variability within the sample. In the patellar compartment, a relatively low BML burden and a narrower scoring range restrict between-case variability; in such contexts, even minor variations in positive findings can disproportionately deflate ICC estimates, despite absolute differences between methods remaining small (see Table 3). Furthermore, the patella's anatomical complexity and smaller size may amplify the visual impact of minor spatial shifts in grid positioning (see Fig. 1.D-F).With respect to feasibility, automated grid placement was found to reduce the total time for KIMRISS scoring by approximately half, and SUS scores demonstrated usability slightly below the average of 68 seen across a wide range of commercial computer tools [20,21]. Experienced users reported higher usability (developers: 73.3 ± 12.3; previous KIMRISS users: 73.8 ± 11.2), whereas novice users reported lower scores (57.0 ± 14.1), suggesting a possible influence of familiarity with the KIMRISS framework on perceived usability.

These findings contrast with a previously published evaluation of the KIMRISS method in which SUS scores were consistently high across readers, with a mean SUS of 85.7 [9]. This discrepancy may reflect differences in study design and system maturity. In the prior study, users evaluated the established manual KIMRISS workflow, whereas in the present study users additionally evaluated automated grid placement, introducing increased cognitive load. The lower usability reported by novice users likely reflects a learning curve associated with simultaneous interpretation of grid placement and scoring tasks, rather than limitations specific to the algorithm itself. Conversely, higher scores among experienced users suggest that automation is better appreciated once users are familiar with the underlying workflow. Targeted improvements to user onboarding, including step-by-step guidance and clearer visual cues during initial use, may help reduce this learning curve and improve usability among novice users. With improved onboarding, novice users may find the tool more acceptable, supporting the integration of iKIMRISS into clinical trial and research workflows.

There is currently no OMERACT-defined domain definition report for BML – that is, no structured document establishing a shared and comprehensive definition of BML as a domain (a distinct and clearly specified aspect of health, disease impact, or overall well-being that is important to measure). As a result, the conceptual boundaries of BML remain variable across studies, and we do not expect this to be fully clarified until new technologies can more clearly distinguish lesions that are inflammatory from those reflecting structural damage [7,24,25]. Key future aims of our work include evaluating other features of iKIMRISS, such as the AI-assisted BML segmentation tool, using the OMERACT Filter. Longitudinal studies assessing responsiveness to therapy and integration into clinical decision-making pipelines to address gaps in a Summary of Measurement Properties (SOMP) table will be essential to fully establish the utility of iKIMRISS in real-world settings. Future work will focus on three main areas: (i) methodological expansion through evaluation of additional iKIMRISS components such as AI-assisted BML segmentation, (ii) technical refinement of grid placement algorithms, particularly in anatomically complex regions such as the intercondylar and patellar compartments, and (iii) usability optimization for first-time users to improve adoption and workflow integration. Importantly, such refinements will be necessary before iKIMRISS can be confidently applied in clinical trial settings.

A limitation of the present work is that joint effusion/synovitis, which forms part of the KIMRISS scoring system, was not included in our automated framework. Rather than replicating semi-quantitative scoring for effusion/synovitis, our approach is directed toward fully quantitative, volumetric assessment of joint fluid, which may provide improved objectivity and sensitivity to change. Ongoing work from our group has demonstrated the feasibility of automated deep learning–based quantification of joint effusions in both the knee and hip, with preliminary validation through OMERACT initiatives [26,27]. Integration of these quantitative fluid measurements with automated structural scoring systems represents an important future direction for comprehensive, fully automated joint assessment.

Taken together, this study provides an important step toward partial automation of the KIMRISS workflow by demonstrating the feasibility and reliability of automated grid placement for BML assessment. Given that grid positioning represents a major time-consuming and operator-dependent component of manual scoring, successful automation of this step has the potential to meaningfully improve efficiency and standardization while preserving the established scoring framework. Importantly, this work was intentionally designed as an intermediate validation step to isolate the contribution of grid automation prior to extending automation to the scoring process itself. Future work will build on these findings by developing and evaluating fully automated end-to-end KIMRISS scoring systems, including lesion identification and quantification, to determine whether additional automation yields further improvements in accuracy, reproducibility, and clinical scalability.

Conclusion

Automated KIMRISS grid placement demonstrates promise in improving the speed and ease of BML scoring in knee OA without compromising quality compared to manual grid placement. However, given that agreement with manual methods is good but not yet optimal,

further algorithmic refinement and validation are required before iKIMRISS can be considered for use in clinical trials. Continued development and broader validation will be important next steps toward establishing iKIMRISS as a reliable and scalable OMERACT-endorsed instrument for clinical trials and OA research.

CRediT authorship contribution statement

Steel M. McDonald: Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Data curation, Conceptualization, Writing – review & editing. **Stephanie Wichuk:** Writing – review & editing, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Rory Gilliland:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis. **Soolmaz Abassi:** Writing – review & editing, Validation, Software, Methodology, Investigation. **Debra M. Baranec:** Writing – review & editing, Validation, Methodology. **Constance M. Lebrun:** Writing – review & editing, Conceptualization. **Joel Paschke:** Writing – review & editing, Software. **Abhilash Hareendranathan:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization. **Dan Durham:** Writing – review & editing, Investigation. **Dorian Nobbee:** Writing – review & editing, Investigation. **Maria S. Stoenoiu:** Writing – review & editing, Visualization, Validation, Investigation, Conceptualization. **Ulrich Weber:** Writing – review & editing, Investigation. **Matthew Budak:** Writing – review & editing, Investigation, Data curation. **Omar Azmat:** Writing – original draft, Investigation, Formal analysis. **Paul Benvenuto:** Writing – review & editing, Validation. **Chris Rothe:** Writing – review & editing, Investigation. **Robert G.W. Lambert:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Walter P. Maksymowych:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Jacob Jaremko:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgements & ethics

SMM is supported by an Alberta Innovates Graduate Student Scholarship. JLJ is supported by a Canada CIFAR AI Chair. Academic time for JLJ is made available by Medical Imaging Consultants (MIC), Edmonton, Canada. We thank all the readers that took part in this exercise. We thank the Patient Research Partners that have contributed to this project through co-designing and pilot testing the exercise as well as reviewing and editing the manuscript. We thank the members of the OMERACT MRI in Arthritis Working Group for their participation and support in this project. We also thank the participants of the MRI in Arthritis Special Interest Group (SIG) at OMERACT 2025 for their insights and feedback.

Ethics approval for this study was obtained from the University of Alberta's Human Research Ethics Board (Pro00147829).

The authors declare no conflicts of interest. All authors reviewed and approved the final manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.semarthrit.2026.153020.

References

[1] Long H, Liu Q, Yin H, et al. Prevalence trends of site-specific osteoarthritis from 1990 to 2019: findings from the global burden of disease study 2019. Arthritis Rheumatol 2022;74:1172–83.
[2] Ge L, Zhang X, Zhu R, Cai G. Bone marrow lesions in osteoarthritis: biomarker or treatment target? A Narrat Rev. Skeletal Radiol 2025;54:175–91.
[3] Moradi K, Mohammadi S, Roemer FW, et al. Progression of bone marrow lesions and the development of knee osteoarthritis: osteoarthritis Initiative data. Radiology 2024;312:e240470.
[4] Sofat N, Howe FA. Bone marrow lesions in osteoarthritis: characterising genetic and histological changes to understand disease pathophysiology. Osteoarthr Cartil Open 2024;6:100531.
[5] Muratovic D, Findlay DM, Cicuttini FM, Wluka AE, Lee YR, Kuliwaba JS. Bone matrix microdamage and vascular changes characterize bone marrow lesions in the subchondral bone of knee osteoarthritis. Bone 2018;108:193–201. https://doi.org/10.1016/j.bone.2018.01.012.
[6] Geurts J, Patel A, Hirschmann MT, Pagenstert GI, Müller-Gerbl M, Valderrabano V, Hügle T. Elevated marrow inflammatory cells and osteoclasts in subchondral osteosclerosis in human knee osteoarthritis. J Orthop Res 2016;34(2):262–9. https://doi.org/10.1002/jor.23009.
[7] D'Agostino MA, Beaton DE, Maxwell LJ, Cembalo SM, Hoens AM, Hofstetter C, Zabalan C, Bird P, Christensen R, de Wit M, Doria AS, Maksymowych WP, Oo WM, Østergaard M, Serban T, Sloan VS, Terslev L, van Rossum MA, Conaghan PG, Boers M. Improving domain definition and outcome instrument selection: lessons learned for OMERACT from imaging. Semin Arthrit Rheum 2021;51(5):1125–33. https://doi.org/10.1016/j.semarthrit.2021.08.004.
[8] Hunter DJ, Guermazi A, Lo GH, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). Osteoarthrit Cartilage 2011;19:990–1002.
[9] Maksymowych WP, Jaremko JL, Pedersen SJ, Eshed I, Weber U, McReynolds A, Bird P, Wichuk S, Lambert RG. Comparative validation of the knee inflammation MRI scoring system and the MRI osteoarthritis knee score for semi-quantitative assessment of bone marrow lesions and synovitis-effusion in osteoarthritis: an international multi-reader exercise. Ther Adv Musculoskelet Dis 2023;15. https://doi.org/10.1177/1759720X231171766. 1759720×231171766.
[10] Roemer FW, Jarraya M, Hayashi D, et al. A perspective on the evolution of semi-quantitative MRI assessment of osteoarthritis: past, present and future. Osteoarthrit Cartilage 2024;32:460–72.
[11] Hayashi D, Roemer FW, Jarraya M, Guermazi A. Imaging in osteoarthritis. Radiol Clin North Am 2017;55:1085–102.
[12] Driban JB, Patarini JC, Liu SH, McAlindon TE, Lapane KL, Lemon SC, Zai AH, Nevitt MC, Hochberg MC, Cauley JA, Eaton CB, Rubin S, Schneider E, Lo GH. The state of the osteoarthritis initiative (OAI): entering a new era. Semin Arthritis Rheum 2025;75:152887. https://doi.org/10.1016/j.semarthrit.2025.152887.
[13] Joseph GB, McCulloch CE, Sohn JH, Pedoia V, Majumdar S, Link TM. AI MSK clinical applications: cartilage and osteoarthritis. Skeletal Radiol 2022;51:331–43.
[14] McDonald SM, Felfeliyan B, Hassan A, Küpper JC, El-Hajj R, Wichuk S, Aneja A, Kwok C, Zhang CXY, Jans L, Herregods N, Hareendranathan AR, Jaremko JL. Evaluating potential for AI automation of quantitative and semi-quantitative MRI scoring in arthritis, especially at the knee: a systematic literature review. Skeletal Radiol 2025;54(11):2339–49. https://doi.org/10.1007/s00256-025-04922-5.
[15] Maxwell LJ, Beaton DE, Boers M, D'Agostino MA, Conaghan PG, Grosskleg S, Shea BJ, Bingham Iii CO, Boonen A, Christensen R, Choy E, Doria AS, Hill CL, Hofstetter C, Kroon FP, Leung YY, Mackie S, Meara A, Touma Z, Tugwell P, Wells GA. The evolution of instrument selection for inclusion in core outcome sets at OMERACT: filter 2.2. Semin Arthritis Rheum 2021;51(6):1320–30. https://doi.org/10.1016/j.semarthrit.2021.08.011.
[16] Eckstein F, Wirth W, Nevitt MC. Recent advances in osteoarthritis imaging–the osteoarthritis initiative. Nat Rev Rheumatol 2012;8(10):622–30. https://doi.org/10.1038/nrrheum.2012.113.
[17] Maksymowych WP, Bessette L, Lambert RG, Carapellucci A, Appleton CT. Osteoarthritis of the knee, inflammation, and the effect of adalimumab (OKINADA): a randomized placebo-controlled trial. Ann. Rheum. Dis. 2022;81:151–2.
[18] He K, Gkioxari G, Dollár P, Girshick R, Mask R-CNN. [Internet]. arXiv; 2018 [internet. Accessed June 18. Available from, http://arxiv.org/abs/1703.06870; 2025.
[19] Maksymowych WP, Jaremko JL, Pedersen SJ, et al. Comparative validation of the knee inflammation MRI scoring system and the MRI osteoarthritis knee score for semi-quantitative assessment of bone marrow lesions and synovitis-effusion in osteoarthritis: an international multi-reader exercise. Ther Adv Musculoskelet Dis 2023;15. 1759720×231171766.
[20] Brooke J. SUS – a quick and dirty usability scale. 1996. p. 189–94.
[21] Hyzy M, Bond R, Mulvenna M, et al. System usability scale benchmarking for Digital Health apps: meta-analysis. JMIR Mhealth Uhealth 2022;10:e37290.
[22] Python Software Foundation. Python [Internet]. Python.Org 2023 [internet. Accessed June 18. Available from, https://www.python.org/psf-landing/; 2025.
[23] Cox CF, Sinkler MA, Black AC, Launico MV, Hubbard JB. Anatomy, bony pelvis and lower limb, knee patella. StatPearls [Internet]. treasure island. FL): StatPearls Publishing; 2023. 2026 Jan–. PMID: 30137819.
[24] Maxwell, et al. OMERACT handbook, chapter 4; defining domains. Semin Arthritis Rheum; 2024.
[25] Maxwell LJ, Beaton DE, Shea BJ, Wells GA, Boers M, Grosskleg S, Bingham 3rd CO, Conaghan PG, D'Agostino MA, de Wit M, Gossec L, March L, Simon LS, Singh JA, Strand V, Tugwell P. Core domain set selection according to OMERACT filter 2.1:

the OMERACT methodology. J Rheumatol 2019;46(8):1014–20. https://doi.org/10.3899/jrheum.181097.

[26] Felfeliyan B, Wichuk S, Hareendranathan AR, Lambert RG, Maksymowych WP, Jaremko J. OMERACT validation of a deep learning algorithm for automated absolute quantification of knee joint effusion versus manual semi-quantitative assessment. Semin Arthritis Rheum 2024;66:152420. https://doi.org/10.1016/j.semarthrit.2024.152420. Epub 2024 Feb 17. PMID: 38422727.

[27] Jaremko JL, Felfeliyan B, Hareendranathan A, Thejeel B, Vanessa QL, Østergaard M, Conaghan PG, Lambert RGW, Ronsky JL, Maksymowych WP. Volumetric quantitative measurement of hip effusions by manual versus automated artificial intelligence techniques: an OMERACT preliminary validation study. Semin Arthritis Rheum 2021;51(3):623–6. https://doi.org/10.1016/j.semarthrit.2021.03.009. Epub 2021 Mar 17. PMID: 33781576.