frontiers in Musculoskeletal Disorders

TYPE Original Research
PUBLISHED 06 November 2026
DOI 10.3389/fmscd.2026.1929203

Check for updates

OPEN ACCESS

EDITED BY
Orazio De Lucia,
Istituto Ortopedico Gaetano Pini, Italy

REVIEWED BY
Carina Obermüller,
University Hospital Zürich, Switzerland
Rasmia Elgohary,
Cairo University, Egypt

*CORRESPONDENCE
Guillén Astete Carlos A
carlosantonio.guillen@salud.madrid.org

RECEIVED 06 July 2026
REVISED 12 October 2026
ACCEPTED 20 October 2026
PUBLISHED 06 November 2026

CITATION
Carlos A GA, María G-E, Eva G-N, Doryan G-O, Gonzalez-Montes A, Sonia J-B, Ana J L-I, Maria A-L, Boris B-C, Marta S-W, África A-S, Aliuska P-V, Marina T-C, Mauro F-S, Juan Carlos N-G and Mónica V-D (2026) Preliminary approach to validate ultrasonographic features of the perientheseal sesamoid fibrocartilage in the diagnosis of early spondyloarthritis. Front. Musculoskelet. Disord. 4:1929203. doi: 10.3389/fmscd.2026.1929203

COPYRIGHT
© 2026 Carlos A, María, Eva, Doryan, Gonzalez-Montes, Sonia, Ana J, Maria, Boris, Marta, África, Aliuska, Marina, Mauro, Juan Carlos and Mónica. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Preliminary approach to validate ultrasonographic features of the perientheseal sesamoid fibrocartilage in the diagnosis of early spondyloarthritis

Guillén Astete Carlos A[1,2,3]*, Gasset-Elices María[4], Gutierrez-Navarro Eva[4], García-Olivas Doryan[5], Alonso Gonzalez-Montes[1], Jiménez-Barrios Sonia[1], Lois-Iglesias Ana J[6], Ahijón-Lana Maria[7], Blanco-Cáceres Boris[1], Serrano-Warleta Marta[1,2], Andreu-Suarez África[8], Palomeque-Vargas Aliuska[9], Tortosa-Cabañas Marina[10], Ferre-Sanfrancisco Mauro[1], Nieto-Gonzalez Juan Carlos[11] and Vázquez-Díaz Mónica[1]

[1]Department of Rheumatology, Hospital Universitario Ramón y Cajal, Madrid, Spain, [2]Department of Rheumatology, Hospital BLUA Sanitas Valdebebas, Madrid, Spain, [3]Department of Medicine, Faculty of Medicine Health and Sports, Universidad Europea de Madrid, Madrid, Spain, [4]Facultad de Medicina, Universidad CEU San Pablo, Madrid, Spain, [5]Department of Rheumatology, Complejo Hospitalario Universitario Insular Materno Infantil de Gran Canaria, Las Palmas de Gran Canaria, Spain, [6]Department of Rheumatology, Complejo Hospitalario Universitario de A Coruña, A Coruña, Spain, [7]Department of Rheumatology, Hospital Universitario de la Princesa, Madrid, Spain, [8]Department of Rheumatology, Hospital Rio Hortega, Valladolid, Spain, [9]Department of Rheumatology, Hospital Clínico Universitario de Valladolid, Valladolid, Spain, [10]Department of Rheumatology, Hospital Reina Sofía de Murcia, Murcia, Spain, [11]Department of Rheumatology, Hospital Universitario Gregorio Marañón, Madrid, Spain

Objective: This study aimed to validate the sonographic features of the sesamoid fibrocartilage (SFC) as an imaging biomarker for spondyloarthritis (SpA) and to assess its diagnostic performance across two groups of rheumatologists with varying levels of ultrasound expertise.

Methods: A cross-sectional test-validation and observer-concordance study was conducted using a large archive of static images and short ultrasound videos of the Achilles tendon obtained between 2018 and 2025. Images corresponded to patients with SpA and to controls (non-SpA patients and healthy volunteers). All images were acquired before biological or local steroid therapy. Two observer tiers (experts and non-experts) with different levels of ultrasound experience evaluated three dichotomous SFC variables—continuity, linearity, and thickness—after a brief training module. Inter-observer agreement was analysed using kappa coefficients, considering patients and controls, and separately by diagnosis.

Results: A total of 75 images or videos from patients and 75 images from controls were collected. Expert observers showed the highest agreement for continuity (κ = 0.963, 95% CI: 0.916–1.010), followed by thickness (κ = 0.954, 95% CI: 0.901–1.007) and linearity (κ = 0.894, 95% CI: 0.855–0.933). Non-expert observers showed similar findings, with continuity showing the greatest reproducibility (κ = 0.924, 95% CI: 0.888–0.960). In the restricted analysis comparing patients with the 20 control records from 19 subjects referred for ultrasound for unrelated reasons, continuity and thickness remained the best-performing features, although perfect separation was no longer observed (expert κ: 0.82–0.90; AUC: 0.92–0.95). These findings confirm that part of the discrimination observed in the

full dataset is attributable to selection constraints applied to patient images. Across disease subtypes, continuity and thickness were positive in 100% of psoriatic arthritis, radiographic axial spondyloarthritis and non-radiographic axial spondyloarthritis cases, while linearity showed slightly lower positivity (93%–97%). These uniformly high positivity proportions explain why pairwise kappa coefficients between subtypes were mathematically constrained and uninformative. Overall, continuity emerged as the most robust and reproducible ultrasonographic feature, while linearity showed greater dependence on observer experience.

Conclusion: Sonographic assessment of the SFC demonstrates strong discriminative power in identifying SpA, even when performed by observers with limited ultrasound experience. Thickness and loss of continuity are the most robust diagnostic features, supporting the hypothesis that SFC reflects enthesic pathological changes. Given its accessibility and reliability, SFC evaluation may be a promising, minimally invasive imaging biomarker for early or preclinical detection of SpA. Further studies should explore its applicability in other enthesic territories.

KEYWORDS

diagnosis, enthesis, fibrocartilage, spondyloarthritis, ultrasound

Introduction

The enthesic organ is the physiopathological target of spondyloarthritis (SpA). It has been demonstrated in murine models that the key process that unleashes IL-17 and IL-22 activity depends on a T cell known as a resident memory T cell, which is located within the enthesis (1). The enthesis itself is the surface of anchorage of any kind of connective tissue on the cortical bone. The enthesic organ includes not only the enthesis but also other organs directly involved in the pathophysiological processes of SpAs: the cortical surface, the tendon, the bursae, and both fibrocartilages, sesamoid and periosteal (2, 3).

Although enthesic inflammation is central to the development of SpAs, some clinical manifestations result from its effects on synovial tissues: erosions, dactylitis, and hyperemia. This relationship is possible because enthesic and synovial territories are topographically close (bursae, tenosinovial tissue, or synovial joints) and are included in the synovioentheseal complex (4). Furthermore, the synovioentheseal complex can be identified easily in large entheses such as the Achilles or patellar tendon and also in as small an enthesis as the one formed in the anchorage of dorsal and flexor plates of the interphalangeal joints (5).

Sesamoid fibrocartilage (SFC) is a thin layer of tissue located between the superficial aspect of the bursae and the epitenon of the Achilles tendon. According to histological preparations, SFC is the projection of the uncalcified layer of the enthesis, which is rich in cells, predominantly chondrocytes (6). Unlike the enthesis tissue, the SFC is readily accessible for histological study due to its proximity to the pre-Achillean bursa. Recently, a single-centre study used the enthesic biopsy of the elbow epicondyle to analyse the biologic response to an anti-IL17 treatment in patients with SpA (7). Their results underlined the importance of enthesic study for precision therapy; however, an enthesic biopsy is, after all, an invasive and potentially painful procedure. If the SFC could provide the same information as the enthesis, it could be considered a very accessible target for preliminary diagnostic and therapeutic response studies.

Several ultrasound scoring systems have been developed to assess enthesitis, including GUESS (7) and MASEI (8). Although widely used, these iices rely predominantly on structural abnormalities such as erosions, calcifications and increased thickness, which may be absent in early spondyloarthritis. As a result, their sensitivity in early disease is limited. The sesamoid fibrocartilage of the Achilles tendon represents an alternative target that may be more consistently detectable in early stages, providing a potential advantage over conventional enthesis scores.

The aim of the present study is to validate the sonographic features of SFC as an image biomarker of SpA with two groups of rheumatologists who practice ultrasound: experts, occasional practitioners, and trainees.

Method

A cross-sectional descriptive study of test validation and observer concordance was conducted. The study population included 75 static images or short (3–5 s) videos of Achilles tendon ultrasound studies recorded from 2018 to 2025 in our rheumatology ultrasound clinic, from patients diagnosed with two forms of SpA: psoriatic arthritis (PsoA) or axial SpA (ax-SpA) who were assessed for symptoms at the Achilles insertion. The dataset consisted of 150 independent imaging records obtained from 150 distinct subjects, as detailed in Appendix 1, 2. Each subject contributed a single record (either a static image or a short dynamic video) of the Achilles enthesis. No subject contributed bilateral studies or multiple acquisitions. Analyses were therefore performed at the record level, with each observation considered independent. We obtained 75 control images or videos from patients not ultimately diagnosed with SpA (11 subjects), from patients without Achilles tendon pain (8 subjects) who were referred for a foot ultrasound for another reason, and from healthy models with no first-degree relative diagnosed with SpA (56 subjects). In all cases, the original ultrasound assessment and a second review for this study

reported no Achilles tendon impairment. To minimise confounding by structural damage, we excluded patient images showing any ultrasonographic sign of acute or chronic Achilles tendon involvement (power Doppler signal, erosions or enthesophytes). No images were excluded in the control group. Control subjects were included based on clinical diagnosis and original ultrasound reports, which did not assess the sesamoid fibrocartilage. The expert review used for anonymisation confirmed the absence of classical enthesitis signs but did not evaluate SFC morphology. Appendix 1 presents the composition of images/videos collected for this study, and Appendix 2 summarises the features of the patients whose images were collected for this study.

All images were obtained before any biological or local steroid treatment; however, we included images from six patients treated with non-steroidal anti-inflammatory drugs (3 patients), methotrexate (1 patient), leflunomide (1 patient), or sulfasalazine (1 patient); in all cases, treatment lasted less than 4 weeks.

Two rheumatologists with more than 10 years of experience in musculoskeletal ultrasound obtained all ultrasound images and videos using three different ultrasound devices: General Electric S8 (58 patients and 24 controls), Canon Aplio-a (12 patients and 34 controls), and ESAOTE MyLab (5 patients and 17 controls). In all cases, a linear probe of 12 to 14 MHz was used to obtain a longitudinal plane (long axis) of the Achilles enthesis that included the pre-achilleal bursae. Each record was labelled with its acquisition device (GE S8, Canon Aplio-a, ESAOTE MyLab). Device distribution across diagnostic classes was quantified and found to be unbalanced and associated with diagnostic category, with a predominance of GE S8 acquisitions among patient images. Device-specific agreement analyses were performed for continuity, thickness and linearity using the four independent readings of each expert. Two researchers extracted the multimedia archives and confirmed the SpA diagnosis or control condition. We processed all images and videos to anonymise them and remove date data.

The observer group included four rheumatologists considered experts (more than 5 years of musculoskeletal ultrasound experience) and seven non-experts observers. The seven non-expert observers were rheumatologists or rheumatology residents with 1–4 years of clinical experience in musculoskeletal ultrasonography. The sample size of 75 records per arm and the inclusion of 11 observers were chosen pragmatically to ensure stable kappa estimates and to allow separate evaluation of expert and non-expert performance. No formal precision target was predefined, but the selected numbers provided narrow observer-level confidence intervals and adequate variability for agreement analyses.

We initially contacted all observers to participate in the exercise and, after they agreed, trained them on SFC location and features with a 5 min educational video. The graded structure corresponded to the hyperechoic interphase line and its immediately subjacent anechoic layer, which together represent the sesamoid fibrocartilage which is located between the epitenon and the bursa (Figure 1). However, because the interphase line follows the shape of the SFC, we used it as a morphological reference (9). Each observer received three groups of images/videos over 45 days (75 from patients and 75 from controls), along with an electronic form to evaluate three dichotomous variables: continuity, linearity, and thickness. Continuity was defined as intact when no interruptions or localised thickenings were present. Any record with one or more interruptions or one or more localised thickenings was classified as disrupted. Thickness was evaluated as a separate attribute and referred to segmental thickening of the sesamoid fibrocartilage, defined as a sustained increase in thickness along a portion of the structure rather than a focal change. Records were classified as "thin" or "thick" according to this segmental criterion. Although a focal thickening constitutes a minor discontinuity and therefore affects continuity, segmental thickening represents a distinct morphological alteration. For this reason, continuity and thickness were analysed as separate features, each capturing different aspects of fibrocartilage structure.

Observers were encouraged to perform the assessments within 15 days of receiving each volume of images/videos. The interval between the training video and the first submission of images or videos was 72 h.

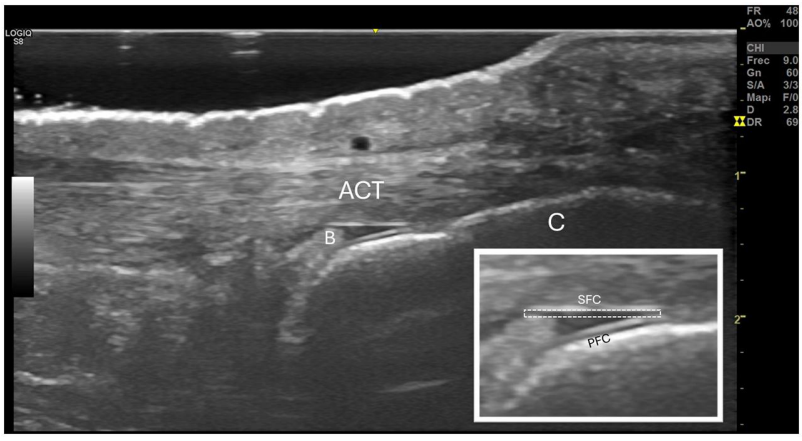

FIGURE 1
Longitudinal aspect of the ultrasound scan of the Achilles tendon enthesis, with a detail of the fibrocartilage sesamoid and periosteal, where the sesamoid fibrocartilage is located within the dotted rectangle. ACT, Achilles tendon body; B, Pre-Achilles Bursa; C, Calcaneus bone cortex; PFC, Periostial fibrocartilage; SFC, Sesamoid fibrocartilage.

10.3389/fmscd.2026.1929203

Statistical analysis

All statistical analyses were performed using Python (Python Software Foundation, Wilmington, DE, USA). Observer ratings were compared against the reference classification (patient vs. control). Agreement between each observer and the reference diagnosis was quantified using Cohen's kappa coefficient (κ), with values interpreted according to the criteria proposed by Landis and Koch. Diagnostic performance was additionally assessed by calculating accuracy, sensitivity and specificity.

The ultrasonographic evaluation of the sesamoid fibrocartilage of the Achilles tendon comprised three predefined features: thickness, continuity and linearity. All these features were assessed by the group of four expert observers. Non-expert observers also performed corresponding assessments. Table 1 and Figure 2 summarise the definitions of thickness, continuity, and linearity.

Inter-observer diagnostic performance was first evaluated separately for each observer. Subsequently, summary analyses were performed for each ultrasonographic feature by calculating the mean kappa coefficient within expert and non-expert groups. Confidence intervals for mean kappa values across observers were calculated using t-based intervals, given the small number of observers per tier. These intervals differ from the per-observer bootstrap confidence intervals reported in Table 2, which quantify uncertainty at the individual-observer level. Inter-observer agreement within each tier (experts and non-experts) was quantified using Fleiss' kappa for each ultrasonographic feature (continuity, thickness and linearity). To determine the overall diagnostic performance of each feature, we generated a consensus classification for expert observers using a majority-vote approach, requiring agreement among at least three of four experts to assign a positive classification. Diagnostic performance of the consensus assessment was evaluated through sensitivity, specificity, accuracy and area under the receiver operating characteristic curve (AUC). Consensus classifications produce a dichotomous test outcome. In this setting, the area under the curve (AUC) is mathematically equivalent to the balanced accuracy of the classifier, computed as (sensitivity + specificity)/2. AUC values are therefore reported for completeness, but should not be interpreted as threshold-independent ROC measures.

To explore the relative discriminative ability of the three ultrasonographic features, we initially planned logistic regression models. However, because expert consensus evaluations of thickness and continuity showed complete separation between cases and controls, we did not consider conventional logistic regression estimates informative, and we based comparative analyses primarily on measures of diagnostic accuracy and agreement.

Subgroup analyses were conducted to evaluate the prevalence of ultrasonographic abnormalities across disease subtypes. For each feature (continuity, thickness and linearity), positivity was defined using expert consensus: an image was considered positive when at least three of the four expert observers classified the feature as interrupted, thick or curved, respectively. Proportions of positive findings were calculated separately for psoriatic spondyloarthritis (PsoA), radiographic axial spondyloarthritis (Rx-SpA) and non-radiographic axial spondyloarthritis (NR-SpA), and exact binomial 95% confidence intervals were obtained. This approach was selected because pairwise kappa coefficients between subtypes are mathematically constrained when features are uniformly positive, making prevalence-based comparisons more appropriate than agreement metrics.

Finally, the ability of the assessed ultrasonographic features to discriminate between inflammatory disease subtypes was investigated through pairwise comparisons between PsoA and Rx-SpA, PsoA and NR-SpA, and Rx-SpA and NR-SpA. We calculated agreement statistics for each observer and summarised them by observer expertise and ultrasonographic feature.

A two-sided p-value < 0.05 was considered statistically significant.

TABLE 1 Definitions used for the three sesamoid fibrocartilage variables to be assessed by the observers, with the two possible attributes. The definitions did not specify whether either attribute was to be considered normal or pathological.

Feature	Definition	Possible attribute 1	Possible attribute 2
Thickness	Appearance of the thickness of the full visible length of the fibrocartilage in the longitudinal plane.	A hyperechoic line of the same or lesser thickness than a fibre of the adjacent tendon itself.	A hyperechoic line or segment of line thicker than a fibre of the adjacent tendon.
Continuity	The integrity of the fibrocartilage, including discontinuities or localised changes in thickness.	No interruptions or localised thickening.	One or more interruptions or one or more localised thickening along its visible length.
Linearity	Straight or curved shape	Straight	Curved or sinuous

Results

Observer agreement and diagnostic performance for distinguishing patients from controls

Diagnostic performance was evaluated using expert-consensus classifications, as detailed below. Consensus-based AUC values are reported in the corresponding section and in Table 4.

To compare the three features, we applied the Friedman test as specified in the Methods. The test showed a significant overall difference ($\chi^2 = 18.4$, $p < 0.001$). *Post-hoc* pairwise comparisons with Bonferroni correction indicated that continuity outperformed both thickness ($p = 0.004$) and linearity ($p < 0.001$), while thickness also outperformed linearity ($p = 0.03$). All reported p-values are two-sided, and statistical significance was defined as $p < 0.05$. Among the six patients receiving treatment, none were included among the discordant linearity classifications.

Expert observers demonstrated excellent agreement with the reference diagnosis. Using t-based confidence intervals appropriate for the small number of observers ($n = 4$), continuity achieved $\kappa = 0.963$ (95% CI: 0.916–1.010), thickness

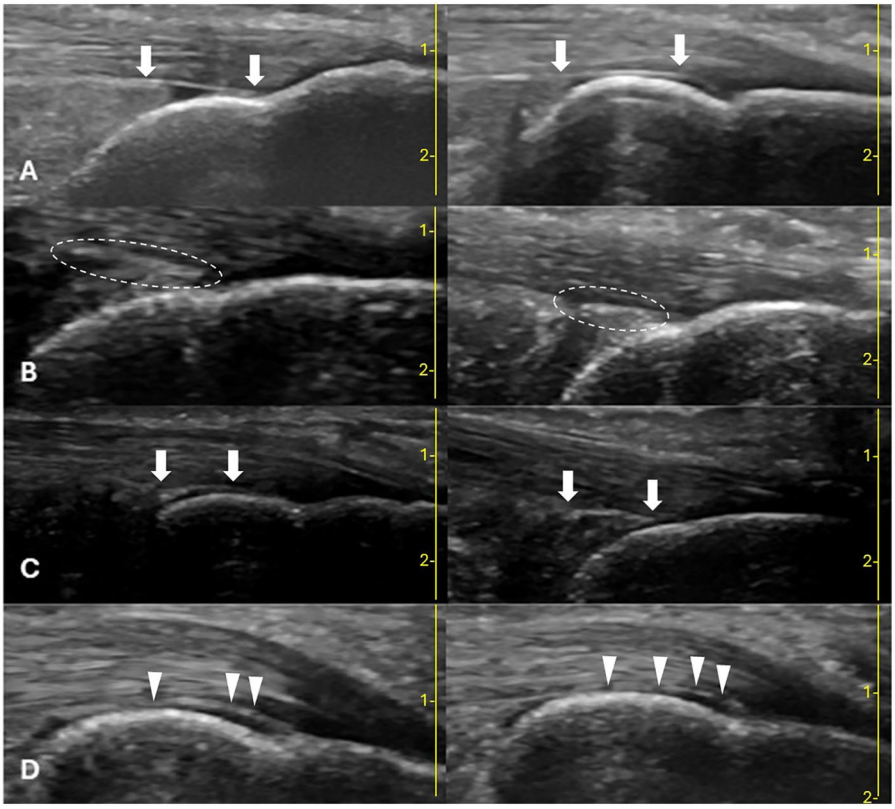

FIGURE 2
Examples of sesamoid fibrocartilage (SFC) used during observers' training. **(A)** Reference images of linear, thin and non-interrupted SFC limited by white arrows. **(B)** SFC considered thick due to enlargement of the complete visible SFC (left) and the proximal aspect (right) of the SFC. **(C)** Non-straight or curved SFC, limited by white arrows. **(D)** Interrupted SFCs pointed by white triangles. All images were obtained from a S8 GE ultrasound device. Depth of study is showed with a proper scale for each panel.

$\kappa = 0.954$ (95% CI: 0.901–1.007), and linearity $\kappa = 0.894$ (95% CI: 0.855–0.933). These intervals are wider than the normal-approximation values reported previously, reflecting the correct use of t-distribution for small-sample inference.

In the non-expert group, continuity remained the best-performing characteristic ($\kappa = 0.924$, 95% CI: 0.888–0.960), followed by thickness ($\kappa = 0.901$, 95% CI: 0.845–0.956) and linearity ($\kappa = 0.798$, 95% CI: 0.733–0.863). Table 2 presents the individual performance of both groups of observers.

Across all three evaluated characteristics, expert raters demonstrated substantial to almost perfect agreement, whereas non-experts showed moderate to substantial agreement. When all raters were combined, κ values remained consistently within the substantial agreement range. For thickness, Fleiss' κ was 0.89 among experts, 0.72 among the non-experts, and 0.83 when all raters were analysed together. For linearity, Fleiss' κ values were 0.86 for experts, 0.69 for non-expert observers, and 0.81 for the combined group. For continuity, agreement was highest, with Fleiss' κ values of 0.93 for experts, 0.75 for non-expert observers, and 0.88 for the combined raters.

Device-specific agreement analyses showed that diagnostic performance was consistent across ultrasound platforms. When stratifying the 150 records by device, continuity yielded kappa values ranging from 0.82 to 0.91, thickness from 0.78 to 0.88, and linearity from 0.74 to 0.86, with no systematic differences among GE S8, Canon Aplio-a, and ESAOTE MyLab. Agreement values stratified by device (κ: 0.82–0.91 for continuity) were consistently lower than the pooled expert estimate (κ: 0.963), indicating that device heterogeneity contributed modestly to variability but did not alter the overall diagnostic pattern.

Consensus-based diagnostic performance

A consensus evaluation generated from the majority vote of expert observers achieved perfect discrimination between patients and controls for both thickness and continuity. Sensitivity, specificity, diagnostic accuracy and AUC were 100% for both variables. Linearity demonstrated a sensitivity of 96.0%, specificity of 100%, accuracy of 98.0% and an AUC of 0.98; 95% CI: 0.96–1.00).

In the restricted analysis comparing patients with the 20 control records from 19 subjects referred for ultrasound for unrelated reasons, continuity and thickness remained the best-performing features, although perfect separation was no longer observed (expert κ: 0.82–0.90; AUC: 0.92–0.95). These findings confirm that part of the discrimination observed in the full dataset is attributable to selection constraints applied to patient images.

Disease subtype analyses

The image collection included 21 images from patients with PsoA, 15 with radiographic SpA, and 39 with non-radiographic SpA, in addition to 75 images considered controls (Appendix 1,

10.3389/fmscd.2026.1929203

TABLE 2 Individual performance of observers in assessing the three predefined variables.

Observer	κ (95% CI)	Sens. (%)	Spec. (%)	PPV (%)	NPV (%)
Thickness					
Expert 1	0.907 (0.838–0.960)	97.3	93.3	93.6	97.2
Expert 2	0.960 (0.906–1.000)	98.7	97.3	97.4	98.6
Expert 3	0.960 (0.907–1.000)	97.3	98.7	98.6	97.4
Expert 4	0.987 (0.959–1.000)	98.7	100.0	100.0	98.7
N-expert 1	0.973 (0.933–1.000)	98.7	98.7	98.7	98.7
N-expert 2	0.960 (0.907–1.000)	98.7	97.3	97.4	98.6
N-expert 3	0.933 (0.867–0.987)	97.3	96.0	96.1	97.3
N-expert 4	0.973 (0.933–1.000)	97.3	100.0	100.0	97.4
N-expert 5	0.827 (0.729–0.907)	89.3	93.3	93.1	89.7
N-expert 6	0.827 (0.732–0.907)	89.3	93.3	93.1	89.7
N-expert 7	0.813 (0.719–0.893)	92.0	89.3	89.6	91.8
Continuity					
Expert 1	0.920 (0.853–0.973)	96.0	96.0	96.0	96.0
Expert 2	0.973 (0.932–1.000)	98.7	98.7	98.7	98.7
Expert 3	0.973 (0.933–1.000)	98.7	98.7	98.7	98.7
Expert 4	0.987 (0.947–1.000)	98.7	100.0	100.0	98.7
N-expert 1	0.987 (0.959–1.000)	100.0	98.7	98.7	100.0
N-expert 2	0.947 (0.892–0.987)	97.3	97.3	97.3	97.3
N-expert 3	0.947 (0.892–0.987)	97.3	97.3	97.3	97.3
N-expert 4	0.960 (0.906–1.000)	97.3	98.7	98.6	97.4
N-expert 5	0.893 (0.813–0.960)	96.0	93.3	93.5	95.9
N-expert 6	0.853 (0.769–0.933)	90.7	94.7	94.4	91.0
N-expert 7	0.880 (0.796–0.947)	94.7	93.3	93.4	94.6
Linearity					
Expert 1	0.920 (0.853–0.973)	93.3	98.7	98.6	93.7
Expert 2	0.907 (0.837–0.960)	93.3	97.3	97.2	93.6
Expert 3	0.867 (0.775–0.933)	89.3	97.3	97.1	90.1
Expert 4	0.880 (0.799–0.947)	94.7	93.3	93.4	94.6
N-expert 1	0.840 (0.746–0.920)	89.3	94.7	94.4	89.9
N-expert 2	0.880 (0.800–0.947)	94.7	93.3	93.4	94.6
N-expert 3	0.893 (0.814–0.960)	92.0	97.3	97.2	92.4
N-expert 4	0.840 (0.747–0.920)	88.0	96.0	95.7	88.9
N-expert 5	0.707 (0.587–0.813)	85.3	85.3	85.3	85.3
N-expert 6	0.760 (0.653–0.853)	86.7	89.3	89.0	87.0
N-expert 7	0.667 (0.547–0.787)	84.0	82.7	82.9	83.8

N-expert, Non-expert; k, kappa coefficient; 95% CI, Bootstrap 95% confidence intervals; Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NPV, negative predictive value.

2). Across all disease subgroups, continuity consistently showed the highest diagnostic agreement, followed by thickness and linearity.

Continuity consistently demonstrated the highest observer agreement across all disease subtypes and levels of expertise. Linearity showed lower agreement and greater variability, particularly among non-expert observers (Table 3).

Diagnostic performance was also analysed separately for each disease subtype (psoriatic arthritis, radiographic axial spondyloarthritis and non-radiographic axial spondyloarthritis). Expert observers achieved complete separation between cases and controls for continuity and thickness across all three subgroups, with sensitivity, specificity and accuracy all reaching 100% and AUC values of 1.00 (95% CI: 1.00–1.00). Non-expert observers showed similarly high performance, with AUC values ranging from 0.97 to 0.99 for continuity and thickness. Linearity demonstrated lower discriminative ability, with expert AUC values of 0.98, 0.97 and 0.99 for PsoA, Rx-SpA and NR-SpA respectively, and non-expert AUC values ranging from 0.94 to 0.96. These results,

summarized in Table 4, confirm that continuity and thickness are highly robust diagnostic markers even in early disease, particularly within the non-radiographic axial SpA subgroup, which constituted the largest diagnostic category.

TABLE 3 Observer agreement for distinguishing individual disease subtypes from healthy controls.

Comparison	Variable	Experts (mean k)	Non-experts (mean k)
PsoA vs. Controls	Thickness	0.935	0.873
	Continuity	0.948	0.896
	Linearity	0.880	0.759
Rx-SpA vs. controls	Thickness	0.917	0.851
	Continuity	0.953	0.866
	Linearity	0.846	0.716
Non-Rx-SpA vs. controls	Thickness	0.942	0.890
	Continuity	0.956	0.917
	Linearity	0.903	0.784

Data are presented as mean Cohen's kappa coefficients for expert and non-expert observers. K, kappa coefficient; PsoA, psoriatic arthritis; Rx-SpA, radiographic spondyloarthritis; Non-Rx-SpA, Non-radiographic spondyloarthritis.

Discrimination between disease subtypes

Because continuity and thickness were consistently classified as positive by all four expert observers across PsoA, Rx-SpA and NR-SpA subtypes, pairwise kappa coefficients between disease groups were mathematically constrained and therefore uninformative. To provide a more appropriate comparison, we quantified the proportion of positive findings for each ultrasonographic feature using expert consensus (≥3 of 4 observers) (Table 5). Continuity and thickness were positive in 100% of PsoA (95% CI: 84%–100%), Rx-SpA (95% CI: 78%–100%) and NR-SpA patients (95% CI: 91%–100%). Linearity showed slightly lower positivity rates, with 95% in PsoA (95% CI: 76%–99%), 93% in Rx-SpA (95% CI: 68%–99%) and 97% in NR-SpA (95% CI: 86%–99%). These results indicate that the assessed features are uniformly present across SpA subtypes, including early disease, and explain the near-zero kappa values observed between subgroups.

Discussion

Ultrasound assessment of entheses in patients suspected of having SpA has been a relevant and recurrent research topic. Many

TABLE 4 Subgroup diagnostic performance of sesamoid fibrocartilage ultrasonographic features, showing sensitivity, specificity, accuracy and AUC (95% CI) for expert and non-expert readers across psoriatic arthritis (PsoA), radiographic axial spondyloarthritis (Rx-SpA) and non-radiographic axial spondyloarthritis (NR-SpA) categories.

Comparison	Observers	Sens.	Spec.	Accu.	AUC (95% CI)
Continuity					
PSOA vs. Control	Experts	100%	100%	100%	1.00 (0.95–1.00)[a]
	Non experts	98%	100%	99%	0.99 (0.97–1.00)
RX-AX vs. Control	Experts	100%	100%	100%	1.00 (0.95–1.00)[a]
	Non experts	97%	100%	98%	0.98 (0.95–1.00)
NR-AX vs. Control	Experts	100%	100%	100%	1.00 (0.95–1.00)[a]
	Non experts	96%	100%	98%	0.98 (0.96–1.00)
Thickness					
PSOA vs. Control	Experts	100%	100%	100%	1.00 (0.95–1.00)[a]
	Non experts	97%	100%	98%	0.98 (0.96–1.00)
RX-AX vs. Control	Experts	100%	100%	100%	1.00 (0.95–1.00)[a]
	Non experts	96%	100%	98%	0.98 (0.95–1.00)
NR-AX vs. Control	Experts	100%	100%	100%	1.00 (0.95–1.00)[a]
	Non experts	95%	100%	97%	0.97 (0.95–1.00)
Lineality					
PSOA vs. Control	Experts	95%	100%	98%	0.98 (0.95–1.00)
	Non experts	90%	100%	95%	0.95 (0.92–0.98)
RX-AX vs. Control	Experts	93%	100%	97%	0.97 (0.93–1.00)
	Non experts	88%	100%	94%	0.94 (0.90–0.97)
NR-AX vs. Control	Experts	97%	100%	99%	0.99 (0.97–1.00)
	Non experts	92%	100%	96%	0.96 (0.93–0.98)

Positive test attributes were defined as follows: thickness = "thick", continuity = "interrupted", and linearity = "curved". Diagnostic performance was calculated against the reference classification (patient vs. control). Subtype analyses compared each disease category (PsoA, Rx-SpA, NR-SpA) exclusively against all controls. AUC values and 95% confidence intervals were derived using non-parametric bootstrap resampling.

[a]Exact interval derived from binomial confidence intervals for sensitivity and specificity.

TABLE 5 Proportion of positive ultrasonographic findings by disease subtype (expert consensus ≥ 3/4). Positive findings were defined as "interrupted" for continuity, "thick" for thickness, and "curved" for linearity.

Feature	PsoA (n = 21)	Rx-SpA (n = 15)	NR-SpA (n = 39)
Continuity (interrupted)	100% (84%–100%)	100% (78%–100%)	100% (91%–100%)
Thickness (thick)	100% (84%–100%)	100% (78%–100%)	100% (91%–100%)
Linearity (curved)	95% (76%–99%)	93% (68%–99%)	97% (86%–99%)

Proportions represent the percentage of patients within each disease subtype exhibiting the positive attribute. Confidence intervals correspond to exact binomial 95% intervals. Controls were not included in this table, as the analysis focuses exclusively on differences between SpA subtypes.

scores, such as GUESS (7) or MASEI (8, 10, 11), assess multiple entheses, and for each enthesis, four or five variables are evaluated, including thickness, structural homogeneity, erosions, calcifications, power Doppler signal, and bursa distention (where applicable). These variables are consistent with the known ultrasonographic features of SpA (12–15), and both scores have been validated and are widely used, especially for diagnostic purposes. Variables with greater weight in terms of scoring are those related to well-established disease, such as erosions, large enthesophytes, or hyperemia; therefore, these scores are less useful in patients with early-stage disease, or even in preclinical settings, such as patients with psoriasis who have suspicious symptoms.

As we have stated previously, SFC appears to be a projection of the enthesis beyond the enthesic territory (3), and although this has not yet been demonstrated, its cellular composition could be comparable. Our results support this hypothesis, given the strong association between morphological changes in the SFC and SpAs. Furthermore, thickness and loss of continuity were the most relevant variable attributes related to the diagnosis. We hypothesise that increased thickness and loss of continuity may relate to heightened T-cell activity. Previous work has shown that enthesitis involves T-cell recruitment and cytokine-mediated remodelling of fibrocartilaginous structures (16). In addition, tissue-resident $\gamma\delta$T-cells within the human enthesis have been demonstrated to produce IL-17A independently of IL-23 signalling (17), and local IL-23 production by myeloid cells has also been identified (18). However, the present study did not include immunological assessments and therefore cannot establish a causal link between these ultrasonographic changes and specific immune pathways. Given that our selected image archive included patients recently diagnosed and not exposed to advanced treatment, characterising the SFC as a method for discriminating between SpA patients and controls seems at least as a promising avenue for further exploration. Taken together, the ultrasonographic abnormalities of the sesamoid fibrocartilage showed a uniform pattern across PsoA, Rx-SpA and NR-SpA subtypes, suggesting that these features represent a shared characteristic within the spectrum of spondyloarthritis rather than a subtype-specific imaging signature. The present study did not include patients with other inflammatory conditions, and therefore cannot determine whether these abnormalities are specific to SpA or reflect a broader inflammatory phenomenon. Establishing such distinctions would require comparison with inflammatory controls, which should be addressed in future studies. Thickness and continuity provided the most reliable discrimination between patients and controls, whereas linearity showed a greater dependence on observer expertise and did not offer additional discriminatory value between disease subtypes.

Another interesting consideration concerns the assessment's simplicity. Even among observers with less ultrasound experience, a brief educational video was sufficient to achieve an excellent concordance coefficient. In our experiment, assessing two variables appears reliable because the linearity assessment does not contribute significantly to the validation test.

This study has several limitations. First, although device allocation was not controlled and the distribution of ultrasound platforms was unbalanced and associated with diagnostic class, device-specific agreement analyses were performed and are reported in the Results. These analyses showed modest variability across platforms but did not alter the overall diagnostic pattern. Second, the study is based on static images and short videos acquired by two experienced ultrasonographers, and therefore validates image interpretation rather than the full acquisition-to-interpretation procedure. Third, because continuity and thickness were uniformly positive across all SpA subtypes, agreement metrics between subgroups were mathematically constrained; subtype comparisons were therefore based on positivity proportions rather than kappa coefficients. Finally, the study focused on the Achilles tendon, and additional entheses should be evaluated in future work.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Comité de ética e investigación científica del hospital universitario Ramón y Cajal. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

GACA: Formal analysis, Writing – review & editing, Project administration, Visualization, Data curation, Resources, Writing – original draft, Methodology, Validation, Investigation, Software, Funding acquisition, Supervision, Conceptualization. G-EM: Writing – original draft, Writing – review & editing. G-NE: Writing – original draft, Writing – review & editing. G-OD: Writing – original draft, Writing – review & editing. AG-M: Writing – original draft, Writing – review & editing. L-IA: Writing – original draft, Writing – review & editing. J-BS: Writing – original draft, Writing – review & editing. A-LM:

Writing – review & editing, Writing – original draft. B-CB: Writing – original draft, Writing – review & editing. S-WM: Writing – original draft, Writing – review & editing. A-SÁ: Writing – review & editing, Writing – original draft. P-VA: Writing – original draft, Writing – review & editing. T-CM: Writing – review & editing, Writing – original draft. F-SM: Writing – original draft, Writing – review & editing. N-GJ: Writing – review & editing, Writing – original draft. V-DM: Writing – review & editing, Writing – original draft.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Lories RJ, McInnes IB. Primed for inflammation: enthesis-resident T cells. *Nat Med*. (2012) 18:1018–9. doi: 10.1038/nm.2854

2. Benjamin M, Moriggl B, Brenner E, Emery P, McGonagle D, Redman S. The "enthesis organ" concept: why enthesopathies may not present as focal insertional disorders. *Arthritis Rheum*. (2004) 50:3306–13. doi: 10.1002/art.20566

3. Benjamin M, McGonagle D. The enthesis organ concept and its relevance to the spondyloarthropathies. *Adv Exp Med Biol*. (2009) 649:57–70. doi: 10.1007/978-1-4419-0298-6_4

4. McGonagle D, Aydin SZ, Tan AL. The synovio-entheseal complex and its role in tendon and capsular associated inflammation. *J Rheumatol Suppl*. (2012) 89:11–4. doi: 10.3899/jrheum.120233

5. Naredo E, Uson J, Olivas-Vergara O, Guillén-Astete C, Del Pozo Pozo G, Mérida-Velasco JR, et al. Sonoanatomy of the finger synovio-entheseal complexes. *Ultrasound Med Biol*. (2024) 50:1903–10. doi: 10.1016/j.ultrasmedbio.2024.08.018

6. McGonagle D, Lories RJU, Tan AL, Benjamin M. The concept of a "synovio-entheseal complex" and its implications for understanding joint inflammation and damage in psoriatic arthritis and beyond. *Arthritis Rheum*. (2007) 56:2482–91. doi: 10.1002/art.22758

7. Balint PV, Kane D, Wilson H, McInnes IB, Sturrock RD. Ultrasonography of entheseal insertions in the lower limb in spondyloarthropathy. *Ann Rheum Dis*. (2002) 61:905–10. doi: 10.1136/ard.61.10.905

8. De Miguel E, Cobo T, Muñoz-Fernández S, Naredo E, Usón J, Acebes JC, et al. Validity of enthesis ultrasound assessment in spondyloarthropathy. *Ann Rheum Dis*. (2009) 68:169–74. doi: 10.1136/ard.2007.084251

9. Wu W-T, Chang K-V, Hsu Y-C, Hsu P-C, Ricci V, Özçakar L. Artifacts in musculoskeletal ultrasonography: from physics to clinics. *Diagnostics*. (2020) 10:645. doi: 10.3390/diagnostics10090645

10. De Miguel E, Muñoz-Fernández S, Castillo C, Cobo-Ibáñez T, Martín-Mola E. Diagnostic accuracy of enthesis ultrasound in the diagnosis of early spondyloarthritis. *Ann Rheum Dis*. (2011) 70:434–9. doi: 10.1136/ard.2010.134965

11. Eder L, Jayakar J, Thavaneswaran A, Haddad A, Chandran V, Salonen D, et al. Is the MAdrid sonographic enthesitis index useful for differentiating psoriatic arthritis from psoriasis alone and healthy controls? *J Rheumatol*. (2014) 41:466–72. doi: 10.3899/jrheum.130949

12. Draghi F, Ferrozzi G, Ballerini D, Bortolotto C. Psoriatic arthritis: ultrasound peculiarities with particular emphasis on enthesitis. *J Clin Ultrasound*. (2022) 50:556–60. doi: 10.1002/jcu.23170

13. Di Matteo A, Filippucci E, Cipolletta E, Martire V, Jesus D, Musca A, et al. How normal is the enthesis by ultrasound in healthy subjects? *Clin Exp Rheumatol*. (2020) 38:472–8.

14. Mata Arnaiz MC, De Miguel Mendieta E. Usefulness of ultrasonography in the assessment of peripheral enthesis in spondyloarthritis. *Reumatol Clin*. (2014) 10:113–9. doi: 10.1016/j.reuma.2013.11.004

15. Agache M, Popescu CC, Popa L, Codreanu C. Ultrasound enthesitis in psoriasis patients with or without psoriatic arthritis, a cross-sectional analysis. *Medicina (B Aires)*. (2022) 58:1557. doi: 10.3390/medicina58111557

16. Schett G, Lories RJ, D'Agostino M-A, Elewaut D, Kirkham B, Soriano ER, et al. Enthesitis: from pathophysiology to treatment. *Nat Rev Rheumatol*. (2017) 13:731–41. doi: 10.1038/nrrheum.2017.188

17. Cuthbert RJ, Watad A, Fragkakis EM, Dunsmuir R, Loughenbury P, Khan A, et al. Evidence that tissue resident human enthesis γδT-cells can produce IL-17A independently of IL-23R transcript expression. *Ann Rheum Dis*. (2019) 78:1559–65. doi: 10.1136/annrheumdis-2019-215210

18. Bridgewood C, Watad A, Russell T, Palmer TM, Marzo-Ortega H, Khan A, et al. Identification of myeloid cells in the human enthesis as the main source of local IL-23 production. *Ann Rheum Dis*. (2019) 78:929–33. doi: 10.1136/annrheumdis-2018-214944

10.3389/fmscd.2026.1929203

Appendix 1. Composition of the records used by the observers for assessment.

Subject category	Subject diagnosis	Subcategory diagnosis	Type of record	Correlation number
Patient	Axial spondyloarthritis	Non-radiographic	Image	01 to 21
			Video	22 to 39
		Radiographic	Image	40 to 49
			Video	50 to 54
	Psoriatic arthritis		Image	55 to 65
			Video	66 to 75
Control	Healthy		Image	76 to 106
			Video	107 to 123
	Other diagnoses than spondyloarthritis		Image	124 to142
			Video	143 to 150

Appendix 2. Characteristics of the patients and controls whose ultrasound studies were used for the validation process.

Images from patients (n = 75)				
Resource (number)	Diagnostic (number of patients)	Time since symptoms onset Mean (SD)	% Male	Age Mean (SD)
Static image (42)	PsoA (11)	22,4 (9,4)	55	36,3 (5,8)
	Ax-SpA (31)	32,5 (10,3)	68	40,6 (5,9)
3–5 s videos (33)	PsoA (10)	24,5 (8,9)	50	37,2 (6,1)
	Ax-SpA (23)	34,7 (8,8)	65	39,3 (5,3)

Images from controls (n = 75)				
Resource (number)	Subgroup (number of subjects)	% Male	Age Mean (SD)	
Static image (40)	Healthy models (30)	61	37.4 (6.3)	
	Non-SpA patients (6)	50	38.9 (5.9)	
	Patients without Achilles pain (4)	75	41.5 (3.9)	
3–5 s videos (35)	Healthy models (26)	55	38.5 (4.0)	
	Non-SpA patients (5)	60	39.1 (4,8)	
	Patients without Achilles pain (4)	50	37.4 (4.6)	

Time is expressed in months. PsoA, Psoriatic arthritis; SD, standard deviation. Ax-SpA, axial spondyloarthritis.