Session: Late-Breaking Abstracts (L15-L20)

L20: Automated Ultrasound System ARTHUR with AI Analysis DIANA Matches Expert Rheumatologist in Hand Joint Assessment of Rheumatoid Arthritis Patients

Tuesday, November 19, 2024

9:15 AM - 9:30 AM Eastern Time

Q Location: Room 146AB







Late-Breaking Abstract Presenter(s)



Søren Andreas Just, MD, PhD

Section of Rheumatology, Department of Medicine, Svendborg Hospital, Odense University Hospital

Svendborg, Denmark

Disclosure information not submitted.

Bill Aplin Frederiksen¹, Mads Ammitzbøll Danielsen², Hilde Berner Hammer³, Benjamin Schultz Overgaard¹, Anders Weber⁴, Lene Terslev², Thiusius Rajeeth Savarimuthu⁵ and **Soren Andreas Just**¹, ¹Section of Rheumatology, Department of Medicine, Svendborg Sygehus OUH, Svendborg, Denmark, ²Center for Rheumatology and Spine Disease, Rigshospitalet, Glostrup, Denmark, ³Center for Treatment of Rheumatic and Musculoskeletal Diseases (REMEDY), Diakonhjemmet Hospital, Oslo, Norway, Faculty of Medicine, University of Oslo, Oslo, Norway, ⁴ROPCA, Odense, Denmark, ⁵Mærsk Mc-Kinney Møller Institute, University of Southern Denmark, Odense, Denmark

Background/Purpose: To evaluate the precision of ARTHUR (Figure 1), a CE-marked, fully automated ultrasound scanning system that captures ultrasound images of 22 hand joints and DIANA, a CE-marked Artificial Intelligence (AI) system that automatically analyzes these images and grades synovial hypertrophy (SH) and Doppler activity in accordance with the global OMERACT-EULAR synovitis score (GLOESS).

Methods: Thirty patients with rheumatoid arthritis (RA) were scanned twice by ARTHUR, followed by scanning by a rheumatologist specialized in musculoskeletal ultrasound. DIANA graded ARTHUR images according to GLOESS. The rheumatologist also graded his images according to GLOESS. A ground truth was established by another human expert who evaluated all images (both ARTHUR's and the rheumatologist's), blinded to the scanning method. Here, the image with the highest disease activity gave the joints ground truth. Agreement with the ground truth was evaluated using weighted Cohen's Kappa, Percent Exact Agreement (PEA), and Percent Close Agreement (PCA) for all four-class assessments. Binary agreement (healthy vs. disease), were also compared at both joint and patient levels.

Results: ARTHUR successfully scanned 564 out of 660 joints (85.45%), with the highest success rates in MCP2, MCP3, MCP4, PIP2, PIP3, and PIP4 joints.

At the joint level (Table 1), ARTHUR+DIANA compared to the ground truth showed for SH: PEA of 49.01% [95% CI: 44.95–53.07%], PCA of 91.23% [95% CI: 88.68–93.36%], and binary agreement of 79.97% [95% CI: 76.55–83.09%]; for Doppler activity: PEA of 62.58% [95% CI: 58.59–66.46%], PCA of 94.37% [95% CI: 92.22–96.07%], and binary agreement of 88.08% [95% CI: 85.22–90.56%].

The rheumatologist compared to the ground truth showed for SH: PEA of 51.52% [95% CI: 47.63–55.39%], PCA of 94.09% [95% CI: 92.01–95.76%], and binary agreement of 82.88% [95% CI: 79.78–85.68%]; for Doppler activity: PEA of 63.73% [95% CI: 59.93–67.41%], PCA of 94.08% [95% CI: 92.00–95.76%], and binary agreement of 90.74% [95% CI: 88.27–92.58%].

The Kappa values (Table 1) comparing ARTHUR+DIANA and the rheumatologist to the ground truth, respectively were for SH: 0.39 [95% CI: 0.32-0.45] vs. 0.46 [95% CI: 0.41-0.52], and for Doppler: 0.48 [95% CI: 0.41-0.55] vs. 0.45 [95% CI: 0.38-0.52].

At the patient level (all joints combined), ARTHUR+DIANA's binary disease assessment (Healthy vs. Disease) showed agreement with the ground truth of 86.67% [95% CI: 69.28–96.24%] for SH and 83.33% [95% CI: 65.28–96.36%] for Doppler activity. The rheumatologist's assessment showed agreements of 53.33% [95% CI: 34.33–71.66%] for SH and 66.67% [95% CI: 47.19–82.71%] for Doppler activity.

Conclusion: The precision of the assessments of RA patients by ARTHUR and DIANA was comparable to that of a rheumatologist specializing in musculoskeletal ultrasound at both the joint and patient levels. Automated systems have the potential to provide rapid and consistent assessments for patients suspected of arthritis and those with established disease, enhancing accessibility to high-quality diagnostic evaluations and potentially improving patient care.



Figure 1. The ARTHUR system connected to an ultrasound scanner.

	Rheumatologist versus ground truth						ARTHUR+DIANA versus ground truth						
	Kappa (+/- 95% CI)	PEA (+/- 95% CI)	PCA (+/- 95% CI)	Sensitivity (+/- 95% CI)	Specificity (+/- 95% CI)	Binary agreement (+/- 95% CI)	Kappa (+/- 95% CI)	PEA (+/- 95% CI)	PCA (+/- 95% CI)	Sensitivity (+/- 95% CI)	Specificity (+/- 95% CI)	Binary agreement (+/- 95% CI)	Sample
Synovial hypertrop hy (0-3)	0.46 (0.41 to 0.52)	51.52 (47.63 to 55.39)	94.09 (92.01 to 95.76)	29.11 (22.17 to 36.86)	99.80 (98.90 to 99.99)	82.88 (79.78 to 85.68)	0.39 (0.32 to 0.45)	49.01 (44.95 to 53.07)	91.23 (88.68 to 93.36)	37.84 (30.00 to 46.17)	93.64 (90.99 to 95.70)	79.97 (76.55 to 83.09)	660
Doppler Activity (0- 3)	0.45 (0.38 to 0.52)	63.73 (59.93 to 67.41)	94.08 (92.00 to 95.76)	34.44 (24.74 to 45.20)	99.65 (98.74 to 99.96)	90.74 (88.27 to 92.85)	0.48 (0.41 to 0.55)	62.58 (58.59 to 66.46)	94.37 (92.22 to 96.07)	41.86 (31.30 to 52.99)	95.75 (93.64 to 97.32)	88.08 (85.22 to 90.56)	659
Combined score (0-3)	0.48 (0.42 to 0.53)	50.15 (46.27 to 54.03)	91.97 (89.63 to 93.93)	30.48 (23.97 to 37.62)	99.79 (98.83 to 99.99)	80.15 (76.90 to 83.13)	0.42 (0.35 to 0.49)	47.68 (43.64 to 51.75)	90.40 (87.76 to 92.63)	42.04 (34.66 to 49.70)	92.76 (89.88 to 95.03)	77.98 (74.46 to 81.22)	660

Table 1. Precision metrics for ARTHUR and of the rheumatologist for joint disease activity for Synovial Hypertrophy, Doppler and the Combined Synovitis Score.

B. Aplin Frederiksen: None; M. Ammitzbøll Danielsen: None; H. Berner Hammer: None; B. Schultz Overgaard: None; A. Weber: ROPCA ApS, 3; L. Terslev: None; T. Rajeeth Savarimuthu: Ropca ApS, 8; S. Just: ROPCA, 2, 8.