RMD Open
Rheumatic & Musculoskeletal Diseases

ORIGINAL RESEARCH

Ultra-high frequency ultrasound of labial salivary glands in Sjögren's disease: diagnostic accuracy and patient stratification

Giovanni Fulvio ,[1,2] Rossana Izzetti,[3] Francesco Ferro,[1] Gaetano La Rocca,[1,2] Beatrice Dei,[1] Silvia Fonzetti,[1] Giuseppe Leto,[1] Francesco Gulia,[3] Valentina Donati,[4] Marta Mosca ,[1] Chiara Baldini[1]

To cite: Fulvio G, Izzetti R, Ferro F, *et al*. Ultra-high frequency ultrasound of labial salivary glands in Sjögren's disease: diagnostic accuracy and patient stratification. *RMD Open* 2026;**12**:e006906. doi:10.1136/rmdopen-2026-006906

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/rmdopen-2026-006906).

Received 9 March 2026
Accepted 31 May 2026

[1]Department of Clinical and Experimental Medicine, Rheumatology Unit, University of Pisa, Pisa, Italy
[2]Department of Translational Research, New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy
[3]Department of Surgical, Medical and Molecular Pathology and Critical Care Medicine, Unit of Dentistry and Oral Surgery, University of Pisa, Pisa, Italy
[4]Unit of Pathological Anatomy 2, Azienda Ospedaliero-Universitaria Pisana, Pisa, Italy

Correspondence to
Dr Giovanni Fulvio;
giovanni.fulvio92@gmail.com

ABSTRACT

Background To evaluate the classification accuracy of ultra-high frequency ultrasound (UHFUS) of labial salivary glands (LSGs) using both Grey Scale (GS) and colour Doppler (CD). We also investigated its role as an alternative to the highest-weighted classification criteria, its classification performance in combination with SSA status, its contribution to patient stratification and age-related classification variability.

Methods A total of 270 patients with suspected Sjögren's disease (SjD) (136 SjD, 134 Sicca) underwent UHFUS. In addition to OMERACT GS (modified to LSG), CD and composite GS+CD Scores, distinct GS morphological lesions were assessed. Diagnostic accuracy parameters were calculated, and additional analyses included correlation, logistic regression and cluster analysis.

Results The optimal classification cut-off was GS ≥2, yielding 66.2% sensitivity and 84.3% specificity. GS Score 3 was highly specific (98.5%), while very hypoechoic areas were exclusively observed in SjD. CD ≥2 showed accuracy comparable to GS, and GS+CD ≥3 enhanced accuracy. Incorporating GS into the American College of Rheumatology/European League Against Rheumatism criteria in place of biopsy achieved excellent accuracy (area under the curve 0.945). Using GS, GS+CD and SSA status, more than 75% of patients could be classified with an accuracy of 94%. Cluster analysis suggests different phenotypes: in Sicca, GS positivity indicated non-immune inflammatory alterations or sialoadenosis; in SjD, GS negativity defined low-grade inflammation. Classification accuracy of GS varied with age, whereas CD remained stable.

Conclusions UHFUS of LSGs is a promising diagnostic tool in SjD when biopsy is not feasible. This study provides novel insights into GS lesions, Doppler assessment, phenotypes and age-related changes of ultrasound.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Ultrasound of the major salivary glands is an established tool in Sjögren's disease (SjD).
⇒ It is also used for patient stratification.

WHAT THIS STUDY ADDS

⇒ Demonstrates the diagnostic value of labial salivary gland (LSG) ultrasound using ultra-high frequency probes.
⇒ Identifies very hypoechoic areas as a specific Grey Scale (GS) lesion for SjD.
⇒ Shows the added diagnostic contribution of Doppler assessment in combination with GS scoring.
⇒ May identify distinct phenotypes associated with ultrasonographic positivity and negativity in both patients with Sicca syndrome and patients with SjD.
⇒ Highlights the influence of age on ultrasonographic findings and classification performance.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Supports the implementation of novel GS and Doppler biomarkers into clinical practice.
⇒ Suggests LSG ultrasound as a feasible alternative to biopsy when histology is not available.
⇒ Recommends incorporating age-related changes and different phenotypes into the interpretation of ultrasonography.

INTRODUCTION

Sjögren's disease (SjD) is a systemic autoimmune disorder affecting the exocrine glands, leading to xerostomia and keratoconjunctivitis sicca.[1] Over the past decade, salivary gland ultrasound (SGUS) has emerged as a valuable imaging tool for the evaluation of patients with suspected SjD.[2] Accumulating evidence indicates that Grey Scale (GS) Score may be integrated into the American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) classification criteria as an additional item contributing to disease classification.[3 4] In addition to its diagnostic role, SGUS positivity has been associated with distinct clinical and biological phenotypes.[5]

Despite these advances, several limitations persist. Although colour Doppler (CD) has recently been standardised, data regarding its value remain limited.[6 7] Moreover, ultrasound-based phenotypical stratification has largely been explored within established SjD cohorts, whereas its relevance in patients without SjD has not been systematically investigated. Finally, the alterations related to age or disease duration may influence GS interpretation and, consequently, affect diagnostic performance.[8 9]

While the role of SGUS in major salivary glands is well established, the application of ultra-high frequency ultrasound (UHFUS) to labial salivary glands (LSGs) has been far less extensively studied. Owing to its high-resolution capabilities, UHFUS enables near-microscopic visualisation of minor salivary gland architecture, potentially enhancing disease identification and patient stratification.[10 11] Preliminary studies have demonstrated a promising reliability of UHFUS applied to LSGs; however, its diagnostic accuracy requires further validation, particularly in relation to classification criteria and across different phenotypes.[12 13]

In this context, we comprehensively evaluated the role of UHFUS of the LSGs in patients with suspected SjD. Specifically, we assessed the diagnostic performance of GS and CD based on the Outcome Measures in Rheumatology (OMERACT) scoring system in distinguishing SjD from Sicca conditions; examined their potential incorporation into classification criteria and their diagnostic relevance in association with anti-SSA antibody positivity; investigated their value in phenotypical stratification of both SjD and Sicca subgroups; and analysed how diagnostic accuracy varies according to age and the factors underlying this variability.

METHODS

Study population

This was a cross-sectional study with prospective enrolment conducted at the Rheumatology Unit, University of Pisa, from January 2018 to December 2023. A total of 375 consecutive patients with suspected SjD underwent echo-assisted LSG biopsy using UHFUS. For each biopsy, at least four LSGs were obtained per patient and the Focus Score (FS) was assessed and considered positive when equal to or greater than 1.[14] Serum samples were collected from all patients and tested for anti-Ro52, anti-Ro60 and anti-SSB/La antibodies using a commercially available immunoblot (EUROIMMUN srl). The level of autoantibodies was analysed using band densitometry. At enrolment, clinical data were also collected on EULAR Sjögren's Syndrome Patient Reported Index, the EULAR Sjögren's Syndrome Disease Activity Index (ESSDAI), ocular tests (Schirmer's test and break-up time (BUT)), unstimulated salivary flow rate (USWR) and routine laboratory parameters.

Of the 375 patients initially assessed, 71 were excluded due to incomplete data on UHFUS evaluation, biopsy results, SSA status or salivary flow. This left 304 patients with complete clinical and diagnostic data sets. Among these, an additional 34 patients were excluded: 26 with other connective tissue diseases and 8 with alternative diagnoses (eg, neoplastic, infectious, amyloidosis, IgG4-related disease or hypereosinophilic syndrome). Following all exclusions, a total of 270 patients were included in the final analysis (online supplemental figure 1). Patients were classified as having either Sicca syndrome (Sicca) or SjD according to the 2016 ACR/EULAR classification criteria.[15]

Ultra-high frequency ultrasound

UHFUS was performed using a 70 MHz probe of the VEVO MD system, a device approved for human use. The lower lip was scanned in three compartments, and the most severely affected compartment was selected for UHFUS assessment and biopsy. Since glands within the same compartment showed comparable ultrasonographic features, UHFUS assessment was performed at the compartment level rather than by scoring each individual gland separately. Examinations were video recorded and independently evaluated by two blinded experts using Horos Imaging software. In cases of disagreement, a third expert was consulted to reach a consensus.

GS evaluation followed a modified OMERACT scoring system (GS OMERACT).[16] Indeed, due to the high frequency of diffuse inhomogeneity in LSGs with UHFUS, this feature alone was not considered sufficient to assign a Score 3. The scoring criteria were as follows: Grade 0, normal echostructure and echogenicity; Grade 1, mild abnormalities including diffuse inhomogeneity alone; Grade 2, hypoechoic areas occupying <50% of the gland; Grade 3, hypoechoic areas occupying >50% of the gland (figure 1A). The presence or absence of specific elementary lesions was systematically recorded. These included: hypoechoic areas, defined as well-demarcated regions that appear or disappear with probe movement; diffuse inhomogeneity, characterised by widespread hypoechogenicity in the absence of clearly delineated hypoechoic areas (though these may coexist); very hypoechoic areas (VHAs), classified either as *'multiple very hypoechoic areas'* when large, very hypoechoic regions involved the entire glandular surface, or as a *'lymphoma pattern'* when six of eight predefined suspicious criteria were met, according to the definitions by Fulvio and Lorenzon;[17 18] and hyperechoic bands, observed as bright echogenic reflections.[19]

Colour Doppler (CD) was scored according to the OMERACT scoring system (CD OMERACT) for major salivary glands: Grade 0, no increased vascular signals; Grade 1, focal dispersed signals; Grade 2, diffuse signals in <50% of the gland; Grade 3, diffuse signals in >50% of the gland[6] (figure 1B). Although specific evidence for minor salivary glands is limited, we applied the same prescanning precautions used for major salivary gland CD assessment. Patients were asked to avoid eating, chewing and smoking for at least 1 hour before the examination. A combined score (GS+CD OMERACT) was also calculated

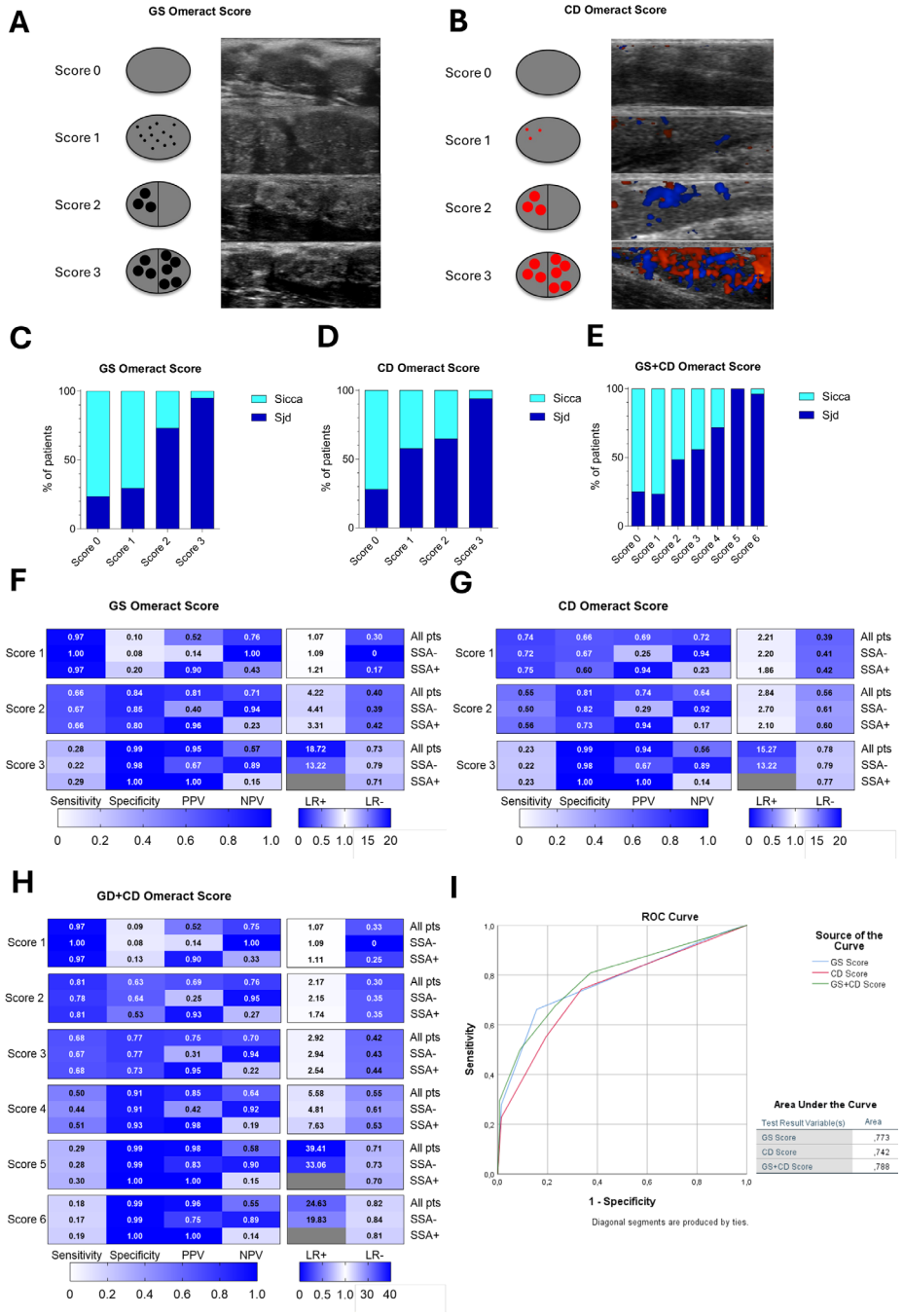

Figure 1 GS, CD and GS+CD scoring systems and classification accuracy. Representative schematic illustrations (left) and UHFUS images (right) of GS (A) and CD (B) Scores in labial salivary glands. (A) GS Score 0, Normal echotexture and echogenicity; Score 1, Inhomogeneity without hypoechoic areas; Score 2, Hypoechoic areas involving <50% of the gland; Score 3, Hypoechoic areas involving >50% of the gland. (B) CD Score: 0, No Doppler signal; Score 1, Focal vascular signals; Score 2, Diffuse signals in <50% of the gland; Score 3, Diffuse signals in >50% of the gland. Distribution of patients with Sicca and Sjögren's disease in GS (C), CD (D), and combined GS+CD Scores (E), showing an increasing prevalence of SjD with higher scores. Classification accuracy heatmaps for GS (F), CD (G), and combined GS+CD (H) Scores. Each heatmap shows, across different score thresholds (greater than or equal to, rows), the main diagnostic parameters (columns): sensitivity, specificity, predictive values (PPV, NPV), and likelihood ratios (LR+, LR–). On the right side of each row are indicated the patient subgroups analysed: all patients, SSA-positive and SSA-negative. Colour intensity reflects the magnitude of each parameter, with scale bars shown below each heatmap: sensitivity, specificity, PPV and NPV on the left; LR+ and LR– on the right. Grey cells indicate cases in which likelihood ratios could not be calculated due to zero values. (I) ROC curve comparing GS, CD, and GS+CD Scores for classification of SjD according to ACR/EULAR 2016 criteria. The GS+CD Score yielded the highest AUC. ACR/EULAR, American College of Rheumatology/European League Against Rheumatism; AUC, area under the curve; CD, colour Doppler; GS, Grey Scale; GS+CD, combined Grey Scale and colour Doppler Score; LR+, positive likelihood ratio; LR–, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; SSA, anti-Ro antibodies; SjD, Sjögren's disease; ROC, receiver operating characteristic; UHFUS, ultra-high frequency ultrasound.

for each patient, based on the sum of the semiquantitative GS Score and the CD Score (range 0–6).

Statistical analysis

Continuous variables were summarised as means with SD, while categorical variables were presented as percentages. Continuous variables were analysed using analysis of variance with the Bonferroni post hoc correction, or the Mann-Whitney or Kruskal-Wallis test, as appropriate. Categorical variables were compared using the chi-square test or Fisher's exact test, as applicable. Correlations between variables were assessed using Spearman's rank correlation coefficient.

Therefore, the analyses assess the performance of UHFUS in terms of classification accuracy rather than its ability to establish a definitive clinical diagnosis. The classification performance of GS, CD and combined GS+CD Scores was evaluated through sensitivity, specificity, positive and negative predictive values (PPV, NPV), likelihood ratios (LR+and LR−), and the area under the receiver operating characteristic (ROC) curve. The optimal cut-off point for each score was determined using Youden's Index, which was then used to define ultrasonographic positivity and negativity.

Binary logistic regression was employed for both univariate and multivariate analyses. An exploratory cluster analysis was also performed using the two-step algorithm. Univariate, multivariate and cluster analyses were repeated in relevant patient subgroups, as appropriate. To evaluate the influence of age on classification performance, sensitivity, specificity and area under the curve (AUC) were calculated across age deciles and plotted using LOESS (locally estimated scatterplot smoothing) in R with the ggplot2 package. All statistical analyses were performed using SPSS V.26, unless otherwise specified. When more advanced plotting was required, GraphPad Prism V.10 and R were employed.

Ethics

Written informed consent was obtained from all participants. The study was conducted in accordance with the Declaration of Helsinki.

RESULTS

Study population

The study population included 136 patients fulfilling SjD classification criteria and 134 sicca subjects not fulfilling SjD ACR/EULAR 2016 classification criteria. The groups were comparable in terms of age and sex distribution, with a similar female predominance (89.6% vs 89.7%). The differences between the two groups are summarised in table 1.

As expected, patients classified as SjD fulfilled a significantly higher number of ACR/EULAR 2016 classification criteria compared with Sicca controls (6.0 vs 1.3). Among the individual classification criteria, histological biopsy demonstrated pronounced differences as reflected by a higher FS (1.7 vs 0.4). Regarding autoantibodies, anti-SSA positivity was observed in 86.8% of patients with SjD compared with 11.2% of those with Sicca. Ro52 and SSB levels were significantly higher in the SjD group versus Sicca (0.8 vs 0.2 and 0.3 vs 0.0, respectively), while Ro60 levels showed a non-significant trend (p = 0.08). Triple-positive serology (Ro52+, Ro60+, SSB+) was more frequent in SjD (33.6% vs 0.7%), whereas seronegative profiles (Ro52−, Ro60−, SSB−) were predominantly observed in patients with Sicca syndrome (73.7% vs 11.2%). Functional glandular impairment was more pronounced in SjD, with a trend toward lower USWR (p = 0.09) and significantly reduced Schirmer's test values (6.2 vs 7.9, p=0.046), although no significant difference was found for tear BUT. Beyond classification criteria, patients with SjD showed higher levels of serum IgG (1467 vs 1182) and greater systemic disease activity, as reflected by higher ESSDAI Scores (5.4 vs 2.1).

UHFUS distinguishes SjD from Sicca and identifies VHA as a disease-specific feature

The distribution of UHFUS Scores and elementary lesions further differentiates the two groups. In patients with Sicca syndrome, the majority exhibited lower GS Scores (0 9.7%, 1 74.6%), whereas higher scores were uncommon. Conversely, patients with SjD showed higher GS Scores, with 38.2% presenting Score 2 and 27.9% Score 3. A similar trend was observed for CD Scores: Score 0 was most common among patients with Sicca syndrome (66.4%), whereas patients with SjD demonstrated a higher frequency of elevated CD Scores, with 32.4% at Score 2 and 22.8% at Score 3.

Across ultrasonographic scores, an inverse relationship was observed between Sicca and SjD, with a progressive shift from Sicca to SjD as the score increased. For the GS Score, patients with Sicca syndrome predominated at Scores 0 and 1 (76.5% and 70.4%, respectively), whereas Scores 2 and 3 were largely composed of patients with SjD (73.2% and 95.0%) (figure 1C). A similar distribution was observed for CD Scores: patients with Sicca syndrome accounted for the majority at Scores 0 (71.8%), while patients with SjD were predominant at Scores 1 (57.8%), 2 (64.7%) and 3 (93.9%) (figure 1D). The combined GS+CD Score further confirmed this trend: patients with Sicca syndrome were most frequently represented at Scores 0, 1 and 2 (75.0%, 76.6% and 51.4%, respectively), whereas patients with SjD became progressively more prevalent at higher scores, comprising 55.8% at Score 3, 71.8% at Score 4, 100% at Score 5 and 96.2% at Score 6 (figure 1E).

Lastly, when examining individual elementary ultrasonographic lesions, significant differences were noted in all features except for diffuse inhomogeneity, where a non-significant trend was observed (p = 0.08) (table 1). Notably, VHA were identified exclusively in patients with SjD (14.8% vs 0%, p < 0.0001; table 1), supporting its role as a lesion specific to the disease.

Table 1 Comparison between patients with Sicca and Sjögren's disease

Study population			
	Sicca (n=134)	SjD (n=136)	P value
Age	52.2±14.0	55.0±14.0	0.10
Sex (M/F)	10.45%	10.29%	0.97
GS Score 0	9.70%	2.94%	
GS Score 1	74.63%	30.88%	
GS Score 2	14.18%	38.24%	
GS Score 3	1.49%	27.94%	
CD Score 0	66.42%	25.74%	
CD Score 1	14.18%	19.12%	
CD Score 2	17.91%	32.35%	
CD Score 3	1.49%	22.79%	**<0.01**
Hypoechoic areas	15.7%	65.9%	**<0.01**
VHA	0.00%	14.81%	**<0.01**
Diffuse inhomogeneity	53.0%	63.7%	0.08
Hyperechoic bands	30.6%	60.0%	**<0.01**
FS	0.4±0.5	1.7±1.6	**<0.01**
FS ≥1	5.22%	72.06%	**<0.01**
SSA+	11.19%	86.76%	**<0.01**
Ro52 level	0.3±0.4	0.8±0.6	**<0.01**
Ro60 level	0.2±0.3	0.5±0.5	0.08
SSB level	0.0±0.1	0.2±0.3	**0.04**
Seronegative	73.68%	11.19%	
Ro60 isolated	3.01%	12.69%	
Ro52 isolated	3.01%	28.36%	
Double positive (Ro52/Ro60)	3.76%	11.94%	
Triple positive (Ro52/Ro60/SSB)	0.75%	33.58%	
Atypical antibodies	15.79%	2.24%	
USWR (mL/min)	2.7±2.3	2.2±2.1	0.09
USWR ≤0.1 mL/min	38.06%	48.53%	0.08
BUT (s)	7.5±3.9	6.9±3.3	0.49
Schirmer (mm/5 min)	7.9±5.1	6.2±4.2	**0.05**
N° ACR/EULAR criteria	1.3±1.1	6.0±1.5	**<0.01**
IgG (mg/dL)	1182.8±336.6	1467.3±634.4	**<0.01**
IgM (mg/dL)	126.7±108.0	130.5±90.5	0.82
RF (IU/mL)	23.3±35.7	24.2±31.2	0.87
ESSDAI	2.1±3.2	5.4±5.8	**<0.01**
VAS dryness	6.3±2.4	6.0±2.6	0.35
VAS eye	6.3±3.0	6.3±2.9	0.84
VAS mouth	6.3±2.6	6.1±2.9	0.59
VAS fatigue	6.6±2.5	6.3±2.8	0.32
VAS pain	6.2±2.8	5.6±3.3	0.14
ESSPRI	6.3±2.2	6.0±2.4	0.25

Bold values indicate statistically significant p-values (p<0.05).

ACR/EULAR, American College of Rheumatology/European League Against Rheumatism; BUT, break-up time; CD, colour Doppler; ESSDAI, EULAR Sjögren's Syndrome Disease Activity Index; ESSPRI, EULAR Sjögren's Syndrome Patient Reported Index; FS, Focus Score; GS, Grey Scale; N° ACR/EULAR criteria, number of classification criteria met according to the ACR/EULAR 2016; N° of foci, number of foci; RF, rheumatoid factor; SjD, Sjögren's disease; USWR, unstimulated salivary flow rate; VAS, Visual Analogue Scale; VHA, very hypoechoic areas.

UHFUS reveals LSG classification accuracy

To determine whether UHFUS, when applied to LSGs, yields classification performance comparable to that reported for major SGUS, we evaluated the accuracy of the GS Score, CD Score and GS+CD Score. We then examined how the accuracy varied according to serological status, specifically the presence or absence of SSA antibodies.

For GS Score, the optimal cut-off was ≥2 with a sensitivity of 66.2%, specificity of 84.3%, PPV of 81.1% and NPV of 71.1% in the overall population (figure 1F). GS Score 3 was highly specific (98.5%) and associated with an elevated PPV (95.0%) and an LR+ of 18.7, indicating strong confirmatory value for SjD. Conversely, GS Score ≥1 showed high sensitivity (97.1%) and the best NPV (76.5%), with an LR− of 0.30, supporting the utility of Score 0 for disease exclusion. When GS Scores were stratified by serological status, the cut-off ≥2 maintained classification performance, with minimal variation in sensitivity (66% overall; 67% SSA−; 66% SSA+) and specificity (84% overall; 85% SSA−; 80% SSA+). However, stratification improved predictive values: in SSA− patients, NPV increased to 94.4%, while in SSA+ patients, PPV increased to 96.3%. Notably, GS Score 0 was exclusively observed in SSA− patients with Sicca, and GS Score 3 occurred only in SSA+ patients with confirmed SjD

CD scoring showed diagnostic accuracy comparable to GS at the optimal cut-off of ≥2, with a sensitivity of 55.1%, specificity of 80.6%, PPV of 74.3%, and NPV of 63.9% (figure 1G). However, at the lower cut-off of ≥1, CD showed markedly higher specificity than GS (66% vs 10%), suggesting a potentially greater specificity of the lower CD Scores.

Combining GS and CD Scores into a composite score further enhanced classification performance. The optimal cut-off was ≥3, yielding a sensitivity of 67.6%, specificity of 76.9%, PPV of 74.8% and NPV of 70.1%. Like GS scoring, the combined GS+CD evaluation demonstrated high predictive values when stratified by serological status: PPV reached 95% in SSA-positive patients, while NPV increased to 94% in SSA-negative individuals. Similar to GS and CD Scores, the GS+CD Score followed a progressive trend with increasing cut-off values: sensitivity and positive likelihood ratios improved, whereas specificity and negative likelihood ratios declined. However, compared with GS and CD Scores, the GS+CD Score demonstrated a more gradual transition, allowing for finer stratification (figure 1H). Finally, ROC curve analysis confirmed the highest AUC for the combined GS+CD Score (AUC 0.788), followed by GS Score alone (AUC 0.773), and CD Score alone (AUC 0.742), thus supporting a synergistic effect when both imaging modalities are used together (figure 1I).

GS Score as a potential alternative to biopsy within the ACR/EULAR 2016 criteria

After confirming the role of UHFUS as a classification tool, we explored its role in modified ACR/EULAR models excluding either FS or anti-SSA antibody positivity. Only patients with complete ocular test data were included and ocular test positivity was defined as at least one abnormal test. For this analysis, we used the cut-offs previously determined: GS ≥ 2, CD ≥ 2, GS+CD Score ≥ 3. Following previous studies,[3 4] we first assessed the weight of each ultrasonographic score to ACR/EULAR classification fulfilment using logistic regression models. As expected, FS and SSA demonstrated the highest ORs, whereas GS ≥2, CD ≥2 and GS+CD ≥3 yielded intermediate ORs, closely aligning with those of minor criteria (figure 2A, online supplemental table 1). This suggests that labial gland UHFUS positivity has a weight equivalent to minor criteria (ie, 1 point).

Next, we compared models excluding either FS or anti-SSA antibody positivity. As expected, correlation analysis revealed the strongest association between GS, CD and GS+CD Scores and FS (Spearman's $\rho = 0.53$–0.64), more than Ro52 and Ro60 level or minor criteria (figure 2B), consistent with the fact that both assessments evaluate LSG involvement. To confirm these correlations, UHFUS variables were included in logistic regression models predicting FS ≥1 and SSA positivity. GS≥2, CD≥2 and GS+CD≥3 emerged as the strongest predictors of FS ≥1 (figure 2C), whereas SSA positivity was best predicted by FS ≥1 (figure 2D). Moreover, the OR for GS ≥2 in predicting FS ≥1 (online supplemental table 2, OR = 13.83, $p < 0.001$) was markedly higher than for SSA+ (online supplemental table 3, OR = 5.56). To further evaluate UHFUS as a potential substitute for either FS or SSA in the ACR/EULAR classification criteria, we performed multivariate regression analyses incorporating GS ≥2, CD ≥2 and GS+CD ≥3 in place of each highest-weighted classification item. Multivariate analysis showed that GS ≥2 was a predictor of criteria fulfilment, showing a stronger association when replacing FS (OR = 4.15, 95% CI 1.45 to 11.93, p=0.01) than when replacing SSA (OR = 3.14, 95% CI 1.19 to 8.29, p=0.02) (online supplemental tables 4 and 5). CD ≥2 was borderline significant only in the FS-replacement model and not significant when SSA was removed. GS+CD ≥3 was not included in final multivariate models, likely due to redundancy with GS or CD alone. Overall, these analyses suggest that UHFUS, particularly GS Score, may perform better when incorporated into the ACR/EULAR criteria as a replacement for biopsy rather than SSA. In contrast, the CD Score may offer additional value only in scenarios where FS is excluded.

Based on these findings, we calculated the accuracy for modified ACR/EULAR criteria, assigning one point to GS ≥2 while removing either FS or SSA. When UHFUS replaced the biopsy, classification performance was excellent (AUC =0.945, figure 2E,F), whereas, as expected, its substitution for SSA yielded lower accuracy (AUC = 0.887, figure 2E, online supplemental table 6). The cut-off of the modified criteria remained unchanged at four points, consistent with the original ACR/EULAR criteria (figure 2F). Adding one point for CD modestly improved

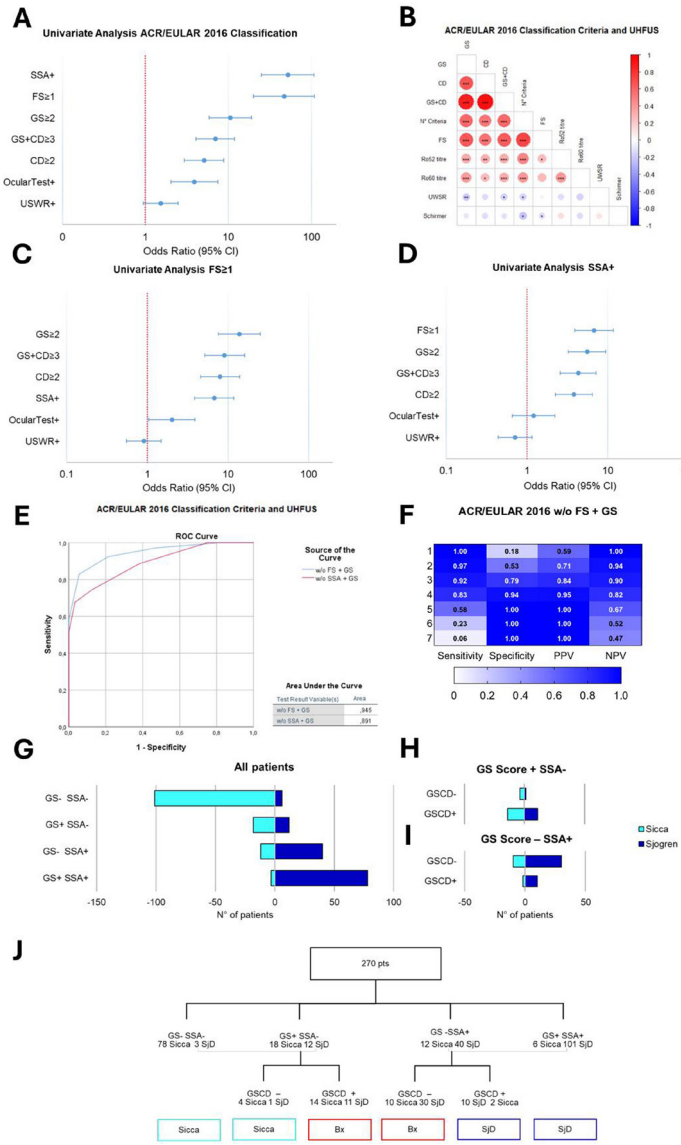

Figure 2 Integration of UHFUS in place of the FS within ACR/EULAR 2016 criteria and its classification value in combination with SSA. (A) Forest plot showing ORs, 95% CI from univariate logistic regression models assessing the weight of individual ACR/EULAR 2016 classification items and UHFUS-derived parameters (GS, CD, GS+CD Scores) to overall classification criteria fulfilment. (B) Spearman correlation plot illustrating the relationships between UHFUS Scores (GS, CD and GS+CD), FS, Ro52 and Ro60 antibody levels, Schirmer's test and unstimulated salivary flow. The strongest correlations were observed between UHFUS Scores and glandular inflammation. The diameter of each circle represents the strength of the correlation, while the colour indicates the corresponding Spearman correlation coefficient (r). Statistically significant correlations are marked with p values displayed inside the circles: *p<0.05, **p<0.01, ***p<0.001. (C, D) Forest plots from univariate logistic regression evaluating the ability of UHFUS parameters to predict histological inflammation (FS ≥1, (C) and SSA positivity (D). GS ≥2 was more strongly predictive of glandular inflammation than of anti-SSA seropositivity. (E) ROC curves comparing classification performance of modified ACR/EULAR algorithms replacing major classification criteria (FS or SSA) with GS. Replacing the biopsy with GS ≥2 achieved the highest AUC. (F) Heatmap illustrating the classification performance of models incorporating GS in place of the FS. Rows represent different thresholds (greater than or equal to), while columns display key metrics: sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). (G) Patient distribution across four subgroups based on GS Score + (positive if GS ≥2) and SSA status. Concordant profiles (GS–SSA–, GS+SSA+) were correctly classified in nearly all cases as Sicca or Sjögren's, respectively. Further refinement of discordant subgroups GS+SSA– (H) and GS–SSA+ (I) using the GS+CD Score (GSCD+, defined as GS+CD ≥3), which improved classification accuracy. (J) Proposed diagnostic algorithm integrating GS, SSA and GS+CD Scores. ACR/EULAR, American College of Rheumatology/European League Against Rheumatism; AUC, area under the curve; Bx, biopsy; CD, colour Doppler; FS, Focus Score; GS, grey scale; GS+CD: combined grey scale and colour Doppler Score; SSA, anti-Ro antibodies; ACR/EULAR 2016 w/o FS+GS, ACR/EULAR criteria modified by replacing the FS with GS; GS+, GS Score≥2; GS-, GS Score<2; GSCD+, GS+CD Score ≥3; GSCD-, GS+CD Score<3; ROC, receiver operating characteristic; SjD, Sjögren's disease; UHFUS, ultra-high frequency ultrasound; USWR+, unstimulated salivary flow rate positivity; w/o FS+GS, modified classification excluding FS and including GS Score; w/o SSA+GS, modified classification excluding SSA and including GS Score.

AUC only in the biopsy-free model (0.948 vs 0.945), with no additional benefit in the absence of SSA (online supplemental figure 2). Lastly, we evaluated whether assigning higher weight (eg, 2 points) to GS improved classification performance, but this approach did not yield superior accuracy (online supplemental figure 3), confirming the results of the logistic regression.

LSG UHFUS and SSA may avoid biopsy in over 75% of patients

Given the excellent classification performance of LSG UHFUS when used in combination with SSA status and minor criteria, we assessed its potential to classify patients using only ultrasonographic findings and SSA serology. As previously demonstrated, both the GS Score and the composite GS+CD Score, in combination with SSA status, exhibit high predictive value, both positive and negative.

Patients were first stratified according to GS Score positivity (GS ≥ 2, ie, GS+) or negativity (GS < 2, that is, GS–) in relation to SSA positivity (SSA+ or SSA–). Among the 270 patients, four principal subgroups were identified. In the two concordant groups, GS+SSA+ (n = 81) and GS–SSA– (n = 107), classification was accurate in 179 out of 188 cases (95%), representing 69% of the total cohort (figure 2G).

In the discordant groups (GS+SSA-and GS-SSA+), GS and SSA alone were not sufficiently discriminative; therefore, we further stratified patients using the composite GS+CD Score, considering it as positive (GSCD+, ie, GS+CD ≥ 3) or negative (GSCD–, ie, GS+CD < 3). In GS+SSA-patients, a negative composite Score (GSCD–) was associated with Sicca, correctly classifying 80% of the patients (4/5) (figure 2H). Conversely, in GS–SSA+ patients, a positive composite Score (GSCD+) was predictive of SjD, correctly classifying 10 of 12 patients (83%) (figure 2I). These findings indicate that when the GS+CD Score is concordant with SSA status, even if GS alone is discordant, it can refine classification. However, in cases where both GS and GS+CD Scores were discordant with SSA status (ie, GS+SSA– GSCD+ and GS–SSA+GSCD–), UHFUS was unable to correctly classify patients.

Based on these observations, we proposed a clinical algorithm: patients with GS–SSA– or GS+SSA–GSCD– could be classified as Sicca; those with GS+SSA+ or GS–SSA+GSCD+ as SjD; while the other patients (GS–SSA+GSCD–, GS+SSA–GSCD+) should be considered for biopsy. Using this approach, 76% of the cohort (205/270) could be classified without biopsy and minor criteria, with a correct classification rate of 94% (193/205) (figure 2J).

In patients with Sicca syndrome, positive GS identified two distinct phenotypes: one characterised by glandular inflammation and the other by age-related damage

After exploring the implications of positive and negative UHFUS Scores in differentiating Sicca from SjD, we next focused on the significance of GS positivity within the Sicca population. As in prior analyses, we adopted a GS Score ≥2 as the threshold for ultrasonographic positivity, thereby distinguishing GS– Sicca from GS+ patients with Sicca syndrome (online supplemental table 7). Among patients with Sicca syndrome, patients with a positive GS Score were significantly older (64.9 vs 49.9), had a higher FS (0.8 vs 0.4), fulfilled more ACR/EULAR criteria (1.8 vs 1.2) and showed a trend towards reduced USWR (1.9 vs 2.8, p=0.08). In addition, GS+ patients with Sicca syndrome exhibited more pronounced UHFUS abnormalities, including higher CD Scores (47.6% vs 14.2%, patients with CD Score ≥2), more frequent diffuse inhomogeneity (81.0% vs 47.8%) and a trend for increased presence of hyperechoic bands (p=0.07).

To better understand the determinants of GS positivity in Sicca, we performed a correlation analysis (figure 3A). The strongest associations were observed with age (ρ=0.42) and FS (ρ=0.31). In a multivariate logistic regression model including all candidate variables, only age (OR=1.11 per year, 95% CI 1.05 to 1.18, p=0.00) and FS (OR=4.03, 95% CI 1.15 to 14.1, p=0.03) remained significant predictors of GS Score ≥2 (figure 3B, online supplemental table 8).

To further explore the heterogeneity within the GS+ Sicca group, we conducted a cluster analysis based on the most significant variables (p<0.10): age, FS, USWR and the number of ACR/EULAR criteria fulfilled. Two distinct clusters were identified (figure 3C, online supplemental figure 4). Cluster 1 was characterised by a higher FS (1.4), more ACR/EULAR criteria (2.9) and a higher CD Score (2.0) compared with GS– Sicca. Despite these differences, the frequency of UHFUS abnormalities such as hyperechoic bands (28.6%) and diffuse inhomogeneity (42.9%) was similar to the GS– group. In contrast, Cluster 2 included older patients (68.8), with lower salivary flow (1.3), confirming the inverse correlation found between age and USWR (ρ = –0.34, figure 3A). Cluster 2 also exhibited more frequent UHFUS abnormalities, which were present in all patients (100%) and a higher frequency of hyperechoic bands (57.1%), although this was not statistically significant. To further assess whether Cluster 1 Sicca GS+ overlapped with classified SjD, we performed a targeted comparison with SjD GS– and SjD GS+ patients using selected classification-related variables (online supplemental table 9). Cluster 1 Sicca GS+ showed glandular inflammatory features, with a mean FS of 1.40 and FS positivity in 42.9% of patients, partially overlapping with SjD GS– patients. However, it had fewer ACR/EULAR criteria and lower SSA, salivary flow and ocular involvement than SjD subgroups, supporting its interpretation as an inflammatory Sicca phenotype.

These findings suggest that GS positivity in patients with Sicca syndrome may identify two distinct subgroups in the absence of overt immunological features: one exhibiting low-grade glandular inflammation, and another defined by advanced age, functional salivary impairment and ultrasonographic glandular damage.

In patients with SjD, negative GS differentiates two phenotypes with low salivary gland inflammation: patients with low immunological activity mostly Ro-52 isolated and young triple positive patients with high immunological activity

After evaluating the relevance of UHFUS in patients with Sicca syndrome, we investigated whether a positive GS Score further stratifies patients within the SjD group. As previously, we adopted a GS Score ≥2 as the threshold to define ultrasonographic positivity. Among patients with SjD, we first compared those with GS positivity to GS negativity (online supplemental table 10). Several significant differences emerged: compared with GS+ patients, SjD GS– patients were younger (51.4 vs 56.9), had lower FS (0.8 vs 2.1), lower Ro60 level (0.3 vs 0.6), more frequent isolated Ro52 positivity (50.0% vs 17.1%) and a lower frequency of triple positive (15.2% vs 43.2%). Moreover, they exhibited a reduced number of ACR/EULAR criteria (5.2 vs 6.4), lower rheumatoid factor (12.6 vs 31.0) and lower ESSDAI (3.5 vs 6.5). These patients also differed in terms of ultrasonographic features: GS– patients less frequently showed CD ≥2 (23.9% vs 71.1%), hyperechoic bands (32.6% vs 74.2%), diffuse inhomogeneity (43.5% vs 74.2%) and VHAs (0.0% vs 22.47%).

To identify variables associated with GS negativity in SjD, we performed correlation and multivariate analyses. Among all parameters, GS Score showed the strongest correlation with FS (ρ=0.64), while the weakest significant correlation was observed with age (ρ=0.23) (figure 4A). However, in multivariate regression analysis, only FS (OR = 5.13, 95% CI 1.93 to 13.57, p < 0.001) and age (OR = 1.06, 95% CI 1.00 to 1.17, p = 0.04) remained independently associated with GS Score positivity (figure 4B, online supplemental table 11).

Following the same approach applied to patients with Sicca syndrome, we next investigated whether GS– patients with SjD could be further stratified into distinct subgroups. A cluster analysis conducted on GS– patients, including only those with complete data, identified two groups (online supplemental figure 5, figure 4C,D). Compared with GS+ patients with SjD, Cluster 1 was characterised by younger age (39.2), higher SSB level (0.4), a similar level of systemic activity measured by ESSDAI (4.8) and a reduced FS (0.8). All patients had a triple-positive

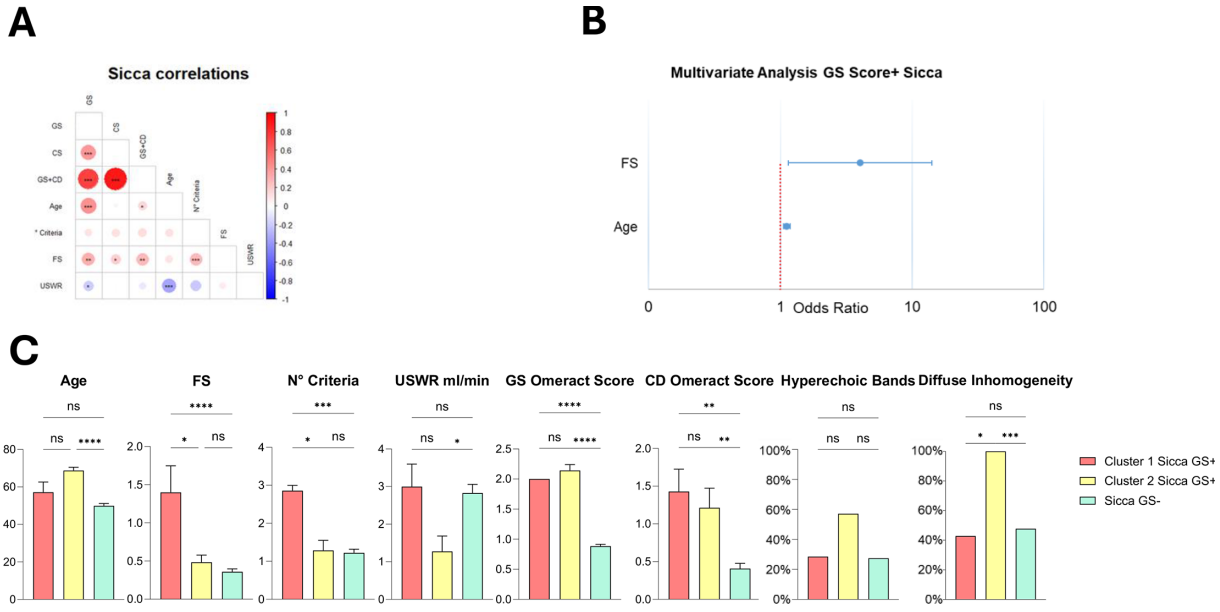

Figure 3 UHFUS identifies subtypes within patients with Sicca syndrome based on Grey Scale Score positivity. (A) Spearman correlation plot in patients with Sicca syndrome showing associations between UHFUS-derived parameters (GS, CD, GS+CD Scores) and age, number of ACR/EULAR classification criteria, FS and unstimulated salivary flow. GS Score correlates most strongly with age and FS. The diameter of each circle represents the strength of the correlation, while the colour indicates the corresponding Spearman correlation coefficient (r). Statistically significant correlations are marked with p values displayed inside the circles: *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. (B) Forest plot from multivariate logistic regression (stepwise method) evaluating the ability of age, number of ACR/EULAR classification criteria, FS and unstimulated salivary flow to predict Grey Scale Score positivity in patients with Sicca syndrome. Age and FS were independently associated with GS positivity, suggesting that both contribute to glandular ultrasound abnormalities. ACR/EULAR criteria and USWR were excluded from the final model due to lack of statistical significance (C) Comparison between the two clusters identified among patients with Sicca syndrome with positive Grey Scale Scores (GS ≥ 2) and patients with Sicca syndrome with negative GS Scores (GS < 2). Cluster 1 exhibited features consistent with low-grade inflammation, including higher FS and increased CD signal. Cluster 2 was characterised by older age, reduced salivary flow and a higher prevalence of UHFUS abnormalities, such as hyperechoic bands and diffuse inhomogeneity, consistent with age-related glandular remodelling. ACR/EULAR, American College of Rheumatology/European League Against Rheumatism; CD, colour Doppler; FS, Focus Score; GS, Grey Scale; GS+CD: combined Grey Scale and colour Doppler Score; UHFUS, ultra-high-frequency ultrasound; USWR, unstimulated salivary flow rate.

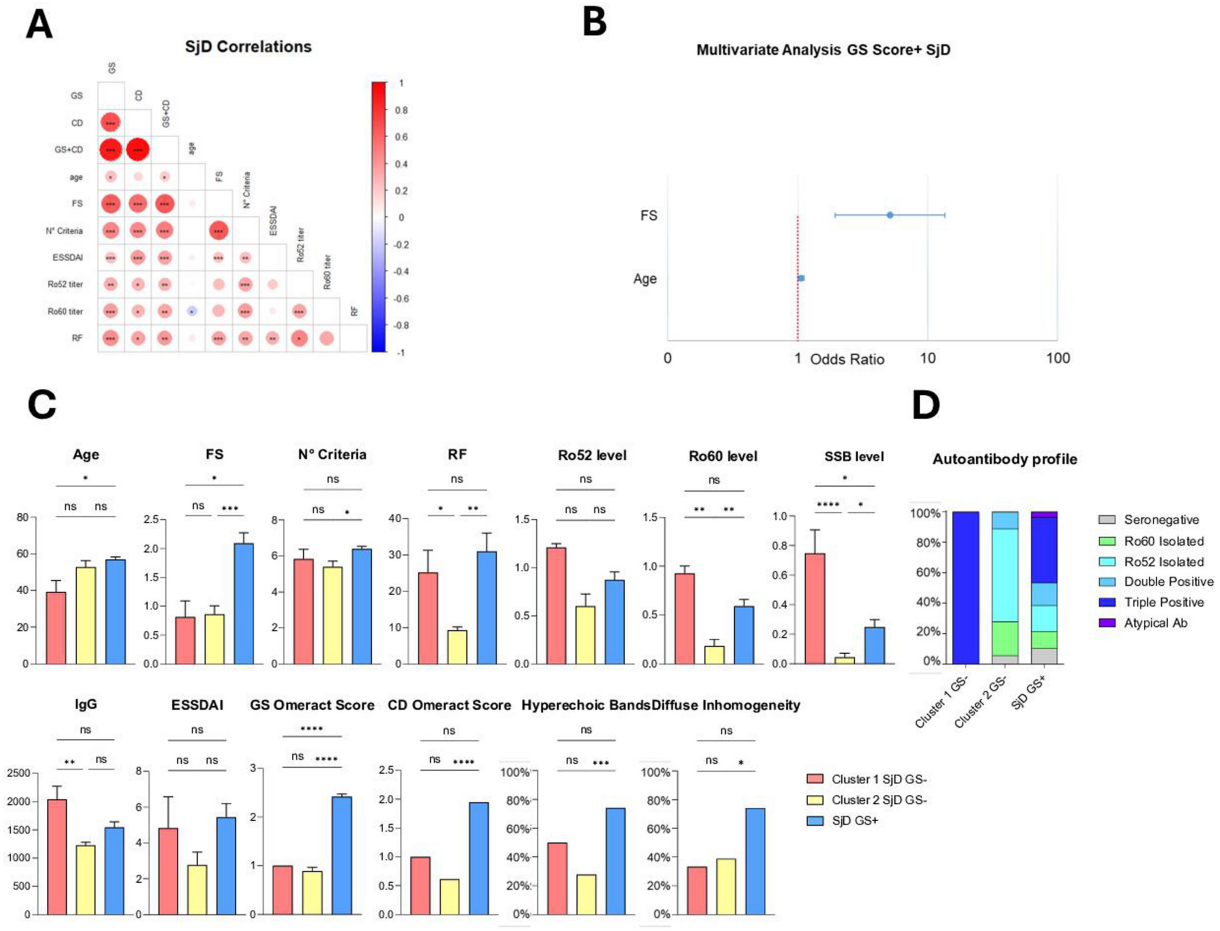

Figure 4 Negative UHFUS Grey Scale Score identifies distinct phenotypes within patients with Sjögren's disease. (A) Spearman correlation plot in patients with SjD displaying relationships between UHFUS-derived parameters (GS, CD, GS+CD Scores) and clinical variables including age, FS, number of ACR/EULAR Classification Criteria fulfilled, serological levels (Ro52, Ro60), rheumatoid factor and systemic activity. Grey Scale Score positivity correlated most strongly with FS. Circle size represents correlation strength, while colour indicates the Spearman coefficient (r). Significant correlations are annotated with p values (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).(B) Forest plot of multivariate logistic regression (stepwise method) assessing independent predictors of GS positivity (GS ≥2) among patients with SjD. Only FS and age emerged as significant predictors. Other variables were excluded from the final model due to lack of statistical significance. (C) Comparison between the two clusters identified among patients with SjD with negative Grey Scale (GS) Scores (GS < 2) and patients with SjD with positive GS Scores (GS ≥ 2). Cluster 1 included younger individuals with high immunological activity despite low GS Score and modest FS, while Cluster 2 was composed of patients with lower disease activity, although differences did not reach statistical significance. (D) Distribution of serological subtypes across GS– clusters and GS+ patients with SjD. Subtypes included: seronegative (no autoantibodies), Ro52 isolated (Ro52+only), Ro60 isolated (Ro60+only), double positive (Ro52+and Ro60+), triple positive (Ro52+, Ro60+ and SSB+) and atypical autoantibodies (associated with other autoimmune conditions). All patients in Cluster 1 exhibited triple seropositivity, whereas Cluster 2 showed a higher prevalence of isolated Ro52+ profiles. ACR/EULAR, American College of Rheumatology/European League Against Rheumatism; CD, colour Doppler; ESSDAI, EULAR Sjögren's Syndrome Disease Activity Index; FS, Focus Score; GS, Grey Scale; GS+CD, combined Grey Scale and colour Doppler Score; RF, rheumatoid factor; SjD, Sjögren's disease; SSA, anti-Ro/SS-A; SSB, anti-La/SS-B; UHFUS, ultra-high frequency ultrasound.

serological profile (Ro52+, Ro60+ and SSB+). Furthermore, Cluster 1 had higher immunoglobulin levels than Cluster 2 (2038±517vs 1225±235). Although not statistically significant, they also showed fewer ultrasonographic abnormalities compared with GS+ patients. Cluster 2, when compared with GS+ patients with SjD, predominantly included patients with isolated Ro52 positivity (61.1%) and was associated with lower FS (0.9), fewer fulfilled ACR/EULAR criteria (5.4) and a trend towards reduced ESSDAI (2.8; p = 0.06). Moreover, this group, compared with both Cluster 1 and GS+ patients with SjD, revealed significantly lower Ro60 (0.1) levels, SSB (0.1) levels, reduced RF levels (9.3), with a non-significant trend towards lower Ro52 levels (0.6; p=0.06). Regarding ultrasound abnormalities, these patients demonstrated lower CD Scores (0.6), fewer hyperechoic bands (27.8%) and diffuse inhomogeneity (38.9%) when compared with GS+ SjD cases.

Overall, these findings may indicate that GS– patients with SjD encompass two distinct phenotypes, both

marked by reduced histopathological and ultrasonographic abnormalities. One subgroup consists of younger, immunologically active individuals with triple antibody positivity and a systemic involvement comparable to GS+ SjD, whereas the second subgroup is composed mainly of patients with isolated Ro52 positivity, showing lower serological activity.

GS Score classification accuracy is influenced by age

Given that FS and age were independently associated with GS Score positivity in both patients with Sicca and patients with SjD, we investigated whether age might influence the classification performance in individuals with suspected SjD. All analyses were conducted using the previously determined cut-offs for GS Score (≥2), CD Score (≥2) and the combined GS+CD Score (≥3). To evaluate interaction effects, logistic regression models were applied, incorporating interaction terms between age and GS Score. In the study population, the interaction term between age and GS Score positivity was significant (OR = 0.94; 95% CI 0.89 to 0.99; p = 0.016), indicating that increasing age reduces the classification contribution of GS Score (online supplemental table 12). Based on this model, we calculated predictive probabilities: with advancing age, the PPV of a positive GS Score progressively declined, while the NPV of a negative score declined only slightly (figure 5A).

To further explore age-related classification variability, we assessed how sensitivity, specificity, and AUC for GS Score changed across age. Classification metrics were computed across age deciles (online supplemental figure

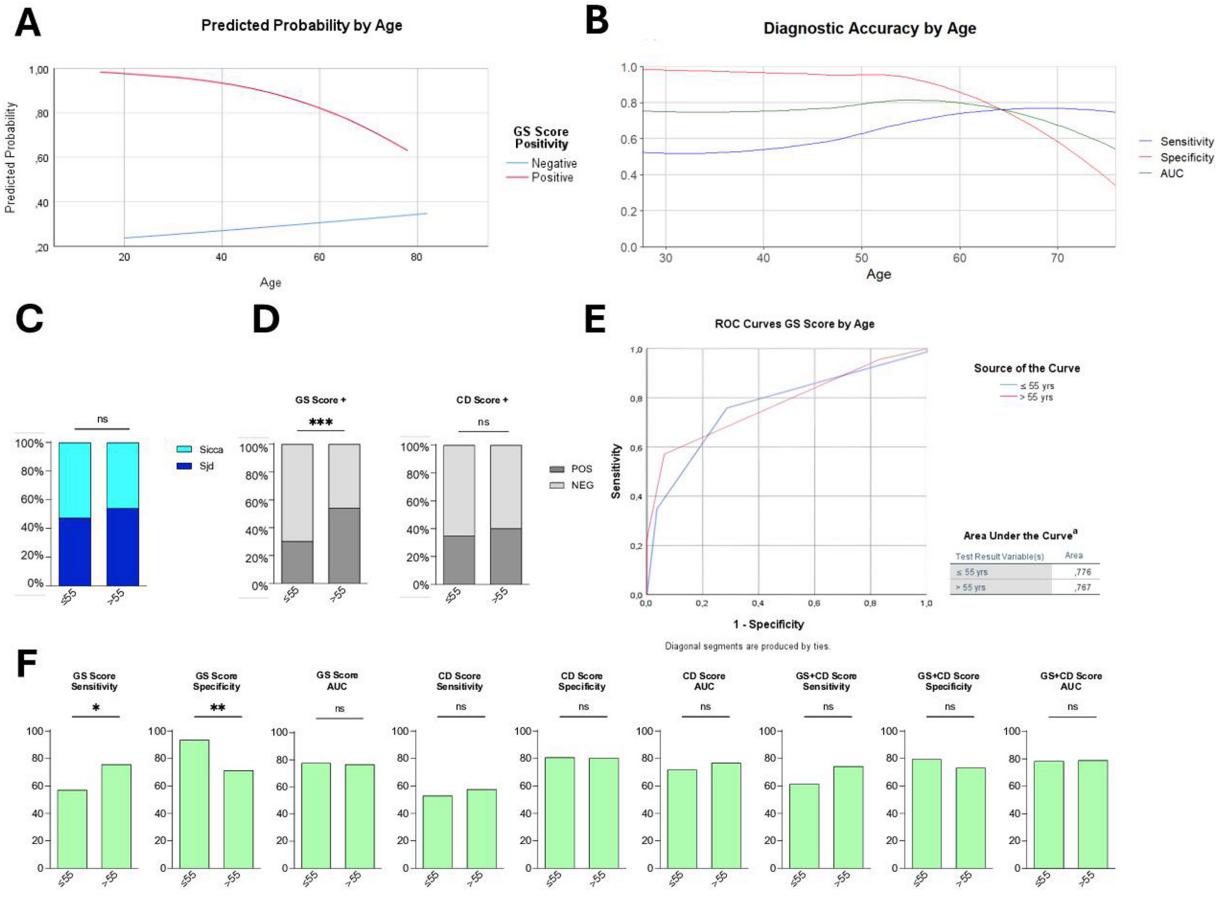

Figure 5 Influence of age on the classification accuracy of UHFUS. (A) Predicted probability of Sjögren's disease classification by age, derived from a multivariate logistic regression model including GS Score positivity (≥2), age (as a continuous variable) and the interaction term GS Score positivity × age. The plot shows two smoothed curves: one for GS Score positive (red) and one for GS Score negative (blue) patients. The model demonstrates a decline in the predictive value of GS Score positivity with increasing age. (B) Classification performance metrics, sensitivity, specificity and area under the ROC curve (AUC), calculated across age deciles and visualised using LOESS regression. Specificity declines with age, whereas sensitivity increases. (C,D) Patients stratified by the median age (≤55 years vs>55 years). No significant difference in the proportion of SjD diagnoses was observed between age groups. However, GS Score + (positivity if ≥2) was significantly more frequent in older patients. (E) ROC curves of GS Score stratified by age group (≤55 years vs>55 years). (F) Comparison of classification accuracy (sensitivity, specificity, AUC) for GS Score, CD Score, and combined GS+CD Score in younger and older patients. GS Score showed age-related differences whereas CD Score + (positivity if ≥2) and GS+CD Score + (positivity if ≥3) remained stable across age groups. *p<0.05, **p<0.01, ***p<0.001. AUC comparisons were performed using DeLong's test in R software. AUC, area under the curve; CD, colour Doppler; POS, positive; NEG, Negative; GS, Grey Scale; LOESS, locally estimated scatterplot smoothing; SjD, Sjögren's disease; ROC, receiver operating characteristic; UHFUS, ultra-high frequency ultrasound.

6) and visualised using LOESS regression (figure 5B). Specificity showed a progressive decline with age, while sensitivity increased. Despite this, the area under the ROC curve (AUC) remained relatively stable.

Given the non-normal age distribution and the observed inflection around the median age, we stratified patients into two groups: ≤55 years and >55 years. No statistically significant difference was observed in the proportion of patients classified as SjD between the younger (≤55 years) and older (>55 years) groups (47.3% vs 54.1%; figure 5C). However, the frequency of GS Score positivity was significantly higher in the older group (30.4% vs 54.1%), while no significant difference was found for CD Score positivity (p = 0.395; figure 5D).

We then compared ROC curves and classification metrics between age groups. For the GS Score, the AUC was comparable in younger and older patients (AUC = 0.776 vs 0.767; figure 5E,F), although curve shapes differed. Indeed, sensitivity was lower in younger individuals (57.1% vs 75.8%), whereas specificity was lower in the older group (71.4% vs 93.6%) (figure 5F). Conversely, CD Score and GS+CD Score performance was stable across age, with no significant differences in sensitivity (CD: 52.9% vs 57.6%; GS+CD: 61.4% vs 74.2%), specificity (CD: 80.8% vs 80.4%; GS+CD: 79.5% vs 73.2%) or AUC (CD: 0.718 vs 0.768; GS+CD: 0.763 vs 0.789) (figure 5F and online supplemental figures 7 and 8).

We also compared clinical, serological and ultrasonographic features between the two age groups. Younger patients had significantly higher levels of IgG (1405.5 vs 1268.5), SSB antibody levels (0.2 vs 0.1), a trend towards higher Ro60 levels (0.5 vs 0.3), and these features were similar to Cluster 1 GS– SjD. In contrast, older patients exhibited reduced unstimulated salivary flow (2.8 vs 2.0), along with more frequent ultrasonographic abnormalities, diffuse inhomogeneity (80.99% vs 39.86%) and hyperechoic bands (58.68% vs 34.46%); these features were similar to Cluster 2 GS+ Sicca (online supplemental table 13).

In summary, while the classification performance based on GS Score appears to be influenced by age, the accuracy of CD Score remains stable across age groups. The lower sensitivity of GS Score in younger patients, who tend to exhibit higher immunological activity, may reflect the presence of SjD Cluster 1 individuals with GS-negative profiles. In contrast, the reduced specificity observed in older patients, with diminished salivary flow and more ultrasonographic abnormalities, may be attributable to the inclusion of Sicca Cluster 2 cases, characterised by GS Score positivity.

DISCUSSION

This study demonstrates that UHFUS of the LSGs is a promising tool for the diagnosis and stratification of patients with SjD.

Histological evaluation of the LSGs remains the gold standard for diagnosis, primarily through the FS. Recent studies have shown that FS can serve as a diagnostic biomarker not only in LSGs but also in other glands, such as the parotid. Moreover, additional diagnostic histological biomarkers have been proposed.[20 21] In this study, we aimed to replicate this approach using ultrasound. First, we demonstrated that the OMERACT GS Score used for major salivary glands can also be applied to LSGs. The optimal cut-off was confirmed to be ≥2, in line with previous data on major salivary glands.[22] This cut-off confirmed that patients with GS ≥2 were more frequently classified as SjD, with high specificity and good accuracy, consistent with previous SGUS data. Second, we expanded our analysis to include additional ultrasonographic biomarkers in GS and CD. In GS, we found that a score of 3 was highly specific for SjD (99%), and that VHAs were found only in patients with SjD. Literature suggests that VHAs are associated with lymphoproliferative risk.[17 18] However, their histopathological counterpart remains to be established, and dedicated imaging–histology correlation studies are needed to clarify their biological meaning.[20] To our knowledge, this is the first study to systematically describe VHA in labial glands using UHFUS, supporting their potential as a novel imaging biomarker of SjD. Regarding CD, we evaluated the OMERACT CD Score and identified an optimal cut-off of ≥2, which demonstrated good specificity. Although the literature on CD scoring in major salivary glands remains limited due to its recent development, our findings indicate that its classification performance is comparable to that of the GS OMERACT Score. Importantly, CD may provide additional information. Interestingly, a CD Score ≥1 outperformed GS Score ≥1 in specificity, in line with early paediatric SjD data.[23] Furthermore, the combination of GS and CD into a composite Score (GS+CD OMERACT), with an optimal cut-off of ≥3, resulted in a higher AUC compared with GS or CD alone. Taken together, our findings suggest for the first time that in SGUS, it is essential not only to consider the GS OMERACT Score, but also to integrate additional GS and Doppler biomarkers. Intriguingly, this is the first study to demonstrate the added diagnostic value of integrating GS and CD Scores in labial glands with UHFUS.

Currently, no diagnostic criteria for SjD are available, and ultrasound has been proposed as a diagnostic tool either to replace certain classification criteria or to be used in combination with SSA status. In line with SGUS literature, we showed that a LSG GS OMERACT Score ≥2 could hypothetically be incorporated into the classification criteria with a value of 1 point.[3 4] Replacing biopsy with GS yielded an AUC of 0.945, whereas replacing SSA resulted in a lower AUC of 0.887. The superior performance as a replacement for biopsy is likely explained by the strong correlation between UHFUS findings and histology.[24] For the first time, we also evaluated CD OMERACT (≥2) and the composite GS+CD OMERACT Score (≥3) as potential replacements in classification criteria, but they did not add substantial value over GS

alone. We also investigated the utility of UHFUS in combination with SSA status alone. Concordance between the GS OMERACT Score and SSA status demonstrated strong predictive value; specifically, GS+SSA+ was highly indicative of SjD, while GS–SSA– of Sicca. While these findings align with previous SGUS studies, our study uniquely adds the inclusion of the CD OMERACT Score.[21 25] In discordant cases (GS+SSA- or GS-SSA+), the composite score (GS+CD OMERACT) helped to further distinguish patients. Notably, concordance between composite score and SSA was strongly predictive: (GS–, GSCD+SSA+) for SjD and (GS+, GSCD– SSA–) for Sicca. We subsequently proposed a flow chart relying solely on UHFUS and SSA status, which allowed accurate classification in approximately 75% of patients, with an overall classification accuracy of 94%.

We also sought to understand why GS OMERACT positivity may be discordant with SjD classification. In patients with Sicca syndrome, positive GS identifies a subgroup with histological glandular abnormalities not associated with immunological alterations. Specifically, we identified one cluster with glandular inflammation but lacking the typical SjD-associated serological profile and reduction in glandular function (Cluster 1 Sicca GS+), and another cluster of older patients with ultrasound and functional signs of glandular damage but no evidence of inflammation (Cluster 2 Sicca GS+). This suggests that in patients with GS +, non-autoimmune causes of sialadenitis or sialadenosis should be considered in the differential diagnosis.[26 27] In SjD, patients with negative GS had lower FS. Two subgroups were identified: one of young, triple-positive patients with higher immunological activity (Cluster 1 SjD GS–), and another mainly composed of Ro52+ patients with low immunological activity (Cluster 2 SjD GS–). Prior studies have suggested that SSA+ patients with negative FS may represent early stage disease, explaining Cluster 1 findings.[28] In contrast, Cluster 2 matches a recently described phenotype of isolated Ro52+ patients with milder disease activity.[29] Analogous findings have been reported in SGUS, where patients with SjD with negative ultrasound correspond to a milder phenotype.[30] However, unlike SGUS studies, which typically assess major glands by ultrasound and minor glands histologically, our study evaluates both imaging and histology on the same gland. Therefore, it can be hypothesised that in such cases, ultrasound closely reflects gland-specific histological changes than global disease classification.

Finally, we assessed classification performance in relation to age. With increasing age, specificity decreased; with younger age, sensitivity decreased. When stratified by median age, younger patients resembled Cluster 1 SjD GS–, while older patients resembled Cluster 2 Sicca GS+. Thus, reduced sensitivity in younger individuals may reflect early SjD with mild glandular inflammation, while reduced specificity in older patients may result from non-immune glandular damage and activity. Importantly, CD performance remained unaffected by age, indicating that it may serve as a diagnostic tool alongside GS across all age groups. Importantly, we show for the first time that classification performance of UHFUS varies with age, a novel observation that may have implications for interpretation of salivary gland ultrasonography in younger versus older patients.

Despite these encouraging findings, several limitations must be acknowledged. First, UHFUS is not yet widely implemented in routine practice, and the use of ultra-high frequency probes remains limited. Nonetheless, recent-generation machines with high-frequency (though not ultra-high) probes are generally sufficient to visualise LSGs.[31] Moreover, UHFUS is under investigation in other rheumatological settings and may become a versatile tool in specialised clinics.[32 33] Second, although VHAs emerged as a novel and highly specific ultrasonographic biomarker, their precise histopathological counterpart remains to be established. Dedicated imaging–histology correlation studies are warranted to confirm this relationship. Third, this was a single-centre study, the proposed OMERACT cut-off values were data-driven, and the cluster analyses were exploratory, factors that may limit generalisability. Therefore, the identified subgroups should be interpreted as hypothesis-generating, and require confirmation in independent cohorts. Nonetheless, the cohort was large and consecutively enrolled, the results overall are consistent with multicentre data on major SGUS, and they confirm preliminary evidence on minor salivary glands assessed by UHFUS. Another limitation is that major SGUS was not systematically performed. Therefore, we could not compare ultrasound abnormalities between major and minor salivary glands. Prospective multicentre studies integrating imaging, serology and histopathology are needed to validate our findings, while dedicated comparative studies across exocrine glands should clarify the relationship between major and minor salivary gland involvement in SjD.

In conclusion, from a research perspective, this is the first study to systematically demonstrate the role of VHA, Doppler signal and age-related changes in SGUS. From a clinical standpoint, UHFUS may serve as an emerging tool within tertiary referral centres, particularly when biopsy is not feasible or available, offering meaningful information for both diagnosis and patient phenotyping.

Contributors GF: conceptualisation, investigation, methodology, writing original draft, writing review and editing. RI: investigation, writing review and editing; FF: investigation, writing review and editing; GLR: writing review and editing. BD: writing review and editing; SF: investigation, writing review and editing; GL: investigation, writing review and editing; FG: investigation, writing review and editing; VD: investigation, writing review and editing; MM: writing review and editing. CB: writing original draft, writing review and editing. GF is the guarantor.

Funding This project is funded under the National Recovery and Resilience Plan (PNRR), Mission 4 Component 2 Investment 1.5, THE Tuscany Health Ecosystem (ECS – Ecosistema dell'innovazione sulle scienze e le tecnologie della vita in Toscana), HUB Tuscany Health Ecosystem Società Consortile a Responsabilità Limitata, SPOKE 7 Innovating Translational Medicine - Call for tender No. 3277 of DATE 30-12-2021 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval The study was approved by the Ethics Committee of the Tuscany Region – Area Vasta Nord Ovest under two protocols: SJOGREN PERSPECT (approval date: 17 October 2024) and Protocol 14540 (approval date: 14 March

2019). All participants provided written informed consent prior to participation. No animal subjects were involved in this study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iDs
Giovanni Fulvio https://orcid.org/0000-0003-0130-4412
Marta Mosca https://orcid.org/0000-0001-5937-4574

REFERENCES

1 Baldini C, Chatzis LG, Fulvio G, *et al*. Pathogenesis of Sjögren's disease: one year in review 2024. *Clin Exp Rheumatol* 2024;42:2336–43.
2 Ramsubeik K, Motilal S, Sanchez-Ramos L, *et al*. Diagnostic accuracy of salivary gland ultrasound in Sjögren's syndrome: A systematic review and meta-analysis. *Ther Adv Musculoskelet Dis* 2020;12:1759720X20973560.
3 Rebel D, de Wolff L, Delli K, *et al*. Added value of the salivary gland ultrasonography OMERACT score in the ACR/EULAR classification criteria for Sjögren's disease. *Semin Arthritis Rheum* 2024;67:152473.
4 Jousse-Joulin S, Gatineau F, Baldini C, *et al*. Weight of salivary gland ultrasonography compared to other items of the 2016 ACR/EULAR classification criteria for Primary Sjögren's syndrome. *J Intern Med* 2020;287:180–8.
5 Mossel E, van Nimwegen JF, Stel AJ, *et al*. Clinical Phenotyping of Primary Sjögren Syndrome Patients Using Salivary Gland Ultrasonography: Data From the RESULT Cohort. *J Rheumatol* 2021;48:717–27.
6 Hočevar A, Bruyn GA, Terslev L, *et al*. Development of a new ultrasound scoring system to evaluate glandular inflammation in Sjögren's syndrome: an OMERACT reliability exercise. *Rheumatology (Oxford)* 2022;61:3341–50.
7 Zabotti A, Manfrè V, Nano A, *et al*. Correlation between colour Doppler activity and parenchymal alterations in salivary glands of patients with primary Sjögren's disease. *Clin Exp Rheumatol* 2025;43:2202–8.
8 Tison A, Jousse-Joulin S, Consigny M, *et al*. Are ultrasound salivary parenchymal lesions more severe in primary Sjögren patients with a longer disease duration? A cross-sectional study. *Rheumatology (Oxford)* 2025;64:3739–46.
9 Choi I, Na DG, Paik W. Ultrasonographic echogenicity of normal salivary glands in adults: comparison of submandibular and parotid glands. *Ultrasonography* 2021;40:342–8.
10 Aringhieri G, Izzetti R, Vitali S, *et al*. Ultra-high frequency ultrasound (UHFUS) applications in Sjogren syndrome: narrative review and current concepts. *Gland Surg* 2020;9:2248–59.
11 Izzetti R, Vitali S, Aringhieri G, *et al*. Ultra-High Frequency Ultrasound, A Promising Diagnostic Technique: Review of the Literature and Single-Center Experience. *Can Assoc Radiol J* 2021;72:418–31.
12 Izzetti R, Fulvio G, Nisi M, *et al*. Reliability of OMERACT Scoring System in Ultra-High Frequency Ultrasonography of Minor Salivary Glands: Inter-Rater Agreement Study. *J Imaging* 2022;8:111.
13 Ferro F, Izzetti R, Vitali S, *et al*. Ultra-high frequency ultrasonography of labial glands is a highly sensitive tool for the diagnosis of Sjögren's syndrome: a preliminary study. *Clin Exp Rheumatol* 2020;38 Suppl 126:210–5.
14 Fisher BA, Jonsson R, Daniels T, *et al*. Standardisation of labial salivary gland histopathology in clinical trials in primary Sjögren's syndrome. *Ann Rheum Dis* 2017;76:1161–8.
15 Shiboski CH, Shiboski SC, Seror R, *et al*. 2016 American College of Rheumatology/European League Against Rheumatism Classification Criteria for Primary Sjögren's Syndrome: A Consensus and Data-Driven Methodology Involving Three International Patient Cohorts. *Arthritis & Rheumatology* 2017;69:35–45.
16 Jousse-Joulin S, D'Agostino MA, Nicolas C, *et al*. Video clip assessment of a salivary gland ultrasound scoring system in Sjögren's syndrome using consensual definitions: an OMERACT ultrasound working group reliability exercise. *Ann Rheum Dis* 2019;78:967–73.
17 Fulvio G, Donati V, Izzetti R, *et al*. Correspondence between minor salivary glands ultra-high frequency ultrasonography and histology: a case report of severe/atypical lymphoid infiltrate in Sjögren's syndrome. *Clin Exp Rheumatol* 2022;40:2474–5.
18 Lorenzon M, Tulipano Di Franco F, Zabotti A, *et al*. Sonographic features of lymphoma of the major salivary glands diagnosed with ultrasound-guided core needle biopsy in Sjögren's syndrome. *Clin Exp Rheumatol* 2021;39 Suppl 133:175–83.
19 Hocevar A, Ambrozic A, Rozman B, *et al*. Ultrasonographic changes of major salivary glands in primary Sjogren's syndrome. Diagnostic value of a novel scoring system. *Rheumatology (Oxford)* 2005;44:768–72.
20 van Ginkel MS, Nakshbandi U, Arends S, *et al*. Increased Diagnostic Accuracy of the Labial Gland Biopsy in Primary Sjögren Syndrome When Multiple Histopathological Features Are Included. *Arthritis & Rheumatology* 2024;76:421–8.
21 Mossel E, Delli K, van Nimwegen JF, *et al*. Ultrasonography of major salivary glands compared with parotid and labial gland biopsy and classification criteria in patients with clinically suspected primary Sjögren's syndrome. *Ann Rheum Dis* 2017;76:1883–9.
22 Fana V, Dohn UM, Krabbe S, *et al*. Application of the OMERACT Grey-scale Ultrasound Scoring System for salivary glands in a single-centre cohort of patients with suspected Sjögren's syndrome. *RMD Open* 2021;7:e001516.
23 Marrani E, Fulvio G, Virgili C, *et al*. Ultra-High-Frequency Ultrasonography of Labial Glands in Pediatric Sjögren's Disease: A Preliminary Study. *Diagnostics* 2023;13:2695.
24 Izzetti R, Nisi M. Imaging the Micron: New Directions in Diagnosis with Ultra-High-Frequency Ultrasound. *Diagnostics (Basel)* 2024;14:735.
25 Al Tabaa O, Gouze H, Hamroun S, *et al*. Normal salivary gland ultrasonography could rule out the diagnosis of Sjögren's syndrome in anti-SSA-negative patients with sicca syndrome. *RMD Open* 2021;7:e001503.
26 Koch M, Sievert M, Iro H, *et al*. Ultrasound in Inflammatory and Obstructive Salivary Gland Diseases: Own Experiences and a Review of the Literature. *J Clin Med* 2021;10:3547.
27 Liu S, Chen C, Zheng Q, *et al*. Non-negligible prevalence of focal lymphocytic sialadenitis in minor salivary glands of non-Sjögren's disease individuals. *Clin Exp Rheumatol* 2024;42:2499–506.
28 Sharma R, Chaudhari KS, Kurien BT, *et al*. Sjögren Syndrome without Focal Lymphocytic Infiltration of the Salivary Glands. *J Rheumatol* 2020;47:394–9.
29 Armağan B, Robinson SA, Bazoberry A, *et al*. Antibodies to Both Ro52 and Ro60 for Identifying Sjögren's Syndrome Patients Best Suited for Clinical Trials of Disease-Modifying Therapies. *Arthritis Care Res (Hoboken)* 2022;74:1559–65.
30 Zandonella Callegher S, Zabotti A, Giovannini I, *et al*. Normal-Appearing Salivary Gland Ultrasonography Identifies a Milder Phenotype of Primary Sjögren's Syndrome. *Front Med* 2020;7.
31 Fulvio G, Izzetti R, Aringhieri G, *et al*. UHFUS: A Valuable Tool in Evaluating Exocrine Gland Abnormalities in Sjögren's Disease. *Diagnostics (Basel)* 2023;13:2771.
32 Naredo E, Pascau J, Damjanov N, *et al*. Performance of ultra-high-frequency ultrasound in the evaluation of skin involvement in systemic sclerosis: a preliminary report. *Rheumatology (Oxford)* 2020;59:1671–8.
33 Di Battista M, Vitali S, Barsotti S, *et al*. Ultra-high frequency ultrasound for digital arteries: improving the characterization of vasculopathy in systemic sclerosis. *Semin Arthritis Rheum* 2022;57:152105.