Naredo E, Pascau J, Damjanov N, Lepri G, Gordaliza PM, Janta I, Ovalles-Bonilla JG, López-Longo FJ, Matucci-Cerinic M. Performance of ultra-high-frequency ultrasound in the evaluation of skin involvement in systemic sclerosis: a preliminary report. Rheumatology (Oxford). 2020 Jul 1;59(7):1671-1678

DOI: 10.1093/rheumatology/kez439

Performance of ultra-high-frequency ultrasound in the evaluation of skin involvement in systemic sclerosis: a preliminary report.

Esperanza Naredo [(1)], Javier Pascau [(2)], Nemanja Damjanov [(3)], Gemma Lepri[(4).], Pedro M. Gordaliza [(2)], Iustina Janta [(5),] Juan Gabriel Ovalles-Bonilla [(5)], Francisco Javier López-Longo [(5)], and Marco Matucci-Cerinic [(4).]

Affiliation

1. Dept of Rheumatology, Joint and Bone Research Unit, Hospital Universitario Fundación Jiménez Díaz. Madrid, Spain.

2. Bioengineering and Aerospace Engineering Department. Universidad Carlos III de Madrid. Instituto de Investigación Sanitaria Gregorio Marañón. Madrid, Spain.

3. Institute of Rheumatology, University of Belgrade Medical School, Belgrade, Serbia

4. Dept of Experimental and Clinical Medicine, University of Florence, & Dept of Geriatric Medicine, Div Rheumatology AOUC, Florence, Italy

5. Dept of Rheumatology. Hospital General Universitario Gregorio Marañon. Madrid, Spain.

Key-words. Ultrasound, sonography, systemic sclerosis, skin, scleroderma, dermis, hypodermis

Running title. Ultra-high-frequency ultrasound in the evaluation of skin involvement in systemic sclerosis

Corresponding author:

Matucci Cerinic Marco

Viale Pieraccini 18, 50139 Florence, Italy

e-mail: marco.matuccicerinic@unifi.it

Author contributions

Study design. Esperanza Naredo, Javier Pascau, Nemanja Damjanov, Francisco Javier López-Longo, Marco Matucci Cerinic

Acquisition of data. Esperanza Naredo, Javier Pascau, Nemanja Damjanov, Pedro M. Gordaliza, Iustina Janta, Juan Gabriel Ovalles-Bonilla, Francisco Javier López-Longo

Analysis and interpretation of data. Esperanza Naredo, Pedro M. Gordaliza, Javier Pascau, Nemanja Damjanov, Gemma Lepri, Iustina Janta, Francisco .Javier López-Longo, Marco Matucci Cerinic

Manuscript preparation. Esperanza Naredo, Javier Pascau, Nemanja Damjanov, Gemma Lepri, Pedro M. Gordaliza, Marco Matucci Cerinic

ABSTRACT

Background. High frequency ultrasound (US) allows visualization of epidermis, dermis and hypodermis, precise measurement of skin thickness, as well as assessment of skin edema, fibrosis and atrophy.

Objective. The aim of this pilot cross-sectional observational study was to assess the performance and multiobserver variability of ultra-high frequency (UHF) (50 MHz) ultrasound (US) in measuring skin thickness as well as the capacity of UHF-derived skin features to differentiate systemic sclerosis (SSc) patients from healthy controls.

Methods. 21 SSc patients (16 limited and 5 diffuse SSc) and 6 healthy controls were enrolled. All subjects underwent US evaluation by three experts at three anatomic sites (forearm, hand and finger). Dermal thickness was measured and two rectangular regions of interest (ROIs), one in dermis and one in hypodermis, were established for texture feature analysis.

Results. UHF US allowed a precise identification and measurement of the thickness of the dermis. The dermal thickness in the finger was significantly higher in patients than in controls ($p<0.05$), while in the forearm it was significantly lower in patients than in controls ($p<0.001$). Interobserver variability for dermal thickness was good to excellent [forearm ICC = 0.754; finger ICC = 0.699; hand ICC = 0.602]. Texture computed analysis of dermis and hypodermis was able to discriminate between SSc and healthy subjects (AUC>0.7).

Conclusions. These preliminary data show that skin UHF US allows a very detailed imaging of skin layers, a reliable measurement of dermal thickness, and a discriminative capacity between dermis and hypodermis texture features in SSc and healthy subjects.

Key words: systemic sclerosis, high frequency ultrasound, sonography, dermis, hypodermis

Abstract words count: 250

INTRODUCTION

Systemic Sclerosis (SSc) is an autoimmune disease with a complex pathogenesis leading to diffuse microangiopathy and tissue fibrosis [1]. Skin involvement is one of the major clinical features of SSc due to an abnormal collagen dermal deposition. The evolution of SSc skin involvement is characterized by three phases in temporal sequence: edematous, fibrotic and atrophic. The most frequently used method for the assessment of skin involvement in SSc is the modified Rodnan skin score (mRSS) which is a sensitive to change outcome measure, being as well predictive of disease outcome and survival [2]. However, intra and inter-observer variability of the mRSS is considerably high, i.e. around 30% [3, 4]. Therefore, a more objective, sensitive and reproducible tool for the measurement of SSc skin involvement is needed, both for clinical trials and daily clinical practice.

High frequency (≥18 MHz) ultrasound (US) with high spatial resolution may visualize the epidermis, dermis and hypodermis, precisely measure skin thickness, as well as assess skin edema, fibrosis and atrophy. A good correlation between US-measured skin thickness and total mRSS has been reported [5, 6]. However, in early SSc, US abnormalities can be observed before clinical signs of skin involvement are detectable [5-7]. US can also detect a decreased thickness and increased echogenicity of the skin due to the progressive skin fibrosis and atrophy in the late phase of SSc [8]. A recent study has shown a positive correlation between dermal thickness and microvascular damage severity assessed by nailfold videocapillaroscopy [9]. Various publications have reported good inter-observer and intra-observer agreement in the US evaluation of skin involvement [5, 10] as well as sensitivity to change of US-measured skin thickness in SSc [11].

The aim of this pilot cross-sectional observational study was twofold. Firstly, to assess the performance of an US probe with a ultra-high frequency (UHF US), i.e. 50-70 MHz, significantly higher than those previously used in SSc, for measuring and differentiating skin thickness in SSc and healthy controls, evaluating also the multi-observer variability of those measures; secondly, we evaluated if the contents of UHF US images were able to predict the condition of the subject (SSc or control). For the second objective, we developed a methodology to extract several features that summarize the spatial distribution of pixel intensities beyond the classical echogenicity. In order to quantify the image intensity information, we obtained texture features measured from the homogeneity, coarseness, regularity and directionality of pixel intensities in regions of interests (ROIs) [12,13]. These values were subsequently employed to build an automatic classifier capable of distinguishing between patients and controls.

PATIENTS AND METHODS

Twenty-one consecutive patients classified according to 2013 American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) SSc criteria [14] were prospectively recruited from the rheumatology outpatient clinic of the Hospital General Universitario Gregorio Marañón (Madrid, Spain). Six healthy controls were also recruited from the hospital staff. The study was conducted in accordance with the Declaration of Helsinki and was approved by the local ethics committee of the Hospital General Universitario Gregorio Marañón (Madrid, Spain). Informed consent was obtained from all patients and controls before study enrollment.

Clinical assessment

Demographics of patients and controls and disease characteristics of SSc patients were obtained. The following clinical and laboratory features were recorded: disease duration, subset of skin involvement [i.e diffuse (dc)/limited (lc)], presence of Raynaud´s phenomenon, arthritis, digital ulcers, interstitial lung disease, pulmonary hypertension and other internal organ involvement, treatment, subtypes of antinuclear antibodies (i.e. anti-Scl70, anticentromere antibodies (abs), others), and capillaroscopic pattern of microvascular damage (i.e. early, active, late). Skin involvement was scored according to the mRSS [2] by two rheumatologists not involved in the US assessment (JGO-B and MM).

Ultrasound assessment

Three rheumatology experts in musculoskeletal US (EN, ND, IJ) examined the SSc patients and the healthy controls consecutively and independently with a 50-70 MHz and 30 microns axial resolution US probe (VEVO MD, Visualsonics Inc., Toronto, Canada). B-mode settings were the following: frequency 50 MHz, gain 30 dB, dynamic range 70 dB, and depth 8.5 mm. The following anatomic areas of the more clinically involved side were scanned using a standardized scanning protocol: 1. (FAR) Dorsal aspect of the forearm, probe placed longitudinal and parasagittal at the mid-third of the forearm; 2. (HAN) Dorsal aspect of the hand, probe longitudinal to the fingers between the second and the third metacarpal bones; 3. (DIG) Dorsal aspect of the second finger, probe placed longitudinally between the metacarpophalangeal joint and the proximal interphalangeal joint. For the US scanning a generous layer of gel was applied that prevented the pressure of the probe on the skin of patients and controls.

The US images acquired by the three investigators were recorded in DICOM format (US modality, 256 gray levels) for further analysis. For every subject in the study we obtained 9 images: one image for each anatomic area and for each expert. The final dataset included a total of 243 images.

A single expert rheumatologist evaluated every US image using ImageJ software [15] and selected two points at the centre of the image, one at the interface between the epidermis and the dermis and the other at the interface between the dermis and the hypodermis to measure dermal thickness and two rectangular regions of interest (ROIs), one in dermis and the other one in hypodermis, with a minimum size of 0.15 mm^2. The ROIs were positioned in an area that characterized the intensity distribution of that tissue and was free of artifacts. The manually drawn ROIs were used to compute several features calculated from the spatial distribution of pixel intensities called texture features. These values were the input for the following analysis which consisted on automatically assessing the condition of the subject (SSc patient/control) from the image contents (Figure 1). The computation of these descriptors is presented in the next section.

Ultrasound assessment using texture features

The intensity distribution inside each ROI was expected to characterize the skin involvement in SSc. We used a set of features that are commonly called texture descriptors that characterize the spatial distribution of the pixel intensities inside the ROI. For instance, different values are obtained depending on the homogeneity, regularity, roughness or directionality of the pattern of intensities. However, it is difficult to establish which features are the correct ones to characterize and determine differences between pathological and healthy tissue, since they depend on the specific application and there are no previous reports on the use of UHF US for SSc characterization. We focused in the feature descriptors calculated from the Gray Level Co-occurrence Matrix (GLCM) [12, 16] due to their proven efficacy in US image analysis [13,17,18.]. The detailed description of this features can be found in the supplementary material. They characterize the pattern of intensities inside the ROI in great detail, so we will be able to detect differences not only in the average value or echogenicity, but on the homogeneity, regularity, roughness or directionality of the distribution of intensities.

Feature Evaluation and Selection

The number of features available for each ROI must be reduced to keep only the ones relevant for our classification problem. For this purpose, the features were ranked with Joint Mutual Information (JMI) technique [19], that ordered them according to their dependency in the data.

The next step was to input a subset of the most N relevant features for each ROI into a Decision Tree classifier, which is a reliable technique for noisy data [20, 21]. One classifier was trained for each anatomical area and layer, obtaining 6 different classifiers. The data available for each one consists of 81 ROIs (27 subjects and three observers). The robustness of our classifiers was ensured with a bootstrap technique and the performance was measured with the Area under the Curve (AUC) (see Supplementary Material). The final 6 classifiers can predict the condition of the subject from a single ROI obtained from a specific layer (dermis/hypodermis) of the UH FUS image, obtaining relevant texture features from the ROI that are measuring the distribution of pixel intensities in that layer.

Statistical Analysis

Inter-observer reliability for dermal thickness measures obtained from images acquired by the experts was tested by calculating the intraclass correlation coefficient (ICC; 2-way mixed effects). A ICC value lower than 0.40 was considered poor, from 0.40 to 0.50 moderate, from 0.50 to 0.70 good and from 0.70 to 1 excellent. Significant differences in skin features between patients and controls were analyzed with likelihood based mixed-effects model, repeated measures approach (MMRM).

RESULTS

Demographics and SSc features

The mean (SD, range) age was 55.0 (11.9, 30-76) for patients and 41.8 (19.1, 18-58) for controls (p-value: 0.047). The mean (SD, range) disease duration was 10.0 (8.4, 1-37) years. Sixteen (76.2 %) patients had limited cutaneous and 5 (23.8 %) had diffuse cutaneous disease. The population included patients with early, fibrotic and atrophic SSc. Seven (33.3%) patients had joint involvement, 6 (28.6%) interstitial lung disease, 3 (14.3%) pulmonary hypertension and 11 (52.4%) gastrointestinal involvement. All patients had Raynaud's phenomenon and 9 (42.9%) had digital ulcers. Twelve (57.1%) patients were positive for anticentromere abs and 3 (14.3%) for antitopoisomerase I abs. Capillaroscopy was available for 12 patients with the following patterns: early pattern in 3 patients, active in 2 patients and late in 7 patients. The mean (SD, range) mRSS for patients was 5.81 (5.20, 0-19) for the first assessor and 7.33 (6.40, 0-22) for the second assessor. The mean (SD, range) mRSS for controls was 0.0 (0.00, 0-0) for the first assessor and 1.00 (1.73, 0-3) for the second assessor. All patients were on treatment mainly with oral corticosteroids, vasodilators and/or immunosuppressants.

UHF US findings

The UHF US allowed the precise identification the dermis (Figure 2). We observed that in patients with puffy fingers the hypodermis was characterized by a marked edema subverting the structure of the cutaneous layer (Figure 2).

Dermal thickness interobserver variability

The interobserver variability results were excellent for the forearm (ICC = 0.754) and good for the finger (ICC = 0.699) and hand (ICC = 0.602). Table 1 shows the ICC obtained for each anatomic area.

Comparison of dermal thickness between patients and controls

The dermal thickness in the finger was significantly higher in patients (0.966 ± 0.181 mm) than in controls (0.805 ± 0.13 mm, $p < 0.05$), while in the forearm it was significantly lower in patients (1.055 ± 0.251 mm) than in controls (1.369 ± 0.226 mm, $p < 0.001$). No statistically significant differences were found for the hand (patients: 0.952 ± 0.264; controls: 0.944 ± 0.287, p=0.735). Figure 3a shows a boxplot of these measures separated for every expert. In all areas, the dispersion of the thickness measure was larger in patients.

Comparison of clinical (mRSS) and UHF US assessment

Model selection for classification

Table 2 shows the AUC number of features for each classifier, corresponding to dermis and hypodermis layers in DIG, FAR and HAN anatomical areas. AUC was always above 0.73, while the number of features selected was never more than 11. Different features were used to classify patients and controls for each anatomical location and dermal layer. In the Supplementary Material we provide a ranked list of the most important features at each anatomical region. Within the list, the feature Information Measure of Correlation1, obtained from GLCMs computed at different distances and directions, always appears in the top 3. Similarly, the Homogeneity feature appears at the top 2 for DIG and FAR, while HAN is characterized by features related with the two-dimensional correlation of the ROIs.

Interobserver agreement

In order to assess the robustness of the classification models we evaluated the previously obtained models with the images corresponding to each expert separately, obtaining the corresponding AUC. The AUC values are shown in Figure 3b: statistically significant differences are found for DIG images between experts, although none of them is below 0.7.

Patient evaluation by classification model *vs* mRSS

In Figure 4, the averaged mRSS of each patient is plotted against the averaged prediction over the three anatomical areas (where each prediction value ranges from 0 when the subject is classified as control and 1 when the subject is classified as patient). The classification model was not trained for scoring the patients, since the small sample size did not allow that kind of training. However, the figure shows how the controls probability is below 0.5 in all cases, while for the patients this output value is always above 0.5 and moderately correlated with the mRSS.

DISCUSSION

Our preliminary data obtained with an UHF US probe show that skin definition obtained with this technique is optimal, allowing us to evaluate all skin layers. In fact, the UHF US probe may specifically and clearly identify the epidermis, the papillary and the reticular dermal layers as well as the hypodermis.

In SSc, pathological studies have always identified the dermal layer as the one targeted by the disease. For this reason, we employed the UHF US to be able to precisely measure the dermal layer separated from the hypodermal layer. Our data demonstrated that UHF US precisely measured the thickness of the dermis in different areas of the body. This may be of significant help in catching the activity and the severity of the disease, and in the future also potentially in tailoring an adequate treatment of the disease.

US has been used in SSc to measure skin thickness and skin abnormalities have been detected in patients, in particular skin thickness and modification of echogenicity in the oedematous phase of the disease [22]. Different studies tried to correlate the results detected by skin US with clinical and histological features in SSc patients. Hesselstrand et al [23], using a 20 MHz probe, confirmed a higher skin thickness in SSc (both dcSSc and lcSSc) and a decrease in skin echogenicity in dcSSc compared to controls. In addition, this study investigated the correlation between skin abnormalities detected by US and the fibroblast production of proteoglycans in vitro showing that patients presenting greater abnormalities seemed to produce more versican while patients with change in echogenicity showed a greater production of biglycan and decorin [23]. Another study showed a correlation of the local skin thickness and echogenicity with the total and local mRSS value and confirmed the inverse correlation of thickness and echogenicity, probably reflecting the presence of oedema in the first phase of the disease. In addition, it suggested a possible role of skin US in in the prediction of a skin involvement also when the

clinical examination is negative (mRSS of 0) [24,25]. Sedky MM and al. compared the skin thickness of forty SSc patients to that of forty controls using 12-5 MHz linear array transducer confirming a correlation between skin thickness and mRSS. In addition, they suggested a potential role of US in identifying disease subsets, as patients with dcSSc had a thicker skin on the chest compared to lcSSc ones, and in identifying different phases of the disease. In fact, patients with a lower duration of the disease had a greater skin thickness compared to that with a long-term disease, suggesting a higher skin thickness in the edematous phase of the disease [26].

Despite many efforts have been made to verify the use of skin US in the assessment of cutaneous involvement in SSc patients, the heterogeneity of the different studies supports the need for further works to validate the use of skin US in the clinical practice [27]. In addition, an important point remains the fact that the majority of the papers do not specify which layer was measured. This obviously may be seen as a limitation because there may be difference when measuring the dermis only in respect to the evaluation of the whole dermal and hypodermal layer together. Therefore, this is an important issue that needs to be addressed in future studies using appropriate probes that allow a clear distinction between skin layers.

In our patients, in the early phase of the disease, characterized by puffy fingers, a significant edema, heavily modifying the layer's structure, was evident in the hypodermis. This unexpected detail merits attention in future studies because it shows that the early disease seems mainly focused into the hypodermis sparing the dermal layer. Moreover, this detail suggests that the evaluation of the skin thickness with the US should measure both skin layers, dermis and hypodermis. In fact, the maneuver of mRSS tests the skin thickness of the whole layer including the dermis and the epidermis. Our model is based on image properties, extracted from regions of interest in dermis and hypodermis, calculated from the spatial distribution of the image intensity. These features can automatically predict the condition of the subject, with small differences between the experts that acquired the images. However, these differences were significant for DIG region, where it is difficult to acquire the same area, although in all cases the average AUC was above 0.75. More importantly, the results show the ability of texture descriptors to characterize the tissue with a finer detail than the single echogenicity. Echogenicity is reflected among the computed texture features (average of each ROI), however our robust statistical analysis exhibits that this descriptor is never one of the features with the biggest impact in the classification task. As it was expected, features as the Information Measure of Correlation and the Homogeneity result much more informative due to their two-dimensional nature. These features allow us to quantify the characteristic speckle noise of an

US image, since they describe patterns of intensity changes inside a region (e.g. due to the collagen layers), while echogenicity cannot characterize these manifestations properly.

The significance of our approach should be considered preliminary as this was a pilot study showing the potential of UHF US probes to measure and characterize damaged tissue in SSc. Therefore, the skin involvement dependence with severity of SSc could not be modelled from the UHF US images at this point of our research. Nevertheless, still from this small pilot sample it is evident that UHF US provides detailed information about the dermal and hypodermal layers which can be used to differentiate patients from controls.

In conclusion, the current results indicate that with a larger and stratified sample we could potentially quantify the specific longitudinal progression and status of SSc thanks to UHF US images. Likely, in the future also other skin diseases could benefit from this imaging technique.

Competing interest

Esperanza Naredo. No competing interests related to the topic.

Javier Pascau. No competing interests related to the topic.

Nemanja Damjanov. No competing interests related to the topic.

Gemma Lepri. No competing interests related to the topic.

Pedro M. Gordaliza. No competing interests related to the topic.

Iustina Janta. No competing interests related to the topic.

Juan Gabriel Ovalles-Bonilla. No competing interests related to the topic.

Francisco Javier López-Longo. No competing interests related to the topic.

Marco Matucci-Cerinic. No competing interests related to the topic.

Funding. This study had no financial support.

References (total 31)

1. Varga J, Trojanowska M, Kuwana M. Pathogenesis of systemic sclerosis: recent insights of molecular and cellular mechanisms and therapeutic opportunities J Scleroderma Relat Disord 2017; 2:137-52.

2. Khanna D, Furst DE, Clements PJ, Allanore Y, Baron M, Czirjak L, et al. Standardization of the modified Rodnan skin score for use in clinical trials of systemic sclerosis. J Scleroderma Relat Disord 2017; 2: 11-18.

3. Steen VD, Medsger TA Jr. Improvement in skin thickening in systemic sclerosis associated with improved survival. Arthritis Rheum 2001; 44: 2828-35.

4. Czirjak L, Nagy Z, Aringer M, Riemekasten G, Matucci-Cerinic M, Furst DE; EUSTAR. The EUSTAR model for teaching and implementing the modified Rodnan skin score in systemic sclerosis. Ann Rheum Dis 2007; 66: 966-9.

5. Clements P, Lachenbruch P, Seibold J, White B, Weiner S, Martin R, et al. Inter and intraobserver variability of total skin thickness score (modified Rodnan TSS) in systemic sclerosis. J Rheumatol 1995; 22: 1281-5.

6. Kaloudi O, Bandinelli F, Filippucci E, Conforti ML, Miniati I, Guiducci S, et al. High frequency ultrasound measurement of digital dermal thickness in systemic sclerosis. Ann Rheum Dis 2010; 69: 1140-3.

7. Hesselstrand R, Scheja A, Wildt M, Akesson A. High-frequency ultrasound of skin involvement in systemic sclerosis reflects oedema, extension and severity in early disease. Rheumatology (Oxford) 2008; 47: 84-7.

8. Akesson A, Hesselstrand R, Scheja A, Wildt M. Longitudinal development of skin involvement and reliability of high frequency ultrasound in systemic sclerosis. Ann Rheum Dis 2004;63:791-6.

9. Sulli A, Ruaro B, Alessandri A, Pizzorni C, Cimmino MA, Zampogna G, et al. Correlation between nailfold microangiopathy severity, finger dermal thickness and fingertip blood perfusion in systemic sclerosis patients. Ann Rheum Dis 2014;73:247-51.

10. Moore TL, Lunt M, McManus B, Anderson ME, Herrick AL. Seventeen-point dermal ultrasound scoring system-a reliable measure of skin thickness in patients with systemic sclerosis. Rheumatology 2003; 42:1559-63.

11. Hesselstrand R, Carlestam J, Wildt M, Sandqvist G, Andreasson K. High frequency ultrasound of the skin in systemic sclerosis –a follow-up study. Arthritis Res Ther 2015; 17:329.

12. Haralick M, K. Shanmugam, and I. H. Dinstein. Textural features for image classification. IEEE Trans. Syst., Man Cybern., vol. 3, no. 6, pp. 610-21, Jun. 1973.

13. Gómez W, Pereira WCA, Infantosi AFC. Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound. IEEE transactions on medical imaging 2012; 31: 1889-99.

14. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. Ann Rheum Dis 2013;72:1747-55.

15. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012 ; 9: 671-5

16. Petrou M , Sevilla P. G , Image Processing. Dealing With Texture. Wiltshire, U.K.: Wiley, 2006

17. Alic L, Niessen WJ, Veenland JF. Quantification of heterogeneity as a biomarker in tumor imaging: a systematic review. PloS one 2014; 9(10):e110300.

18. Thijssen JM. Ultrasonic speckle formation, analysis and processing applied to tissue characterization. Pattern Recognition Letters 2003; 24: 659-75.

19. Chow TW, Huang D. Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information. IEEE Transactions on Neural networks 2005; 16:213-24.

20. Frénay B, Verleysen M. Classification in the presence of label noise: a survey. IEEE transactions on neural networks and learning systems 2014; 25: 845-69.

21. Nettleton DF, Orriols-Puig ., Fornells A. A study of the effect of different types of noise on the precision of supervised learning techniques. Artificial intelligence review 2010; 33: 275-306.

22. Polańska A, Dańczak-Pazdrowska A, Jałowska M, Żaba R, Adamski Z. Current applications of high-frequency ultrasonography in dermatology. Adv Dermatol Allergol 2017; 34: 535-42.

23. Hesselstrand R, Westergren-Thorsson G, Scheja A, Wildt M, Åkesson A. The association between changes in skin echogenicity and the fibroblast production of biglycan and versican in systemic sclerosis. Clinical and Experimental Rheumatology 2002; 20: 301-8.

24. Hesselstrand R, Scheja A, Wildt M, Akesson A. High-frequency ultrasound of skin involvement in systemic sclerosis reflects oedema, extension and severity in early disease. Rheumatology 2008; 47: 84-87.

25. Sulli A, Ruaro B, Smith V, Paolino S, Pizzorni C, Pesce G, Cutolo M. Subclinical dermal involvement is detectable by high frequency ultrasound even in patients with limited cutaneous systemic sclerosis. Arthritis Res Ther 2017: 19:61.

26. Sedky MM[1], Fawzy SM, El Baki NA, El Eishi NH, El Bohy Ael M. Systemic sclerosis: an ultrasonographic study of skin and subcutaneous tissue in relation to clinical findings. Skin Res Technol 2013 Feb;19(1):e78-84.

27. Santiago T, Santiago M, Ruaro B, Salvador MJ, Cutolo M, da Silva JAP. Ultrasonography for the Assessment of Skin in Systemic Sclerosis: A Systematic Review. Arthritis Care Res (Hoboken). 2019 Apr;71(4):563-574.

28. Davidson R, MacKinnon JG. Bootstrap tests: How many bootstraps?. Econometric Reviews 2000;19: 55-68.

29. Wilcox RR. Fundamentals of modern statistical methods: Substantially improving power and accuracy. Springer 2010.

30. Efron B, Tibshirani R. Improvements on cross-validation: the 632+ bootstrap method. Journal of the American Statistical Association 1997; 92:548-60.

31. Sahiner B, Chan HP, Hadjiiski L. Classifier performance prediction for computer-aided diagnosis using a limited dataset. Medical Physics 2008; 35:1559-70.

Tables (2)

Table I: ICC values for the different anatomical areas, DIG=finger; FAR=forearm; HAN= hand			
Area	**ICC**	**p**	**CI (95 %)**
DIG	0.699	<0.001	0.519 to 0.836
FAR	0.754	<0.001	0.594 to 0.868
HAN	0.602	<0.001	0.389 to 0.775
DIG: digital-finger; FAR: forearm; HAN: hand; ICC: intraclass correlation coefficient; CI: confidence interval.			

Table II: Best mean AUC over 500 bootstrap samples obtained using the best N features (N.F.) for each layer and anatomical area						
	DIG		**FAR**		**HAN**	
	Dermis	Hypodermis	Dermis	Hypodermis	Dermis	Hypodermis
AUC	0.804	0.824	0.738	0.793	0.782	0.749
N.F.	3	6	4	3	11	3
DIG, finger; FAR, forearm; HAN, hand						

Figure legends:

Figure 1. Example of UHF US image of the forearm location: a) Acquisition display in the US device; b) US image with calipers used to measure dermis thickness in yellow and ROIs extracted to characterize texture features (blue rectangle for dermis and red for hypodermis); c.1) zoomed ROI in dermis; c.2) zoomed ROI in hypodermis.

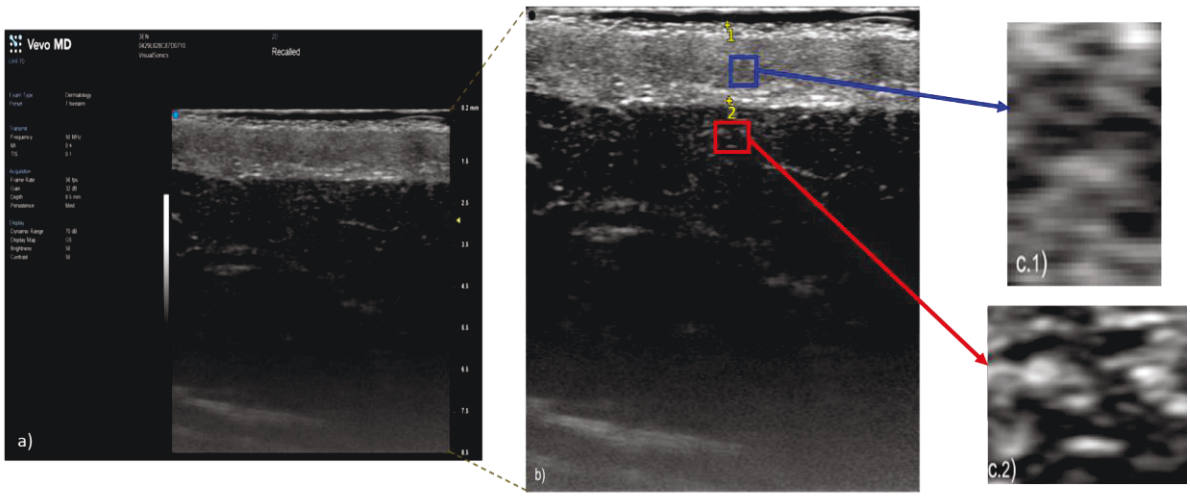


Figure 2. UHF US of HAN region showing dermis and hypodermis in a healthy control (first row) and in a patient with mRRS = 6,6 (second row). Each column corresponds to the image acquired by one of the experts. The yellow landmarks are the calipers used to measure dermis thickness.

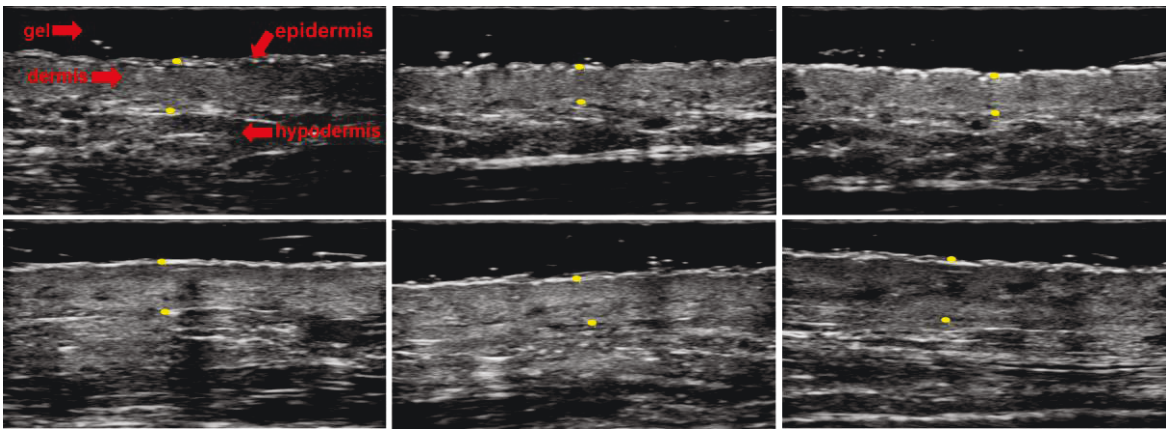
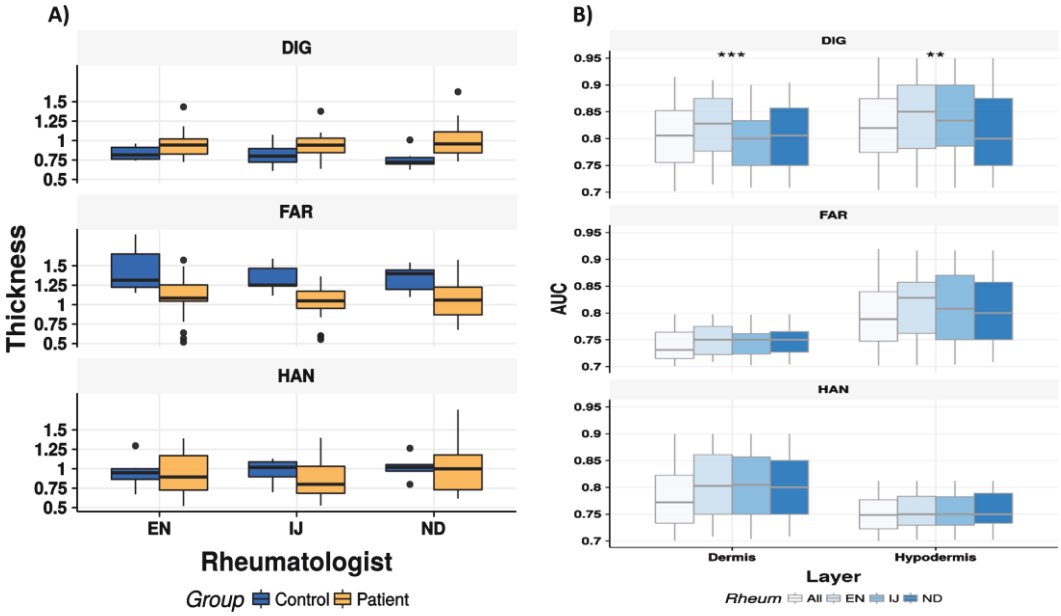


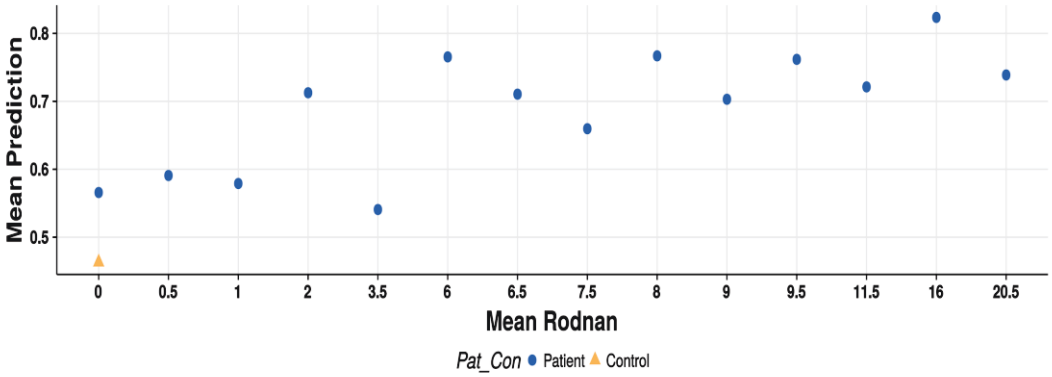
Figure 3. A) Each boxplot represents the dermal thickness measures obtained from the image acquired by the indicated Rheumatologist in the 27 subjects (21 patient, 6 controls). Each row represents an anatomical area. Each level in the Rheum axis represents controls (blue) or patients (orange) measures. DIG=finger; FAR=forearm; HAN= hand. **B)** Model Selection- Each boxplot shows the AUCs obtained from

all the bootstrap samples in each anatomic area (in rows) and for each skin layer (in columns). In each boxplot, the first box from the left corresponds to all experts together, while the other three are the ones obtained from individual expert images (from left to right EN, IJ and ND). Significant differences tested with likelihood based mixed-effects.



DIG=finger; FAR=forearm; HAN= hand - model, repeated measures approach: $p < 0.05$ "*", $p < 0.01$"**" and $p < 0.001$ "***"

Figure 4: The plot shows the average prediction using the three experts. The average value for patients is depicted as ● and the average for control as ▲



Supplementary Material:

The GLCM provides a mathematical description of all pairwise combinations of gray levels separated by a particular distance and orientation inside the ROI. From each 256 gray-levels image and for every ROI we calculated GLCM in four directions (0º, 45º, 90º, 135º) and at five distances (1, 2, 3, 4, 5 pixels) obtaining a total of 20 GLCMs per ROI. From each one of these matrices we extracted 22 well-known feature descriptors. These descriptors can be understood as the two-dimensional counterpart of the descriptors usually employed for histograms, i.e: average (or echogenicity), standard deviation, variance, skewness, kurtosis, etc. We extracted 22 feature descriptors for each GLCM, so since we have GLCM in 4 different orientations and 5 distances, each ROI is described by a total of 440 features.

When training a classifier with a small number of samples it is easy to overfit the model, so we created 500 bootstrap samples [28, 29] from the 81 original ROIs to train the classifier with the small sample size in this pilot study. Every training set included 63 random ROIs and the resulting classifier was tested with the remaining 18 ROIs, for which we obtained the Area Under the Curve (AUC) using the 0.632+ error estimation [13,30,31]. The model with N features that resulted in the best mean AUC was selected for SSc tissue characterization using the UHF US images in that anatomical area. This means that from the 440 original texture features, we selected the N ones that better classify the patients and controls. These features will be related only with the distribution of intensities, not with dermal thickness, since they have been calculated from a rectangular ROI in each layer.

The features selected for each classifier are detailed in the following paragraphs. Each classifier corresponds to one anatomical location and layer:

Anatomical location: **DIG**

Layer: **Dermis**

Number of features used for classification: **3**

10 most important features (in bold the ones used for classification):

Homogenity1_5_90, InformationMeasureCorrelation1_3_135, InformationMeasureCorrelation2_5_0, InformationMeasureCorrelation1_5_135, InformationMeasureCorrelation1_5_135, InformationMeasureCorrelation1_5_90, Correlation1_1_45, ClusterShade_5_45, ClusterShade_3_45, InformationMeasureCorrelation1_1_135

Anatomical location: **DIG**

Layer: **Hypodermis**

Number of features used for classification: **6**

10 most important features (in bold the ones used for classification):

InformationMeasureCorrelation1_5_90, Homogenity2_5_135, InformationMeasureCorrelation1_5_45, InformationMeasureCorrelation1_5_45, InformationMeasureCorrelation1_5_45, Correlation1_3_45, Correlation2_3_45, ClusterShade_3_45, ClusterShade_5_45, Correlation1_5_90

Anatomical location: **FAR**

Layer: **Dermis**

Number of features used for classification: **4**

10 most important features (in bold the ones used for classification):

Homogenity2_5_90, Homogenity2_5_45, InformationMeasureCorrelation1_5_90, Homogenity2_3_45, Homogenity1_3_45, InformationMeasureCorrelation1_3_0, Correlation1_1_90, Correlation2_1_90, InformationMeasureCorrelation1_5_90, InformationMeasureCorrelation1_1_0

Anatomical location: **FAR**

Layer: **Hypodermis**

Number of features used for classification: **3**

10 most important features (in bold the ones used for classification):

InformationMeasureCorrelation1_5_90, Homogenity2_3_45, DifferenceVariance_5_0, InformationMeasureCorrelation1_5_45, SumAverage_3_135, InformationMeasureCorrelation1_5_0, InformationMeasureCorrelation1_3_0, ClusterShade_3_90, ClusterShade_5_90, Correlation1_3_45

Anatomical location: **HAN**

Layer: **Dermis**

Number of features used for classification: **11**

10 most important features (in bold the ones used for classification):

Correlation1_5_45, Correlation2_5_45, InformationMeasureCorrelation1_5_45, InformationMeasureCorrelation1_5_45, Correlation1_5_45, Correlation2_5_45, Correlation1_5_45, Correlation2_5_45, InformationMeasureCorrelation1_5_90, InformationMeasureCorrelation1_5_90

Anatomical location: **HAN**

Layer: **Hypodermis**

Number of features used for classification: **3**

10 most important features (in bold the ones used for classification):

DifferenceEntropy_3_45, InformationMeasureCorrelation1_5_135, InformationMeasureCorrelation1_5_135, SumVariance_1_0, InformationMeasureCorrelation1_3_90, ClusterProminance_5_135, Correlation1_5_90, Correlation2_5_90, InformationMeasureCorrelation1_5_135, ClusterShade_5_45