frontiers in Medicine

ORIGINAL RESEARCH
published: 01 March 2021
doi: 10.3389/fmed.2021.589197

Artificial Intelligence for Ultrasound Informative Image Selection of Metacarpal Head Cartilage. A Pilot Study

Edoardo Cipolletta[1†], Maria Chiara Fiorentino[2†], Sara Moccia[2,3†], Irene Guidotti[2], Walter Grassi[1], Emilio Filippucci[1†] and Emanuele Frontoni[2†]*

[1] Rheumatology Unit, Department of Clinical and Molecular Sciences, Polytechnic University of Marche, Ancona, Italy, [2] Department of Information Engineering, Polytechnic University of Marche, Ancona, Italy, [3] Department of Advanced Robotics, Italian Institute of Technology, Genoa, Italy

OPEN ACCESS

Edited by:
Helena Canhao,
New University of Lisbon, Portugal

Reviewed by:
Andreas P. Diamantopoulos,
Martina Hansens Hospital, Norway
Sandra Salvador Falcao,
New University of Lisbon, Portugal
Fernando Saraiva,
Centro Hospitalar Universitário Lisboa Norte - Lisbon, Portugal

***Correspondence:**
Edoardo Cipolletta
edoardocipolletta@gmail.com

†ORCID:
Edoardo Cipolletta
orcid.org/0000-0002-6881-8197
Maria Chiara Fiorentino
orcid.org/0000-0002-7397-803X
Sara Moccia
orcid.org/0000-0002-4494-8907
Emilio Filippucci
orcid.org/0000-0002-7251-7784
Emanuele Frontoni
orcid.org/0000-0002-8893-9244

Specialty section:
This article was submitted to Rheumatology, a section of the journal Frontiers in Medicine

Received: 30 July 2020
Accepted: 19 January 2021
Published: 01 March 2021

Citation:
Cipolletta E, Fiorentino MC, Moccia S, Guidotti I, Grassi W, Filippucci E and Frontoni E (2021) Artificial Intelligence for Ultrasound Informative Image Selection of Metacarpal Head Cartilage. A Pilot Study. Front. Med. 8:589197. doi: 10.3389/fmed.2021.589197

Objectives: This study aims to develop an automatic deep-learning algorithm, which is based on Convolutional Neural Networks (CNNs), for ultrasound informative-image selection of hyaline cartilage at metacarpal head level. The algorithm performance and that of three beginner sonographers were compared with an expert assessment, which was considered the gold standard.

Methods: The study was divided into two steps. In the first one, an automatic deep-learning algorithm for image selection was developed using 1,600 ultrasound (US) images of the metacarpal head cartilage (MHC) acquired in 40 healthy subjects using a very high-frequency probe (up to 22 MHz). The algorithm task was to identify US images defined informative as they show enough information to fulfill the Outcome Measure in Rheumatology US definition of healthy hyaline cartilage. The algorithm relied on VGG16 CNN, which was fine-tuned to classify US images in informative and non-informative ones. A repeated leave-four-subject out cross-validation was performed using the expert sonographer assessment as gold-standard. In the second step, the expert assessed the algorithm and the beginner sonographers' ability to obtain US informative images of the MHC.

Results: The VGG16 CNN showed excellent performance in the first step, with a mean area (AUC) under the receiver operating characteristic curve, computed among the 10 models obtained from cross-validation, of 0.99 ± 0.01. The model that reached the best AUC on the testing set, which we named "MHC identifier 1," was then evaluated by the expert sonographer. The agreement between the algorithm, and the expert sonographer was almost perfect [Cohen's kappa: 0.84 (95% confidence interval: 0.71–0.98)], whereas the agreement between the expert and the beginner sonographers using conventional assessment was moderate [Cohen's kappa: 0.63 (95% confidence interval: 0.49–0.76)]. The conventional obtainment of US images by beginner sonographers required 6.0 ± 1.0 min, whereas US videoclip acquisition by a beginner sonographer lasted only 2.0 ± 0.8 min.

Conclusion: This study paves the way for the automatic identification of informative US images for assessing MHC. This may redefine the US reliability in the evaluation of MHC integrity, especially in terms of intrareader reliability and may support beginner sonographers during US training.

Keywords: hyaline cartilage, ultrasonography, metacarpal head, artificial intelligence, deep learning, convolutional neural network, rheumatoid arthritis, osteoarthritis

INTRODUCTION

Hyaline cartilage is a highly specialized connective tissue characteristic of synovial joints. Its principal function is to provide low-friction articular surfaces and to act as a shock absorber during the joint movement (1). Hyaline cartilage lacks blood vessels, thus it has a limited capacity for intrinsic healing and repair. In this regard, the integrity of this noble tissue is essential to joint health. The chondrocyte, the unique cell type in adult hyaline cartilage, maintains a stable equilibrium between the synthesis and the degradation of extracellular matrix components. With age and/or in the presence of various rheumatic diseases, such as rheumatoid arthritis and osteoarthritis, this balance is undermined, and the catabolic activity exceeds the anabolic one, thus leading to dehydration, degeneration, and thinning of the cartilage layer (2).

Although conventional radiography is the most adopted imaging method for the assessment of joint damage in daily clinical practice, it provides only an indirect visualization of the hyaline cartilage through the evaluation of the joint space narrowing. The accuracy of conventional radiography has been questioned in non-weight-bearing joints such as the ones of hands and wrists, which are commonly involved in different rheumatic diseases (3, 4). Moreover, in several studies, conventional radiography was found to be less sensitive than ultrasonography (US) in the detection of joint damage (5–11).

Recently, US has been suggested as a reliable and reproducible tool for the assessment of the hyaline cartilage of the small joints of the hand (4, 11–17). One of the main drawbacks of US is its subjectivity in the interpretation of US findings and the consequent variable inter- and intraobserver reliability (18–20). This issue is particularly relevant for the beginner musculoskeletal sonographer (18–20). Thus, the development of a tool that can enhance the US learning process is noteworthy (21).

In the last few years, artificial intelligence has been gaining importance in US, and a number of advantages have been claimed for this alternative method over the conventional acquisition and interpretation of US images, including faster performance, higher reliability and better standardization of image acquisition (22–27). Only a few studies have applied artificial intelligence in the field of musculoskeletal diseases (28–32), and no studies explored the artificial intelligence (AI) in the US assessment of hyaline cartilage, except our previous preliminary work (33).

To date, deep learning (DL) has shown its value in the healthcare domain for computer-assisted medical image analysis (24). DL is a branch of AI, and its algorithms are inspired by human brain, being able to learn from a large amount of data by itself. In fact, DL has the advantage of directly learning image features from raw data, avoiding the need to design hand-crafted features as for traditional machine learning approaches (23). In particular, Convolutional Neural Networks (CNNs), one of the most popular DL algorithms, are widely used for medical image analysis tasks such as classification, segmentation, detection, and biometric measurements, with applications in diagnosis and image-guided interventions and therapy (23). CNNs are able to recognize complex structures in an image by applying convolutional filters (each defined by a kernel matrix) whose weights are iteratively learned during CNN training. Training a CNN relies on labeled data, which are used to learn the mapping from the input to the desired output. Training a CNN from scratch requires a large amount of labeled data, which may not always be available, especially in the medical image analysis domain where image labeling by expert clinicians is a resource-expensive procedure. Moreover, training a CNN from scratch may often lead to overfitting issues (i.e., CNN inability to generalize on new sets of data). A feasible alternative is to exploit transfer learning. Transfer learning consists of extracting knowledge from one task (where large annotated datasets are available) and using the extracted knowledge for a second one. It has been demonstrated that using transfer learning is particularly useful for medical image analysis since limited training data are usually available (34, 35).

Driven by this last consideration, we decided to use transfer learning for training a CNN for US informative-frame selection. We investigated CNNs pretrained on ImageNet, a large image database that includes more than 1 million of annotated natural images (e.g., cats, dogs, cars).

The main aim of this study was to develop an automatic DL algorithm for US informative-image selection of hyaline cartilage at metacarpal head level. An image was defined informative when it shows enough information to fulfill the Outcome Measure in Rheumatology US definition of healthy hyaline cartilage (36). The algorithm performance and that of three beginner sonographers were compared with an expert assessment, which was considered the gold standard.

MATERIALS AND METHODS

Study Design

The study was conducted from January 2019 to March 2020. The study was divided in two steps. In the first one, a CNN algorithm for informative image selection was developed and trained using 1,600 static US images. The US images were acquired by an expert

(E.F.) in musculoskeletal US who evaluated the metacarpal head cartilage (MHC) from the 2nd to the 5th digit bilaterally in 40 healthy subjects.

In the second step of the study, the CNN output was compared with the conventional assessment of the MHC carried out independently by three beginner (E.M., F.F., and J.D.B.) and the expert (E.F.) sonographers. A beginner sonographer was defined as a sonographer with limited experience (<1 year) in the US assessment of hyaline cartilage. MHC from the 2nd to the 5th digit of both hands of eight healthy subjects was independently evaluated on the same day by the beginner and the expert sonographers. The beginner sonographers were asked to provide a set of eight US images per subject (one US image per each metacarpal head) for a total of 192 static images and a set of eight 10-s videoclips on the same healthy subjects for a total of 192 videoclips. A random sample of 128 static images and 64 videoclips were used in the reliability analysis. The US images evaluated by the beginners (each beginner judged a third of the US images) and the videoclips assessed by the algorithm were tested against the expert opinion who had to state whether the US images were informative or not. Expert was blinded to the US images authorship (i.e., beginner sonographer or artificial intelligence algorithm). The time required for each US examination was registered.

Subjects

Healthy subjects were selected from relatives visiting or accompanying in- and out-patients, friends of the authors, and medical students attending the "Carlo Urbani" Hospital (Jesi, Ancona, Italy). Healthy subjects were enrolled because they had the lowest probability to present US abnormalities indicative of cartilage damage. In fact, pathologic findings may generate a bias in the interpretation of US images by both the beginner sonographers and the algorithm.

Exclusion criteria were as follows: (i) previous diagnosis of inflammatory/degenerative arthropathies; (ii) joint pain [visual analog scale (VAS) ≥10/100] and/or analgesic or non-steroidal anti-inflammatory drugs' intake in the 4 weeks preceding the visit; (iii) age <18 years old; and (iv) hard tissue enlargement or deformity of the metacarpophalangeal, proximal, or distal interphalangeal joints suggestive of hand osteoarthritis and/or joint inflammation at physical examination. The following demographic data were recorded: sex, age, handedness, height, weight, and body mass index.

US Image Acquisition and Interpretation

US examinations were carried out using a MyLab Class C (Esaote SpA, Genoa, Italy), equipped with a very high-frequency broadband linear probe (10–22 MHz). A grayscale standard setting was adopted (B-mode frequency: 22 MHz, master gain: 70%, mechanical index: 0.3, dynamic range: 12, depth 15 mm, focus position at the area of interest).

The hands were placed on the table, with the metacarpophalangeal joints in maximal flexion (>60°) (14, 36, 37). The metacarpal head from the 2nd to the 5th digit of both hands was scanned on the dorsal aspect from radial to ulnar and from proximal to distal sides to ensure the maximal exploration of the hyaline cartilage using the EULAR standard scans (38). Particular attention was paid to ensure a perpendicular insonation of the cartilage surface (14, 36, 37).

The normal appearance of the hyaline cartilage is characterized by a homogenous hypo-anechoic layer delimited by two regular, sharp, and bright margins where insonated orthogonally (36, 37). An US image displaying such characteristics was defined informative. The detection in the videoclip by the algorithm or by the expert of at least a frame showing US features of healthy hyaline cartilage was a sufficient criterion to classify it as informative. **Figure 1** provides a pictorial evidence of a healthy MHC.

CNN Algorithm for Informative Frame Selection

A VGG16 CNN, pre-trained on the ImageNet dataset, was used for transfer learning. The VGG16 architecture was chosen for two main reasons: (1) its sequential architecture results to be particularly suitable for small-size training set and low image variability (2, 39) its shallow architecture (only 3 × 3 convolutional layers stacked on top of each other in increasing depth) is associated with low computation cost.

In the original VGG16 architecture implementation, the ImageNet images are processed through 13 convolutional (conv) layers to perform feature extraction. The filters used in each conv layer have very small receptive field (3 × 3) (the smallest size to capture notion about left/right, up/down and center),

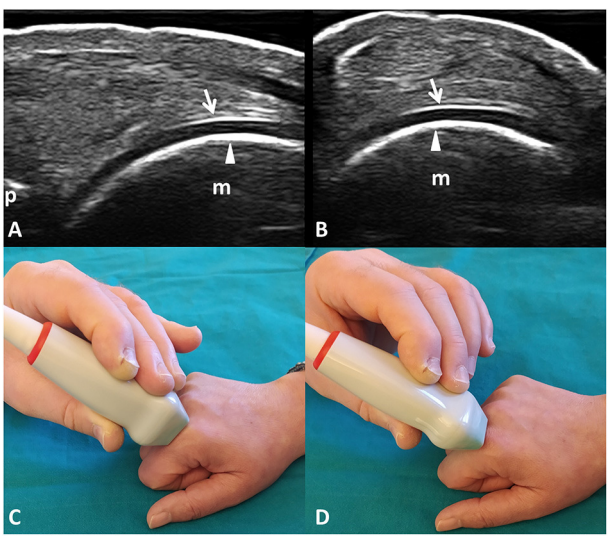

FIGURE 1 | Healthy hyaline cartilage. Dorsal longitudinal **(A)** and transverse **(B)** scans of the hyaline cartilage of the 2nd metacarpal head in a healthy subject. Hyaline cartilage appears as a homogeneously hypo-anechoic layer delimited by two regular, sharp, continuous, and hyperechoic interfaces where insonated orthogonally. Arrows indicate the outer margin (i.e., the chondrosynovial interface); arrowheads indicate the inner margin (i.e., the osteochondral interface). **(C,D)** The position of the probe in the dorsal longitudinal **(C)** and transverse **(D)** scans. m, metacarpal head; p, base of the proximal phalanx.

followed by a rectified linear unit (ReLU) activation function. After the convolutional layers, three fully connected (FC) layers (4,096 neurons in the first two layers and 1,000 neurons in the last one) followed by a softmax layer are used to predict class probability.

In this work, the ImageNet pretrained weights were used as a starting point for the CNN training process. We modified the three FC layers using 1,024, 512, and 2 neurons in the first, second, and third FC layers, respectively, to adapt the architecture to our binary classification problem (informative—non-informative image classification) (**Figure 2**). The architecture was trained freezing the first 10 conv layers and tuning the remaining ones. In this way, we managed to exploit the knowledge encoded in the VGG16 trained on the large ImageNet dataset.

Training Strategy

Prior training the VGG16, US images were resized to 224 × 224 pixels and converted to RGB images, repeating the grayscale channel for three times, to match the image input dimension required by the pre-trained VGG16. Intensity mean was removed from each image.

The mini-batch gradient descent, with a momentum of 0.9, was used as optimizer using the categorical cross-entropy as loss function. The batch size was set to 64 as a balance between training speed and gradient convergence. Training was performed for 100 epochs with a learning rate of 0.0001.

A leave-four-subject out cross-validation was performed for testing the classification. The dataset was divided into 10 subsets of subjects. In turn, one of the 10 subsets (containing four subjects) was used as the test set while the other nine subsets were used as training set. The validation was performed selecting randomly four subjects from the training set, obtaining in such a way 10 models.

The analyses were performed using Keras with TensorFlow library as backend on Google Colaboratory (https://colab.research.google.com/).

Performance Metrics

To measure the performance of our approach, we computed the mean area under the curve (AUC) of the operating characteristic curve (ROC) and the mean classification Precision (*Prec*), Recall (*Rec*), and f1-score (*f1*) for the *i*th class, with i ∈ C (informative, non-informative), where TP_i,FP_i,FN_i were the true positives, false positives, and the false negatives, respectively.

$$Prec_i = \frac{TP_i}{TP_i + FP_i} \quad (1)$$

$$Rec_i = \frac{TP_i}{TP_i + FN_i} \quad (2)$$

$$f1_i = \frac{2\ x\ Prec_i\ x\ Rec_i}{Prec_i + Rec_i} \quad (3)$$

Statistical Analysis

Results are expressed as number and/or corresponding percentage for qualitative variables and as mean and standard deviation (SD) for quantitative variables. The Chi-square test and the Mann-Whitney test were used to compare the qualitative and quantitative variables, respectively. The agreement in the informative image selection between the expert (i.e., the gold standard) and the algorithm, and between the expert and the conventional assessment of the beginners was calculated using an unweighted Cohen's kappa and interpreted according to Landis and Koch (40).

Two-tailed *p*-values <0.05 were considered significant. Statistical analysis was performed using Statistical Package for the Social Sciences (SPSS) software (version 26.0 for Windows, Chicago, Illinois, USA).

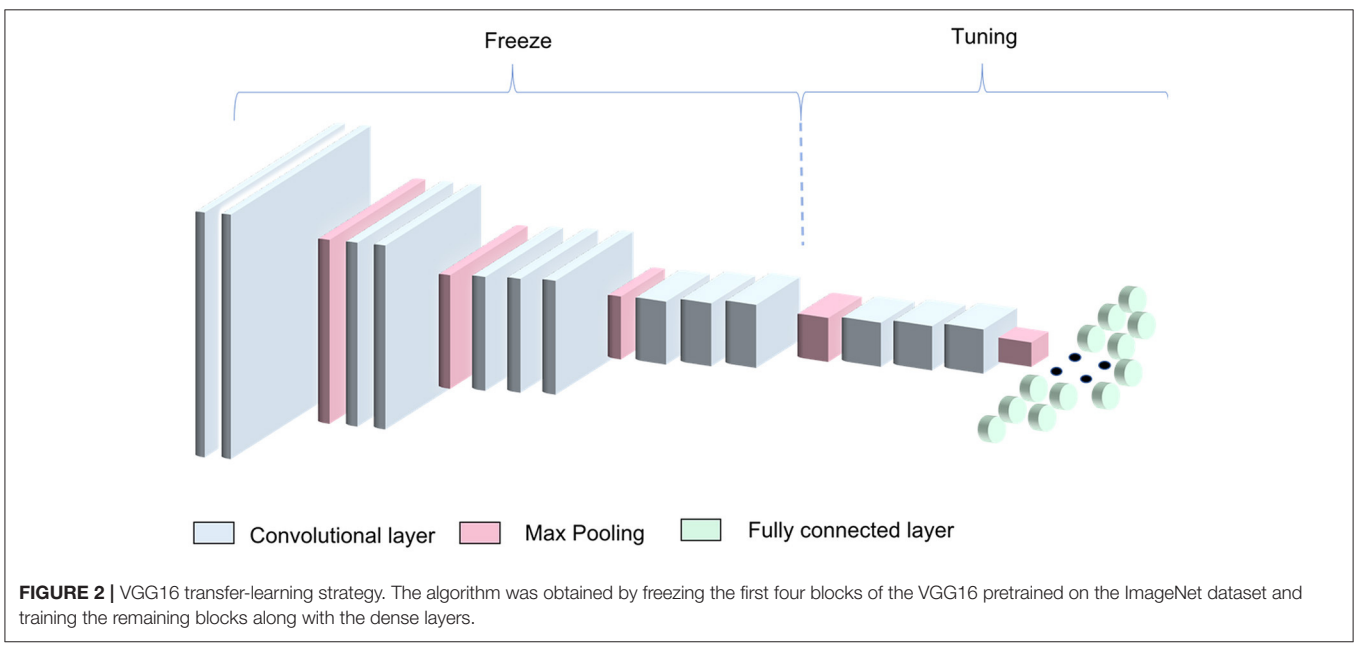

FIGURE 2 | VGG16 transfer-learning strategy. The algorithm was obtained by freezing the first four blocks of the VGG16 pretrained on the ImageNet dataset and training the remaining blocks along with the dense layers.

RESULTS

Subjects

A total of 48 healthy subjects were included in this monocentric and cross-sectional study: 40 in the training and testing of the algorithm (first step) and 8 in the reliability analysis (second step). **Table 1** shows the main demographic characteristics of the participants.

Artificial-Intelligence Algorithm: Training and Testing

In the first step of the study, the VGG16 CNN showed excellent performance in the informative image selection task, with an AUC of 0.99 ± 0.01 (**Figure 3**) computed among the 10 models.

Table 2 shows the classification results for both informative and non-informative frame obtained from the cross-validation procedure.

Figure 4 shows an example of informative and non-informative frames selected by the model that reached the best AUC, which we named "MHC identifier 1."

TABLE 1 | Demographic characteristics of the healthy participants.

Sex (female/male)	33/15
Age (years, mean ± SD)	54.6 ± 5.6
Handedness (right/left)	40/8
Height (cm, mean ± SD)	170.7 ± 9.5
Weight (kg, mean ± SD)	72.4 ± 6.4
Body mass index (kg/m^2, mean ± SD)	24.8

SD, standard deviation.

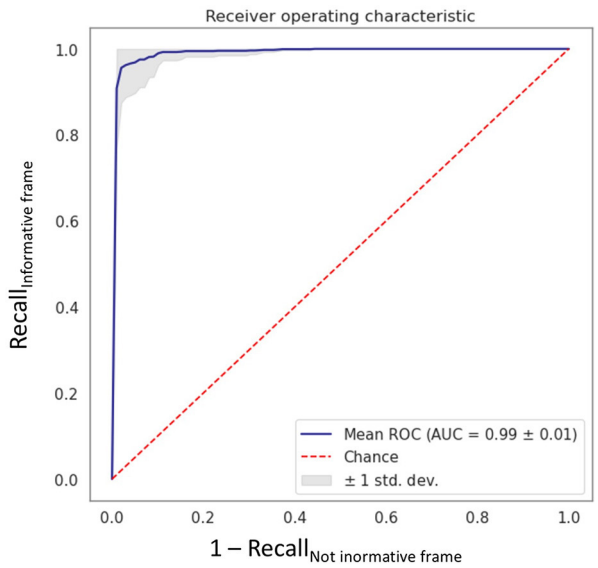

FIGURE 3 | Receiving operating characteristic curve obtained with the fine-tuned VGG16 CNN algorithm "MHC identifier 1." AUC, area under the curve; ROC, receiving operating characteristics; std dev, standard deviation.

Feasibility

The average time required to complete the conventional US assessment was 6.0 ± 1.0 and 4.0 ± 0.5 min for beginners and expert sonographer, respectively ($p < 0.01$). On the other hand, the time spent to acquire the videoclips was 2.0 ± 0.8.

Reliability Analysis

The agreement between the automatic algorithm and the expert sonographer was almost perfect (Cohen's kappa: 0.84, 95% confidence interval: 0.71–0.98); whereas, the agreement between the expert and the beginners using conventional assessment was moderate (Cohen's kappa: 0.63, 95% confidence interval: 0.49–0.76) ($p < 0.01$) (**Table 3**) without significant difference in the interobserver agreement between the expert and each beginner sonographer ($p = 0.14$).

TABLE 2 | Classification metrics for precision, recall, and f1-score, with i ∈ C (informative, non-informative).

	i	Precision	Recall	f1-score
VGG16 CNN algorithm "MHC identifier 1"	Informative	0.94 (0.07)	0.98 (0.02)	0.97 (0.05)
	Non-informative	0.98 (0.02)	0.96 (0.06)	0.97 (0.04)

Values in brackets are the standard deviation.
MHC, metacarpal head cartilage.

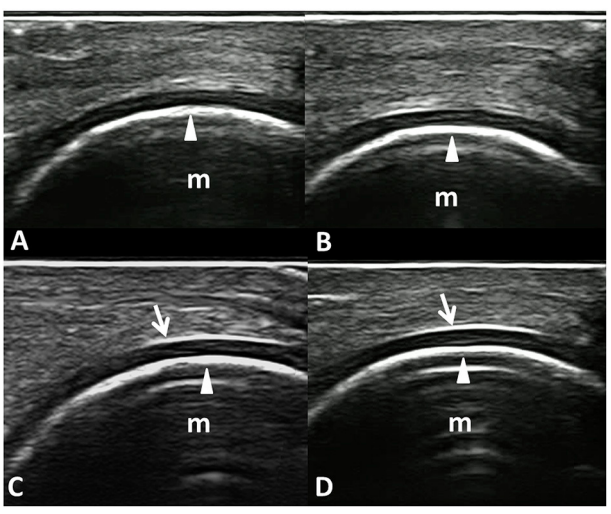

FIGURE 4 | Informative and non-informative US frame selected by the VGG16 CNN algorithm "MHC identifier 1." Dorsal longitudinal scans acquired at the metacarpal head level in healthy subjects. **(A,B)** Examples of non-informative frames; **(C,D)** Examples of informative frames. In fact, while the inner margin (arrowheads) is evident in all the panels, the chondrosynovial (arrows) interface is clearly visible only in the lower images **(C,D)**. m, metacarpal head; arrows, chondrosynovial interface; arrowheads, osteochondral interface.

TABLE 3 | Interobserver agreement.

		Beginners	
US static images ($n = 128$)		**Informative**	**Non-informative**
Expert	Informative	56 (43.8)	8 (6.3)
	Non-informative	16 (12.5)	48 (37.5)
Total agreement: 81.3%			
		MHC Identifier 1	
10-s videoclips ($n = 64$)		**Informative**	**Non-informative**
Expert	Informative	31 (48.4)	2 (3.1)
	Non-informative	3 (4.7)	28 (43.8)
Total agreement: 92.2%			

Values in brackets are percentages.
MHC, metacarpal head cartilage; US, ultrasound.
The beginner sonographers and the VGG16 CNN algorithm "MHC identifier 1" agreement with the expert sonographer considered the gold standard.

DISCUSSION

Last-generation US systems allow the real-time identification of otherwise undetectable musculoskeletal abnormalities, which have a growing impact in the management of many rheumatic diseases (41, 42). However, US is a highly operator-dependent technique, and sonographer skills and experience may affect both acquisition and interpretation processes (19, 20). Several international initiatives were undertaken to ensure the standardization of US assessment and to increase its reproducibility in rheumatological setting (36, 38, 43, 44). The use of artificial intelligence in musculoskeletal US may further increase its reproducibility and may save sonographers time as shown in cardiological setting (45).

The correct acquisition of an US image is the essential step to ensure an accurate and reliable assessment of the image itself (31, 39, 46). Thus, we believe that the availability of an algorithm facilitating the identification of the region of interest during the acquisition process of US images represents a further step toward the standardization of US examination.

This study describes the first steps taken to develop an algorithm that can identify informative US images for the assessment of the MHC. The application of such an algorithm, that we called "MHC identifier 1," may redefine the US reliability in the evaluation of the MHC integrity, especially in terms of intrareader reproducibility. MHC identifier 1 showed an almost perfect agreement with the expert sonographer. Here are some possible explanations of disagreement between the algorithm and the expert: while processing an US videoclip, the algorithm may detect even just one frame to define it as informative, which may be missed or considered not relevant by the expert assessment. Conversely, the expert may consider sufficiently assessable US images even if rejected by the algorithm for not strictly fulfilling all the morphostructural criteria. The choice to use both videoclips and static images could be considered a limit of the present study. However, we decided to test the performance of the algorithm using videoclips instead of pictures, since its task will be to support the sonographer during a real-time US examination and not only in the interpretation of static images. Conversely, the sonographer selects a representative US image which conveys the message of a part of the US examination. Finally, according to the data we recorded, the use of this algorithm may shorten the time required to obtain informative US images up to one-third. However, it should be borne in mind that, to date, MHC identifier 1 cannot be routinely applied to clinical practice. In fact, the images must be manually exported from the US system and transferred in a computer where the algorithm can analyze them. A future development may include the incorporation of MHC identifier1 into an US machine to test the clinical value and the feasibility of this method in real-life setting. Thus, the feasibility of this method, including easiness of its use, costs, and availability are yet to be determined.

Our study presents some limitations. First, the same expert sonographer, who served as the imaging "gold standard," was also the teacher and tutor of the beginner sonographers. This fact may imply a possible risk of systematic bias. Second, only a relatively low number of US images were used in the reliability exercise. However, this limit should be read in light of the fact that it is a pilot study. Third, its monocentric design may limit the generalizability of our results. Fourth, the impact of machine-assisted acquisition of US images was not evaluated. Furthermore, the intrareader reliability of both the sonographers and the algorithm was not tested. Finally, the impact of using different US systems needs to be tested, as images in this study were all obtained with the same US system.

The assessment of MHC status is progressively gaining a relevant role in the management of patients with different chronic arthropathies. In fact, in 2019, the Outcome Measure in Rheumatology US Working Group proposed the US definitions of cartilage damage in patients with rheumatoid arthritis and osteoarthritis (36, 44). The same group of experts is currently carrying out a Delphi exercise to define and quantify the structural joint damage (including cartilage lesion) in rheumatoid arthritis by US.

Although preliminary, our results open up new horizons in the use of artificial intelligence in the US evaluation of hyaline cartilage. In fact, the algorithm MHC identifier 1 could enhance the learning process improving the awareness of the beginner sonographer regarding the probe positioning required to obtain images conveying essential information to assess the MHC. In addition, the ability of this algorithm to identify informative frame on videoclips may suggest its use as a tool that could assist the sonographer during the real time US examination.

Further implementation may allow to measure the MHC thickness and to evaluate the hyaline cartilage of other sites commonly involved by rheumatic diseases such as the knee and the hip (16, 17, 47).

The possibility of identifying, evaluating, and measuring the hyaline cartilage in a reliable and faster way may reduce the US examination time, shorten the learning curve of beginner sonographers by taking advantage

of AI feedbacks, and promote new studies in this field (e.g., aimed to compare the semiquantitative scoring system of cartilage damage and the quantitative assessment, and to follow-up the progression of cartilage damage).

In conclusion, this study describes the first steps in the development of an algorithm identifying informative US images for assessing the MHC. The automatic selection of US images acquired by beginner sonographers resulted reliable and feasible, as shown by the comparison with an expert sonographer. The application of such an algorithm may redefine the US reliability in the evaluation of the MHC integrity, especially in terms of intrareader reliability and may support beginner sonographers during US training. However, this algorithm needs further validation before its use in clinical practice.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

EC, EFi, MF, and SM substantially contributed to study conception and design, acquisition, analysis, and interpretation of data. IG substantially contributed to acquisition, analysis, and interpretation of data. EFr and WG substantially contributed to analysis and interpretation of data. All the authors revised the paper and approved the final version of the article to be published.

ACKNOWLEDGMENTS

We would like to thank Dr. Jacopo Di Battista, Dr. Francesca Francioso, and Dr. Erica Moscioni for acting as beginner sonographers in the agreement exercise.

REFERENCES

1. Fox SAJ, Bedi A, Rodeo SA. The basic science of articular cartilage: structure, composition, and function. *Sports Health*. (2009) 1:461–8. doi: 10.1177/1941738109350438
2. Pap T, Korb-Pap A. Cartilage damage in osteoarthritis and rheumatoid arthritis-two unequal siblings. *Nat Rev Rheumatol*. (2015) 11:606–15. doi: 10.1038/nrrheum.201595
3. Navarro-compán V, Landewé R, Provan SA, Ødegård S, Uhlig T, Kvien TK, et al. Relationship between types of radiographic damage and disability in patients with rheumatoid arthritis in the EURIDISS cohort: a longitudinal study. *Rheumatology*. (2014) 54:83–90. doi: 10.1093/rheumatology/keu284
4. Mandl P, Supp G, Baksa G, Radner H, Studenic P, Gyebnar J, et al. Relationship between radiographic joint space narrowing, sonographic cartilage thickness and anatomy in rheumatoid arthritis and control joints. *Ann Rheum Dis*. (2015) 74:2022–7. doi: 10.1136/annrheumdis-2014-205585
5. Døhn UM, Ejbjerg BJ, Court-Payen M, Hasselquist M, Narvestad E, Szkudlarek M, et al. Are bone erosions detected by magnetic resonance imaging and ultrasonography true erosions? A comparison with computed tomography in rheumatoid arthritis metacarpophalangeal joints. *Arthritis Res Ther*. (2006) 8:R110. doi: 10.1186/ar1995
6. Wakefield RJ, Gibbon WW, Conaghan PG., O'Connor P, McGonagle D, Pease C, et al. The value of sonography in the detection of bone erosions in patients with rheumatoid arthritis: a comparison with conventional radiography. *Arthritis Rheum*. (2000) 43:2762–70. doi: 10.1002/1529-0131(200012)43:12<2762::AID-ANR16>3.0.CO;2
7. Szkudlarek M, Klarlund M, Narvestad E., Court-Payen M, Strandberg C, Jensen KE, et al. Ultrasonography of the metacarpophalangeal and proximal interphalangeal joints in rheumatoid arthritis: a comparison with magnetic resonance imaging, conventional radiography and clinical examination. *Arthritis Res Ther*. (2006) 8:R52. doi: 10.1186/ar1904
8. Scheel AK, Hermann KGA, Ohrndorf S, Werner C, Schirmer C, Detert J, et al. Prospective 7 year follow up imaging study comparing radiography, ultrasonography, and magnetic resonance imaging in rheumatoid arthritis finger joints. *Ann Rheum Dis*. (2006) 65:595–600. doi: 10.1136/ard.2005041814
9. Funck-Brentano T, Etchepare F, Joulin SJ, Gandjbakch F, Pensec VD, Cyteval C, et al. Benefits of ultrasonography in the management of early arthritis: a cross-sectional study of baseline data from the ESPOIR cohort. *Rheumatology*. (2009) 48:1515–9. doi: 10.1093/rheumatology/kep279
10. Wiell C, Szkudlarek M, Hasselquist M., Møller JM, Vestergaard A, Nørregaard J, et al. Ultrasonography, magnetic resonance imaging, radiography, and clinical assessment of inflammatory and destructive changes in fingers and toes of patients with psoriatic arthritis. *Arthritis Res Ther*. (2007) 9:R119. doi: 10.1186/ar2327
11. Hurnakova J, Filippucci E, Cipolletta E, Matteo Di. A, Salaffi F, Carotti M, et al. Prevalence and distribution of cartilage damage at the metacarpal head level in rheumatoid arthritis and osteoarthritis: an ultrasound study. *Rheumatology*. (2019) 58:1206–13. doi: 10.1093/rheumatology/key443
12. Möller B, Bonel H, Rotzetter M, Villiger PM, Ziswiler H. Measuring finger joint cartilage by ultrasound as a promising alternative to conventional radiograph imaging. *Arthritis Rheum*. (2009) 61:435–41. doi: 10.1002/art24424
13. Filippucci E, da Luz KR, Di Geso L, Salaffi F, Tardella M, Carotti M, et al. Interobserver reliability of ultrasonography in the assessment of cartilage damage in rheumatoid arthritis. *Ann Rheum Dis*. (2010) 69:1845–8. doi: 10.1136/ard.2009125179
14. Cipolletta E, Filippucci E, Matteo Di. A, Tesei G, Cosatti M, Di Carlo M, et al. The reliability of ultrasound in the assessment of hyaline cartilage in rheumatoid and healthy metacarpal heads. *Ultraschall Med*. (2020). doi: 10.1055/a-1285-4602. [Epub ahead of print].
15. Iagnocco A, Conaghan PG, Aegerter P, Moller I, Bruyn GAW, Chary-Valckenaere I, et al. The reliability of musculoskeletal ultrasound in the detection of cartilage abnormalities at the metacarpo-phalangeal joints. *Osteoarthritis Cartilage*. (2012) 20:1142–6. doi: 10.1016/j.joca.2012.07003
16. Cipolletta E, Filippucci E, Matteo Di A, Hurnakova J, Di Carlo M, Pavelka K, et al. FRI0634 standard reference values of metacarpal head cartilage thickness measurement by ultrasound in healthy subjects. *Ann Rheum Dis*. (2019) 78:1014–5. doi: 10.1136/annrheumdis-2019-eular5807
17. Cipolletta E, Hurnakova J, Matteo Di A, Di Carlo M, Pavelka K, Grassi W, et al. Prevalence and distribution of cartilage and bone damage at metacarpal head in healthy subjects. *Clin Exp Rheumatol* (2020). [Epub ahead of print].
18. Hammer HB, Iagnocco A, Mathiessen A, Filippucci E, Gandjbakhch F, Kortekaas MC, et al. Global ultrasound assessment of structural lesions in osteoarthritis: a reliability study by the OMERACT ultrasonography group

on scoring cartilage and osteophytes in finger joints. *Ann Rheum Dis*. (2016) 75:402–7. doi: 10.1136/annrheumdis-2014-206289
19. Gutiérrez M, Di Geso L, Rovisco J, Di Carlo M, Ariani A, Filippucci E, et al. Ultrasound learning curve in gout: a disease-oriented training program. *Arthritis Care Res*. (2013) 65:1265–74. doi: 10.1002/acr22009
20. Gutierrez M, Filippucci E, Ruta S, Salaffi F, Blasetti P, di Geso L, et al. Inter-observer reliability of high-resolution ultrasonography in the assessment of bone erosions in patients with rheumatoid arthritis: experience of an intensive dedicated training programme. *Rheumatology*. (2011) 50:373–80. doi: 10.1093/rheumatology/keq320
21. Filippucci E, Meenagh G, Ciapetti A, Iagnocco A, Taggart A, Grassi W. E-learning in ultrasonography: a web-based approach. *Ann Rheum Dis*. (2007) 66:962–5. doi: 10.1136/ard.2006064568
22. Akkus Z, Cai J, Boonrod A, Zeinoddini A, Weston AD, Philbrick KA., et al. A survey of deep-learning applications in ultrasound: artificial intelligence–powered ultrasound for improving clinical workflow. *J Am Coll Radiol*. (2019) 16:1318–28. doi: 10.1016/j.jacr.2019.06004
23. Liu S, Wang Y, Yang X, Lei B, Liu L, SX Li, et al. Deep learning in medical ultrasound analysis: a review. *Engineering*. (2019) 5:261–75. doi: 10.1016/j.eng.2018.11020
24. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. (2017) 42:60–88. doi: 10.1016/j.media.2017.07005
25. Stoel B. Use of artificial intelligence in imaging in rheumatology – current status and future perspectives. *RMD Open*. (2020) 6:e001063. doi: 10.1136/rmdopen-2019-001063
26. Zaffino P, Moccia S, De Momi E, Spadea MF. A review on advances in intra-operative imaging for surgery and therapy: imagining the operating room of the future. *Ann Biomed Eng*. (2020) 48:2171–91. doi: 10.1007/s10439-020-02553-6
27. Moccia S, De Momi E, El Hadji S, Mattos LS. Blood vessel segmentation algorithms - review of methods, datasets and evaluation metrics. *Comput Methods Programs Biomed*. (2018) 158:71–91. doi: 10.1016/j.cmpb.2018.02001
28. Chang RF, Lee CC, Lo CM. Computer-aided diagnosis of different rotator cuff lesions using shoulder musculoskeletal ultrasound. *Ultrasound Med Biol*. (2016) 42:2315–22. doi: 10.1016/j.ultrasmedbio.2016.05016
29. Klauser AS, Franz M, Bellmann WR, Gruber J, Hartig F, Mur E, et al. Contrast-enhanced ultrasonography for the detection of joint vascularity in arthritis - subjective grading versus computer-aided objective quantification. *Ultraschall der Medizin*. (2011). 32:E31–7. doi: 10.1055/s-0031-1281671
30. Chang RF, Lee CC, Lo CM. Quantitative diagnosis of rotator cuff tears based on sonographic pattern recognition. *PLoS ONE*. (2019) 14:e0212741. doi: 10.1371/journal.pone0212741
31. Andersen JKH, Pedersen JS, Laursen MS, Holtz K, Grauslund J, Savarimuthu TR, et al. Neural networks for automatic scoring of arthritis disease activity on ultrasound images. *RMD Open*. (2019) 5:e000891. doi: 10.1136/rmdopen-2018-000891
32. Christensen ABH, Just SA, Andersen JKH, Savarimuthu TR. Applying cascaded convolutional neural network design further enhances automatic scoring of arthritis disease activity on ultrasound images from rheumatoid arthritis patients. *Ann Rheum Dis*. (2020) 79:1189–93. doi: 10.1136/annrheumdis-2019-216636
33. Fiorentino MC, Moccia S, Cipolletta E, Filippucci E, Frontoni E. A learning approach for informative-frame selection in US rheumatology images. In: Autexier S, Campbell J, Rubio J, Sorge V, Suzuki M, Wiedijk F, editors. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. New York, NY: Springer Verlag (2019). p. 228–36. doi: 10.1007/978-3-030-30754-7_23
34. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. (2017) 542:115–8. doi: 10.1038/nature21056
35. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging*. (2016) 35:1299–312. doi: 10.1109/TMI.2016253 5302
36. Mandl P, Studenic P, Filippucci E, Bachta A, Backhaus M, Bong D, et al. Development of semiquantitative ultrasound scoring system to assess cartilage in rheumatoid arthritis. *Rheumatology*. (2019) 58:1802–11. doi: 10.1093/rheumatology/ke z153
37. Torp-Pedersen S, Bartels E, Wilhjelm J, Bliddal H. Articular Cartilage thickness measured with US is not as easy as it appears: a systematic review of measurement techniques and image interpretation. *Ultraschall der Medizin - Eur J Ultrasound*. (2010) 32:54–61. doi: 10.1055/s-0029-1245386
38. Möller I, Jan I, Backhaus M, Ohrndorf S, Bong DA, Martinoli C, et al. The 2017 EULAR standardised procedures for ultrasound imaging in rheumatology. *Ann Rheum Dis*. (2017) 76:1974–9. doi: 10.1136/annrheumdis-2017-211585
39. Patrini I, Ruperti M, Moccia S, Mattos LS, Frontoni E, De Momi E. Transfer learning for informative-frame selection in laryngoscopic videos through learned features. *Med Biol Eng Comput*. (2020) 58:1225–38. doi: 10.1007/s11517-020-02127-7
40. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. (1977) 33:159–74. doi: 10.2307/2529310
41. Filippucci E, Cipolletta E, Mashadi Mirza R, Carotti M, Giovagnoni A, Salaffi F, et al. Ultrasound imaging in rheumatoid arthritis. *Radiol Med*. (2019) 124:1087–100. doi: 10.1007/s11547-019-01002-2
42. Ruta S, Reginato AM, Pineda C, Gutierrez M. Pan-American League Against Rheumatisms (PANLAR) Ultrasound Study Group. General applications of ultrasound in rheumatology: why we need it in our daily practice. *J Clin Rheumatol*. (2015) 21:133–43. doi: 10.1097/RHU0000000000000230
43. Naredo E, D'Agostino MA, Conaghan PG, Backhaus M, Balint P, Bruyn GAW, et al. Current state of musculoskeletal ultrasound training and implementation in Europe: results of a survey of experts and scientific societies. *Rheumatology*. (2010) 49:2438–43. doi: 10.1093/rheumatology/keq243
44. Bruyn GA, Iagnocco A, Naredo E, Balint PV, Gutierrez M, Hammer HB, et al. OMERACT definitions for ultrasonographic pathologies and elementary lesions of rheumatic disorders 15 years on. *J Rheumatol*. (2019) 46:1388–93. doi: 10.3899/jrheum181095
45. Davis A, Billick K, Horton K, Jankowski M, Knoll P, Marshall JE, et al. Artificial intelligence and echocardiography: a primer for cardiac sonographers. *J Am Soc Echocardiogr*. (2020) 33:1061–6. doi: 10.1016/j.echo.2020. 04025
46. Moccia S, Vanone GO, Momi E, Laborai A, Guastini L, Peretti G, et al. Learning-based classification of informative laryngoscopic frames. *Comput Methods Programs Biomed*. (2018) 158:21–30. doi: 10.1016/j.cmpb.2018.01030
47. Martel-Pelletier J, Barr AJ, Cicuttini FM, Conaghan PG, Cooper C, Goldring MB, et al. Osteoarthritis. *Nat Rev Dis Primers*. (2016) 2:16072. doi: 10.1038/nrdp.201672

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.