



UNIVERSITÀ POLITECNICA DELLE MARCHE
Repository ISTITUZIONALE

A deep-learning framework for metacarpal-head cartilage-thickness estimation in ultrasound rheumatological images

This is the peer reviewed version of the following article:

Original

A deep-learning framework for metacarpal-head cartilage-thickness estimation in ultrasound rheumatological images / Fiorentino, M.C., Cipolletta, E., Filippucci, E., Grassi, W., Frontoni, E., Moccia, S..
- In: COMPUTERS IN BIOLOGY AND MEDICINE. - ISSN 0010-4825. - 141:(2022).
[10.1016/j.combiomed.2021.105117]

Availability:

This version is available at: 11566/294105 since: 2024-09-17T13:11:36Z

Publisher:

Published

DOI:10.1016/j.combiomed.2021.105117

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. The use of copyrighted works requires the consent of the rights' holder (author or publisher). Works made available under a Creative Commons license or a Publisher's custom-made license can be used according to the terms and conditions contained therein. See editor's website for further information and terms and conditions.

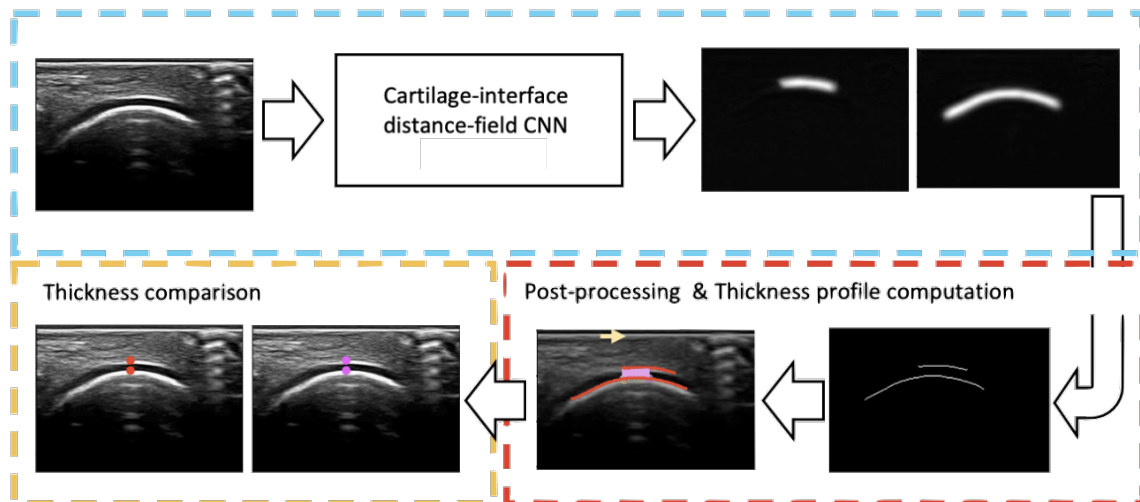
This item was downloaded from IRIS Università Politecnica delle Marche (<https://iris.univpm.it>). When citing, please refer to the published version.

(Article begins on next page)

Graphical Abstract

A deep-learning framework for metacarpal-head cartilage-thickness estimation in ultrasound rheumatological images

Maria Chiara Fiorentino, Edoardo Cipolletta, Emilio Filippucci, Walter Grassi, Emanuele Frontoni, Sara Moccia



Highlights

A deep-learning framework for metacarpal-head cartilage-thickness estimation in ultrasound rheumatological images

Maria Chiara Fiorentino, Edoardo Cipolletta, Emilio Filippucci, Walter Grassi, Emanuele Frontoni, Sara Moccia

- This paper addresses the problem of automatic cartilage-thickness estimation in US rheumatological images.
- The problem of cartilage-thickness estimation is tackled by regressing coordinates of cartilage interface by means of distance field.
- The dataset is acquired in the actual clinical practice, consisting of 38 subjects, with a mean of 24 (standard deviation ± 7.4) images per subject. Each image is annotated by an expert clinician with the value of cartilage thickness.
- The framework achieved a mean absolute difference (*ADF*) of 0.032 (± 0.026) mm that was comparable to the intra-observer variability, proving the potential of DL in the field.

A deep-learning framework for metacarpal-head cartilage-thickness estimation in ultrasound rheumatological images

Maria Chiara Fiorentino^{a,*}, Edoardo Cipolletta^b, Emilio Filippucci^b, Walter Grassi^b, Emanuele Frontoni^a and Sara Moccia^c

^aDepartment of Information Engineering, Università Politecnica delle Marche, Italy

^bDepartment of Rheumatology, Università Politecnica delle Marche, Italy

^cThe BioRobotics Institute and Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna

ARTICLE INFO

Keywords:

Ultrasound

Rheumatoid arthritis

Hyaline cartilage

Distance fields

Convolutional neural network

ABSTRACT

Objective: Rheumatoid arthritis (RA) is a chronic disease characterised by erosive symmetrical polyarthritis. Bone and cartilage are the main joint targets of this disease. Cartilage damage is one of the most relevant determinants of physical disability in RA patients. Cartilage damage is nowadays assessed by clinicians, which manually measure cartilage thickness in ultrasound (US) imaging. This poses issues relevant to intra-and inter-observer variability. Relying on the acquisition of metacarpal-head US images from 38 subjects, this work addresses the problem of automatic cartilage-thickness measurement by designing a new deep-learning (DL) framework. **Methods:** The framework consists of a Convolutional Neural Network (CNN), responsible for regressing cartilage-interface distance fields, followed by a post-processing step to delineate the two cartilage interfaces from the predicted distance fields and compute the cartilage thickness. **Results:** Our framework achieved encouraging results with a mean absolute difference (*ADF*) of $0.032 (\pm 0.026)$ mm against manual thickness annotation by an expert clinician. When evaluating the intra-observer variability, we obtained an *ADF* = $0.036 (\pm 0.028)$ mm. **Conclusion:** The proposed framework achieved an *ADF* against manual annotation that was comparable to the intra-observer variability, proving the potential of DL in the field. **Significance:** This work is the first to address the problem of automatic cartilage-thickness estimation in US rheumatological images with DL, paving the way for future research in the field. From a clinical perspective, the proposed framework proved to be a valuable tool to support the clinical routine increasing the reproducibility of cartilage thickness measurements.

1. Introduction

Rheumatoid arthritis (RA) is a common systemic inflammatory autoimmune disease characterized by erosive symmetrical polyarthritis, which may severely impair physical function and quality of life McWilliams and Walsh (2017). RA is frequently associated with joint destruction, involving both cartilage and bone damage, even though irreversible physical disability has been more clearly associated with cartilage damage than bone erosion Aletaha, Funovits and Smolen (2011); Torp-Pedersen, Bartels, Wilhjelm and Bliddal (2011); Smolen, van der Heijde, Keystone, van Vollenhoven, Goldring, Guérette, Cifaldi, Chen, Liu and Landewé (2013). Currently, conventional radiography is the most widely adopted imaging technique to assess joint damage in the clinical practice Colebatch, Edwards, Østergaard, van der Heijde, Balint, D'Agostino, Forslind, Grassi, Haavardsholm, Haugeberg et al. (2013). However, it has been shown that other techniques, such as ultrasound (US) and magnetic resonance imaging, are more accurate to detect joint damage in RA Døhn, Ejbjerg, Hasselquist, Narvestad, Szkudlarek, Møller, Thomsen, Østergaard et al. (2006). Being radiation free and cost-effective, US is in fact one of the most promising techniques for the early detection and monitoring of both bone erosion and cartilage damage in RA Mancarella, Addimanda, Cavallari and Meliconi (2017); Oo, Linklater, Daniel, Saarakkala, Samuels, Conaghan, Keen, Deveza and Hunter (2018); Möller, Bonel, Rotzetter, Villiger and Ziswiler (2009).

In US images, the normal appearance of hyaline cartilage is characterized by a homogenous hypoanechoic layer delimited by two regular, sharp, and bright margins (the outer or chondrosynovial one, and the inner or osteochondral one). To date, two main methods have been proposed for the assessment of cartilage damage by US: semi-quantitative

*Corresponding author: m.c.florentino@pm.univpm.it (M.C. Fiorentino)

ORCID(s): 0000-0002-7397-803X (M.C. Fiorentino); 0000-0002-6881-8197 (E. Cipolletta); 0000-0002-7251-7784 (E. Filippucci); 0000-0002-8893-9244 (E. Frontoni); 0000-0002-4494-8907 (S. Moccia)

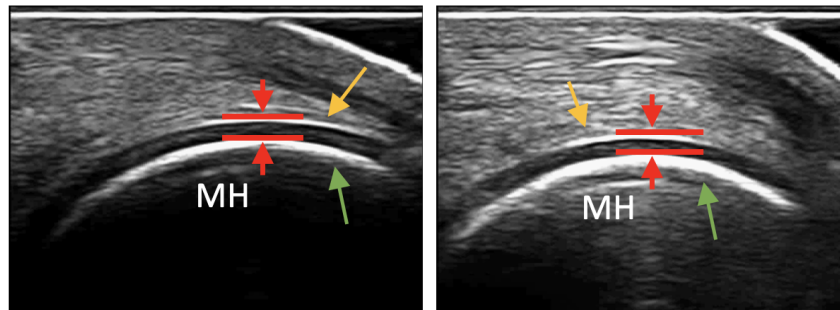


Figure 1: Two samples of thickness measurement of metacarpal head (MH) cartilage by US. Yellow arrow: chondrosynovial interface. Green arrow: osteochondral interface.

and quantitative methods. Semi-quantitative scoring systems are based on a qualitative and subjective evaluation of the hyaline cartilage Cipolletta, Filippucci, Di Matteo, Tesei, Cosatti, Di Carlo and Grassi (2020); Cipolletta, Hurnakova, Di Matteo, Di Carlo, Pavelka, Grassi and Filippucci (2021); Filippucci, da Luz, Di Geso, Salaffi, Tardella, Carotti, Natour and Grassi (2010). In 2019, the Outcome of Measure in Rheumatology Ultrasound Taskforce published a three-semiquantitative score: grade 0 = normal cartilage, grade 1 = minimal change such as focal thinning or incomplete loss of cartilage, grade 2 = severe change such as vast thinning or complete loss of cartilage Mandl, Studenic, Filippucci, Bachta, Backhaus, Bong, Bruyn, Collado, Damjanov, Dejaco et al. (2019). The major advantage of a semi-quantitative scoring system is its simplicity and speed. Quantitative-based scoring systems allow on the other hand to accurately measure the cartilage thickness and to monitor its changes Cipolletta et al. (2020, 2021); Möller et al. (2009). During a quantitative-based scoring evaluation, cartilage thickness is measured by including the chondrosynovial interface but not the osteochondral one while keeping the line of measurement perpendicular to the subchondral bone surface. This process is shown in Fig. 1 for two metacarpal-head images Mandl, Supp, Baksa, Radner, Studenic, Gyebrnar, Kurucz, Niedermayer, Aletaha, Balint et al. (2015). This is a time consuming procedure (e.g., a comprehensive US evaluation from the II and V fingers of the hands takes at least 6-7 minutes), and may be prone to inter- and intra-observer variability, especially for observers with limited experience Cipolletta et al. (2020). The availability of a software that allows automatic and real-time measurements during the US assessment could overcome this limitation.

In order to automatize the measurement process and improve inter- and intra-observer reliability, computer-assisted approaches have been developed in the years (Sec. 1.1). The most common challenges that have to be tackled in the observation process include presence of speckle noise, intensity in-homogeneity and high intra- and inter-subject variability due to individual variation of anatomical structures. These particular challenges have hampered the translation of the developed methodology in the actual clinical practice so far. During the last decade, artificial intelligence (AI) has become a tool in translational medicine, changing the way clinicians help patients Mediouni, R. Schlatterer, Madry, Cucchiari and Rai (2018). In fact, the objective evaluation provided by AI algorithms may derive useful insights in order to better understand the mechanisms of a disease, by identifying features or details that may seem insignificant to the human eye Toh, Dondelinger and Wang (2019). For the aforementioned reasons, deep learning (DL) algorithms have been recently gaining relevance in aiding in US image analysis Liu, Wang, Yang, Lei, Liu, Li, Ni and Wang (2019); Stoel (2020); van Sloun, Cohen and Eldar (2019) since they are able to directly extract complicated information from input images. Convolutional neural networks (CNNs) have already shown promising results for cartilage segmentation Kompella, Antico, Sasazawa, Jeevakala, Ram, Fontanarosa, Pandey and Sivaprakasam (2019); Antico, Sasazawa, Takeda, Jaiprakash, Wille, Pandey, Crawford, Carneiro and Fontanarosa (2020); Dunnhofer, Antico, Sasazawa, Takeda, Camps, Martinel, Micheloni, Carneiro and Fontanarosa (2020). This suggests that DL could potentially help tackling the challenges of automatic cartilage-thickness estimation. Nonetheless, to our knowledge, the potential of DL in the field hasn't been unlocked yet. This may be probably due to the lack of publicly-available annotated datasets in the field.

1.1. Related work

Considering cartilage thickness as an indicator for cartilage health, a number of approaches has been developed over the year with the intent of automatizing the thickness measurement and support clinicians in the clinical practice.

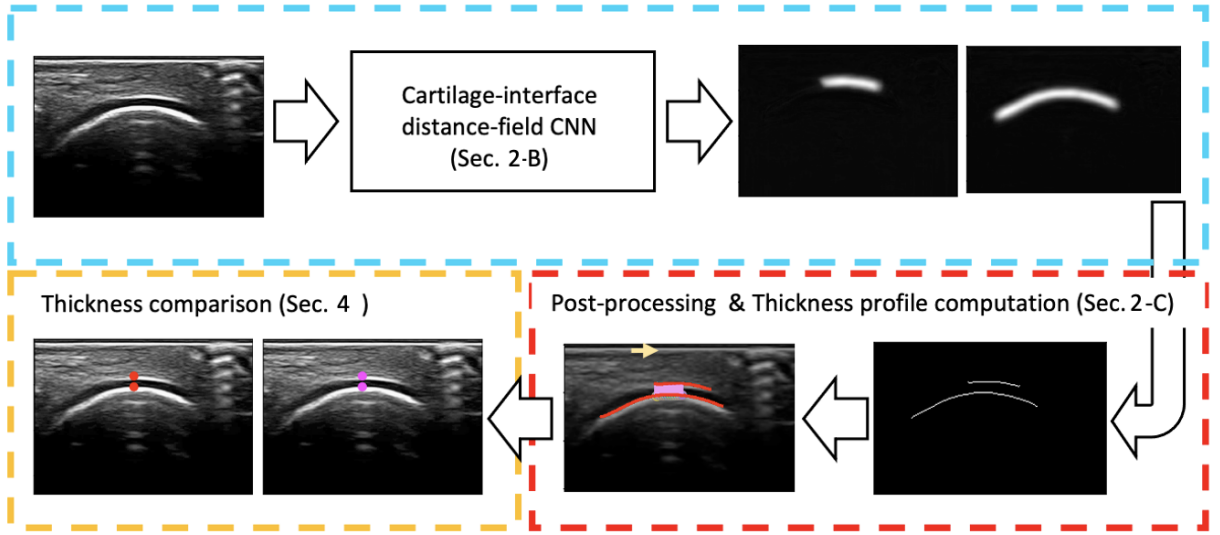


Figure 2: Proposed framework pipeline. A convolutional neural network is trained to predict the distance field from each of margin of the interfaces taken into account for thickness computation (lightblue box). A threshold operation is computed to extract the superficial margin of the chondrosynovial and the superficial margin of the osteochondral profiles from distance field. Cartilage thickness is then computed between corresponding points in the upper and lower interfaces, from the left-end (yellow arrow) to the right-end of the upper interface (red box). The obtained thickness measurement (purple points) is compared to that obtained by an expert clinician (red points).

In Hossain, Lai, Pingguan-Murphy, Hum, Salim and Liew (2014), knee joint cartilage thickness is evaluated by enhancing cartilage contours by means of a multipurpose beta optimized recursive bi-histogram equalization method. The work is intended to overcome the limitation of traditional adaptive histogram method by integrating information coming from brightness shift and contrast enhancement. However, the approach strongly relies on the chosen setting which, in turn, cause the enhancement of soft-tissue interfaces, which could affect cartilage segmentation and thickness measurement. As a prior for cartilage thickness estimation, the work in Faisal, Ng, Goh and Lai (2018) performs knee cartilage segmentation using the locally statistical level set method which is based on shape and appearance terms failing to robustly model all variation present in the visual object of interest Huang, Zhang and Li (2018). Also the framework proposed in Desai and Hacıhaliloglu (2018) relies on cartilage segmentation. The segmentation, performed using a random-walker algorithm, takes place after an enhancement which is done by means of local-phase image features and dynamic programming. Cartilage thickness is measured manually from the segmented region. The work in Desai and Hacıhaliloglu (2018) is expanded upon in Desai and Hacıhaliloglu (2019) in order to lower the need for manual intervention, by calculating the thickness as the Euclidean distance map from the segmented cartilage region. Despite the promising results, the segmentation of the cartilage strongly depends on hand-crafted features which vary as the anatomy of the subject varies, posing issues relevant to intra- and inter-subject variability.

DL may potentially offer a solution to tackle the challenges of cartilage-thickness estimation in US images since its algorithm training doesn't depend on domain expertise or hard core feature extraction, but the lack of annotated images has hampered the development of DL algorithms, so far. Following up on this consideration, in this work, we propose a framework based on CNNs for performing automatic cartilage-thickness measurement from US images acquired in the actual clinical practice.

1.2. Contribution

In this paper, we address the problem of automatic cartilage-thickness estimation in US rheumatological images, by designing and developing a new DL framework. The problem of cartilage-thickness estimation is tackled at first as a cartilage-interface delineation problem. Then, thickness is computed as the distance between points belonging to the superficial margin of chondrosynovial interface and the superficial margin of the osteochondral one.

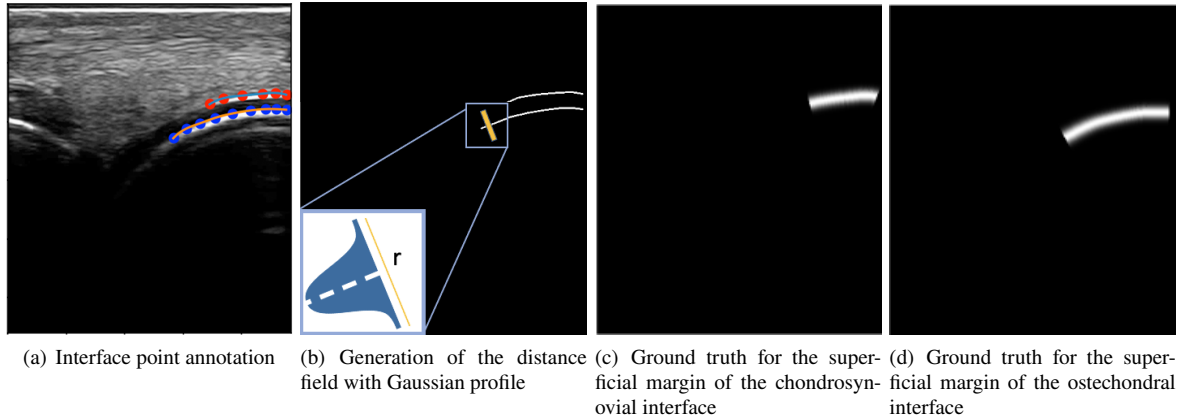


Figure 3: Ground-truth generation for distance fields. The raw image is annotated by drawing points along the interfaces of interest. (a) Cartilage interfaces delineation is obtained by cubic spline interpolation. The distance field is built by mounting a Gaussian distribution (length = r) in the line that is normal to the tangent line at each cartilage point (b). This is done for both the upper (c) and lower (d) interfaces, obtaining a double channel ground.

To delineate the cartilage interfaces, one possible strategy could be to classify each image pixel as background or interface. However, the area covered by the interfaces is small and narrow, and this would pose overfitting issues when training a classification network. Recently, researchers in other fields of medical-image analysis Liu et al. (2019); Fiorentino, Moccia, Capparuccini, Giamberini and Frontoni (2020); Hattab, Arnold, Strenger, Allan, Arsentjeva, Gold, Simpfendorfer, Maier-Hein and Speidel (2020) have shown how edge delineation problems may be modelled as heatmap-regression tasks. In this work we follow this paradigm and regress the cartilage-interface distance fields via CNNs.

The contribution of this paper could be summarized as follow:

- Development of a new framework that is intended to estimate cartilage thickness from US images automatically. The framework relies on a CNN that regresses the superficial margin of the chondrosynovial interface and the superficial margin of the osteochondral interface via distance fields.
- Collection of the first dataset for DL applications in the field of metacarpophalangeal US image analysis for RA diagnosis. This dataset, acquired in the actual clinical practice, is collected from 38 subjects, with a mean of 24 (standard deviation ± 7.4) images per subject. The dataset is annotated by an expert clinician with the value of cartilage thickness.
- Comprehensive experimental validation, including the estimation of intra-observer variability.

Compared to what is currently done in the clinical practice, where the cartilage thickness is measured via only two points, the proposed framework produces a cartilage-thickness profile, i.e. the thickness is computed for each point-pair in the two interfaces. This provides the clinicians with additional quantitative data that can be used to perform early-stage detection of cartilage damage.

In this work, we focus our experiments on US metacarpal-head images, since metacarpal head is one of the most involved areas in RA patients Aletaha, Neogi, Silman, Funovits, Felson, Bingham III, Birnbaum, Burmester, Bykerk, Cohen et al. (2010). Nonetheless, we believe that the proposed framework could be equally applied to other anatomical locations, such as femur cartilage.

2. Methods

The proposed framework for cartilage-thickness estimation from US images is shown in Fig. 2. A CNN is responsible for the regression of the cartilage interfaces via distance fields (Sec. 2.2). The CNN is followed by a post processing step to highlight the cartilage surfaces from the predicted distance fields. Cartilage thickness is the computed by measuring the distance between each point in the upper interface and its corresponding point in the

Table 1

Cartilages' regression network architecture with Conv2D = convolution, MaxPooling = max pooling, Upconv = up convolution, Concat = concatenation

	Operator	Kernel dimension	Dilation rate	Filters
Encoding path	Conv2D	3x3	1	64
	Conv2D	3x3	2	64
	MaxPooling	2x2	-	-
	Conv2D	3x3	1	128
	Conv2D	3x3	2	128
	MaxPooling	2x2	-	-
	Conv2D	3x3	1	256
	Conv2D	3x3	2	256
	MaxPooling	2x2	-	-
	Conv2D	3x3	1	512
	Conv2D	3x3	2	512
	MaxPooling	2x2	-	-
Bottleneck 2x	Conv2D	3x3	1	1024
Decoding path	Upconv	2x2	-	512
	Conv2D	2x2	1	512
	Concat	-	-	-
	Conv2D	3x3	1	512
	Conv2D	3x3	1	512
	Upconv	2x2	-	256
	Conv2D	2x2	1	256
	Concat	-	-	-
	Conv2D	3x3	1	256
	Conv2D	3x3	1	256
	Upconv	2x2	-	128
	Conv2D	2x2	1	128
	Concat	-	-	-
	Conv2D	3x3	1	128
	Conv2D	3x3	1	128
	Upconv	2x2	-	64
	Conv2D	2x2	1	64
	Concat	-	-	-
	Conv2D	3x3	1	64
	Conv2D	3x3	1	64
	Conv2D	3x3	1	2

lower one (Sec. 2.2.1). This approach follows what is currently being done in clinical practice, where, however, only two points, one for the superficial margin of the chondrosynovial interface and one for the superficial margin of the osteochondral one, are taken into account by the clinicians.

2.1. Data preparation

To develop our framework for cartilage thickness estimation, we rely on two ground-truth annotations:

1. Distance-field ground-truth images, which are used to train our regression CNN
2. Cartilage thickness annotations, which are made just like in the actual clinical practice (see Fig 2, Thickness comparison). Specifically, an expert clinician selects a point in the superficial margin of the chondrosynovial margin and draws the line perpendicular to the superficial margin of the osteochondral interface from that point.

To build the ground-truth images, we start from the manual delineation of the cartilage interfaces. The interfaces are annotated at first by manually selecting points along the boundaries of each interface (Fig. 3a). The annotation is performed under the clinician supervision using LabelMe Wada (2016). Cubic spline interpolation is used to obtain the cartilage manual delineation from the annotated points. In order to build the distance-field ground truths from the manual delineation, we were inspired by Fiorentino et al. (2020). For each point in each of the two interfaces, the normal line to the interface is found (Fig. 3b). A Gaussian profile (with length r and standard deviation $r/2$) is drawn along the normal line. A two-channel (one for the superficial margin of the chondrosynovial interface and one for the superficial margin of the osteochondral one) ground truth is built to avoid overlap among the interface distance fields and to separate the two surfaces without the need of any post-processing (Fig. 3c-d).

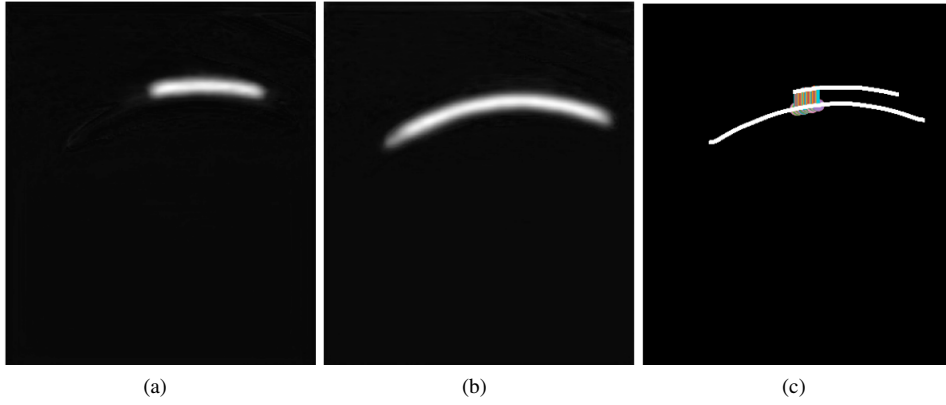


Figure 4: Post-processing procedure. The predicted distance fields for the upper (a) and lower (b) interfaces are thresholded and skeletonized (c). Cartilage thickness is calculated as the Euclidean distance between the corresponding points of the skeletonized interfaces (c).

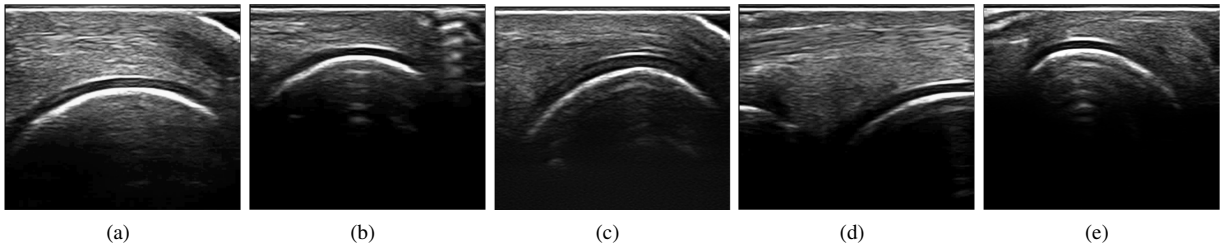


Figure 5: Challenges in the dataset include US artifacts (b-e), and anatomical (a-c) and cartilage position (d) variability.

2.2. Cartilage-surface distance-field CNN

To regress the cartilage-interface distance fields, we rely on a dilated-UNet network, as atrous convolutions have been demonstrated to be effective in boundary distance regression to segment kidney boundaries in US image Yin, Peng, Li, Zhang, You, Fischer, Furth, Tasian and Fan (2020). Moreover, Unet is proven to be particularly effective with a limited dataset Ronneberger, Fischer and Brox (2015).

The CNN architecture is detailed in Table 1. It follows the Unet-structure of Ronneberger et al. (2015), which consists of a encoding, bottleneck and decoding paths. The encoding path is made up of a sequence of one convolution (kernel size = 3x3 with dilation rate = 1, padding = 1), one convolution (kernel size = 3x3 with dilation rate = 2, padding = 1) and max pooling (stride = 2). This sequence is repeated 4 times. The number of filter is doubled at each iteration so as to learn more and more complex features. The bottleneck, which has the role to force the model to focus only on relevant information, connects the encoding path with the decoding path by means of two convolutional layers (kernel size = 3x3, dilation rate = 1). The number of filters doubles from 512 to 1024. The decoding path consists of a sequence of one up convolution (kernel size = (2x2)) that halves the numbers of feature channels, followed by a convolution layer (filter = 2x2), which is linked via skip connections to the downsampling feature maps. Two convolutions in a row (kernel size = (3x3), dilation rate = 1, padding = 1) are successively used. This sequence is repeated 4 times. Every convolutional layer is activated by a rectified linear unit function and its output is batch normalized in order to stabilize the learning process. Finally, a convolution layer is used to reduce the number of filters to 2 in order to obtain the distance-field predictions.

Our CNN is trained using the stochastic gradient descent with momentum as optimizer, minimizing the mean squared error (MSE) loss function:

Table 2

Demographic characteristics of the subjects included in the study.

age (years, mean $\pm$ standard deviation)	54.6 $\pm$ 5.6
height (cm, mean $\pm$ standard deviation)	170.7 $\pm$ 9.5
handedness (right/left)	40/8
weight (kg, mean $\pm$ standard deviation)	72.4 $\pm$ 6.4
sex (female/male)	33/15
body mass index (kg/m ² , mean $\pm$ standard deviation)	24.8 $\pm$ 1.7

$$MSE(S, \hat{S}) = \frac{1}{N} \sum_{i=1}^N (S_i - \hat{S}_i)^2 \quad (1)$$

where S_i and $\hat{S}_i$ are the ground-truth and predicted distance field for the i -th training image, respectively, and N is the number of training samples.

2.2.1. Thickness profile computation

After obtaining the 2-channel distance-field prediction, Otsu's thresholding is used. The binarized images are skeletonized by means of Zhang's method Zhang and Suen (1984) and re-sampled by means of a cubic spline interpolation to better smooth the interfaces. Then, a Gaussian filter is used to remove noise. To compute the cartilage thickness, for each point in the upper interface, the perpendicular to the surface is drawn until it intersects with a point in the lower interface. The cartilage thickness is then computed as the Euclidean distance between each point pairs (Fig. 4(c)).

3. Experimental setup

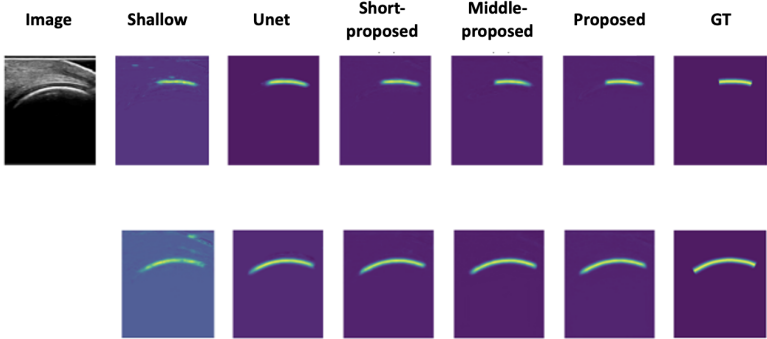
3.1. Dataset

To develop and evaluate the proposed framework, we collected a dataset from 38 subjects, approved by Comitato Etico Regione Marche id CERM 383/2019. In Table 2, information about subjects' age, height, handedness, weight, sex and body mass index is reported. Inclusion criteria were:

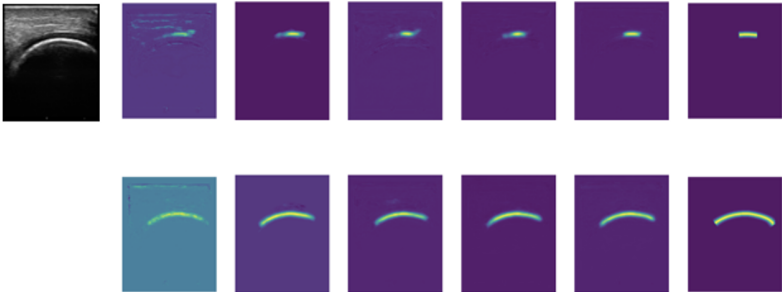
- No diagnosis of inflammatory/degenerative arthropathies
- No joint pain [visual analogue scale (VAS) $\leq 10/100$] and/or analgesic or non-steroidal anti-inflammatory drugs' intake in the four weeks prior to the visit
- Age > 18-year-old
- No signs of osteoarthritis of metacarpophalangeal, proximal or distal interphalangeal joints at physical examination

Being the first attempt at exploring this kind of approach for cartilage thickness estimation, only healthy people were considered for this experiment, with the full intent to later extend this new suggested approach to subjects with RA.

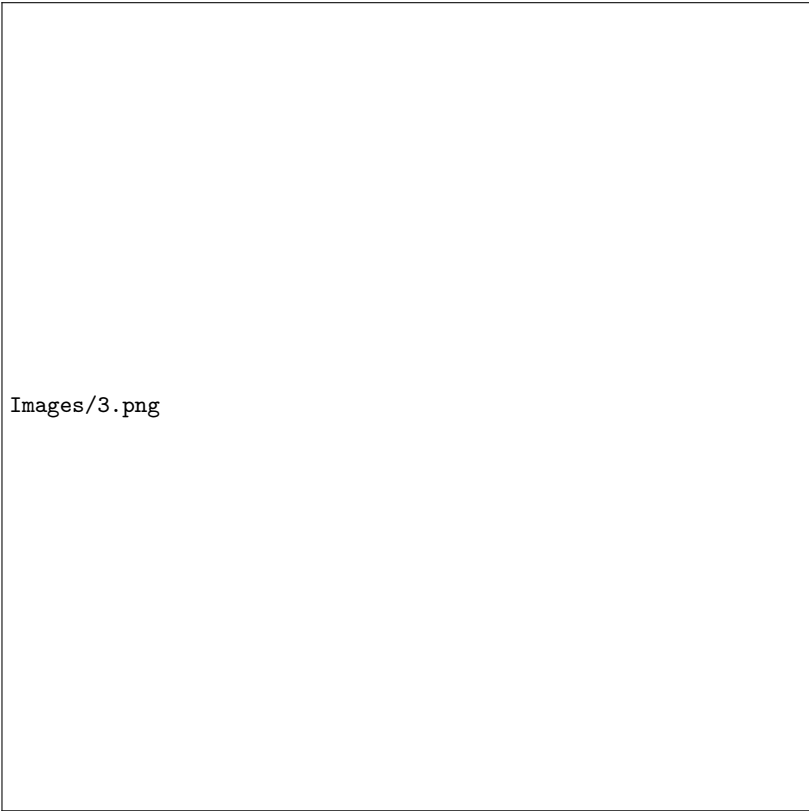
US examinations were performed with a MyLab Class C (Esaote, Genoa, Italy), equipped with a very high frequency broadband linear trasducer (10-22 MHz) by a sonographer with five years of experience. To increase the reproducibility, grey scale setting parameters were standardised (B-mode frequency: 22 MHz, master gain: 70%, mechanical index: 0.3, dynamic range: 12, depth 15 mm, focus position at the area of interest) and maintained unchanged during all the US examinations. Cipolletta et al. (2020). The hands were placed on a table, with the metacarpophalangeal joints in maximal flexion ($>60^\circ$) Cipolletta et al. (2020); Mandl et al. (2019); Torp-Pedersen et al. (2011). The metacarpal heads from the 2nd to the 5th finger of both hands were 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 Möller, Janta, Backhaus, Ohrndorf, Bong, Martinoli, Filippucci, Sconfienza, Terslev, Damjanov et al. (2017). Particular attention was paid to ensure a perpendicular insonation of the cartilage surface Cipolletta et al. (2020); Mandl et al. (2019); Torp-Pedersen et al. (2011).



(a) Sample 1

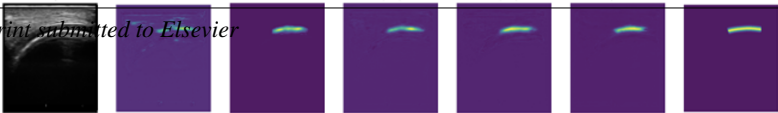


(b) Sample 2



Images/3.png

(c) Sample 3



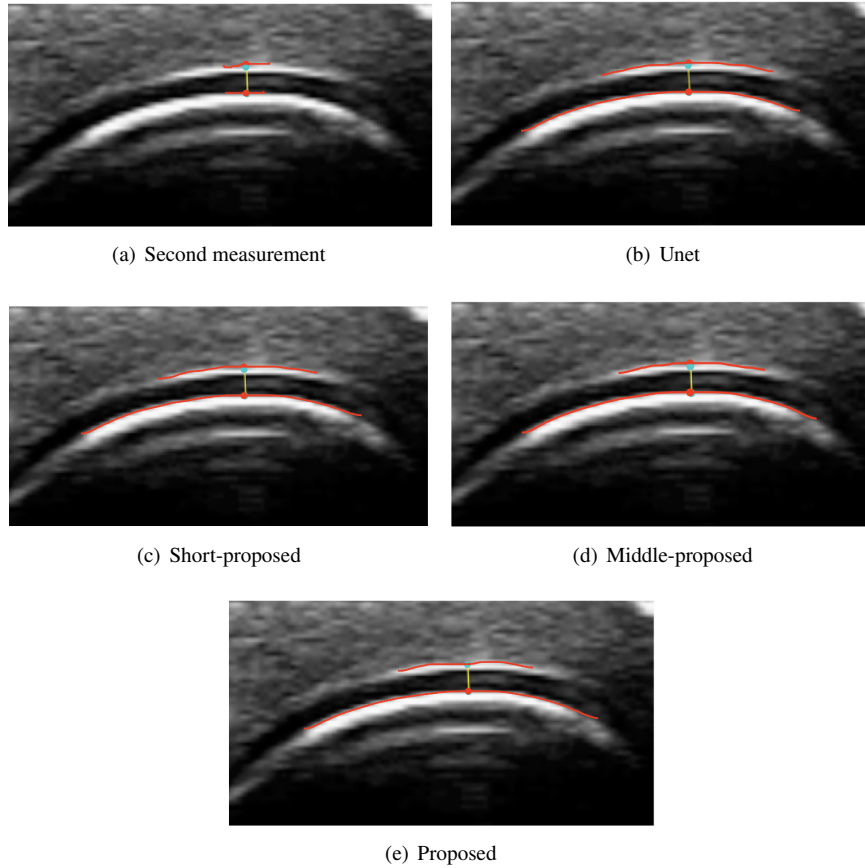


Figure 7: Visual examples of thickness measurement obtained by the architectures of our ablation study and by an expert clinician performing the measurement for the second time. The sea water points are the ones annotated by the clinician and used as ground true. Red points are the closer points to the sea water point, as estimated by each architecture of the ablation study and second clinician annotation. Images are zoom in to better appreciate results.

A mean of 24 (standard deviation = ± 7.4) images for subject was acquired, obtaining 904 images in total. Image size was 480x560 pixels, with a pixel size of 0.0310 [mm]. Challenges in the dataset (Fig. 5) included both typical challenges of US image analysis (reverberations, speckles, noise, unclear contours) and peculiar challenges of the specific anatomical district (i.e., different areas of metacarpal head cartilage, shape and appearance).

Training, validation and test sets consisted of 25 (number of images = 615), 5 (118) and 8 (173) patients, respectively. Dataset split for training and testing was performed at subject's level.

3.2. Parameter setting

Prior to feeding the CNN to obtain the distance-field regression, all training and validation images were resized to 240x280 pixels in order to limit both training time and memory requirements. Images were zero-padded from 240x280 to 240x288, to be compatible with the operations performed by the CNN. Mean intensity was removed from each image. The ground-truth images used for CNN's training were obtained with an r empirically set to 100 pixels and a Gaussian standard deviation of $r/2$. This allowed to fully cover the interfaces of interest. Once obtained the testing image predictions, the latter images were scaled back to 480x560 pixels to evaluate the proposed approach in the real image dimension.

The CNN was trained minimizing the mini-batch stochastic gradient descent with momentum = 0.98. A learning rate = 0.001 for 1500 epochs was used. Batch size was set to 8, as a trade off between memory requirements and training convergence. On-the-fly data augmentation was performed during training by horizontal flipping and rotating (range $\pm 45^\circ$) images. The transformations were randomly applied at each training iteration. The best model was selected

Table 3

Ablation study summary. Encoder-decoder structure, atrous convolutions and depth (number of layers) are the model characteristics taken into account.

	Encoder-decoder	Atrous convolutions	Depth
Shallow		X	1
Unet	X		4
Short-proposed	X	X	2
Middle-proposed	X	X	3
Proposed	X	X	4

according to the lowest mean absolute error (MAE) obtained in the validation set:

$$MAE(M, \hat{M}) = \frac{1}{N} \sum_{i=1}^N |M_i - \hat{M}_i| \quad (2)$$

where M_i and $\hat{M}_i$ are the regression mask and predicted outcome of i -th training image, respectively, and N is the number of training samples. All the analyses were performed using Keras¹ on a NVIDIA RTX 2080TI, with a Xeon e5 CPU and 128 GB RAM.

3.3. Performance metrics

To assess the performance of our CNN network, we computed the MAE between the predicted distance-field and the distance-field ground truth in the test images.

In order to mimic what is currently done in clinical practice, obtaining a more comparable understanding of the potential of our findings, an expert clinician measured cartilage thickness in each of the testing image, as described in Sec 2.1. The following metrics were computed:

- Difference (DF) [mm]:

$$DF = CT_A - CT_B \quad (3)$$

- Absolute difference (ADF) [mm]:

$$ADF = |CT_A - CT_B| \quad (4)$$

with CT_A and CT_B are the cartilage thickness obtained by the proposed framework and computed by the clinician, respectively. ADF is used to determine accuracy of predictions, on the other hand DF is used to assess the over/underestimation of the thickness.

3.4. Ablation study & intra-clinician variability assessment

We carried out an ablation study to identify the best configuration of our framework. The effectiveness of the encoder-decoder structure of our CNN was investigated by training a shallow CNN (*Shallow*). The shallow CNN consisted of a sequence of one convolution (kernel size = 3x3 with dilation rate = 1, padding = 1) and one convolution (kernel size = 3x3 with dilation rate = 2, padding = 1), each followed by a batch normalization layer. The sequence was repeated 4 times (with 64, 128, 128 and 64 filters for the first, second, third and forth repetition, respectively).

The effect of the CNN depth on the distance-field regression performance was investigated, too. We trained two additional architectures, which were obtained by removing the deeper layers of the proposed architectures, stopping at 512 filters for *Middle – proposed* architecture and 256 filters for *Short – proposed* architecture. Dilation rate was also evaluated by training a standard *Unet* architecture without atrous convolutions. Table 3 shows a summary of our ablation study. For fair comparison, we always used the same dataset split and training settings, as well as the same computational hardware.

To assess the intra-observer variability, the expert sonographer who provided the ground truth thickness measurement repeated the measurement a second time.

¹<https://keras.io>

Table 4

Ablation study comparison in terms of mean absolute error between distance field architectures predictions and distance field ground truth images.

	Mean Absolute error
Shallow	0.014 ($\pm$ 0.003)
Unet	0.0041 ($\pm$ 0.001)
Short-proposed	0.0054 ($\pm$ 0.001)
Middle-proposed	0.0051 ($\pm$ 0.001)
Proposed	0.0050 ($\pm$ 0.001)

Table 5

Ablation study & intra-clinician variability assessment in terms of absolute difference and difference. We report the mean for each metric, with standard deviation in brackets. Metrics are calculated between the cartilage thickness obtained by the architecture (or for the second time by the same clinician) and clinician manual annotation.

	Difference [mm]	Absolute Difference [mm]
Clinician (2nd time)	-0.009 (0.045)	0.036 (0.028)
Unet	0.010 (0.048)	0.039 (0.029)
Short-proposed	-0.002 (0.045)	0.035 (0.028)
Middle-proposed	0.003 (0.041)	0.035 (0.026)
Proposed	-0.004 (0.041)	0.032 (0.026)

4. Results

The performance comparison in terms of MAE is reported in Table 4. The highest $MAE = 0.014 (\pm 0.003)$ is obtained by Shallow architecture, followed by the Short-proposed one. Middle-proposed and proposed architectures performed comparably well with a $MAE = 0.0051 (\pm 0.001)$ and $0.0050 (\pm 0.003)$, respectively. The best $MAE = 0.0041 (\pm 0.003)$ is instead reached by Unet. In Fig. 6, samples of cartilage regression distance field predictions are shown. Results are obtained by shallow (first column), Unet (second column), short-proposed (third column), middle-proposed (fifth column) and proposed (fifth column) architectures, respectively. The distance field ground truth is shown too. The first and second row correspond to upper and lower cartilage surface, respectively.

The performance comparison in terms of DF and ADF is reported in Table 5. We decide to not show Shallow architecture performances because of the poor results obtained in MAE calculation. The best performance in term of ADF is obtained by the proposed architecture, followed by its shorter versions (short-proposed and middle-proposed). An $ADF=0.039 (\pm 0.029)$ mm is instead obtained by Unet, which was the best in MAE calculation. As regard DF , the proposed architectures obtained the lowest values as opposed to unet and clinician. Figure 7 shows visual examples of thickness measurement performed by ablation study architectures, which are compared to the one performed by the sonographer. Intra-variability of the measurement is also shown. The sea water color points are the ones annotated by the clinician used as ground true optimal location for thickness calculation. The red points are the closest to the sea-water colored ones which were first annotated by the clinician. The same red points were estimated by ablation study architectures and second clinician annotation.

5. Discussions

Preventing or halting cartilage damage is of paramount importance in RA. Cartilage thickness measurement using US is a promising method to early detect and monitor cartilage damage. In daily clinical practice, such measurement is performed manually by sonographer, posing issues related to intra- and inter-observer variability. Different computer-assisted approaches have been developed over time with the aim of attenuating such issues, but none of them has fully exploited the potential of DL, yet.

To unlock the potential of DL in the field, we collected a dataset of 38 subjects (with a mean of 24 images each) that underwent a US assessment of the metacarpal head cartilage, facing challenges related to US artifacts (such as reverberations and speckles) and anatomical variability of the investigated anatomical structures. In light of the considerable effort from the clinician side in providing, not just images acquisition, but more importantly the corresponding image-level annotations, we had to find a trade off between such effort and dataset size. With this

considerations in mind, data were acquired taking into account as much variability as possible (Fig. 5). Relying on this dataset, we developed a DL framework for automatic cartilage thickness estimation. The framework consisted of a CNN that regresses distance field from the cartilage interfaces. The cartilage thickness was computed operating a post-processing on the predicted distance fields. As a result, a cartilage thickness profile is provided to the clinician. The proposed framework achieved encouraging results with a mean $ADF = 0.032$ (standard deviation = ± 0.026) mm, $DF = -0.004 (\pm 0.041)$ mm, proving to be an accurate method for cartilage thickness measurement.

From our ablation study, Unet, short-proposed, middle-proposed and the proposed architecture performed quite well in estimating the distance field of the upper cartilage surface in US images with contrasted cartilage surfaces (Fig 6, upper-left image), with a long, wide and curve upper interface cartilage. The shallow architecture performed slightly less precisely, although the error in the distance-field estimation did not affect the cartilage-thickness measurement.

As the upper left interface was less visible (Fig. 6a), shallow architecture predictions deteriorated. Short-proposed architecture predictions deteriorate too, but not drastically, as opposed to Unet, middle-proposed and proposed that still performed comparably well. This may be attributed to the power of encoder decoder structure which has the ability to extract use-full features, whose expressibility increases exponentially with the depth Ye and Sung (2019). For this reason, the shallow architecture, which consists of a sequence of convolution, fails to identify surface boundaries with precision, and the depth of the short-proposed is not as such to precisely locate distance field pixels on image.

Images in which the upper cartilage surface appeared flat, thus not characterized by the typical curvature, were particularly difficult for any architecture (Fig. 6c). This is due to the fact that this kind of images are few compared to the whole dataset, typically characterized by curved upper surfaces.

Regarding lower interface predictions, Unet, short-proposed, middle-proposed and proposed architectures performed quite well with some exception by short-proposed (Fig. 6c-2d). Shallow architecture had difficulties even in the lower interface regression.

The best MAE was reached by the standard Unet (Table 4). However, even though it is able to regress the distance field in the side portions of the cartilage interfaces as opposed to the proposed architecture, upper surface predictions appear less precise (see Figure 7b-c). Atrous convolutions offers, in fact, a way to control the field-of-view, finding the best trade-off between accurate localization (small field-of-view) and context assimilation (large field-of-view), integrating therefore local and general information together. This permits the architecture to better focus on edges. As a matter of fact, Unet architecture covers a larger portion of the surfaces but seems to be less precise in delineating it, obtaining an $ADF = 0.039 (\pm 0.028)$ mm (Tab. 5). Short-proposed and Middle-proposed perform comparably well with a $ADF = 0.035 (\pm 0.028)$ mm and $0.035 (\pm 0.026)$ mm, respectively. As the depth increases, the proposed network becomes more and more able to discriminate cartilage boundaries. This is particularly visible by looking at Fig 7: points annotated by clinician and points estimated by proposed architecture overlap completely, proving to be a valuable tool to cartilage thickness calculation. The lowest DFs are obtained by proposed architectures proving to be less prone to over/underestimate the measurement.

In fact, whatever the choice of the network is, the pipeline of the entire work seems to be particularly promising, except for those architectures that do not have a encoder-decoder structure. Evaluating other recent architectures (attentions, multi-scale feature concatenation, siamese learning) is out of the scope of this work, which is indeed focused on presenting the entire pipeline for estimating cartilage thickness in US images.

The ADF of $0.036 (\pm 0.028)$ mm obtained during the intra-clinician variability analysis was higher than the ADF obtained with the proposed framework ($0.032 (\pm 0.026)$ mm). This suggests that our framework may be a true support in the clinical routine with a view to make RA assessment more reproducible. As opposed to state-of-the-art approaches, our framework doesn't rely on domain expertise or hard core feature extraction, making our framework very flexible to be adapted to new data.

Limitations in our experimental protocol include 1) poor dataset size 2) acquisition from the same device and annotation performed by the same clinician 3) inclusion of healthy subjects. With a view to translate the proposed pipeline in the actual clinical practice, these limitations must be overcome. However, we believe that this dataset, consisting of 38 subjects, whose size is similar to other datasets in literature ², is a valuable initial step to stimulate further research in the field. In fact, US image challenges are getting of increasing interest among researches, enough to be considered almost every year in the Gran Challenge of MICCAI (e.g. A-AFMA ³ and HC18 challenge van den Heuvel, de Bruijn, de Korte and Ginneken (2018)).

²<https://fetreg2021.grand-challenge.org/>

³<https://a-afma.grand-challenge.org/>

A way to directly delineate the upper and lower surfaces could also be exploited in order to not rely on thresholding methods. This could be done by adding a CNN to train cartilage surfaces end-to-end with distance-field CNN Yin et al. (2020). In addition, the proposed work contributes to increase the research in the field of cartilage damage and places itself among the latest approaches to the matter Miramini, Smith, Zhang and Gardiner (2017); Smith, Gardiner, Zhang and Grodzinsky (2019); Tanska, Julkunen and Korhonen (2018).

6. Conclusions

In this work, we suggest an innovative framework for cartilage-thickness measurement in US metacarpophalangeal images by acquiring the first metacarpophalangeal dataset suited for DL analysis. Our framework relies on distance fields to accurately delineate cartilage interfaces. With our framework, we achieved an $ADF = 0.032 (\pm 0.026)$ mm from the annotation performed by an expert clinician, while our experiment on intra clinician variability led to an $ADF = 0.036 (\pm 0.028)$ mm. This proved our framework to be a valuable tool to support clinicians. This work is among the first attempts at exploiting DL in cartilage thickness measurement and has great potential to be translated in the clinical practice.

Compliance With Ethical Standards Disclosures

The authors have no conflict of interest to disclose.

Ethical Standards

The procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000.

References

- Aletaha, D., Funovits, J., Smolen, J.S., 2011. Physical disability in rheumatoid arthritis is associated with cartilage damage rather than bone destruction. *Annals of the rheumatic diseases* 70, 733–739.
- Aletaha, D., Neogi, T., Silman, A.J., Funovits, J., Felson, D.T., Bingham III, C.O., Birnbaum, N.S., Burmester, G.R., Bykerk, V.P., Cohen, M.D., et al., 2010. 2010 rheumatoid arthritis classification criteria: an american college of rheumatology/european league against rheumatism collaborative initiative. *Arthritis & rheumatism* 62, 2569–2581.
- Antico, M., Sasazawa, F., Takeda, Y., Jaiprakash, A.T., Wille, M.L., Pandey, A., Crawford, R., Carneiro, G., Fontanarosa, D., 2020. Bayesian CNN for segmentation uncertainty inference on 4d ultrasound images of the femoral cartilage for guidance in robotic knee arthroscopy. *IEEE Access*.
- Cipolletta, E., Filippucci, E., Di Matteo, A., Tesi, G., Cosatti, M.A., Di Carlo, M., Grassi, W., 2020. The reliability of ultrasound in the assessment of hyaline cartilage in rheumatoid arthritis and healthy metacarpal heads. *European Journal of Ultrasound*.
- Cipolletta, E., Hurnakova, J., Di Matteo, A., Di Carlo, M., Pavelka, K., Grassi, W., Filippucci, E., 2021. Prevalence and distribution of cartilage and bone damage at metacarpal head in healthy subjects. *Clin Exp Rheumatol*.
- Colebatch, A.N., Edwards, C.J., Østergaard, M., van der Heijde, D., Balint, P.V., D'Agostino, M.A., Forslund, K., Grassi, W., Haavardsholm, E.A., Haugeberg, G., et al., 2013. Eular recommendations for the use of imaging of the joints in the clinical management of rheumatoid arthritis. *Annals of the rheumatic diseases* 72, 804–814.
- Desai, P., Hacıhaliloglu, I., 2019. Knee-cartilage segmentation and thickness measurement from 2D ultrasound. *Journal of Imaging* 5, 43.
- Desai, P.R., Hacıhaliloglu, I., 2018. Enhancement and automated segmentation of ultrasound knee cartilage for early diagnosis of knee osteoarthritis, in: *IEEE 15th International Symposium on Biomedical Imaging*, IEEE. pp. 1471–1474.
- Døhn, U.M., Ejbjerg, B.J., Hasselquist, M., Narvestad, E., Szkudlarek, M., Møller, J.M., Thomsen, H.S., Østergaard, M., et al., 2006. Are bone erosions detected by magnetic resonance imaging and ultrasonography true erosions? a comparison with computed tomography in rheumatoid arthritis metacarpophalangeal joints. *Arthritis research & therapy* 8, 1–9.
- Dunnhofer, M., Antico, M., Sasazawa, F., Takeda, Y., Camps, S., Martinel, N., Micheloni, C., Carneiro, G., Fontanarosa, D., 2020. Siam-U-Net: encoder-decoder siamese network for knee cartilage tracking in ultrasound images. *Medical Image Analysis* 60, 101631.
- Faisal, A., Ng, S.C., Goh, S.L., Lai, K.W., 2018. Knee cartilage segmentation and thickness computation from ultrasound images. *Medical & Biological Engineering & Computing* 56, 657–669.
- Filippucci, E., da Luz, K.R., Di Geso, L., Salaffi, F., Tardella, M., Carotti, M., Natour, J., Grassi, W., 2010. Interobserver reliability of ultrasonography in the assessment of cartilage damage in rheumatoid arthritis. *Annals of the rheumatic diseases* 69, 1845–1848.
- Fiorentino, M.C., Moccia, S., Capparuccini, M., Giamberini, S., Frontoni, E., 2020. A regression framework to head-circumference delineation from US fetal images. *Computer Methods and Programs in Biomedicine* 198, 105771.
- Hattab, G., Arnold, M., Strenger, L., Allan, M., Arsentjeva, D., Gold, O., Simpfendorfer, T., Maier-Hein, L., Speidel, S., 2020. Kidney edge detection in laparoscopic image data for computer-assisted surgery. *International Journal of Computer Assisted Radiology and Surgery* 15, 379–387.
- van den Heuvel, T.L., de Bruijn, D., de Korte, C.L., Ginneken, B.v., 2018. Automated measurement of fetal head circumference using 2d ultrasound images. *PloS one* 13, e0200412.

- Hossain, M.B., Lai, K.W., Pingguan-Murphy, B., Hum, Y.C., Salim, M.I.M., Liew, Y.M., 2014. Contrast enhancement of ultrasound imaging of the knee joint cartilage for early detection of knee osteoarthritis. *Biomedical Signal Processing and Control* 13, 157–167.
- Huang, Q., Zhang, F., Li, X., 2018. Machine learning in ultrasound computer-aided diagnostic systems: a survey. *BioMed Research International* 2018.
- Kompella, G., Antico, M., Sasazawa, F., Jeevakala, S., Ram, K., Fontanarosa, D., Pandey, A.K., Sivaprakasam, M., 2019. Segmentation of femoral cartilage from knee ultrasound images using mask R-CNN, in: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE. pp. 966–969.
- Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S.X., Ni, D., Wang, T., 2019. Deep learning in medical ultrasound analysis: a review. *Engineering* 5, 261–275.
- Mancarella, L., Addimanda, O., Cavallari, C., Meliconi, R., 2017. Synovial inflammation drives structural damage in hand osteoarthritis: a narrative literature review. *Current Rheumatology Reviews* 13, 43–50.
- Mandl, P., Studenic, P., Filippucci, E., Bacht, A., Backhaus, M., Bong, D., Bruyn, G.A., Collado, P., Damjanov, N., Dejacó, C., et al., 2019. Development of semiquantitative ultrasound scoring system to assess cartilage in rheumatoid arthritis. *Rheumatology* 58, 1802–1811.
- Mandl, P., Supp, G., Baksa, G., Radner, H., Studenic, P., Gyebrnar, J., Kurucz, R., Niedermayer, D., Aletaha, D., Balint, P., et al., 2015. Relationship between radiographic joint space narrowing, sonographic cartilage thickness and anatomy in rheumatoid arthritis and control joints. *Annals of the Rheumatic Diseases* 74, 2022–2027.
- McWilliams, D.F., Walsh, D.A., 2017. Pain mechanisms in rheumatoid arthritis. *Clinical and Experimental Rheumatology* 35, 94–101.
- Mediouni, M., R. Schlatterer, D., Madry, H., Cucchiari, M., Rai, B., 2018. A review of translational medicine. the future paradigm: how can we connect the orthopedic dots better? *Current medical research and opinion* 34, 1217–1229.
- Miramini, S., Smith, D.W., Zhang, L., Gardiner, B.S., 2017. The spatio-temporal mechanical environment of healthy and injured human cartilage during sustained activity and its role in cartilage damage. *Journal of the mechanical behavior of biomedical materials* 74, 1–10.
- Möller, B., Bonel, H., Rotzetter, M., Villiger, P.M., Ziswiler, H., 2009. Measuring finger joint cartilage by ultrasound as a promising alternative to conventional radiograph imaging. *Arthritis Care & Research: Official Journal of the American College of Rheumatology* 61, 435–441.
- Möller, I., Janta, I., Backhaus, M., Ohrndorf, S., Bong, D.A., Martinoli, C., Filippucci, E., Sconfienza, L.M., Terslev, L., Damjanov, N., et al., 2017. The 2017 EULAR standardised procedures for ultrasound imaging in rheumatology. *Annals of the Rheumatic Diseases* 76, 1974–1979.
- Oo, W.M., Linklater, J.M., Daniel, M., Saarakkala, S., Samuels, J., Conaghan, P.G., Keen, H.I., Devez, L.A., Hunter, D.J., 2018. Clinimetrics of ultrasound pathologies in osteoarthritis: systematic literature review and meta-analysis. *Osteoarthritis and Cartilage* 26, 601–611.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.
- van Sloun, R.J., Cohen, R., Eldar, Y.C., 2019. Deep learning in ultrasound imaging. *Proceedings of the IEEE* 108, 11–29.
- Smith, D.W., Gardiner, B.S., Zhang, L., Grodzinsky, A.J., 2019. Articular cartilage dynamics. Springer.
- Smolen, J.S., van der Heijde, D.M., Keystone, E.C., van Vollenhoven, R.F., Goldring, M.B., Guérette, B., Cifaldi, M.A., Chen, N., Liu, S., Landewé, R.B., 2013. Association of joint space narrowing with impairment of physical function and work ability in patients with early rheumatoid arthritis: protection beyond disease control by adalimumab plus methotrexate. *Annals of the rheumatic diseases* 72, 1156–1162.
- Stoel, B., 2020. Use of artificial intelligence in imaging in rheumatology—current status and future perspectives. *RMD open* 6.
- Tanska, P., Julkunen, P., Korhonen, R.K., 2018. A computational algorithm to simulate disorganization of collagen network in injured articular cartilage. *Biomechanics and modeling in mechanobiology* 17, 689–699.
- Toh, T.S., Dondelinger, F., Wang, D., 2019. Looking beyond the hype: Applied ai and machine learning in translational medicine. *EBioMedicine* 47, 607–615.
- Torp-Pedersen, S., Bartels, E., Wilhjelm, J., Bliddal, H., 2011. Articular cartilage thickness measured with us is not as easy as it appears: a systematic review of measurement techniques and image interpretation. *European Journal of Ultrasound* 32, 54–61.
- Wada, K., 2016. labelme: Image Polygonal Annotation with Python. <https://github.com/wkentaro/labelme>.
- Ye, J.C., Sung, W.K., 2019. Understanding geometry of encoder-decoder CNNs, in: International Conference on Machine Learning, PMLR. pp. 7064–7073.
- Yin, S., Peng, Q., Li, H., Zhang, Z., You, X., Fischer, K., Furth, S.L., Tasian, G.E., Fan, Y., 2020. Automatic kidney segmentation in ultrasound images using subsequent boundary distance regression and pixelwise classification networks. *Medical Image Analysis* 60, 101602.
- Zhang, T.Y., Suen, C.Y., 1984. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM* 27, 236–239.