

Research paper

Cascade learning in multi-task encoder–decoder networks for concurrent bone segmentation and glenohumeral joint clinical assessment in shoulder CT scans

Luca Marsilio^a, Davide Marzorati^b, Matteo Rossi^a, Andrea Moglia^a, Luca Mainardi^a,
Alfonso Manzotti^d, Pietro Cerveri^{a,c}^{*}

^a Department of Electronics, Information and Bioengineering, Politecnico di Milano University, via Ponzio 34/5, Milan, 20133, Italy

^b Institute of Digital Technologies for Personalised Healthcare, Department of Technology and Innovation, University of Applied Sciences and Arts of Southern Switzerland, Via la Santa 1, Lugano, CH-6962, Switzerland

^c Università di Pavia, Via A. Ferrata, 5, Pave, 27100, Italy

^d Hospital ASST FBF-Sacco, piazzale Brescia, 20, Milan, 20149, Italy

ARTICLE INFO

Keywords:

Glenohumeral joint diagnostics

Bone CT segmentation

Shoulder arthroplasty

Multi-task deep learning

PSI-based intervention

ABSTRACT

Osteoarthritis is a degenerative condition that affects bones and cartilage, often leading to structural changes, including osteophyte formation, bone density loss, and the narrowing of joint spaces. Over time, this process may disrupt the glenohumeral (GH) joint functionality, requiring a targeted treatment. Various options are available to restore joint functions, ranging from conservative management to surgical interventions, depending on the severity of the condition. This work introduces an innovative deep learning framework to process shoulder CT scans. It features the semantic segmentation of the proximal humerus and scapula, the 3D reconstruction of bone surfaces, the identification of the GH joint region, and the staging of three common osteoarthritic-related conditions: osteophyte formation (OS), GH space reduction (JS), and humeroscapular alignment (HSA). Each condition was stratified into multiple severity stages, offering a comprehensive analysis of shoulder bone structure pathology. The pipeline comprised two cascaded CNN architectures: 3D CEL-UNet for segmentation and 3D Arthro-Net for threefold classification. A retrospective dataset of 571 CT scans featuring patients with various degrees of GH osteoarthritic-related pathologies was used to train, validate, and test the pipeline. Root mean squared error and Hausdorff distance median values for 3D reconstruction were 0.22 mm and 1.48 mm for the humerus and 0.24 mm and 1.48 mm for the scapula, outperforming state-of-the-art architectures and making it potentially suitable for a PSI-based shoulder arthroplasty preoperative plan context. The classification accuracy for OS, JS, and HSA consistently reached around 90% across all three categories. The computational time for the entire inference pipeline was less than 15 s, showcasing the framework's efficiency and compatibility with orthopedic radiology practice. The achieved reconstruction and classification accuracy, combined with the rapid processing time, represent a promising advancement towards the medical translation of artificial intelligence tools. This progress aims to streamline the preoperative planning pipeline, delivering high-quality bone surfaces and supporting surgeons in selecting the most suitable surgical approach according to the unique patient joint conditions.

1. Introduction

Osteoarthritis (OA) is a degenerative condition affecting bones and cartilage, usually ending up in changes of the bony surfaces such as the formation of osteophytes, bone density loss, and the narrowing of joint spaces [1–4]. In the shoulder, the glenohumeral (GH) joint is formed by the glenoid (or humeral socket) and the humeral head,

physiologically concentric to the glenoid. When primary OA develops, cartilage deterioration leads to a sensible reduction of the GH space, ultimately causing direct contact between the humeral head and its socket. This condition can progress up to complete impingement inducing inflammation, pain, and the reduction of the joint's range of motion. As OA severity increases, osteophytes might arise in the antero-inferior portion of the humeral head and extend downward [2].

* Correspondence to: Politecnico di Milano University, Milan 20133, Italy.

E-mail address: pietro.cerveri@polimi.it (P. Cerveri).

<https://doi.org/10.1016/j.artmed.2025.103131>

Received 12 September 2024; Received in revised form 7 April 2025; Accepted 9 April 2025

Available online 22 April 2025

0933-3657/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

The constant bone rubbing can also lead the glenoid to flatten, along with further osteophyte progression to its boundaries [5]. On a long-term basis, the overall process disrupts the GH joint functionality [5]. Different treatment options can be pursued based on the severity of the condition and the osteophyte location. They span from rehabilitation and drug therapy to joint resurfacing and total arthroplasty, offering a spectrum of interventions tailored to individual needs [6–8]. Moreover, a pathologic humeroscapular alignment (HSA) further contributes to joint instability and osteoarthritis development [9,10]. This condition, also called humeral eccentricity, involves the humeral head shift to the glenoid surface in the superoinferior and anteroposterior planes. Typically, it requires the surgical repositioning of the humeral head within its socket, following rehabilitation [11]. If the eccentricity is associated with osteoarthritis or rotator cuff laxity, shoulder arthroplasty may be considered an effective therapeutic option [12,13]. In this situation, the assessment of the HSA before surgery improves the efficacy of the clinical procedure, and the rehabilitation plan after the operation can be tailored accordingly [14,15]. Therefore, GH osteophyte severity, intra-articular space staging, and HSA are among the ultimate predictive factors for shoulder joint treatment. The ensemble of these conditions is crucial to identifying the most suitable surgical implant. Anatomical and reverse are the two main prosthetic devices used in shoulder replacement surgeries to restore joint function. An anatomical shoulder implant is designed to mimic the natural shoulder anatomy. Conversely, the reverse is specialized for individuals with severe osteoarthritis combined with irreparable rotator cuff damage. Since a strong humeral head eccentricity is strongly correlated to rotator cuff damage [16,17], an eccentric HSA with a high degree of osteophyte size and GH joint space narrowing suggest a reverse approach, while concentric HSA and a physiological GH distance an anatomical prosthesis [18]. Furthermore, an accurate alignment in primary arthroplasty intervention is related to a reduced need for revision procedures, enhancing the outcomes of such techniques [19]. Personalized Surgical Instruments (PSIs) recently revealed decreased surgical time and a more repeatable neutral postoperative alignment, reducing also intra-operative complications [20–22]. PSIs for shoulder arthroplasty intervention are patient-specific cutting jigs reproducing the patient's proximal humerus and scapula edges. Their production and deployment require access to digital three-dimensional (3D) surfaces of the humerus and scapula, obtained through image segmentation [23,24]. These surfaces are needed to properly manufacture the disposable jigs, choose the prosthesis size, estimate the optimal cutting planes. It was showcased that this technique reduces the risk of poor fitting and associated loosening of the implants [25]. Diagnostics and pre-operative computed tomography (CT) or magnetic resonance imaging (MRI) of the upper thorax are mandatory to identify osseous morphology, stage articular surface degeneration, characterize pathological deformations, and assess bone misalignment. Nevertheless, the presence of GH osteoarthritic-related pathological conditions results in highly irregular bone profiles, making the delineation of surface boundaries challenging even for expert radiologists [26]. In addition, automated image processing and analysis still represent a critical issue in the orthopedic pipeline [4,27,28]. Deep learning-based tools, such as convolutional neural networks (CNNs), were explored in this field [29–31]. Encoder–decoder (E-D) architectures, such as the UNet and nnUNet, were effective in the identification of osseous regions and soft tissues in two-dimensional (2D) X-ray and volumetric (3D) scans [29,30,32–37]. CNNs were also studied to stage osteoarthritis [38–40] and optimal treatment prediction, offering a valuable tool to assist clinicians in preoperative planning [8,41]. Summarizing, accurate bone segmentation is challenging due to irregular profiles and pathological changes, necessitating advanced image processing tools. While PSIs enhance outcomes, their dependence on manual or semi-automated workflows introduces inefficiencies and variability. CNNs excel in segmentation and osteoarthritis staging but often lack multi-task capabilities for comprehensive GH

joint analysis. That being said, we here propose an innovative CNN-based framework designed for processing shoulder CT scans, aiming to streamline a PSI-based shoulder preoperative planning pipeline by delivering high-quality bone surfaces and support surgeons in selecting the most suitable surgical approach according to the unique patient joint conditions. It includes the semantic segmentation of the proximal humerus and scapula with the CEL-UNet model [35], the 3D reconstruction of bone surfaces, the identification of the GH joint region, and the staging of three common osteoarthritic-related conditions: osteophyte formation (OS), glenohumeral joint space reduction (JS), and humeroscapular misalignment (HSA).

1.1. Literature background

1.1.1. Automated CT segmentation methods for PSI-based preoperative plan

The success of PSI-based surgery relies on achieving sub-millimetric alignment between the patient's bone surface and the jig footprint [42,43]. Consequently, the accuracy of bone segmentation methods is critical to ensuring precise instrumentation alignment with the patient's anatomy. Traditional automated segmentation techniques, such as statistical shape models (SSMs) and region-growing algorithms, have been investigated for this purpose. A hierarchical acetabulum and femoral head SSM was proposed for the automatic segmentation of diseased hip [44]. In [45,46], 3D SSMs were deployed for pelvic bones segmentation from CT data, while in [47] the pre-morbid proximal humeral anatomy was predicted in patients with severe degenerative osteoarthritis or a fracture of the proximal humerus. Region-growing algorithms have also been utilized for surgical planning applications [48–50]. However, both SSMs and region-growing algorithms face significant challenges when dealing with complex anatomical structures and the variability often present in damaged joints. CNNs, such as the UNet [51] and the nn-UNet [29], demonstrated the ability to overcome these limitations by effectively capturing intricate patterns and accommodating anatomical variability. They were successful in knee bone CT scans [34,35,52], vertebral bodies in CT images [53–55], and automatically produced tibia and femur volumes compliant for a PSI-based total knee arthroplasty intervention [56]. 3D UNet architectures also systematically outperformed their 2D counterparts in processing volumetric scans [34,37,53].

1.1.2. Deep learning approaches to image-based joint clinical evaluation

CNNs were proposed to stage cartilage osteonecrosis in the knee from X-ray images according to the Kellgren-Lawrence scoring, reporting results similar to intra-expert repeatability [31,38]. In [57], different pre-trained models, including ResNet and DenseNet, were compared to assess fracture/non-fracture conditions in X-ray images. A multi-task deep learning model was proposed for grading radiographic (weight-bearing anterior–posterior pelvic) hip osteoarthritis in 4368 patients [58], disregarding other clinical conditions. In the shoulder, rotator cuff muscle degeneration was assessed using a CNN that analyzed CT scans from 95 patients, achieving accuracy comparable to that of expert raters [39]. A regression CNN was used to estimate the critical shoulder angle [59] from anteroposterior shoulder X-ray images, with outcomes compatible with the clinical setting. A similar pipeline was also proposed to assess the most common causes of shoulder pain, such as proximal humeral fractures, joint dislocation, periarticular calcification, osteoarthritis, osteosynthesis, and joint endoprosthesis, on 2700 X-ray images [60]. However, enhancing the shoulder evaluation for a patient-specific surgical plan requires a more comprehensive joint analysis, including osteophyte staging and humeroscapular alignment, often underestimated in the current literature.

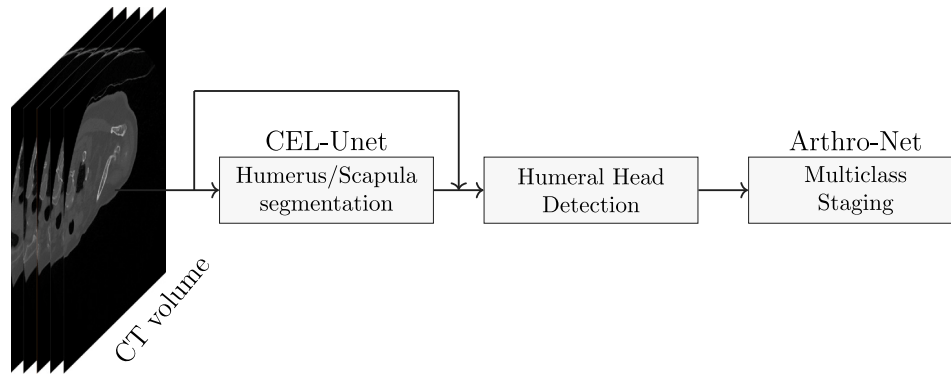


Fig. 1. The proposed processing pipeline involving the volume segmentation to extract the humerus and the scapula (CEL-UNet model), the detection of the humeral head region in the original CT volume exploiting the two reconstructed surfaces, and finally the multiclass staging of the GH joint (Arthro-Net model), in terms of osteophyte staging (3 classes of increasing osteophyte size), GH joint space (3 classes of increasing severity) and humeral scapular alignment (2 classes).

1.2. Contributions

Considering the preceding context, this study proposes a novel automatic deep learning pipeline (Fig. 1), combining in cascade the semantic segmentation of proximal humerus and scapula, the 3D reconstruction of the two bones, and a multi-task multi-class classifier, predicting osteophyte staging, GH osteoarthritis severity, and HSA, concurrently. The paper's contributions can be summarized as:

- the CEL-UNet segmentation quality assessment against state-of-the-art nnUNet architectures over a 571 CT scan dataset featuring several degrees of bone morphological deformations;
- the automatic extraction of the GH joint region in the CT volume to speed up the identification of the region where relevant joint pathological conditions are detected;
- the design of a novel convolutional neural networks (CNN) architecture (Arthro-Net) tailored for staging three GH osteoarthritic-related pathological conditions, concurrently;
- the deployment of an automated deep learning-based pipeline coupling two cascaded CNN architectures to simultaneously perform CT segmentation, 3D bone reconstruction, and GH pathology classification.

2. Methodology

2.1. Dataset description

A dataset of 607 axial CT scans, acquired in DICOM standard format in the context of PSI-based preoperative planning for total or reverse shoulder arthroplasty intervention, was provided by MEDACTA International SA (Castel San Pietro, Switzerland). Patients, aged 74 ± 11 years, were diagnosed with various clinical conditions like osteoarthritis, inflammatory arthritis, osteonecrosis involving the glenoid, and post-traumatic degenerative GH joint disease. Preliminary image analysis confirmed varying degrees of osteophytes, cartilage defects, misalignment of the humeral head, and GH impingement. All patients underwent shoulder arthroplasty between 2021 and 2022 (about 20% anatomical implants and the remaining inverse implants). CT scans were acquired with various imaging equipment. Most scans had a resolution of 512×512 pixels and an average of 330 slices. The pixel size varied from 0.30 to 0.98 mm, and axial slicing ranged from 0.30 to 2.5 mm.

In addition to CT images, the dataset included 3D surface models of the proximal humerus and scapula in STL format. For each patient, two distinct humeral surfaces were provided: the original proximal humerus morphology and a modified version with osteophytes manually removed to represent a physiological humeral head anatomy (Fig. 2).

These two surfaces were created to enhance the PSI-based preoperative planning process. The original morphology is essential for designing custom-cut implants and planning the contact areas between the cutting jigs. The osteophyte-cleared model is crucial for determining the appropriate prosthesis size and optimal positioning. Bony surfaces were generated by expert radiological operators from semi-automated CT segmentation with Mimics software (v. 16.0) by Materialise (Leuven, Belgium). To enhance the reliability of these annotations, each surface was revised by one additional expert. STL surfaces were then processed with 3D Slicer (v. 5.2.2) to extract the segmentation labels to train, validate, and test the performance of the models.

2.2. Training set preparation

Out of the initial 607 cases, 36 were excluded due to the presence of inner metal parts such as screws, implants, and plaques, which fell outside the scope of this study. The remaining 571 cases (297 males and 274 females, comprising 271 left shoulders and 300 right shoulders) were considered for the present investigation. Since CT scans originated from different scanning equipment, voxel intensities were normalized to improve the network's learning performance. Values, acquired in the Hounsfield Unit (HU) scale, were first clipped between -1024 and 2500 units, representing air to dense cortical bone [61], then shifted to positive ranges, and normalized between 0 and 1 for consistency. From this dataset, 85% of the cases, totaling 481, were randomly selected for training and validation sets, including 410 and 71 volumes, respectively. The remaining 90 images defined the test set to evaluate the network performances. The splitting ensured an equal distribution of different pathologies across the test set to provide a robust evaluation of model performance across various conditions. To address the two tasks, the dataset was duplicated into two sets: one for segmentation (DSeg), and one for classification (DClass) (Fig. 3). CT scans in the DSeg set were cropped in the axial, coronal, and sagittal view, to remove all the slices where the proximal humerus and scapula were unavailable. A patch-based method was then applied to increase training and validation instances, while preserving the original voxel resolution. Each cropped CT was patched into a different number of $160 \times 160 \times 160$ sub-volumes, depending on their initial dimension and degree of overlapping (around 25% of the size). For the DClass set, CT scans were cropped to a single $160 \times 160 \times 160$ sub-volume, ensuring that both the humeral head and the glenoid surface were included and centered on the GH joint area. This focused extraction was essential to minimize classification biases from surrounding tissues and organs and to reduce computational overhead. The GH-centered bounding box was extracted through an automated algorithm. This builds upon the technique described in [23], originally focused on the proximal femoral head. It employs morphological analysis of the proximal humerus to delineate the desired region. The number of CT

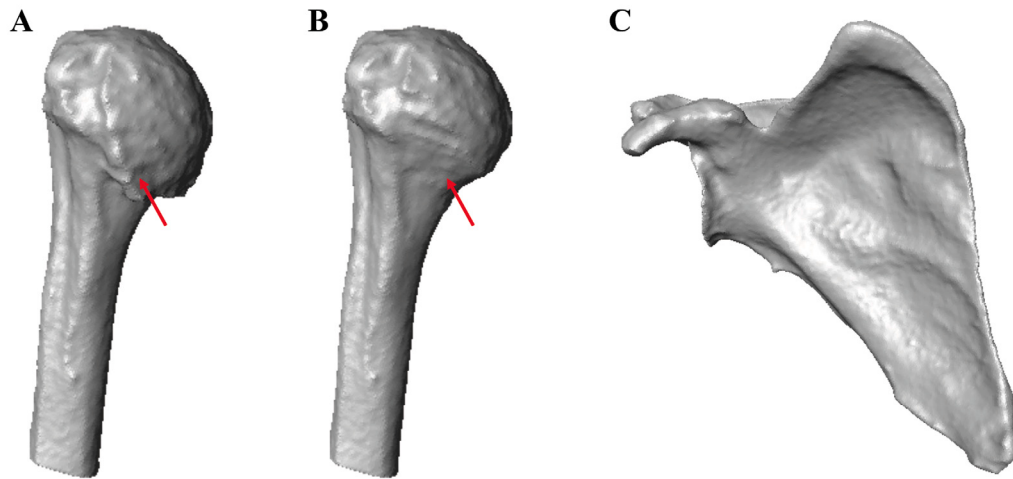


Fig. 2. Example of bone volumes provided in the dataset (A) original humerus morphology with osteophytes pointed out by the red arrow (B) osteophyte-cleared humerus (C) scapula.

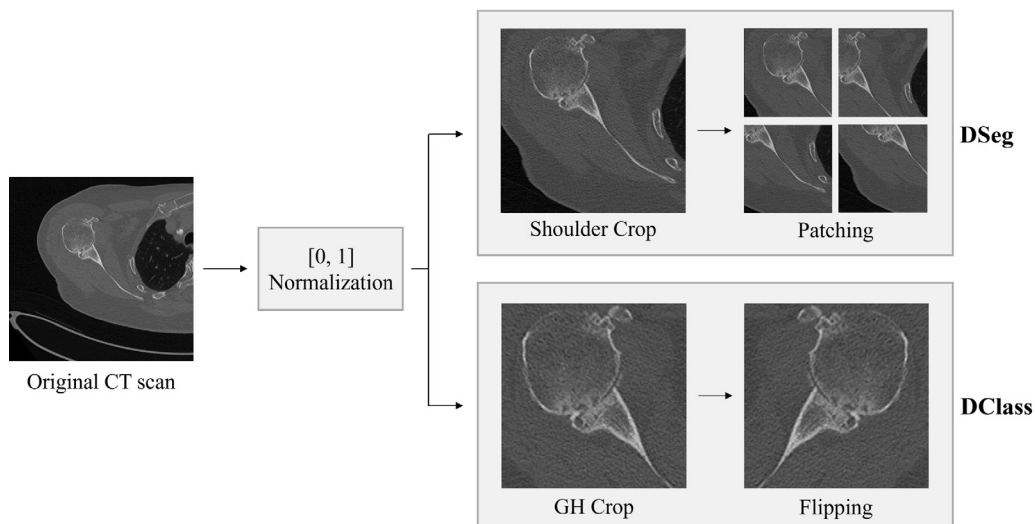


Fig. 3. Segmentation (DSeg) and classification (DClass) training set generation pipeline. After CT normalization, DSeg is produced by cropping the volume to focus on the shoulder joint and patching the extracted sub-volumes, while for DClass a GH-centered crop was performed, and data augmentation was obtained by flipping the sub-volumes on the sagittal plane.

scans in the DClass set was finally doubled by flipping the originated sub-volumes in the sagittal plane. This technique effectively mirrored each joint, generating complementary and physiological counterparts, increasing the dataset number to enhance model generalization.

2.3. GH pathological conditions staging and data labeling

The severity of humeral head osteophytes (OS) was staged into three classes based on the Samilson-Prieto grading system [62,63], featuring increasing osteophyte size. Grade 0 denoted small-size ($s_o < 3$ mm), grade 1 medium-size ($3 < s_o < 7$ mm), and grade 2 large-size ($s_o > 7$ mm) osteophytes. Preliminary labeling of OS was performed automatically. The process involved computing the maximum distance (in millimeters) between the morphologic and osteophyte-cleared humerus reference surfaces for each case in the dataset. The corresponding classification label was then assigned. A radiological expert with more than 20 years of orthopedic clinical practice (A.M.) revised the final labeling of the OS. The GH joint space was tagged manually according to the Kellgren-Lawrence grading system [1] under the supervision of the same radiological expert. Three classes were identified for each case

depending on the residual GH joint space (JS), featuring physiological joint space (grade 0), slightly narrowed joint space (grade 1), and non-detectable joint space (grade 2). HSA was assessed by analyzing the displacement of the humeral head relative to the glenoid surface defining either eccentric-pathological or concentric-physiological humeral head alignments [15]. The frequency across classes within each diagnostic task is reported in Table 1. Fig. 4 presents three shoulder joints from the dataset with different pathological conditions of the GH joint. In case A, acromiohumeral distance (i.e., the space between the humeral head and the acromion) was absent, indicating significant eccentricity in the HSA. However, joint space was still visible and only small-size osteophytes were present in the anteroinferior portion of the humeral head. In both cases B and C, the intra-articular spaces were missing, with the humeral head and glenoid making direct contact. This bone-on-bone rubbing led to the development of medium and large-size osteophytes, respectively. The humeral head in joint B was still concentric to its socket, while joint C showed an eccentric condition.

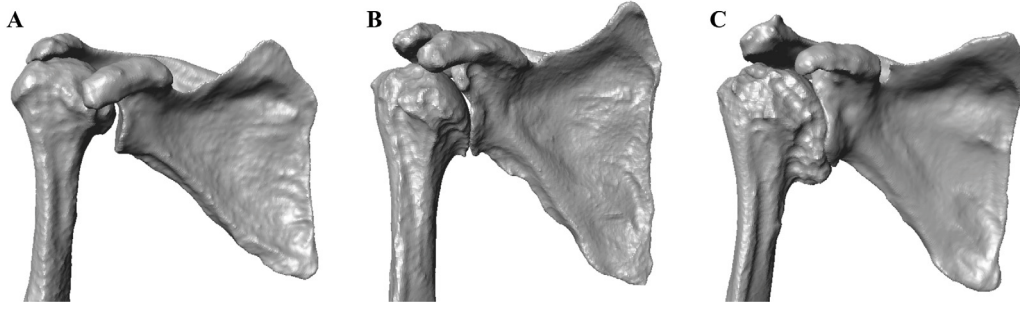


Fig. 4. Different pathological conditions. (A) Extreme eccentricity of the humeral head, (B) Mild GH osteoarthritis, (C) Severe GH osteoarthritis with large osteophyte presence.

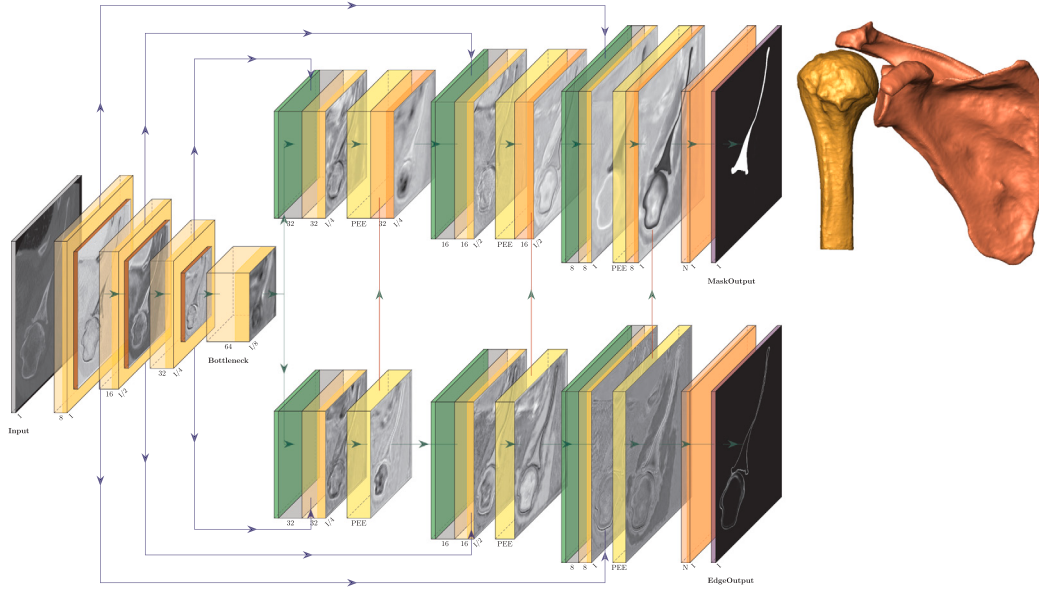


Fig. 5. CEL-UNet architecture. The encoding path, on the left, extracts segmentation features at different resolutions, ending with a bottleneck. The decoder, on the right, is split into two parallel branches, devoted to region segmentation (RA) and edge detection (CA). The output of each CA decoding block is combined with the corresponding RA layer through vertical skip connections (red lines) to enhance semantic segmentation. Pyramidal edge extraction (PEE) modules (yellow boxes) are added after decoding blocks. The segmentation mask output is then used to reconstruct the bone surfaces using traditional matching cubes algorithm.

Table 1

GH osteoarthritic-related condition labeling. The first column describes the classification tasks, namely osteophyte size (OS), GH joint space (JS), and humeroscapular alignment (HSA). The second and third columns show the labeling criteria for each multi-class task and the provided index. The last column reports the frequency of each class in the corresponding task.

Classification task	Criteria	Index	Frequency [%]
Osteophyte staging (OS)	<3 mm	0	31.1
	3–7 mm	1	36.1
	>7 mm	2	32.8
GH joint space (JS)	Physiological	0	38.2
	Narrowed	1	27.2
	Non-detectable	2	34.6
HSA	Concentric	0	56.1
	Eccentric	1	43.9

2.4. Segmentation network: CEL-UNet

The CEL-UNet model was prior [35,56] designed and developed by our group to boost the segmentation quality on knee CT scans and enable preoperative planning in PSI-based total knee arthroplasty. The identification of severe pathological anatomy deformations and narrow joint space was enhanced by combining traditional semantic mask segmentation with boundary identification (Fig. 5). In detail, the decoder was split into two parallel branches, one devoted to region

segmentation (RA, region-aware branch) and the other addressing edge detection (CA, contour-aware branch). Vertical unidirectional skip connections were added between each CA branch processing block and its corresponding one in the RA branch. This design integrates region and boundary-related features at progressively higher resolutions. A specialized processing module was also implemented after each CA branch layer, using the pyramidal edge extraction (PEE) approach inspired by [40]. A custom loss function dynamically weighting RA and CA branch losses throughout the training process was deployed to better combine each contribution. It exploited the importance of the region boundary, by incorporating the distance-weighted map (DWM) factor into the formula. DWM was computed as the negative exponential function of the Euclidean distance transform (EDT) as:

$$DWM_c = \left(1 + \gamma \cdot \exp \left(-\frac{EDT_c}{\sigma} \right) \right) \quad (1)$$

where c was the current label, while γ and σ were heuristic parameters. EDT assigned to each voxel the value of its distance from the closest voxel belonging to the boundary of the corresponding mask label. Hence, the distance-weighted cross-entropy loss C_w embedded the DWM factor to raise the significance of contour and near-surface voxels as:

$$C_w = - \sum_c \left(\sum_c (DWM_c \cdot y_c \cdot \log(\hat{y}_c)) \right) \quad (2)$$

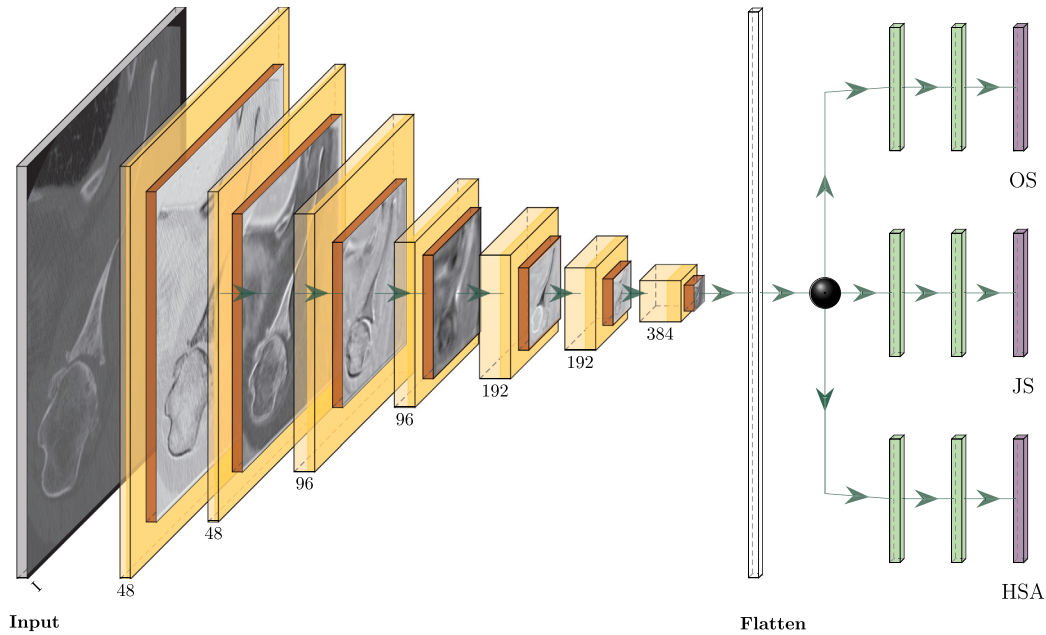


Fig. 6. Arthro-Net architecture. Each processing block (yellow box) is featured by two sequences of convolutional with linear activation, batch normalization, and ReLU activation layers, mimicking the CEL-UNet encoder. Downsampling is performed through max pooling (orange box). The block ends with a flatten layer followed by three separate fully connected branches to specialize the network for the three classification tasks, namely osteophyte staging (OS), GH joint space (JS), and humeroscapular alignment (HSA).

where y_c and $\hat{y}_c$, respectively the true and segmented voxels for the label c , were computed over the N voxels. Thus, the loss of the region-aware branch was:

$$\mathcal{L}_{RA} = 1 - (\alpha \cdot D + (1 - \alpha) \cdot C_w) \quad (3)$$

where α was a tunable parameter to modulate differently the two contributions during the training process and D represents the dice formula. In this work, CEL-UNet segmentation results were compared to the results obtained with the nnUNet proposed by [29] alternatively trained with two different loss functions, namely distance cross-entropy and focal.

2.5. Classification network: Arthro-Net

The Arthro-Net (Fig. 6) is a novel CNN architecture designed for multi-task, multi-class staging of GH osteoarthritic-related conditions. The network processes GH-centered patches and classifies the severity of each pathology according to predefined categories: three classes for osteophyte size (OS), three classes for joint space narrowing (JS), and two classes for the humeroscapular alignment (HSA). The Arthro-Net encoder is a sequence of convolutional/downsampling blocks to extract features at decreasing resolution up to the bottleneck. Each processing block embeds two sequences of convolutional with linear activation, batch normalization, and ReLU activation layers. The filter size and stride length are $3 \times 3 \times 3$ and $1 \times 1 \times 1$, respectively. This configuration is consistent with the CEL-UNet feature extractor, which captures relevant osseous characteristics from CT images [35]. However, the number of feature maps in the Arthro-Net doubles every two processing blocks (Fig. 6), to avoid unnecessary over-parametrization. Downsampling is performed through max-pooling layers with filter size and stride length of $2 \times 2 \times 2$. According to state-of-the-art classification architecture [64], a flatten layer was placed after the bottleneck to reshape the data into a one-dimensional tensor while preserving the total number of elements. Three separate fully connected branches, characterized by two consecutive dense layers with input sizes of 256 and 32, and a ReLU activation function, were added to specialize the network for each classification task. Each branch was preceded by a dropout layer to prevent overfitting and reduce sparsity, due to the

large number of unique connections at this stage. The dropout rate was set to 0.6. Lastly, the three outputs consisted of one softmax layer each, with three, three, and two neurons, respectively. Categorical cross-entropy was chosen as the loss function. To balance the uneven training label frequency for each task, shown in Table 1, each loss function was weighted with a parameter K_c to balance the representation of every class as:

$$K_c = \frac{\frac{1}{N_c}}{\sum_{i=1}^C (\frac{1}{N_i})} \quad (4)$$

where c , N_c , and C were the current class for the specific classification task, the number of total cases for each class, and the number of classes, respectively. To further validate the novel architecture performances, an ablation test was performed to evaluate the dependency of the classification scores on the number of convolutional layers and feature maps. Two networks of increasing depth were analyzed, by training the model with either seven or eight processing blocks. Moreover, the initial number of feature maps was alternatively set to 16, 32, and 48 (the source code of the overall module is available at <https://github.com/LucaMarsilio/AI-Shoulder.git>).

2.6. Network training and prediction pipeline

The CEL-UNet and Arthro-Net were trained separately for the two tasks using Tensorflow framework (v. 2.14). The learning rate for the Adam optimizer was set to $1e-04$. Nevertheless, segmentation and classification were sequentially performed in the prediction step, combining the two networks with a cascade approach (Fig. 1). The CEL-UNet region identification output was processed by two custom algorithms. The first one was a marching cube-based method generating the 3D humerus and scapula surfaces from the corresponding segmentation maps [23]. The second method created a bounding box around the GH portion of the CT given the shoulder joint side (left or right) and the bone coordinates obtained in the previous step. The extracted sub-volume served as the input for the Arthro-Net to classify osteoarthritic-related conditions. Training and testing procedures were performed on a 32-core CPU and NVIDIA A100-PCIE GPU with 40 GB RAM. Due to memory constraints, the batch size in the training was

Table 2

Dice, Jaccard, precision, and recall median and IQR scores for humerus (*above*) and scapula (*below*) comparing the performances of the three networks (CEL-UNet, and DCE- and FOC-nnUNet).

	Humerus			
	Dice	Jaccard	Precision	Recall
DCE-nnUNet	0.98 (0.97–0.99)	0.97 (0.94–0.98)	0.98 (0.97–0.99)	0.98 (0.97–0.99)
FOC-nnUNet	0.98 (0.97–0.99)	0.96 (0.94–0.98)	0.99 (0.98–0.99)	0.97 (0.96–0.98)
CEL-UNet	0.99* (0.98–0.99)	0.98* (0.97–0.99)	0.99 (0.98–0.99)	0.99 (0.98–0.99)
	Scapula			
	Dice	Jaccard	Precision	Recall
DCE-nnUNet	0.97 (0.96–0.98)	0.95 (0.92–0.96)	0.97 (0.95–0.98)	0.97* (0.96–0.98)
FOC-nnUNet	0.97 (0.96–0.98)	0.94 (0.92–0.96)	0.97 (0.96–0.98)	0.97 (0.95–0.98)
CEL-UNet	0.98* (0.97–0.98)	0.95* (0.94–0.96)	0.99* (0.99–0.99)	0.96 (0.95–0.97)

* Defines a statistically significant difference between the outcome distributions against the two other networks.

set to six. Early stopping was used to prevent overfitting, stopping the training after 50 epochs with no improvement in validation loss. As an example, one training took 183 epochs.

2.7. Segmentation and classification quality assessment

The humerus and scapula segmentation performance on the test set was evaluated by comparing the CEL-UNet results (Dice and Jaccard indices) with those obtained by the state-of-the-art nnUNet [29]. The nnUNet full-resolution model was trained with two different loss functions, namely distance cross entropy (DCE) and focal (FOC) losses. Precision and recall were further computed to detect over- and under-segmentation errors. To evaluate 3D surface reconstruction accuracy, root mean squared error (RMSE) and Hausdorff distance were measured, capturing average and maximum distances, respectively, between predicted and reference volumes. For classification tasks, accuracy, precision, recall, F1-score, and confusion matrices were utilized to determine the optimal Arthro-Net architecture. Time evaluations were conducted across the automated pipeline steps, encompassing CT segmentation, 3D reconstruction, GH identification, and pathological condition staging. The statistical analysis of segmentation and 3D reconstruction outcomes was conducted using the Friedman test, followed by Wilcoxon Signed-Rank tests with Bonferroni correction for post-hoc analysis. This choice was driven by the features of the results: three paired sets with non-Gaussian distributions. Asterisk symbol (*) in Table 2 defines a statistically significant difference between the outcome distributions against the two other networks. Fig. 7 reports the unique differences between the CEL-UNet and the other models, a single asterisk (*) refers to a p-value < 0.05, while three (***) to a p-value < 0.001.

3. Results

3.1. Segmentation result comparison

Focusing on the humerus (Table 2 and Fig. 7), Dice and Jaccard distributions achieved with the CEL-UNet were statistically higher than the other two networks. CEL-UNet also outscored DCE-nnUNet on precision and FOC-nnUNet on recall, respectively. Considering the scapula, statistically significant differences demonstrated the superiority of the CEL-UNet over state-of-the-art architectures in terms of Dice, Jaccard, and precision indices. The highest recall value was found with the DCE-nnUNet instead. A qualitative comparison of humerus and scapula segmentation across the three networks (Fig. 8) highlighted the CEL-UNet's ability to capture osteophyte contours (A, B) and narrowed scapular anatomy (C) accurately. In this example, DCE- and FOC-nnUNet did not recognize the bone spur (A, B), and FOC-nnUNet exhibited under-segmentation issues (black arrow).

3.2. Humerus and scapula 3D reconstruction

The CEL-UNet achieved global RMSEs of less than 1 mm for both 3D reconstructed surfaces (Fig. 9). Median and IQR values were 0.22(0.15–0.38), 0.37(0.22–0.77), and 0.55(0.29–0.78) mm for the humerus, and 0.16(0.12–0.27) mm, 0.19(0.13–0.36) mm, and 0.24(0.15–0.54) mm for the scapula with CEL-UNet, DCE-nnUNet, and FOC-nnUNet, respectively. CEL-UNet statistically outscored the other networks. Statistical evidence was confirmed with Hausdorff distance errors, reporting median and IQR of 1.48(0.97–2.26), 2.56(1.47–4.51), and 4.68(2.50–6.73) mm for the humerus, and 1.48(1.13–2.56), 2.22(1.31–3.48), and 2.63(1.84–3.60) mm for the scapula. The qualitative examination of the 3D humerus and scapula reconstructed surfaces for a test set case example (Fig. 10) revealed the CEL-UNet proficiency in the reconstruction of irregular humeral head osteophytes (circled black area A), missed by DCE- and FOC-nnUNet. All three networks faced difficulties in accurately reconstructing the scapula, particularly around the superior angle. This region is highly variable and often sub-millimeter thickness poses a challenge for segmentation at standard voxel resolutions. Despite this, the CEL-UNet delivered superior reconstruction quality, especially in the subscapular fossa. Even in the most challenging cases, where the humeral head was in direct contact with the glenoid surface, the reconstruction remained feasible, successfully maintaining a clear separation between the two structures (Fig. 11).

3.3. Arthro-Net: classification results of the three pathologic conditions

The ablation study on Arthro-Net's key hyperparameters (number of processing blocks and feature maps), revealed two optimal design configurations (Table 3). The first one, named A(7, 48), featuring seven processing blocks and 48 initial feature maps (crf. Fig. 6), showed globally the best accuracy, precision, recall, and F1-score scores, ranging from 0.88 to 0.95 across the three classification tasks. Only the F1-score for HSA (0.88) was less than the corresponding score (0.90) in the design with 32 feature maps. The second one, B(8, 32), showed scores ranging from 0.83 to 0.94, greater than the other two competing configurations, while comparable with the previous design A. The further analysis, performed on the normalized confusion matrices computed for A and B designs (Fig. 12), was resolute in choosing the optimal Arthro-Net configuration. The osteophyte size staging was better achieved with design A, correctly predicting each small-size bone spur and outscoring design B for the other sizes. Likewise, a perfect GH physiological joint space classification was obtained with design A, and non-detectable joint space was better identified as well. Design B instead prevails in the identification of narrow joint space and the detection of eccentric humeroscapular alignment.

Concentric condition staging was better achieved with design A. According to such results, the final Arthro-Net was configured according to design A. As expected the OS recall was more accurate when detecting small (<3 mm) and large (>7 mm) osteophytes than osteophytes of intermediate size (3–7 mm) (Fig. 12). Likewise, the

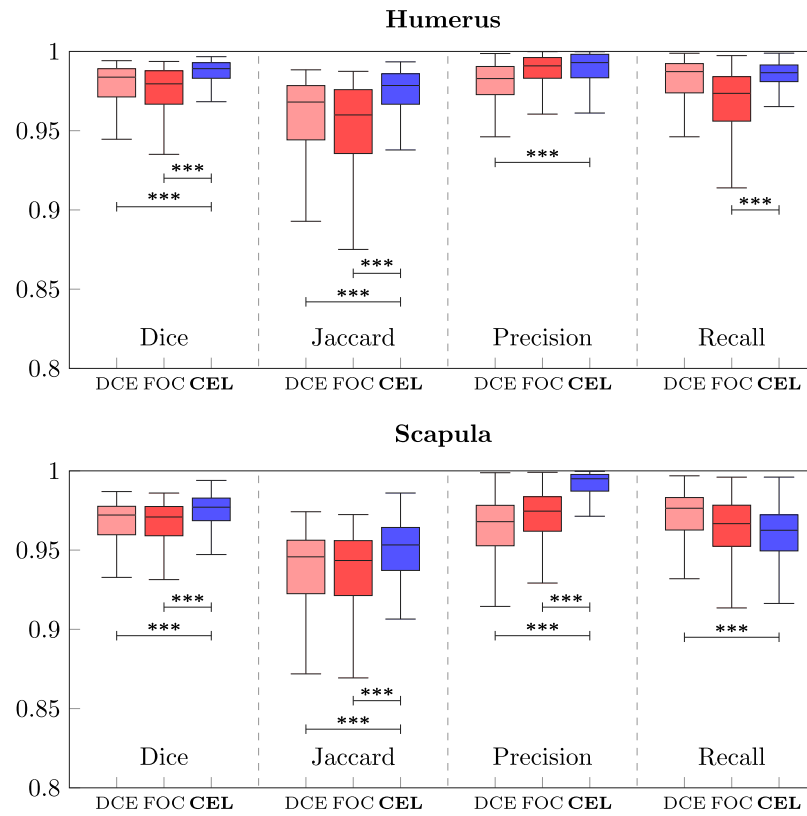


Fig. 7. Boxplots of Dice, Jaccard, precision, and recall for humerus (above) and scapula (below). Blue boxes depict the CEL-UNet results, while light red and dark red boxes show the nnUNet [29] architecture outcomes after training with Distance Cross Entropy (DCE) and Focal (FOC) losses, respectively. Asterisks (***) report a statistically significant difference between the CEL-UNet and the other networks ($p < 0.001$).

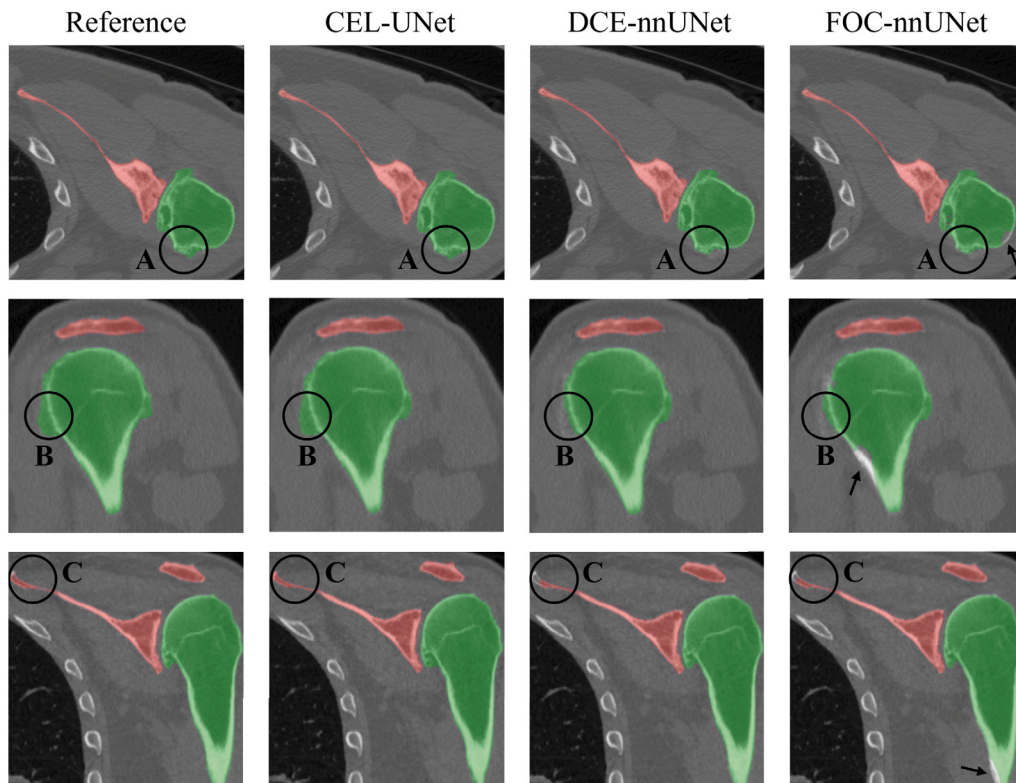


Fig. 8. Qualitative comparison of the humerus (green) and scapula (red) segmentation for a test set case example against the reference label. Reference labels are in the first column, CEL-UNet outputs in the second, while DCE- and FOC-nnUNet segmentations are in the third and fourth, respectively. Circled areas display segmentation errors of the UNet models in identifying humeral head osteophytes (A and B) and the narrow scapular superior angle (C).

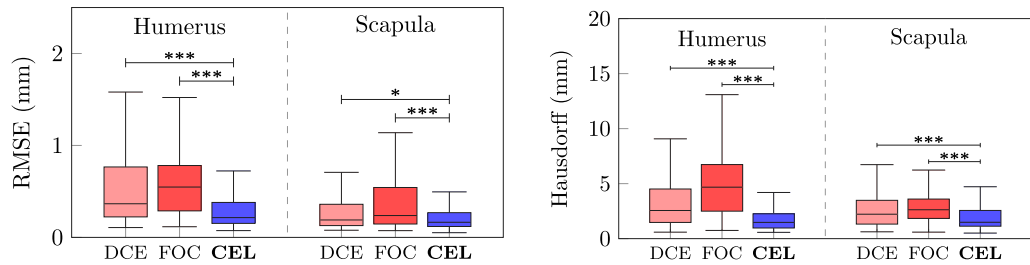


Fig. 9. Humerus and scapula 3D reconstruction error comparison in terms of RMSE (left) and Hausdorff distance (right). Blue boxes depict the CEL-UNet architecture metric distributions, while light and dark red boxes display DCE- and FOC-nnUNet scores. Asterisks (*, ***) report a statistically significant difference between the CEL-UNet and the other networks ($p < 0.05$ and $p < 0.001$, respectively).

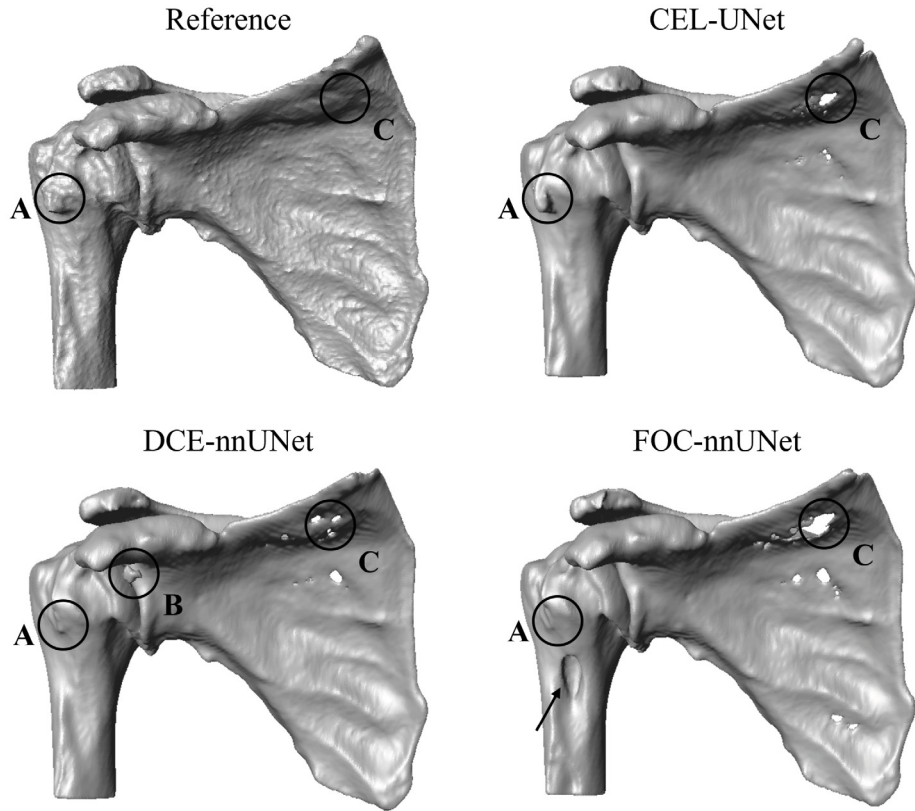


Fig. 10. Qualitative 3D reconstruction comparison of the humerus and scapula for a test set case example against the reference label. (A) irregular humeral head osteophyte reconstruction comparison (B) scapula osseous portion wrongly identified by the DCE-nnUNet (C) scapula superior angle reconstruction. The black arrow points out the error in the humerus shaft reconstruction from the FOC-nnUNet.

Table 3

Accuracy, precision, recall, and F1-score results for the three GH conditions classified, including osteophyte size (OS), joint space (JS), and humeroscapular alignment (HSA) with an increasing number of feature maps (FM) in the first processing block (16, 32, and 48), and comparing two Arthro-Net architectures featuring seven (above) and eight (below) total processing blocks.

FM	7 Processing blocks											
	Accuracy			Precision			Recall			F1-Score		
	OS	JS	HSA	OS	JS	HSA	OS	JS	HSA	OS	JS	HSA
16	0.70	0.75	0.84	0.68	0.76	0.92	0.69	0.74	0.75	0.69	0.72	0.83
32	0.84	0.89	0.90	0.84	0.90	0.91	0.83	0.89	0.90	0.83	0.89	0.91
48	0.89	0.91	0.91	0.89	0.92	0.95	0.88	0.91	0.88	0.89	0.91	0.91
FM	8 Processing blocks											
	Accuracy			Precision			Recall			F1-Score		
	OS	JS	HSA	OS	JS	HSA	OS	JS	HSA	OS	JS	HSA
16	0.78	0.73	0.87	0.79	0.71	0.97	0.78	0.72	0.77	0.78	0.69	0.86
32	0.84	0.88	0.92	0.85	0.89	0.94	0.83	0.88	0.92	0.84	0.88	0.93
48	0.80	0.81	0.88	0.81	0.83	0.93	0.80	0.81	0.83	0.80	0.80	0.88

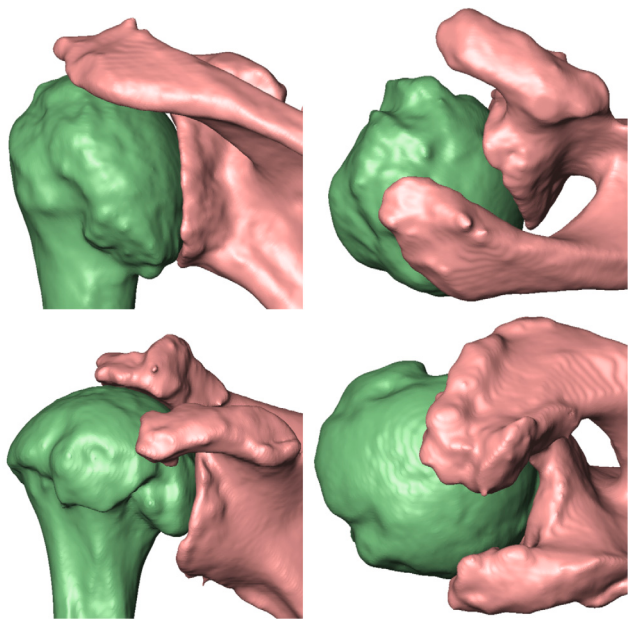


Fig. 11. Frontal (on the left) and top (on the right) views for two CEL-UNet-based 3D reconstructions in test set where the humeral head was in physical contact with the glenoid surface.

Table 4

Time spent for each step in the pipeline, including CT segmentation, humerus and scapula 3D reconstruction, and GH related condition staging. Tests were run on a 32-core CPU, and NVIDIA A100-PCIE GPU with 40 GB RAM. The same tasks were tested on a 12th Gen Intel(R) Core(TM) i7-12850HX, 2.10 GHz, 32.0 GB of RAM and took on average 95 s.

	Segmentation	3D Reconstruction	Classification	Overall
Time [s]	9.2 (7.7–11.6)	4.6 (3.6–6.4)	0.9 (0.6–1.2)	14.8 (11.7–17.1)

recall results of both physiological (100%) and non-detectable (93%) JS conditions were more reliable than the recall value of narrowed (80%) joint space. The classification of humeral head eccentricity was a little poorer (88%) than concentricity (95%).

3.4. Deployment of the overall pipeline and inference speed

To validate the potential of translating the developed pipeline in clinical practice, the computational time of each operation was measured across the whole test set (Table 4). This included CT segmentation using the CEL-UNet, 3D reconstruction, glenohumeral region identification, and diagnostic classification using the Arthro-Net. Among these, segmentation, encompassing CT loading, preprocessing, and postprocessing, emerged as the most time-consuming step, with a median time of 9.2 s. Proximal humerus reconstruction time outpaced that of scapula reconstruction, taking 1.0 s (range: 0.65–1.7) versus 3.6 s (range: 2.9–4.6), owing to the smaller size of the analyzed bone segment. The humeral head patching algorithm time to identify the glenohumeral region was neglectable, while the classification took about 1 s. The cumulative median time for all processes amounted to 14.8 s, confirming a high-speed throughput underscoring the efficiency of the proposed cascaded multi-task approach making it potentially suitable for clinical diagnostic practice.

4. Discussion

In recent years, the use of deep learning algorithms for biomedical image processing has gained considerable attention in several key areas, including semantic segmentation for 3D anatomy reconstruction,

morphological lesion analysis, and disease classification for clinical staging to aid diagnostics. However, much of the current literature is highly specialized, frequently focusing on just one of these domains at the time. In this work, a comprehensive automated framework for shoulder CT scan processing was developed and tested on a 571 retrospective dataset. It integrates the proximal humerus and scapula segmentation and 3D reconstruction, with the classification of three GH joint pathological conditions at various severity degrees. The outcomes showcased median root mean squared errors less than 0.5 mm for both bones, even in case of severe osteoarthritis, extensive morphological deformation, and inter-surface impingement (cfr. Fig. 7). The CEL-UNet segmentation results outscored state-of-the-art nnUNet models (cfr. Fig. 9) [29], revealing an advantage on severe pathological joints (Fig. 8, 10). A novel CNN (Arthro-Net) was further proposed to achieve an automated staging of the GH osteophyte size, joint space, and the relative position between the humeral head and the glenoid surface, following the CEL-UNet-based segmentation. The ablation study allowed us to identify the optimal balance between network depth and feature map size. Configurations with fewer feature maps and depth often lack the complexity needed to accurately capture detailed anatomical features, leading to poorer performance. Conversely, deeper networks with more feature maps may suffer from overfitting. The optimal results from the ablation study delivered a configuration that strikes a balance between model complexity and accuracy, mitigating both under and overfitting. The classification accuracy was around 90% for all three class sets. A deterministic automated algorithm for the GH centered patch extraction was implemented following shoulder segmentation, diverging from the conventional approach of integrating an additional deep learning model such as Mask-R-CNN, YOLO, or analogous object detection networks [65,66]. This approach streamlined the computational workflow and minimized potential errors introduced by inter-model dependencies. In addition, maintaining the original CT resolution was a key consideration in this study. While pixel size standardization is commonly applied in small datasets, resampling can introduce distortions in pixel values, potentially compromising model performance. Given the sufficiently large dataset available, keeping the native resolution ensured that the model operated on unaltered image data, leading to more reliable predictions. Moreover, generating outputs at the same resolution as the input facilitated direct clinical interpretation without requiring additional post-processing steps.

4.1. Literature overview and limitations

Despite the recent advancements in automated image processing, several limitations persist in the existing literature, impacting the robustness, scalability, and clinical translation of the proposed methods. Many studies rely on relatively small datasets [37,67,68], raising concerns about the generalizability of their approaches, not adequately representing the variability found in real-world clinical scenarios. Other works [69,70] analyze CT images from patients who are not engaged in preoperative planning contexts, resulting in datasets with less severe joint damage. This can lead to reduced scalability when applied to cases with significant pathological conditions. Additionally, a reported mean 3D reconstruction error exceeding 1 mm, as in [67,71], renders a method less suitable for PSI-based preoperative planning frameworks, where high reconstruction accuracy is crucial [42,43]. In [71], the GH joint was reconstructed by deploying a SSM yielding to a mean surface RMSE of the humerus and scapula around 2 mm. A UNet-based glenoid segmentation model was trained and tested with a 237 CT scan dataset featuring patients with unilateral shoulder dislocations achieving an average intersection-over-union score of 0.96 [72]. Nearly 100 MRIs were processed by AlexNet and UNet-based segmentation models to extract humerus, humeral head, and articular bone segmentation, reaching a positive predictive value of dice coefficient between 0.94 and 0.97 [73].

The comparison of the Arthro-Net with other models in the literature was not straightforward because two main reasons. First, most of

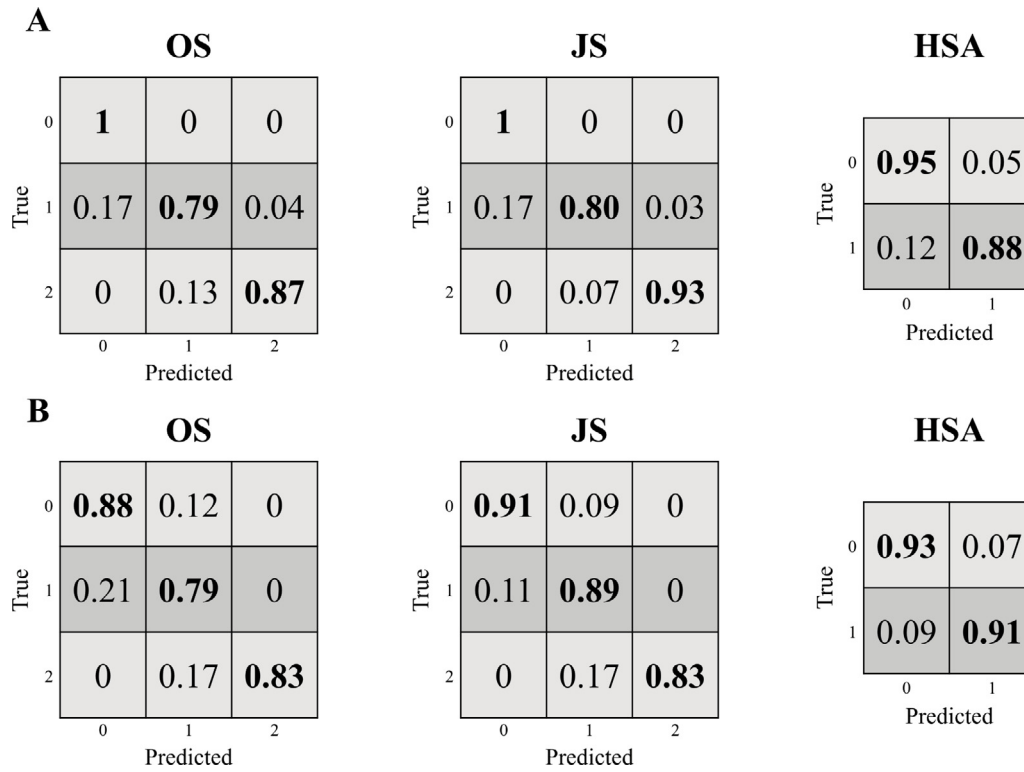


Fig. 12. Confusion matrices of the two best Arthro-Net configurations presented in Table 3 for the three classification tasks (OS, JS, and HSA). (A) Arthro-Net with seven processing blocks and 48 starting feature maps. (B) Arthro-Net with eight processing blocks and 32 starting feature maps.

the works, using deep learning for detecting and stages articular joint pathologies, faced binary classification tasks. Second, while addressing multi-label classification, very few works focused on the glenohumeral joint. More in general, the GH joint deep learning-based analysis has been reported to address different acquisition modalities, pathological conditions, and shoulder morphological parameters. In [74], the authors reported a CNN to automatically extract discriminative features from proton density MRI sequences, achieving about 98% success rate in classifying normal, edematous, and Hill-Sachs lesion humeral heads. However, the geometric relation between the humeral head and the glenoid surface was not investigated. A set of bone morphological parameters, including the glenoid inclination and the shoulder critical angle, were automatically assessed on about 1250 anteroposterior shoulder radiographs using a UNet model [75]. Nonetheless, the staging of potential osteoarthritic conditions of the GH was not computed. Finally, ensemble deep learning models were proposed to process X-ray images and assess fracture/non-fracture conditions in the shoulder area, featuring an accuracy of nearly 84% [57].

4.2. Translational contribution and work limitations

From a translation point of view, we highlight two main potential contributions of this work. Firstly, the achieved median root mean squared errors below 0.5 mm renders the segmentation tool potentially suitable to be integrated into a PSI-based planning pipeline, delivering radiologists high-quality bone surfaces within a computation time-frame that aligns with clinical requirements [56]. Secondly, the three GH joint condition staging may support surgeons selecting the most suitable implant solution among the available ones. For instance, the proximal migration of the humeral head could suggest a large rotator cuff tear discouraging an anatomical prosthesis as well as advocating a reverse total shoulder replacement at the same time. The absence of osteophytes and a concentric humeral head could suggest more conservative solutions, including the GH joint resurfacing [16,17]. Likewise, based on the predicted osteophyte staging and the presence

of a concentric humeral head, there arises a compelling argument in favor of GH joint resurfacing as a potentially superior alternative to complete anatomical shoulder replacement [18].

It is essential nonetheless to interpret the findings of this study within the context of its limitations. Although traditional descriptions of humeral head eccentricity (subluxation) include posterior, anterior, inferior, and superior directions [15], this study focused on a single binary HSA condition, restricting the comprehensive assessment of glenohumeral conditions. This choice was due to the retrospective nature of the study, involving a dataset with multiple cases of degenerative shoulder arthritis. Additionally, while the dataset included nearly 600 patients, it was biased towards an older demographic, with a median age of 74 years. This reflects a common trend in studies of degenerative shoulder conditions, where older patients, who are more susceptible to bone deterioration and decreased turnover [76], are often over-represented. To enhance the scalability and applicability of the segmentation module, future work should aim to include a more diverse patient population, but also cases with implants, plates, and screws to allow the deployment in preparation for revision procedures. As far as reference segmentation is concerned, we remark, like all bone segmentation efforts in CT scans, that the proposed methodology faced the challenge of an uncertain ground truth, particularly in contact regions where expert-driven segmentation remains subjective. This limitation is critical for applications like PSI design, where small errors in the surface reconstruction can have significant consequences. Concurrently, we consider that all AI-based tools for automatic reconstruction of digital anatomical surfaces still constitute a support to the clinical decision, meaning that the final quality check is in charge of the clinician. Nonetheless, the obtained results, especially at contact areas, supported the conclusion that the deep-learning-based methodology was able to ensure accuracy in the millimeter range, thus reasonably compatible with PSI-based surgery. The high throughput presented in Table 4 is achieved with high performance and dedicated hardware, not always available in clinics. Nevertheless, the same pipeline was tested on a laptop featuring 12th Gen Intel(R) Core(TM) i7-12850HX,

2.10 GHz, 32.0 GB of RAM, and equipped with an NVIDIA RTX A200 8 GB video card and completed in around 95 s. Finally, validating this framework with clinicians and surgeons will be essential to assess its practical advantages and effectiveness in real-world clinical scenarios.

5. Conclusions

We have introduced a comprehensive AI-based automated pipeline for reconstructing the two primary bones of the shoulder and predicting three distinct diagnostic conditions of the GH joint. While existing literature offers valuable insights, our study stands out for two key innovative contributions: firstly, our fully automated framework enables the concurrent proximal humerus and scapula 3D reconstruction and the staging of three major glenohumeral joint pathological conditions. This tool has the potential to be further validated within a PSI-based preoperative planning pipeline. It delivers high-quality bone surface models and a comprehensive assessment of the joint condition, that may assist surgeons in the selection of the most suitable surgical approach (arthroplasty or resurfacing) and implant type (anatomical or reverse), thereby streamlining the decision-making process. Secondly, our results derive from a larger test set compared to those in similar studies [67,69,77,78], enhancing the robustness and reliability of our findings.

CRedit authorship contribution statement

Pietro Cerveri: Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Luca Marsilio:** Conceptualization, Software, Data curation and analysis, Writing – original draft. **Davide Marzotti:** Formal analysis, Methodology. **Matteo Rossi:** Software, Visualization. **Andrea Moglia:** Investigation, Writing – review & editing. **Luca Mainardi:** Resources, Writing – review & editing. **Alfonso Manzotti:** Data curation, Investigation.

Ethical background

Image data were retrospectively available in anonymized form by Medacta International SA (Castel San Pietro, Switzerland).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the program PON-FESR 2014–2020 - European Regional Development Fund, ARIA-2020-403, ID 138.360.852, and by PNRR-PE, Future Artificial Intelligence Research (FAIR) - Italian Ministry of University and Research. The authors want to thank Medacta International SA (CH) for providing data.

References

- [1] Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthritis. *Ann Rheum Dis* 1957;16:494–502.
- [2] Ogawa Kiyohisa, Yoshida Atsushi, Ikegami Hiroyasu. Osteoarthritis in shoulders with traumatic anterior instability: Preoperative survey using radiography and computed tomography. *J Shoulder Elb Surg* 2006;15(1):23–9.
- [3] Thomas Michael, Bidwai Amit, Rangan Amar, Rees Jonathan L, Brownson Peter, Tennent Duncan, Connor Clare, Kulkarni Rohit. Glenohumeral osteoarthritis. *Shoulder Elb* 2016;8:203–14.
- [4] Khazzam Michael, Gee Albert O, Pearl Michael. Management of glenohumeral joint osteoarthritis. *J Am Acad Orthop Surg* 2020;28:781–9.
- [5] Lo Lawrence, Koenig Scott, Leong Natalie L, Shiu Brian B, Ashfaq Hasan S, Gilotra Mohit N, Wang Kenneth C. Glenoid bony morphology of osteoarthritis prior to shoulder arthroplasty: what the surgeon wants to know and why. *Skelet Radiol* 2021;50:881–94.
- [6] Junker Susann, Krumbholz Grit, Frommer Klaus W, Rehart Stefan, Steinmeyer Jürgen, Rickert Markus, Schett Georg, Müller-Ladner Ulf, Neumann Elena. Differentiation of osteophyte types in osteoarthritis - proposal of a histological classification. *Jt Bone Spine* 2016;83:63–7.
- [7] Ward Brian E, Dines Joshua S. Patient-specific guides/instrumentation in shoulder arthroplasty. *Am J Orthop* 2018;47(2).
- [8] Leung Kevin, Zhang Bofei, Tan Jimin, Shen Yiqiu, Geras Krzysztof J, Babb James S, Cho Kyunghyun, Chang Gregory, Deniz Cem M. Prediction of total knee replacement and diagnosis of osteoarthritis by using deep learning on knee radiographs: Data from the osteoarthritis initiative. *Radiol* 2020;296:584–93.
- [9] Sassoon Adam, Schoch Bradley, Rhee Peter, Schleck Cathy D, Harmsen William S, Sperling John W, Cofield Robert H. The role of eccentric and offset humeral head variations in total shoulder arthroplasty. *J Shoulder Elb Surg* 2013;22:886–93.
- [10] Sabesan Vani J, Callanan Mark, Youderian Ari, Iannotti Joseph P. 3D ct assessment of the relationship between humeral head alignment and glenoid retroversion in glenohumeral osteoarthritis. *J Bone Jt Surg Am* 2014;96:e64.
- [11] Kohles SS, Gregorczyk KN, Phillips TC, Brody LT, Orwin JF, Vanderby R. Concentric and eccentric shoulder rehabilitation biomechanics. *Proc Inst Mech Eng Part H: J Eng Med* 2007;221(3):237–49.
- [12] Buck Florian M, Jost Bernhard, Hodler Juerg. Shoulder arthroplasty. *Eur Radiol* 2008;18(12):2937–48.
- [13] Kim Hyun-Min Mike, Chacon Alexander C, Andrews Seth H, Roush Evan P, Cho Edward, Conaway William K, Kunselman Allen R, Lewis Gregory S. Biomechanical benefits of anterior offsetting of humeral head component in posteriorly unstable total shoulder arthroplasty: A cadaveric study. *J Orthop Res : Off Publ Orthop Res Soc* 2016;34:666–74.
- [14] Petsavage-Thomas Jonelle. Preoperative planning and postoperative imaging in shoulder arthroplasty. *Semin Musculoskelet Radiol* 2014;18:448–62.
- [15] Klein Benjamin D, Hinz Maximilian, Geyer Stephanie, Scheiderer Bastian, Imhoff Andreas B, Siebenlist Sebastian. A 3-dimensional classification for degenerative glenohumeral arthritis based on humeroscapular alignment. *Orthop J Sport Med* 2022;10(8):23259671221110512.
- [16] Saupé Nadja, Pfirrmann Christian WA, Schmid Marius R, Jost Bernhard, Werner Clément ML, Zanetti Marco. Association between rotator cuff abnormalities and reduced acromiohumeral distance. *Am J Roentgenol* 2006;187(2):376–82.
- [17] Goutallier D, Le Guilloux P, Postel J-M, Radier C, Bernageau J, Zilber S. Acromio humeral distance less than six millimeter: its meaning in full-thickness rotator cuff tear. *Orthop Traumatol: Surg Res* 2011;97(3):246–51.
- [18] Goetti Patrick, Denard Patrick J, Collin Philippe, Ibrahim Mohamed, Maz-zolari Adrien, Lädermann Alexandre. Biomechanics of anatomic and reverse shoulder arthroplasty. *EFORT Open Rev* 2021;6(10):918–31.
- [19] Watters Tyler Steven, Mather 3rd Richard C, Browne James A, Berend Keith R, Lombardi Jr Adolph V, Bolognesi Michael P. Analysis of procedure-related costs and proposed benefits of using patient-specific approach in total knee arthroplasty. *J Surg Orthop Adv* 2011;20(2):112–6.
- [20] Seon Jong-Keun, Park Hyeon-Won, Yoo Seung-Hyun, Song Eun-Kyoo. Assessing the accuracy of patient-specific guides for total knee arthroplasty. *Knee Surg Sport Traumatol Arthrosc* 2016;24(11):3678–83.
- [21] Noble Jr John W, Moore Chris A, Liu Ning. The value of patient-matched instrumentation in total knee arthroplasty. *J Arthroplast* 2012;27(1):153–5.
- [22] Mattei Lorenzo, Pellegrino Pietro, Calò Michel, Bistolfi Alessandro, Castoldi Filippo. Patient specific instrumentation in total knee arthroplasty: a state of the art. *Ann Transl Med* 2016;4(7). 126–126.
- [23] Cerveri Pietro, Sacco Costanza, Olgiati Gianluca, Manzotti Alfonso, Baroni Guido. 2D/3d reconstruction of the distal femur using statistical shape models addressing personalized surgical instruments in knee arthroplasty: A feasibility analysis. *Int J Med Robot + Comput Assist Surg : MRCAS* 2017;13(4).
- [24] Lombardi Jr Adolph V, Berend Keith R. Patient-specific approach in total knee arthroplasty. *Orthop (Online)* 2008;31(9):927.
- [25] Mandolini Marco, Brunzini Agnese, Facco Giulia, Mazzoli Alida, Forcellese Archimede, Gigante Antonio. Comparison of three 3D segmentation software tools for hip surgical planning. *Sensors* 2022;22(14):5242.
- [26] Goud Ajay, Segal Dmitri, Hedayati Pejman, Pan John J, Weissman Barbara N. Radiographic evaluation of the shoulder. *Eur J Radiol* 2008;68(1):2–15.
- [27] Maffulli Nicola, Rodriguez Hugo C, Stone Ian W, Nam Andrew, Song Albert, Gupta Manu, Alvarado Rebecca, Ramon David, Gupta Ashim. Artificial intelligence and machine learning in orthopedic surgery: a systematic review protocol. *J Orthop Surg Res* 2020;15(1):478.
- [28] Ladd Lauren M, Crews Marlee, Maertz Nathan A. Glenohumeral joint instability: A review of anatomy, clinical presentation, and imaging. *Clin Sports Med* 2021;40:585–99.
- [29] Isensee Fabian, Jaeger Paul F, Kohl Simon AA, Petersen Jens, Maier-Hein Klaus H. Nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Commun* 2021;18(2):203–11.

- [30] Wang Guangbin, Han Yaxin. Convolutional neural network for automatically segmenting magnetic resonance images of the shoulder joint. *Comput Methods Programs Biomed* 2021;200:105862.
- [31] Qing Yeoh Pauline Shan, Lai Khin Wee, Goh Siew Li, Hasikin Khairunisa, Hum Yan Chai, Tee Yee Kai, Dhanalakshmi Samiappan. Emergence of deep learning in knee osteoarthritis diagnosis. *Comput Intell Neurosci* 2021;2021:4931437.
- [32] Norman Berk, Padoia Valentina, Majumdar Sharmila. Use of 2d u-net convolutional neural networks for automated cartilage and meniscus segmentation of knee mr imaging data to determine relaxometry and morphometry. *Radiol* 2018;288(1):177–85.
- [33] Ambellan Felix, Tack Alexander, Ehlke Moritz, Zachow Stefan. Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative. *Med Image Anal* 2019;52:109–18.
- [34] Marzorati Davide, Sarti Mattia, Mainardi Luca, Manzotti Alfonso, Cerveri Pietro. Deep 3d convolutional networks to segment bones affected by severe osteoarthritis in ct scans for psi-based knee surgical planning. *IEEE Access* 2020;8:196394–407.
- [35] Rossi Matteo, Marsilio Luca, Mainardi Luca, Manzotti Alfonso, Cerveri Pietro. Cel-unet: Distance weighted maps and multi-scale pyramidal edge extraction for accurate osteoarthritic bone segmentation in ct scans. *Front Signal Process* 2022;2:857313.
- [36] Florkow Mateusz C, Willemsen Koen, Mascarenhas Vasco V, Oei Edwin HG, Stralen Marijn van, Seevinck Peter R. Magnetic resonance imaging versus computed tomography for three-dimensional bone imaging of musculoskeletal pathologies: A review. *J Magn Reson Imaging : JMRI* 2022;56(1):11–34.
- [37] Schnider Eva, Huck Antal, Toranelli Mireille, Rauter Georg, Müller-Gerbl Magdalena, Cattin Philippe C. Improved distinct bone segmentation from upper-body ct using binary-prediction-enhanced multi-class inference. *Int J Comput Assist Radiol Surg* 2022;17:2113–20.
- [38] Thomas Kevin A, Kidziński Łukasz, Halilaj Eni, Fleming Scott L, Venkataraman Guhan R, Oei Edwin HG, Gold Garry E, Delp Scott L. Automated classification of radiographic knee osteoarthritis severity using deep neural networks. *Radiol: Artif Intell* 2020;2(2):e190065.
- [39] Taghizadeh Elham, Truffer Oskar, Becce Fabio, Eminian Sylvain, Gidoin Stacey, Terrier Alexandre, Farron Alain, Büchler Philippe. Deep learning for the rapid automatic quantification and characterization of rotator cuff muscle degeneration from shoulder ct datasets. *Eur Radiol* 2021;31:181–90.
- [40] Wang Ruxin, Chen Shuyuan, Ji Chaojie, Fan Jianping, Li Ye. Boundary-aware context neural network for medical image segmentation. *Med Image Anal* 2022;78:102395.
- [41] Potty Anish G, Potty Ajish SR, Maffulli Nicola, Blumenschein Lucas A, Ganta Deepak, Justin Mistovich R, Fuentes Mario, Denard Patrick J, Sethi Paul M, Shah Anup A, Gupta Ashim. Approaching artificial intelligence in orthopaedics: Predictive analytics and machine learning to prognosticate arthroscopic rotator cuff surgical outcomes. *J Clin Med* 2023;12(6):2369.
- [42] Ogura Takahiro, Le Kiet, Merkely Gergo, Bryant Tim, Minas Tom. A high level of satisfaction after bicompartamental individualized knee arthroplasty with patient-specific implants and instruments. *Knee Surg Sport Traumatol Arthrosc* 2019;27:1487–96.
- [43] Anderl Werner, Pauzenberger Leo, Kölblinger Roman, Kiesselbach Gabriele, Brandl Georg, Laky Brenda, Krieglleder Bernhard, Heuberger Philipp, Schwameis Eva. Patient-specific instrumentation improved mechanical alignment, while early clinical outcome was comparable to conventional instrumentation in tka. *Knee Surg Sport Traumatol Arthrosc* 2016;24:102–11.
- [44] Yokota Futoshi, Okada Toshiyuki, Takao Masaki, Sugano Nobuhiko, Tada Yukio, Tomiyama Noriyuki, Sato Yoshinobu. Automated ct segmentation of diseased hip using hierarchical and conditional statistical shape models. In: *Medical image computing and computer-assisted intervention–MICCAI 2013: 16th international conference, Nagoya, Japan, September 22–26, 2013, proceedings, part II* 16. Springer; 2013, p. 190–7.
- [45] Lamecker Hans, Seebass Martin, Hege Hans-Christian, Deuffhard Peter. A 3d statistical shape model of the pelvic bone for segmentation. In: *Medical imaging 2004: Image processing*. vol. 5370, SPIE; 2004, p. 1341–51.
- [46] Seim Heiko, Kainmueller Dagmar, Heller Markus, Lamecker Hans, Zachow Stefan, Hege Hans-Christian. Automatic segmentation of the pelvic bones from ct data based on a statistical shape model. *VCBM* 2008;8:93–100.
- [47] Poltaretskyi Sergii, Chaoui Jean, Mayya Mohammad, Hamitouche C, Bercik MJ, Boileau Pascal, Walch G. Prediction of the pre-morbid 3d anatomy of the proximal humerus based on statistical shape modelling. *Bone Jt J* 2017;99(7):927–33.
- [48] Mendoza Carlos S, Acha Begoña, Serrano Carmen, Gómez-Cía Tomás. Fast parameter-free region growing segmentation with application to surgical planning. *Mach Vis Appl* 2012;23(1):165–77.
- [49] Sivewright GJ, Elliott PJ. Interactive region and volume growing for segmenting volumes in mr and ct images. *Med Inform* 1994;19(1):71–80.
- [50] Sekiguchi Hiroyuki, Sano Koichi, Yokoyama Tetsuo. Interactive 3-dimensional segmentation method based on region growing method. *Syst Comput Japan* 1994;25(1):88–97.
- [51] Ronneberger Olaf, Fischer Philipp, Brox Thomas. U-net: Convolutional networks for biomedical image segmentation. In: *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III* 18. Springer; 2015, p. 234–41.
- [52] Zhang Xiaolu, Zheng Yiqiang, Bai Xiaoming, Cai Liquan, Wang Liangming, Wu Shiqiang, Ke Qingfeng, Huang Jianlong. Femoral image segmentation based on two-stage convolutional network using 3d-dmfnet and 3d-resunet. *Comput Methods Programs Biomed* 2022;226:107110.
- [53] Klein André, Warszawski Jan, Hillengaß Jens, Maier-Hein Klaus H. Automatic bone segmentation in whole-body ct images. *Int J Comput Assist Radiol Surg* 2019;14(1):21–9.
- [54] Suri Abhinav, Jones Brandon C, Ng Grace, Anabaraony Nancy, Beyrer Patrick, Domi Albi, Choi Grace, Tang Sisi, Terry Ashley, Lechner Thomas, Fathali Iman, Bastin Nikita, Chesnais Helene, Rajapakse Chamith S. A deep learning system for automated, multi-modality 2d segmentation of vertebral bodies and intervertebral discs. *Bone* 2021;149:115972.
- [55] Schmidt David, Ulén Johannes, Enqvist Olof, Persson Erik, Trägårdh Elin, Leander Peter, Edenbrandt Lars. Deep learning takes the pain out of back breaking work - automatic vertebral segmentation and attenuation measurement for osteoporosis. *Clin Imaging* 2022;81:54–9.
- [56] Marsilio Luca, Moglia Andrea, Rossi Matteo, Manzotti Alfonso, Mainardi Luca, Cerveri Pietro. Combined edge loss unet for optimized segmentation in total knee arthroplasty preoperative planning. *Bioengineering* 2023;10(12):1433.
- [57] Uysal Fatih, Hardalaç Fırat, Peker Ozan, Tolunay Tolga, Tokgöz Nil. Classification of shoulder x-ray images with deep learning ensemble models. *Appl Sci* 2021;11(6):2723.
- [58] von Schacky Claudio E, Sohn Jae Ho, Liu Felix, Ozhinsky Eugene, Jungmann Pia M, Nardo Lorenzo, Posadzky Magdalena, Foreman Sarah C, Nevitt Michael C, Link Thomas M, Padoia Valentina. Development and validation of a multitask deep learning model for severity grading of hip osteoarthritis features on radiographs. *Radiol* 2020;295(1):136–45.
- [59] Minelli Marco, Cina Andrea, Galbusera Fabio, Castagna Alessandro, Savevski Victor, Sconfienza Luca Maria. Measuring the critical shoulder angle on radiographs: an accurate and repeatable deep learning model. *Skelet Radiol* 2022;51:1873–8.
- [60] Grauhan Nils F, Niehues Stefan M, Gaudin Robert A, Keller Sarah, Vahldiek Janis L, Adams Lisa C, Bressen Keno K. Deep learning for accurately recognizing common causes of shoulder pain on radiographs. *Skelet Radiol* 2022;51(2):355–62.
- [61] Schneider Uwe, Pedroni Eros, Lomax Antony. The calibration of ct hounsfield units for radiotherapy treatment planning. *Phys Med Biol* 1996;41(1):111–24.
- [62] Elsharkawi Mohammed, Cakir Balkan, Reichel Heiko, Kappe Thomas. Reliability of radiologic glenohumeral osteoarthritis classifications. *J Shoulder Elb Surg* 2013;22(8):1063–7.
- [63] Habermeyer Peter, Magosch Petra, Weiß Christel, Hawi Nael, Lichtenberg Sven, Tauber Mark, Ipach Bastian. Classification of humeral head pathomorphology in primary osteoarthritis: a radiographic and in vivo photographic analysis. *J Shoulder Elb Surg* 2017;26(12):2193–9.
- [64] Traore Boukaye Boubacar, Kamsu-Foguem Bernard, Tangara Fana. Deep convolution neural network for image recognition. *Ecol Inform* 2018;48:257–68.
- [65] Anantharaman Rajaram, Velazquez Matthew, Lee Yugyung. Utilizing mask r-cnn for detection and segmentation of oral diseases. In: *2018 IEEE international conference on bioinformatics and biomedicine. BIBM, IEEE; 2018.*
- [66] Liu Zhi-Qiang, Hu Zi-Jian, Wu Tian-Qiong, Ye Geng-Xin, Tang Yu-Liang, Zeng Zi-Hua, Ouyang Zhong-Min, Li Yuan-Zhe. Bone age recognition based on mask r-cnn using xception regression model. *Front Physiol* 2023;14.
- [67] Wong Victoria, Calivá Francesco, Su Favian, Padoia Valentina, Lansdown Drew. Comparing bone shape models from deep learning processing of magnetic resonance imaging to computed tomography-based models. *JSES Int* 2023;7(5):861–7.
- [68] Kim Sae Hoon, Yoo Hye Jin, Yoon Soon Ho, Kim Yong Tae, Park Sang Joon, Chai Jee Won, Oh Jiseon, Chae Hee Dong. Development of a deep learning-based fully automated segmentation of rotator cuff muscles from clinical mr scans. *Acta Radiol* 2024;02841851241262325.
- [69] Liu Yung-Chun, Lin Yung-Chieh, Tsai Pei-Yin, Iwata Osuke, Chuang Chuew-Chuen, Huang Yu-Han, Tsai Yi-Shan, Sun Yung-Nien. Convolutional neural network-based humerus segmentation and application to bone mineral density estimation from chest x-ray images of critical infants. *Diagnostics* 2020;10(12):1028.
- [70] Noguchi Shunjiro, Nishio Mizuho, Yakami Masahiro, Nakagomi Keita, Togashi Kaori. Bone segmentation on whole-body ct using convolutional neural network with novel data augmentation techniques. *Comput Biol Med* 2020;121:103767.
- [71] Huang Yichen, Robinson Dale L, Pitocchi Jonathan, Sin Lee Peter Vee, Ackland David C. Glenohumeral joint reconstruction using statistical shape modeling. *Biomech Model Mechanobiol* 2022;21(1):249–59.
- [72] Zhao Qingqing, Feng Quanlong, Zhang Jianlun, Xu Jingxu, Wu Zifeng, Huang Chencui, Yuan Huishu. Glenoid segmentation from computed tomography scans based on a 2-stage deep learning model for glenoid bone loss evaluation. *J Shoulder Elb Surg* 2023;32(12):e624–35.

- [73] Wang Guangbin, Han Yaxin. Convolutional neural network for automatically segmenting magnetic resonance images of the shoulder joint. *Comput Methods Programs Biomed* 2021;200:105862.
- [74] Sezer Aysun, Sezer Hasan Basri. Convolutional neural network based diagnosis of bone pathologies of proximal humerus. *Neurocomputing* 2020;392:124–31.
- [75] Moein Shariatnia M, Ramazanian Taghi, Sanchez-Sotelo Joaquin, Kremers Hilal Maradit. Deep learning model for measurement of shoulder critical angle and acromion index on shoulder radiographs. *JSES Rev Rep Tech* 2022;2(3):297–301.
- [76] Rüdén Christian Von, Augat Peter. Failure of fracture fixation in osteoporotic bone. *Injury* 2016;47:S3–10.
- [77] Carl Michael, Lall Kaustubh, Pai Darren, Chang Eric Y, Statum Sheronda, Brau Anja, Chung Christine B, Fung Maggie, Bae Won C. Shoulder bone segmentation with deeplab and u-net. *Osteology* 2024;4(2):98–110.
- [78] Zhao Qingqing, Feng Quanlong, Zhang Jianlun, Xu Jingxu, Wu Zi xuan, Huang Chencui, Yuan Hui zong. Glenoid segmentation from ct scans based on a two-stage deep learning model for glenoid bone loss evaluation. *J Shoulder Elb Surg* 2023;32(12):e624–35.