Check for updates

Proceedings of the ASME 2025
International Design Engineering Technical Conferences and
Computers and Information in Engineering Conference
IDETC-CIE2025
August 17-20, 2025, Anaheim, CA

DETC2025-168564

Downloaded from http://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/IDETC-CIE2025/89213/V02BT02A022/7557339/v02bt02a022-detc2025-168564.pdf by Universita Degli Studi Di Bergamo user on 31 October 2025

DEEP LEARNING APPROACHES FOR AUTOMATIC SEGMENTATION OF SHOULDER BONES USING MULTI-SEQUENCE MRI

Anna Ghidotti*, Paola Pinetti, Daniele Regazzoni, Caterina Rizzi

University of Bergamo, Bergamo, Italy

ABSTRACT

Shoulder pain is a common complaint in clinical practice, often requiring MRI for diagnosis and treatment planning. Accurate segmentation of shoulder bones from MRI scans is crucial for 3D anatomical assessment. Deep learning (DL) models have shown promise in automating segmentation, yet previous studies have primarily focused on the humerus and scapula, with limited investigation of the clavicle. This study compares DL models for automatic segmentation of the humerus, scapula, and clavicle from MRI. We explore multiple MRI sequences, including PDW SPAIR (coronal and sagittal) and PD-TSE, to assess their impact on segmentation accuracy. DL architectures, including 3D U-Net and 3D Attention U-Net, were trained, and their hyperparameters were optimized. Performance was assessed using the Dice Similarity Coefficient (DSC), Hausdorff Distance (HD), and volume variation. Additionally, morphological measurements were analyzed to quantify segmentation errors. The best segmentation performance was achieved using a combination of coronal and sagittal PDW SPAIR sequences with Attention U-Net, yielding DSC values of 0.878 (clavicle), 0.843 (scapula), and 0.931 (humerus), and minimal HD. Our study highlights the importance of multimodal MRI input and model optimization for shoulder segmentation. The integration of morphological assessments enhances clinical applicability, providing reliable automated segmentation for surgical planning and medical training.

Keywords: Deep Learning; Convolutional Neural Network; Image Segmentation; Shoulder; Magnetic Resonance Imaging

1. INTRODUCTION

Shoulder pain is the third most common reason for consultation in primary care, often requiring imaging for diagnosis and treatment planning [1]. Magnetic Resonance Imaging (MRI) is an important tool in evaluating shoulder pathologies because it provides high-resolution soft tissue contrast without ionizing radiation [2]. Traditionally, morphological assessment has relied on two-dimensional (2D) analysis of selected MRI slices or radiographic projections. However, these approaches present limitations: single-slice evaluations may not fully represent the complex three-dimensional (3D) anatomy of the shoulder, and radiographs are prone to projection errors that can distort anatomical relationships [3]. To overcome these limitations, 3D anatomical analysis, based on segmentation of computed tomography (CT) or MRI scans, has been explored in smaller studies [4-5]. However, manual segmentation of the shoulder anatomy is time-consuming and limits large-scale applications in clinical practice.

Deep learning (DL), particularly convolutional neural networks (CNNs), has emerged as a powerful tool for automatic medical image segmentation [6-9]. Several studies have applied DL models to segment shoulder structures, focusing mainly on the humerus and scapula. Rodrigues et al.[10] successfully employed CNNs for shoulder bone segmentation on MRI scans, enabling 3D reconstructions for assessing normal anatomy and glenoid bone loss. Similarly, Satir et al.[2] developed an automatic scapula segmentation method using CT images, though it was limited to scapular structures without considering the humerus or clavicle. Hess et al.[11] introduced a deep learning model for MRI-based segmentation of the humerus and scapula. However, their study highlighted the challenge of decreased accuracy in cases with significant anatomical variations. Marsilio et al.[12] applied DL to CT-based segmentation of the proximal humerus and scapula. Their method was particularly effective for preoperative planning in shoulder arthroplasty but was limited by its focus on CT images, restricting its applicability to MRI-based workflows. Despite these advancements, several research gaps remain. Firstly, most

* Corresponding author: anna.ghidotti@unibg.it

Copyright © 2025 by ASME; reuse license CC-BY 4.0

studies have focused on the segmentation of the scapula and humerus, with limited research on the clavicle, an essential component of shoulder biomechanics. In addition, many segmentation studies do not assess the clinical applicability of their models, limiting their integration into routine medical practice.

MRI provides a variety of acquisition sequences, each designed to optimize visualization of specific tissue contrasts and anatomical structures [13]. Among these, proton density-weighted (PDW) sequences are widely used in musculoskeletal imaging due to their ability to provide high-resolution soft tissue contrast while preserving fine anatomical details [14]. Two key PDW sequences for shoulder imaging are PDW SPAIR (Spectral Attenuated Inversion Recovery) and PD-TSE (Turbo Spin Echo), each with distinct advantages. PDW SPAIR employs an inversion pulse to suppress fat signal, enhancing the visibility of subtle pathological changes such as rotator cuff tears, joint effusions, and bone marrow edema. By reducing signal contamination from surrounding adipose tissue, it improves lesion conspicuity, making it a preferred sequence for detecting inflammatory and degenerative conditions. On the other hand, PD-TSE is a proton density-weighted sequence acquired without fat suppression, providing high signal-to-noise ratio and superior anatomical detail. PD-TSE is particularly beneficial for evaluating fibrocartilaginous structures such as the glenoid labrum and articular cartilage, where fat suppression may introduce signal heterogeneity. The combination of PDW SPAIR and PD-TSE allows for a comprehensive assessment of shoulder pathology, ensuring both anatomical fidelity and enhanced pathological contrast.

This study aims to compare deep learning models optimized for the segmentation of the humerus, scapula, and clavicle in MRI images, leveraging multi-sequence MRI inputs to enhance anatomical representation and segmentation accuracy. To achieve this, different combinations of MRI sequences, including PDW SPAIR in coronal and sagittal orientations as well as PD-TSE, will be explored to assess their impact on segmentation performance.

Deep learning architectures will be evaluated, including 3D U-Net [15] and 3D Attention U-Net, which are widely used in medical image segmentation due to their ability to capture anatomical structures. To identify the most effective model, hyperparameter optimization will be performed, focusing on key parameters such as the number of training epochs, learning rate, and batch size. This optimization process aims to enhance segmentation accuracy and generalizability across different imaging conditions. By systematically analyzing the impact of sequence selection, network architecture, and hyperparameter tuning, this study seeks to establish an optimal deep learning framework for automated MRI-based shoulder segmentation, facilitating its potential application in clinical practice.

2. PROPOSED METHODOLOGY

Figure 1 illustrates the workflow of the proposed methodology, beginning with MRI acquisition and the selection of different sequence combinations for training DL models. Following data preprocessing, DL architectures were selected (3D U-Net and 3D Attention U-Net). Hyperparameter optimization was performed to refine model performance, adjusting parameters such as the number of training epochs, learning rate, and batch size for each network. Finally, the trained models were evaluated using quantitative metrics, including the Dice Similarity Coefficient, Hausdorff Distance, percentage volume variation, and morphological parameter error, to assess segmentation accuracy and robustness.

FIGURE 1: OVERVIEW OF THE APPLIED METHODOLOGY

As part of this study, MONAI (Medical Open Network for Artificial Intelligence) [16] and 3D Slicer [17] were utilized. MONAI is a widely used open-source framework designed for DL applications in medical imaging. Specifically, MONAI Label, an open-source tool for image labeling and interactive learning, was employed to generate annotated datasets and develop AI-driven annotation models for clinical assessment. This software enables researchers to construct, train, and deploy labeling applications in a serverless environment. Additionally, MONAI Label can be integrated with 3D Slicer, an open-source, cross-platform software package widely used for biomedical imaging research.

2.1 MRI acquisition and preparation

This study utilized multiple MRI protocols to assess the impact of different input configurations on DL-based segmentation. The dataset consisted of 33 MRI scans, divided into four input configurations: (1) PDW SPAIR coronal, (2) PDW SPAIR sagittal, (3) a combined input of PDW SPAIR coronal and sagittal, and (4) a multimodal approach integrating all PDW SPAIR and PD-TSE sequences. These configurations were selected to analyze how variations in spatial resolution and contrast mechanisms influence the neural network's ability to segment anatomical structures accurately. The differences between these sequences are illustrated in Figure 2. The volumes used in this study had a slice thickness of 3.3 mm in 55% of the dataset and 4.4 mm in 45% of the dataset on the acquisition plane.

To enhance model performance and improve generalization, data augmentation techniques were applied, increasing both the quantity and diversity of training samples. Specifically, rotation-based augmentation was implemented, resulting in the following dataset expansions:

(1) PDW SPAIR coronal: expanded from 25 to 75 volumes
(2) PDW SPAIR sagittal: expanded from 25 to 75 volumes

Downloaded from http://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/IDETC-CIE2025/89213/V02BT02A022/7557339/v02bt02a022-detc2025-168564.pdf by Universita Degli Studi Di Bergamo user on 31 October 2025

Copyright © 2025 by ASME; reuse license CC-BY 4.0

(3) PDW SPAIR coronal and sagittal (combined input): expanded from 50 to 150 volumes
(4) Multimodal approach (all PDW SPAIR and PD-TSE sequences): expanded from 66 to 198 volumes.

Following augmentation, the dataset was split into three subsets: training, validation, and testing. The training set was used to optimize the deep learning models, while the validation set helped fine-tune hyperparameters and prevent overfitting. Finally, the testing set was reserved for independent evaluation, ensuring an unbiased assessment of segmentation performance.

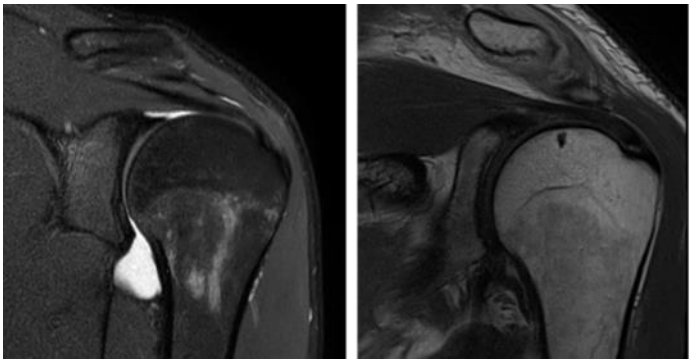

FIGURE 2: MR IMAGES OF A SHOULDER IN SAGITTAL PLANE USING PDW-SPAIR AND PD-TSE SEQUENCES, RESPECTIVELY

2.2 DL models selection and fine tuning

To address the segmentation task, DL architectures were employed. Specifically, 3D U-Net and 3D Attention U-Net were utilized to process volumetric data, in contrast to traditional 2D networks that operate on individual slices [18]. In the 3D U-Net architecture, there are five layers, comprising four downsampling layers, a bottom layer, and four corresponding upsampling layers. It consists of a series of convolutional layers, with the number of channels increasing as the network deepens from 16 to 256. In each layer, the network applies 3D convolutions with filter sizes of 3x3x3, followed by batch normalization and ReLU activation. A step size of 2 is used in each dimension of downsampling, effectively halving the spatial dimensions. By using skip links, the encoders and decoders are connected, which facilitates the reuse of features and enhances the flow of gradients during training. By combining two residual convolutional units, each encoder block maintains the ability to learn both local and global features, and mitigates vanishing gradient effects [15]. The second model, Attention U-Net is an extension of U-Net that integrates attention mechanisms, enabling the model to focus on the most relevant areas of the image and improve segmentation performance, particularly in complex regions [19].

To ensure optimal performance, a systematic approach to hyperparameter tuning was adopted. Several key hyperparameters were adjusted, including the number of epochs, learning rate, and batch size. The number of epochs was varied to balance model convergence with the risk of overfitting. The learning rate was carefully tuned to maintain stable gradient updates throughout training, ensuring that the model would converge without diverging. Batch size was selected to optimize memory usage and computational efficiency while maintaining stable training dynamics. The hyperparameter space was explored, allowing for the identification of the most effective parameter combinations for each model.

2.3 Evaluation metrics

The performance of the DL models was assessed using a range of evaluation metrics to provide a comprehensive understanding of segmentation accuracy and anatomical consistency. The Dice Similarity Coefficient (DSC) was used to quantify the spatial overlap between the predicted segmentations and the ground truth, with higher DSC values indicating better segmentation performance. The DSC ranges from 0 to 1, where 1 represents perfect overlap. The Hausdorff Distance (HD) was employed to measure the maximum surface distance between the predicted and reference segmentations, which helps capture any irregularities in boundary alignment [8][20]. In this study, we considered the average HD as a robust metric for evaluating segmentation accuracy. Additionally, the percentage volume variation was calculated to assess the relative difference in the segmented volume compared to the ground truth, providing insights into the accuracy of volume estimation.

Finally, morphological parameters were evaluated to determine the accuracy of segmented models in relation to anatomical structures. This analysis helped to identify the areas that show the most similarity and areas where discrepancies occur. These morphological parameters were identified in the literature [21-25]. Among the most relevant parameters, five main measures were identified for the humerus [21-23]: the transverse diameter of the humeral head (HHTD), the vertical diameter of the humeral head (HHVD), the diameter of the greater tuberosity, the diameter of the lesser tuberosity and the HB distance (distance between the most proximal point of the humeral head and the distal point on the surgical neck). The morphological parameters measured on 3D model of the humerus are shown in Figure 3.

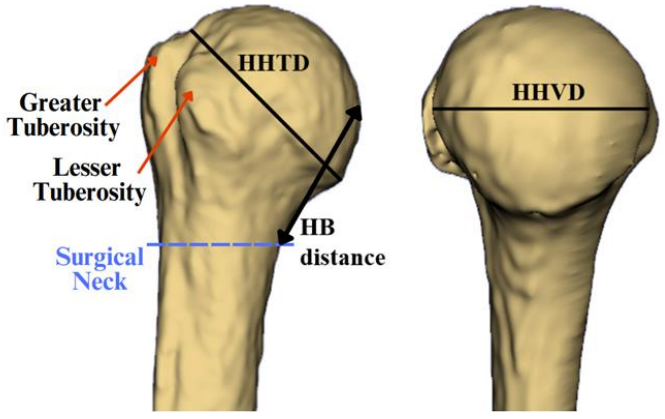

FIGURE 3: MORPHOLOGICAL PARAMETERS MEASURED ON 3D MODEL OF THE HUMERUS

Downloaded from http://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/IDETC-CIE2025/89213/V02BT02A022/7557339/v02bt02a022-detc2025-168564.pdf by Universita Degli Studi Di Bergamo user on 31 October 2025

Copyright © 2025 by ASME; reuse license CC-BY 4.0

Regarding the clavicle, three morphological parameters were identified (Figure 4): achromial diameter (AD), diameter of the conoid tubercle (CD) and achromioclavicular space (AC) [24].

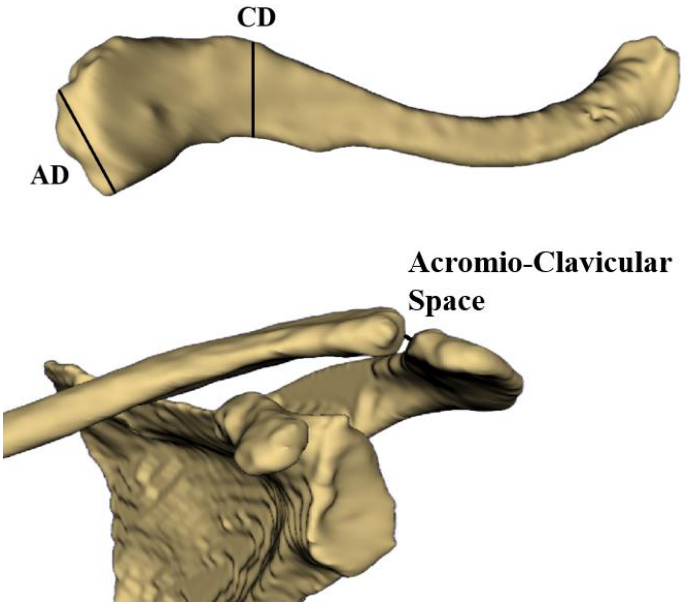

FIGURE 4: MORPHOLOGICAL PARAMETERS RELATED TO THE CLAVICLE MEASURED ON 3D MODELS

The analysis of the morphology of the scapula focuses on four parameters: the Width and Height of the Glenoid Fossa, the Coracoclavicular Distance and the Length of the Coracoid Process, as shown in Figure 5 [25]. The percentage error in morphological parameters was used to quantify the discrepancies in the extraction of anatomical shape features, providing an understanding of the model's ability to accurately represent the structure of the segmented objects.

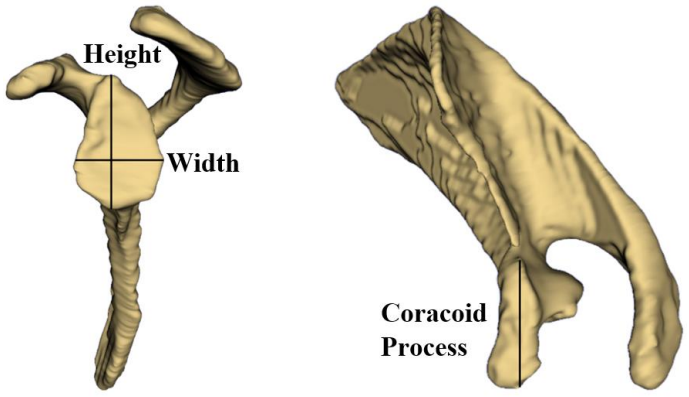

FIGURE 5: MORPHOLOGICAL PARAMETERS MEASURED ON 3D MODEL OF THE SCAPULA

3. RESULTS

The segmentation performance of the DL models was evaluated using different MRI sequence combinations: (1) PDW SPAIR coronal, (2) PDW SPAIR sagittal, (3) a combined input of PDW SPAIR coronal and sagittal, and (4) a multimodal approach incorporating PDW SPAIR and PD-TSE. The evaluation metrics included DSC, HD, and Volume Variation(%), providing insights into segmentation accuracy across the clavicle, scapula, and humerus.

The dataset size, after data augmentation, varied across the configurations, with the single-sequence datasets including 48 patients for training, 12 for validation and 5 for testing, while the combined coronal-sagittal input included 96 patients for training, 24 for validation and 5 test cases. The multimodal approach, incorporating PDW SPAIR and PD-TSE sequences, had the largest dataset, with 148 patients for training, 38 for validation and 6 patients for testing. Each configuration was optimized using DL architectures.

All dataset configurations achieved better performance with Attention U-Net compared to U-Net. The most significant difference between the two DL models (Attention U-Net and U-Net) was observed in clavicle segmentation, where the DSC for the PDW SPAIR sagittal input improved by 12%. The DSC for the scapula segmentation with the combined PDW SPAIR coronal-sagittal input showed a 7% increase. For the remaining configurations, DSC variations were below 3.5%, indicating a relatively minor performance difference between the two architectures.

The best segmentation performance was obtained using the combined coronal and sagittal PDW SPAIR sequences. This configuration achieved the highest DSC values, with 0.878 for the clavicle, 0.843 for the scapula, and 0.931 for the humerus. It also produced the lowest HD values, with 0.40 mm for the clavicle, 0.77 mm for the scapula, and 0.53 mm for the humerus. The multimodal approach using PDW SPAIR and PD-TSE resulted in comparable DSC values but had slightly higher HD, indicating that the addition of the PD-TSE sequence did not significantly enhance segmentation accuracy. Among the single-sequence inputs, PDW SPAIR coronal provided better segmentation for the clavicle and scapula, while the sagittal sequence performed slightly better for the humerus, achieving a DSC of 0.918. The volume variation analysis does not indicate a clear superiority of one sequence over the others. Using the PDW SPAIR sequence in a combined coronal and sagittal approach, the clavicle exhibits the lowest variation, while higher variation is observed for the scapula and humerus. Table 1 presents the segmentation performance in terms of DSC, HD and volume variation for the three bones across the different MRI sequence combinations.

To further assess segmentation accuracy, morphological parameters were extracted and compared between the ground truth and the automated segmentation. This analysis was conducted for the best-performing configuration in terms of DSC and HD, which utilized the combined PDW SPAIR coronal and sagittal sequences. By evaluating these parameters, we aimed to quantify segmentation errors and provide a more detailed

Downloaded from http://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/IDETC-CIE2025/89213/V02BT02A022/7557339/v02bt02a022-detc2025-168564.pdf by Universita Degli Studi Di Bergamo user on 31 October 2025

Copyright © 2025 by ASME; reuse license CC-BY 4.0

assessment of the model's reliability. This approach enhances the clinical interpretability of the results, offering objective metrics to assess the accuracy of automated segmentation in real-world applications. The results of this analysis are presented in Table 2.

TABLE 1: SEGMENTATION PERFORMANCE ACROSS DIFFERENT MRI SEQUENCE COMBINATIONS, INCLUDING DSC, HD, AND VOLUME VARIATION (%) FOR THE CLAVICLE, SCAPULA, AND HUMERUS. THE BOLDED VALUES REPRESENT THE BEST PERFORMANT CONFIGURATION

		Clavicle	Scapula	Humerus
DSC	1 PDW SPAIR coronal	0.832	0.768	0.929
	2 PDW SPAIR sagittal	0.787	0.763	0.918
	3 PDW SPAIR coronal and sagittal	**0.878**	**0.843**	**0.931**
	4 PDW SPAIR and PD-TSE	0.879	0.822	0.927
HD	1 PDW SPAIR coronal	0.81	1.32	0.60
	2 PDW SPAIR sagittal	0.76	2.31	0.74
	3 PDW SPAIR coronal and sagittal	**0.40**	**0.77**	**0.53**
	4 PDW SPAIR and PD-TSE	0.50	0.89	0.63
Volume variation (%)	1 PDW SPAIR coronal	-9.50	7.48	2.42
	2 PDW SPAIR sagittal	3.92	-12.46	10.74
	3 PDW SPAIR coronal and sagittal	-0.12	11.12	4.42
	4 PDW SPAIR and PD-TSE	12.94	10.28	5.54

TABLE 2: MORPHOLOGICAL PARAMETER COMPARISON BETWEEN THE GROUND TRUTH AND THE AUTOMATED SEGMENTATION, UTILIZING PDW SPAIR CORONAL AND SAGITTAL SEQUENCES

Parameters	Clavicle	Scapula	Humerus
HD (mm)	0.40	0.77	0.53
DSC	0.878	0.843	0.931
Volume Variation (%)	-0.12	11.12	4.42
Maximun Humeral Head Diameter (%)	N/A	N/A	2.02
Transverse Humeral Head Diameter (%)	N/A	N/A	-0.38
HB distance (%)	N/A	N/A	1.08
Achromial Diameter (%)	7.52	N/A	N/A
Conoid Tubercle Diameter (%)	0.80	N/A	N/A
Acromio-Clavicular Space (%)	-7.27	N/A	N/A
Glenoid Fossa Width (%)	N/A	9.52	N/A
Glenoid Fossa Height (%)	N/A	-0.13	N/A
Coracoclavicular Distance (%)	N/A	3.36	N/A
Coracoid Length (%)	N/A	1.45	N/A

The analysis showed that the clavicle had the lowest HD value (0.40 mm) and an intermediate DSC (0.878), indicating precise segmentation with minimal surface discrepancies. Among the morphological parameters, the Conoid Tubercle Diameter exhibited low variation (0.80%), while the Acromio-Clavicular Space showed a higher percentage difference (7.52%), likely due to its inherently small anatomical dimensions (1–7 mm), where even minor discrepancies result in large relative variations. The scapula presented the highest HD (0.77 mm) and the lowest DSC (0.843), suggesting greater segmentation challenges. The largest variation was observed in the Glenoid Fossa Width (9.52%), highlighting difficulties in accurately defining the glenoid margins. Other lower discrepancies included the Coracoclavicular Distance (3.36%) and Coracoid Length (1.45%), although these were less pronounced. The humerus achieved the highest DSC (0.931) and a relatively low HD (0.53 mm), confirming strong agreement between the automated and ground-truth segmentations. Morphological measurements further supported these findings, with the Maximum and Transverse Humeral Head Diameters showing minimal variations (2.02% and 0.38%, respectively), indicating the model's ability to capture the humeral head's shape with high accuracy. These results indicate that the humerus was segmented more accurately than the scapula and clavicle, likely due to its relatively simple and regular anatomical structure. Unlike the scapula, which has complex contours, thin regions, and sharp edges, the humerus is a more voluminous and smooth bone, making it easier for the deep learning model to reconstruct. Similarly, the clavicle's elongated and curved shape introduces challenges that are less pronounced in the humerus.

The HD was calculated by measuring the maximum Euclidean distance from each point on the automated segmentation mesh to the nearest point on the ground truth mesh, with distances visualized using a color map to highlight spatial discrepancies, as shown for a representative test patient in Figure 6. The

Downloaded from http://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/IDETC-CIE2025/89213/V02BT02A022/7557339/v02bt02a022-detc2025-168564.pdf by Universita Degli Studi Di Bergamo user on 31 October 2025

Copyright © 2025 by ASME; reuse license CC-BY 4.0

humerus exhibited minimal discrepancies, while the largest deviations were observed in the scapula, particularly at the glenoid margins, consistent with the quantitative findings. In Figure 6, a stepwise appearance is clearly visible. This effect is caused by the slice thickness used during image acquisition. When slices are relatively thick (e.g., 3-4 mm), interpolation between them can create discontinuities in the reconstructed 3D structure, leading to a "stair-step" artifact, particularly in the sagittal plane. This phenomenon is more pronounced in regions with curved anatomical structures, such as the scapula and clavicle.

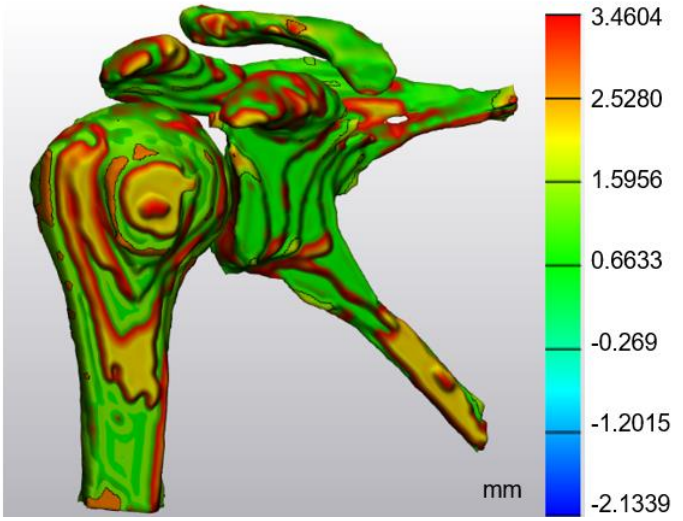

FIGURE 6: COLOR-CODED DISTANCE MAP ILLUSTRATING THE DIFFERENCES BETWEEN THE AUTOMATED SEGMENTATION AND THE GROUND TRUTH FOR A REPRESENTATIVE TEST PATIENT

4. DISCUSSION

Our study demonstrates that DL model performance in shoulder bone segmentation depends on both the input data and the algorithm applied. By evaluating different MRI sequences, DL architectures, and hyperparameter configurations, we identified the optimal approach for accurate segmentation. The hyperparameter optimization was guided by monitoring the validation loss and the mean DSC across all target anatomical structures. The best results were achieved using the combined coronal and sagittal PDW SPAIR sequences, particularly with the 3D Attention U-Net model optimized with 150 epochs, a learning rate of 10^{-6}, and a batch size of 4. This configuration produced the highest DSC values (0.878 for the clavicle, 0.843 for the scapula, and 0.931 for the humerus) alongside the lowest HD values.

Attention U-Net outperforms traditional U-Net for shoulder bone segmentation from MRI due to its ability to focus on the most relevant anatomical structures while suppressing less important background information. This is particularly beneficial in MRI shoulder segmentation, where bones often appear with low contrast against surrounding soft tissues, and artifacts or noise may interfere with boundary detection.

When compared to existing literature, our results align with previous studies. Similarly, Hess et al.[11] achieved DSC values of 0.91 for the humerus and 0.92 for the scapula. Our humerus segmentation (DSC = 0.931) surpasses this study in accuracy, while our scapula segmentation (DSC = 0.843) is slightly lower in DSC. Additionally, our results exceed those of Dai et al.[26], who reported a DSC of 0.93 for the shoulder joint in a sample of 52 swimming patients, as well as those of Mu et al.[27], who achieved a DSC of 0.91. Rodrigues et al.[10] reported a DSC of 0.95 for the humerus and 0.86 for the glenoid, using convolutional neural networks for MRI-based segmentation. However, their study included a small sample size to assess the accuracy of fully automated 3D models. In CT-based studies, Satir et al.[2] achieved a DSC of 0.97 for scapula segmentation; however, the average HD values were 2.3 mm for normal shoulders and 5.1 mm for pathological cases. Similarly, Marsilio et al.[12] reported median HD values of 1.48 mm for the humerus and scapula, suggesting that our method may offer improved surface accuracy. Though direct comparisons between MRI and CT are complex due to modality differences, our approach demonstrates competitive performance. Recognizing bones in CT is easier because there is better contrast between bones and surrounding tissues.

Our study also identifies research gaps in the literature. While previous studies primarily focused on the humerus and scapula, our inclusion of the clavicle represents a novel contribution. Additionally, clinical applicability remains under-evaluated, with only a limited number of studies translating segmentation results into practical medical applications. To address this limitation, we incorporated morphological measurements to quantify segmentation errors. This approach provides clinicians with objective data to assess the reliability of the segmentation and better understand where errors occur. By offering detailed error quantification, our method enhances trust in automated segmentation, supporting its integration into clinical workflows.

The varying number of images across different MRI protocols likely influenced the segmentation performance. The combined coronal-sagittal input demonstrated the best results, which may be partially attributed to the larger dataset size (120 patients for training and validation, compared to 60 in the single-sequence protocols). A greater number of images provides more diverse anatomical variations, helping the deep learning model generalize better. Additionally, the multimodal approach, which incorporated both PDW SPAIR and PD-TSE sequences, had the largest dataset (186 patients for training and validation). However, this configuration did not outperform the combined sagittal-coronal approach. This may suggest that while multimodal information is valuable, the number of training images still plays a crucial role in model performance. If more PD-TSE images had been available, results might have improved further. Given the data available in this study, the combined sagittal-coronal approach emerged as the most effective configuration. Future studies should investigate whether

Downloaded from http://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/IDETC-CIE2025/89213/V02BT02A022/7557339/v02bt02a022-detc2025-168564.pdf by Universita Degli Studi Di Bergamo user on 31 October 2025

Copyright © 2025 by ASME; reuse license CC-BY 4.0

increasing the number of PD-TSE images or exploring alternative multimodal strategies could enhance segmentation accuracy. In addition, the sample size will be increased to include more patients in the test set to increase the statistical power of the results.

A critical limitation of conventional 2D multislice MRI acquisitions is anisotropic resolution, where voxel dimensions differ across imaging planes. This variation affects image quality and introduces artifacts when visualizing structures in planes orthogonal to the acquisition plane [28]. For example, in a coronal PDW SPAIR sequence, the coronal plane exhibits high resolution, while the sagittal and axial views may show step-like discontinuities due to lower through-plane resolution. Conversely, a sagittal PDW SPAIR sequence provides optimal resolution in the sagittal plane but introduces artifacts in coronal and axial reconstructions. These limitations arise because 2D multislice acquisitions have thicker slice spacing in the through-plane direction, reducing spatial accuracy in reconstructed views.

The slice thickness of the MRI volumes used in this study ranged from 3 mm to 4 mm, which may have influenced the segmentation performance. Thicker slices can introduce partial volume effects, reducing spatial resolution and making it more challenging for the deep learning model to accurately delineate fine anatomical structures. This limitation is particularly evident in the sagittal plane, where a stepwise appearance ("stair-step" artifact) may occur, potentially affecting the segmentation of curved or complex bone surfaces such as the scapula and clavicle. A more uniform and thinner slice thickness could improve segmentation accuracy by providing higher-resolution input data. Future studies could explore interpolation techniques or specialized deep learning approaches to mitigate the effects of varying slice thickness and enhance segmentation performance.

5. CONCLUSION

Our findings confirm the effectiveness of deep learning for MRI-based shoulder bone segmentation. The realistic segmentations obtained in our study may be valuable for surgical planning, patient education, and medical training. Future research should focus on multimodal imaging, larger datasets, and real-world clinical applications to further enhance segmentation reliability and clinical impact.

REFERENCES

[1] Mitchell C, Adebajo A, Hay E, Carr A. Shoulder pain: diagnosis and management in primary care. BMJ 2005;331:1124–8. https://doi.org/10.1136/BMJ.331.7525.1124.

[2] Satir OB, Eghbali P, Becce F, Goetti P, Meylan A, Rothenbühler K, et al. Automatic quantification of scapular and glenoid morphology from CT scans using deep learning. Eur J Radiol 2024;177. https://doi.org/10.1016/J.EJRAD.2024.111588.

[3] Suter T, Gerber Popp A, Zhang Y, Zhang C, Tashjian RZ, Henninger HB. The influence of radiographic viewing perspective and demographics on the critical shoulder angle. J Shoulder Elbow Surg 2015;24:e149–58. https://doi.org/10.1016/J.JSE.2014.10.021.

[4] Suter T, Henninger HB, Zhang Y, Wylie JD, Tashjian RZ. Comparison of measurements of the glenopolar angle in 3D CT reconstructions of the scapula and 2D plain radiographic views. Bone and Joint Journal 2016;98-B:1510–6. https://doi.org/10.1302/0301-620X.98B11.37800/LETTERTOEDITOR.

[5] Mah D, Chamoli U, Smith GC. Usefulness of computed tomography based three-dimensional reconstructions to assess the critical shoulder angle. World J Orthop 2021;12:301. https://doi.org/10.5312/WJO.V12.I5.301.

[6] Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X. A Review of Deep Learning based Methods for Medical Image Multi-Organ Segmentation. Phys Med 2021;85:107. https://doi.org/10.1016/J.EJMP.2021.05.003.

[7] Li M, Jiang Y, Zhang Y, Zhu H. Medical image analysis using deep learning algorithms. Front Public Health 2023;11:1273253. https://doi.org/10.3389/FPUBH.2023.1273253.

[8] Ghidotti A, Vitali A, Regazzoni D, Cohen MW, Rizzi C. Comparative Analysis of Convolutional Neural Network Architectures for Automated Knee Segmentation in Medical Imaging: A Performance Evaluation. J Comput Inf Sci Eng 2024;24. https://doi.org/10.1115/1.4064450/1193886.

[9] Aibinder D, Weisberg M, Ghidotti A, Weiss Cohen M. Enhanced Attention Res-Unet for Segmentation of Knee Bones. Mathematics 2024, Vol 12, Page 2284 2024;12:2284. https://doi.org/10.3390/MATH12142284.

[10] Rodrigues TC, Deniz CM, Alaia EF, Gorelik N, Babb JS, Dublin J, et al. Three-dimensional MRI Bone Models of the Glenohumeral Joint Using Deep Learning: Evaluation of Normal Anatomy and Glenoid Bone Loss. Radiol Artif Intell 2020;2:1–9. https://doi.org/10.1148/RYAI.2020190116.

[11] Hess H, Ruckli AC, Bürki F, Gerber N, Menzemer J, Burger J, et al. Deep-Learning-Based Segmentation of the Shoulder from MRI with Inference Accuracy Prediction. Diagnostics 2023, Vol 13, Page 1668 2023;13:1668. https://doi.org/10.3390/DIAGNOSTICS13101668.

[12] Marsilio L, Marzorati D, Rossi M, Moglia A, Mainardi L, Manzotti A, et al. Cascade learning in multi-task encoder-decoder networks for concurrent bone segmentation and glenohumeral joint assessment in shoulder CT scans 2024.

[13] Florkow MC, Willemsen K, Mascarenhas V V., Oei EHG, van Stralen M, Seevinck PR. Magnetic Resonance Imaging Versus Computed Tomography for Three-Dimensional Bone Imaging of Musculoskeletal Pathologies: A Review. Journal of Magnetic Resonance Imaging 2022;56:11. https://doi.org/10.1002/JMRI.28067.

Downloaded from http://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/IDETC-CIE2025/89213/V02BT02A022/7557339/v02bt02a022-detc2025-168564.pdf by Universita Degli Studi Di Bergamo user on 31 October 2025

Copyright © 2025 by ASME; reuse license CC-BY 4.0

[14] Ashir A, Lombardi A, Jerban S, Ma Y, Du J, Chang EY. Magnetic resonance imaging of the shoulder. Pol J Radiol 2020;85:e420. https://doi.org/10.5114/PJR.2020.98394.
[15] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2015;9351:234–41. https://doi.org/10.1007/978-3-319-24574-4_28.
[16] MONAI - Home n.d. https://monai.io/.
[17] 3D Slicer image computing platform | 3D Slicer n.d. https://www.slicer.org/.
[18] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: Learning dense volumetric segmentation from sparse annotation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2016;9901 LNCS:424–32. https://doi.org/10.1007/978-3-319-46723-8_49/TABLES/3.
[19] Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, et al. Attention gated networks: Learning to leverage salient regions in medical images. Med Image Anal 2019;53:197–207. https://doi.org/10.1016/J.MEDIA.2019.01.012.
[20] Sim KS, Nia ME, Tso CP, Kho TK. Brain Ventricle Detection Using Hausdorff Distance. Emerging Trends in Applications and Infrastructures for Computational Biology, Bioinformatics, and Systems Biology: Systems and Applications 2016:523–31. https://doi.org/10.1016/B978-0-12-804203-8.00034-1.
[21] Kabakci A, Buyukmumcu ;, Yilmaz ;, Cicekcibasi ;, Akın ;, Cihan &. An Osteometric Study on Humerus Estudio Osteométrico del Húmero. Int J Morphol 2017;35:219–26.
[22] Alslehat RA, Samara OA, Alshatarat AT, Badran DH. Morphometry of the Humeral Head, Greater and Lesser Tuberosities: A Radiographic Study. National Journal of Clinical Anatomy 2024;13:56–62. https://doi.org/10.4103/NJCA.NJCA_10_24.
[23] Somesh MS, Prabhu L V., Shilpa K, Pai MM, Krishnamurthy A, Murlimanju B V. Morphometric Study of the Humerus Segments in Indian Population. International Journal of Morphology 2011;29:1174–80. https://doi.org/10.4067/S0717-95022011000400017.
[24] Sehrawat JS, Pathak RK. Variability in anatomical features of human clavicle: Its forensic anthropological and clinical significance. Translational Research in Anatomy 2016;3–4:5–14. https://doi.org/10.1016/J.TRIA.2016.08.001.
[25] Jia Y, He N, Liu J, Zhang G, Zhou J, Wu D, et al. Morphometric analysis of the coracoid process and glenoid width: A 3D-CT study. J Orthop Surg Res 2020;15:1–7. https://doi.org/10.1186/S13018-020-01600-1/TABLES/5.
[26] Dai L, Md Johar MG, Alkawaz MH. The diagnostic value of MRI segmentation technique for shoulder joint injuries based on deep learning. Scientific Reports 2024 14:1 2024;14:1–14. https://doi.org/10.1038/s41598-024-80441-y.
[27] Mu X, Cui Y, Bian R, Long L, Zhang D, Wang H, et al. In-depth learning of automatic segmentation of shoulder joint magnetic resonance images based on convolutional neural networks. Comput Methods Programs Biomed 2021;211:106325. https://doi.org/10.1016/J.CMPB.2021.106325.
[28] Kulseng CPS, Nainamalai V, Grøvik E, Geitung JT, Årøen A, Gjesdal KI. Automatic segmentation of human knee anatomy by a convolutional neural network applying a 3D MRI protocol. BMC Musculoskelet Disord 2023;24:1–12. https://doi.org/10.1186/S12891-023-06153-Y/FIGURES/6.

Downloaded from http://asmedigitalcollection.asme.org/IDETC-CIE/proceedings-pdf/IDETC-CIE2025/89213/V02BT02A022/7557339/v02bt02a022-detc2025-168564.pdf by Universita Degli Studi Di Bergamo user on 31 October 2025

 Copyright © 2025 by ASME; reuse license CC-BY 4.0