Santos-Blázquez, P., Vázquez-Ingelmo, A., García-Holgado, A., García-Peñalvo, F. J., Sánchez-Puente, A., & Sánchez, P. L. (2026). AI-Powered DICOM Image Segmentation: A Collaborative Platform for Continuous Expert Feedback. In A. Rocha, F. J. García-Peñalvo, C. J. Costa, & R. Gonçalves (Eds.), Proceedings of 20th Iberian Conference on Information Systems and Technologies (CISTI 2025) (Vol. 1, pp. 42–51). Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-10929-3_4

AI-Powered DICOM Image Segmentation: A Collaborative Platform for Continuous Expert Feedback

Pablo Santos-Blázquez[1,2,3], Andrea Vázquez-Ingelmo[3](✉), Alicia García-Holgado[3], Francisco José García-Peñalvo[3], Antonio Sánchez-Puente[4], and Pedro L. Sánchez[4]

[1] University of Tübingen, Tübingen, Germany
pablosbl@usal.es
[2] Konrad Zuse School of Excellence in Learning and Intelligent Systems (ELIZA), Darmstadt, Germany
[3] GRIAL Research Group, Instituto Universitario de Ciencias de la Educación (IUCE), Universidad de Salamanca, Salamanca, Spain
{andreavazquez,aliciagh,fgarcia}@usal.es
[4] Cardiology Department, Hospital Universitario de Salamanca, SACyL. IBSAL, Facultad de Medicina, Universidad de Salamanca, and CIBERCV (ISCiii), Salamanca, Spain
{sanchezpu,plsanchez}@saludcastillayleon.es
https://ror.org/02f40zc51

Abstract. This work presents the development of an interactive web platform that integrates deep learning techniques for the segmentation of cardiac ultrasound (echocardiogram) images. The platform incorporates a Picture Archiving and Communication System (PACS) to facilitate the seamless visualization, annotation, and automated processing of DICOM images. The web platform features an intuitive interface that allows healthcare professionals to interactively annotate medical images, providing feedback that directly informs model improvements. The system's retraining workflow ensures that AI-driven segmentation remains adaptable to real-world clinical needs. These findings underscore the importance of iterative AI model refinement through expert feedback, paving the way for more reliable and personalized medical image analysis.

Keywords: DICOM · Image Analysis · Deep Learning · PACS

1 Introduction

Medical image analysis and diagnosis are fundamental pillars of modern medicine. Techniques such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and ultrasound (US) provide doctors with detailed visual information about the human body, enabling early disease detection, treatment monitoring, and surgical planning. These techniques have transformed the approach to many diseases, increasing diagnostic precision and significantly improving patient outcomes.

The advancement of artificial intelligence (AI) technologies has added a new dimension to this field, revolutionizing medical image interpretation [1, 2]. Machine learning algorithms, based on deep neural networks, can analyze vast amounts of image data, identify complex patterns, and offer accurate and rapid diagnoses that often surpass human capabilities. The integration of these algorithms in radiology has substantially improved the detection of lung nodules, tumor identification, and retinal disease analysis.

This project builds on previous work carried out through the CARTIER-IA [3] platform, a program focused on the research and development of advanced AI technologies applied to healthcare. CARTIER-IA aims to foster collaboration between academic institutions, hospitals, and tech companies to drive innovations that improve disease diagnosis and treatment using AI.

The main goal of this work is to develop an interactive web platform designed to assist healthcare professionals in the visualization, annotation, and analysis of DICOM images, thereby enhancing multidisciplinary collaboration and medical feedback. This interactive web platform also acts as a Picture Archiving and Communication System (PACS) [4, 5], where health professionals can directly send the images to the node to be automatically processed by the available AI scripts.

The specific objectives of this project include (a) creating an intuitive interface that allows doctors to efficiently access and analyze DICOM images; (b) implementing annotation tools that facilitate feedback and collaboration among healthcare professionals; and (c) integrating an AI model based on the UNET architecture [6], specifically trained for cardiac image segmentation to accurately identify and segment anatomical structures such as the left ventricle, left atrium, and myocardium in cardiac ultrasound images.

To evaluate the performance of the UNET model and its adaptability to new annotations, three different retraining experiments were conducted. The results of these experiments demonstrated that including a greater number of images from the original dataset along with new annotations improves the model's generalization ability, significantly reducing overfitting.

The rest of this paper is organized as follows. Section 2 details the methodology followed to implement the web platform, the PACS server, and the AI integration. Section 3 presents the platform and the results of the experiments. Finally, Sect. 4 discusses and concludes this work.

2 Methodology

The methodology is structured in several sections covering the implementation of the web interface, the procedures for connecting with the PACS server, and the training and retraining of the UNET neural network.

2.1 Web Development

To implement the web platform, a framework for developing web applications in Python (Django) was used to provide easy compatibility with Python algorithms, allowing for the effective integration of artificial intelligence models and image-processing techniques used in this project.

For the visualization and annotation of DICOM images, the Cornerstone[1] framework, specialized in medical image visualization, was integrated. This framework enables interactive visualization of DICOM images as well as their annotation and manipulation. However, working with this framework presented significant challenges due to its scarce and complex documentation and incompatibilities between its versions.

The web platform has been designed to be intuitive and easy to use, ensuring that doctors can quickly access relevant information and make annotations efficiently. The interface was developed using HTML, CSS, and JavaScript to ensure a smooth and responsive user experience. Django's framework handles the backend of the application, managing requests and serving the necessary data for each view.

2.2 DICOM Image Processing

To allow integration with medical imaging workflows, a custom PACS (Picture Archiving and Communication System) server was implemented using the pynetdicom and pydicom Python libraries. The PACS server supports core DICOM operations including C-ECHO, C-STORE, and A-ASSOCIATE, enabling secure image storage and retrieval.

The PACS server acts as a centralized receiver and repository for DICOM files, allowing physicians to upload images from ultrasound machines or other diagnostic systems. The communication flow between image acquisition and storage in the PACS follows the standard DICOM protocol, involving additional steps that can be consulted in previous research [7].

2.3 Neural Network Configuration

The primary goal of the neural network is to segment the left ventricle, left atrium, and myocardium in 2D and 4D thoracic ultrasounds (echocardiograms). The UNET architecture was chosen for this task due to its proven capability to perform precise segmentations in biomedical images. UNET can identify complex patterns such as edges, circular segments, and textures, making it particularly suitable for the segmentation of anatomical structures in medical images.

The initial training of the network was conducted using images from the CAMUS database[2], which focuses on using deep learning techniques for the segmentation of thoracic ultrasounds (echocardiograms) in 2D, particularly the left ventricle (LV), left atrium (LA), and myocardium (MYO). A total of 600 images were used, split as follows:

- 70% for training
- 15% for validation
- 15% for testing

Images were resized to 384×384 pixels, normalized to a [0,1] range, and segmented masks were generated for multi-class output. Images were selected based on quality and availability of annotations, ensuring diversity in patient cases.

The network uses a symmetrical encoder-decoder structure with skip connections:

1 //docs.cornerstonejs.org/

2 https://www.creatis.insa-lyon.fr/Challenge/camus/

- The input resolution is 384 × 384 pixels.
- The encoder contains six convolutional blocks, with increasing filter sizes up to 320.
- The decoder mirrors this structure with upsampling and concatenation steps.
- The final layer applies a 1 × 1 convolution with softmax activation to classify pixels into four classes: background, LV, LA, and MYO.

To quantify the fit of the prediction generated by the network to the ground truth segmentation, several metrics can be used. The most common ones include precision, recall, F1-score, and Dice coefficient. In this project, the Dice coefficient was chosen as the main metric because it is especially useful in image segmentation problems. This metric quantifies the overlap between the predicted and real segmentation, providing a clear and direct evaluation of the segmentation quality.

Additionally, to achieve the best network performance, it is necessary to select an optimal optimizer algorithm. The most common optimizer algorithms include Stochastic Gradient Descent (SGD), RMSprop, and Adam, with the latter being selected for this project due to its robustness, efficiency, and adaptability.

2.4 Retraining

The retraining of the neural network is a crucial phase in the system's development, allowing the model to be adjusted and improved based on new annotations made by doctors. The network must be capable of learning using the images and annotations provided by physicians as ground truth, as shown in Fig. 1. This continuous learning process ensures that the model remains updated and optimized for the specific needs of the clinical environment.

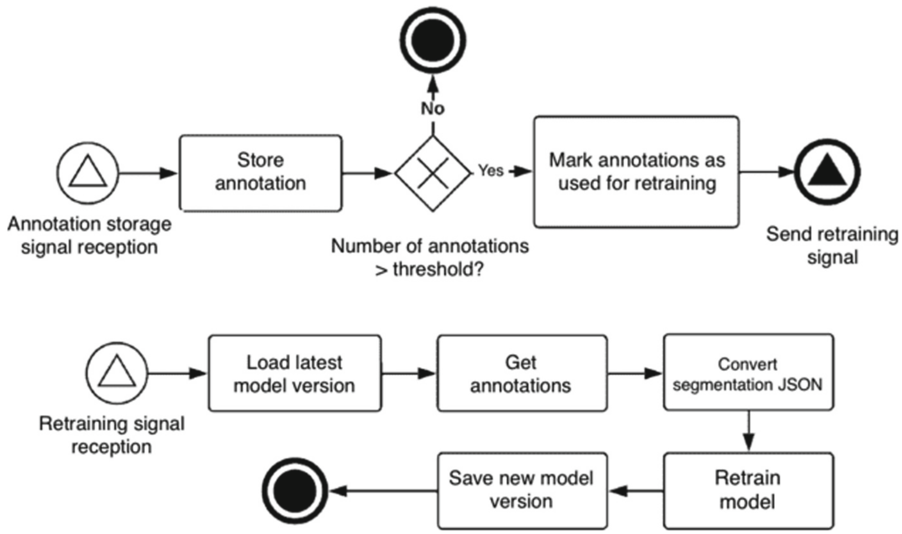

Fig. 1. Workflow of the retraining process.

To evaluate the performance and adaptability of the UNET model to new annotations, three distinct retraining experiments were conducted. Each experiment used different proportions of original data and new annotations to improve the model's generalization capability.

- **First Experiment**: Involves retraining the model solely with 25 annotations made on the same image. The objective is to observe how the model adjusts to a small, highly specific dataset and evaluate the risk of overfitting.
- **Second Experiment**: Retrains the model using the 25 annotations mentioned above along with an equal number of images from the original training dataset. This aims to verify whether the inclusion of original data helps maintain the model's prior knowledge while integrating new information.
- **Third Experiment**: Similar to the second experiment, but using twice the annotations along with the original training dataset images. This experiment seeks to evaluate if a greater number of annotations improves accuracy without compromising the model's prior knowledge.

These experiments demonstrated that including images from the original dataset along with new annotations significantly improves the UNET model's generalization ability, reducing overfitting.

3 Results

The platform allows users to explore different images following the DICOM structure. Once a study is selected, the editor view, represented in Fig. 2, presents the main data of the patient and the study in question. Here, two versions of the DICOM image are shown: the original and the one processed by the UNET network.

The view allows interactive annotation on both images, thus facilitating the segmentation task for the doctor. The interface also supports the visualization and annotation of multi-DICOM images, navigating through different images using a slider. The right side of the screen shows the annotations taken on that image.

The retraining experiments were conducted on three different DICOM images, labeled 'E', 'F', and 'G'. Figure 3 shows these images in their original format along with the segmentations predicted by the original model version.

The batch of 25 annotations used for the experiments corresponds to the DICOM image 'G'. Therefore, images 'E' and 'F' were never used for training or retraining and were instead part of the validation set. Figure 4 illustrates the specific annotation used in these experiments.

As seen in Fig. 4, the segmentation has been modified so that the left ventricle (red segment) is slightly larger, the right atrium (purple segment) is also larger and fits the actual shape, making it less circular, and the myocardium (green segment) is noticeably narrower. These annotations are stored in JSON format, which is later adapted to match the input of the Cornerstone tools that generate the segmentation.

For each experiment, three figures were obtained (as shown in Figs. 5, 6, and 7, corresponding to experiment 2), each corresponding to one of the images. These figures represent four independent segmentations showing the predictions made by the system after retraining with different batch size and epoch values. Each segmented image

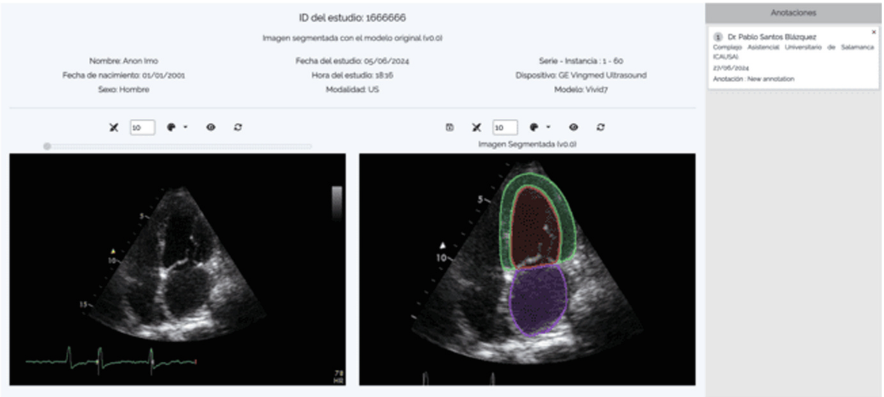

Fig. 2. DICOM viewer and editor

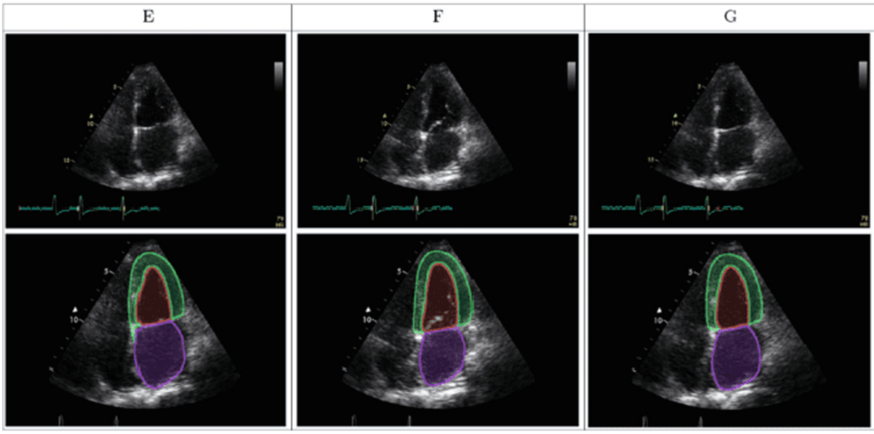

Fig. 3. Original model prediction.

obtained with each configuration shows the final retraining value characterized by the multiclass Dice metric.

The retraining experiments demonstrated the effectiveness of including new annotations to improve the UNET model's generalization ability:

- **Experiment 1**: Used a limited dataset, revealing overfitting issues. Predictions for images 'E', 'F', and 'G' showed significant discrepancies from the real segmentations.
- **Experiment 2**: Included a greater number of images from the original dataset along with new annotations, improving the model's generalization. Predictions for images 'E', 'F', and 'G' were more accurate.
- **Experiment 3**: With a 2:1 proportion of images from the original dataset to annotations, it achieved a balanced adaptation, significantly reducing overfitting. Predictions for images 'E', 'F', and 'G' showed high concordance with real segmentations.

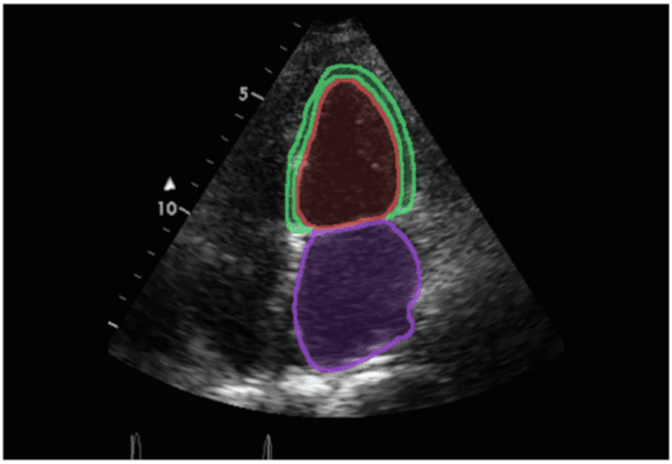

Fig. 4. Annotation used for the different experiments.

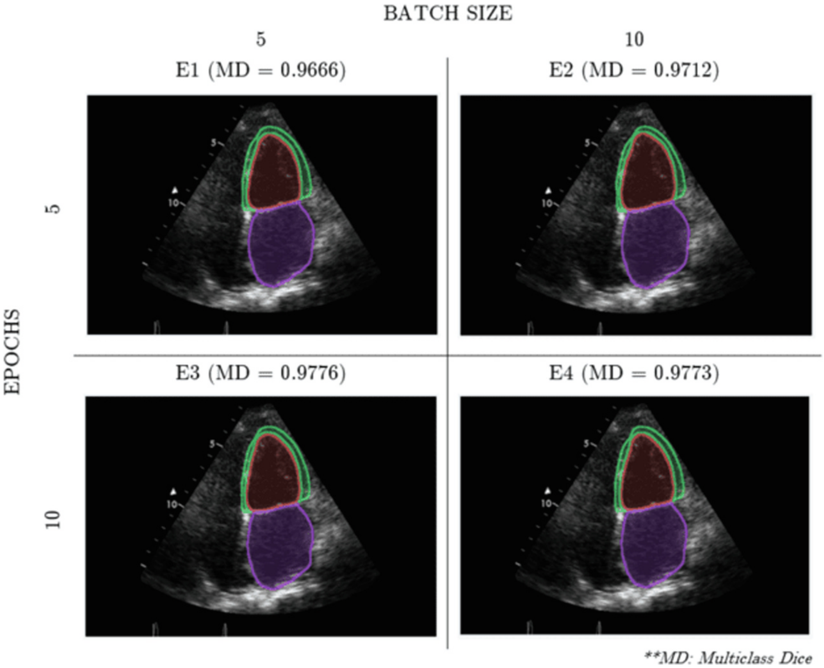

Fig. 5. Predictions obtained for image 'E' in experiment 2.

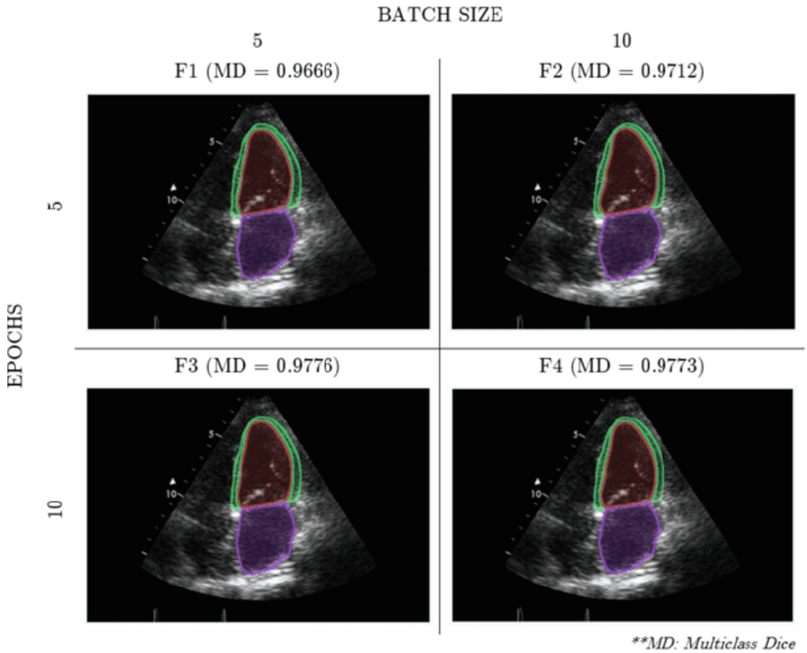

Fig. 6. Predictions obtained for image 'F' in experiment 2.

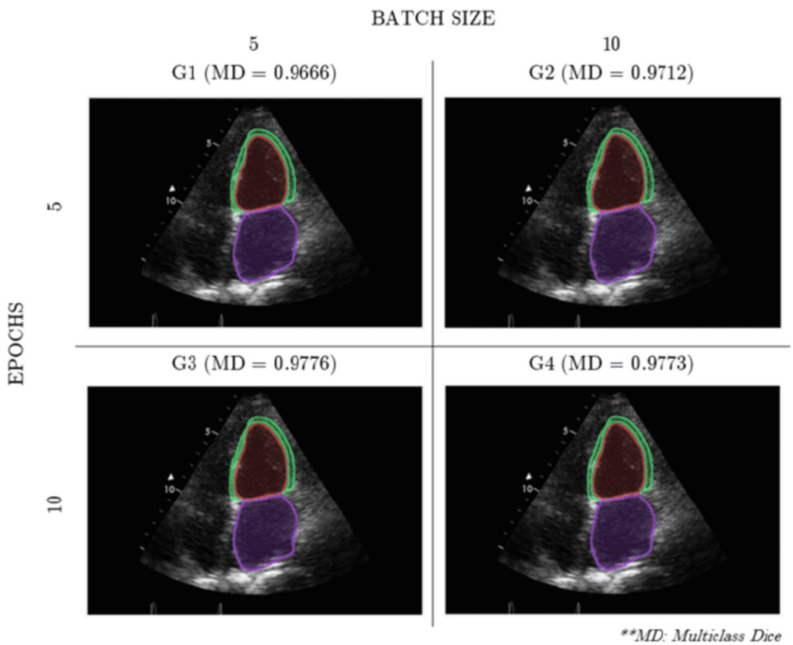

Fig. 7. Predictions obtained for image 'G' in experiment 2.

4 Discussion and Conclusions

The UNET architecture has proven to be highly effective in medical image segmentation tasks due to its ability to capture both local and global features of images. In this study, the results have shown high precision in the segmentation of specific cardiac structures, with Dice coefficient values close to 1.0 in many cases. More than eighty previous studies identified during the mapping stage have also used the UNET architecture and its variants for the segmentation of MRI and CT images, achieving similar results in terms of precision and generalization capacity. Our results are consistent with these studies and confirm the robustness of the UNET architecture for cardiac ultrasound image segmentation.

The three experiments conducted in this study have been fundamental in understanding the influence of different training data configurations and parameters on the model's precision. Experiment 1 demonstrated an excessively high precision and a clear case of overfitting. Experiment 2 improved the model's generalization capacity and showed more realistic precision values. Experiment 3 achieved a balanced adaptation, significantly reducing overfitting. These experiments demonstrated that the inclusion of images from the original dataset allows for better adaptation of the model to new annotations.

The retraining of the model with different batch size and epoch configurations showed that these hyperparameters have a significant influence on segmentation precision. A larger batch size allows the model to learn more robust patterns and stabilize the training process, although it also increases the risk of overfitting if not properly managed. A greater number of epochs allows the model to better adjust to the training data but also introduces the risk of overfitting. The results obtained demonstrate that retraining the UNET network with additional annotations and a larger number of images from the original dataset significantly improves segmentation precision.

These findings underline the importance of using varied datasets and appropriate training configurations to develop robust and precise segmentation models. The choice of retraining approach should be decided when the project enters the production phase, depending on the medical specialists' evaluation of the initial segmentation. If the initial segmentation is not very precise, faster training will be required, and the strategy from experiment 2 will be implemented; if it is quite precise, the strategy from experiment 3 will be implemented.

In terms of future directions, it is suggested to explore the integration of other artificial intelligence models and visualization frameworks that can complement and improve the current capabilities of the platform. Additionally, more extensive validation in real clinical environments will be essential to ensure that this platform can meet the requirements and expectations of healthcare professionals in daily practice. Continuously improving the model's precision and robustness through the use of more varied datasets and advanced retraining techniques will be crucial for the project's continued success.

Acknowledgement. Pablo Santos Blázquez is supported by the Konrad Zuse School of Excellence in Learning and Intelligent Systems (ELIZA) through the DAAD programme Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Education and Research. This research was partially funded by the Spanish Ministry of Science and Innovation through the AVisSA project grant number (PID2020-118345RB-I00). This work was also supported by competitive community grants (GRS 2033/A/19, GRS 2030/A/19, GRS 2031/A/19, GRS 2032/A/19) from the SACYL, Junta Castilla y León; by competitive national grants (PI14/00695, PIE14/00066, PI17/00145, DTS19/00098, PI19/00658, PI19/00656, PI21/00369) from the Institute of Health Carlos III, Spanish Ministry of Science and Innovation and co-funded by ERDF/ESF, "Investing in your future" and; by the CIBERCV (CB16/11/00374) from the Institute of Health Carlos III, Spanish Ministry of Science and Innovation.

References

1. Najjar, R.: Redefining radiology: a review of artificial intelligence integration in medical imaging. Diagnostics **13**, 2760 (2023)
2. Pinto-Coelho, L.: How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications. Bioengineering **10**, 1435 (2023)
3. García-Peñalvo, F.J., et al.: Application of artificial intelligence algorithms within the medical context for non-specialized users: the CARTIER-IA platform. Int. J. Interactive Multimed. Artif. Intell. **6**, 46–53 (2021)
4. Mansoori, B., Erhard, K.K., Sunshine, J.L.: Picture archiving and communication system (PACS) implementation, integration & benefits in an integrated health system. Acad. Radiol. **19**, 229–235 (2012)
5. Choplin, R.H., Boehme, J., 2nd., Maynard, C.D.: Picture archiving and communication systems: an overview. Radiographics **12**, 127–129 (1992)
6. Ronneberger, O., Fischer, P., Brox, T.: 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, pp. 234–241. Springer (Year)
7. Vazquez-Ingelmo, A., Garcia-Holgado, A., Garcia-Peñalvo, F.J., Pérez-Sánchez, P., Sánchez-Puente, A., Vicente-Palacios, V., Ignacio Dorado-Díaz, P., Luis Sánchez, P.: D-AI-Com: A Dicom Reception Node to Automate the Application of Artificial Intelligence Scripts to Medical Imaging Data. WorldCIST 2023, Pisa, Italy (2023)