IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 September 2023, accepted 18 October 2023, date of publication 23 October 2023, date of current version 30 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3326882

RESEARCH ARTICLE

SAM-UNETR: Clinically Significant Prostate Cancer Segmentation Using Transfer Learning From Large Model

JESUS ALEJANDRO ALZATE-GRISALES[1,2], ALEJANDRO MORA-RUBIO[1,2], FRANCISCO GARCÍA-GARCÍA[3], REINEL TABARES-SOTO[2,4,5], AND MARIA DE LA IGLESIA-VAYÁ[1]

[1]Fundación para el Fomento de la Investigación Sanitario y Biomédica de la Comunidad Valenciana, Unidad Mixta de Imagen Biomédica FISABIO-CIPF, 46020 Valencia, Spain
[2]Department of Electronics and Automation, Universidad Autónoma de Manizales, Manizales 170001, Colombia
[3]Bioinformatics and Biostatistics Unit, Principe Felipe Research Center (CIPF), 46012 Valencia, Spain
[4]Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Santiago 7941169, Chile
[5]Departamento de Sistemas e Informática, Universidad de Caldas, Manizales 170001, Colombia

Corresponding authors: Jesus Alejandro Alzate-Grisales (jesus.alzate@fisabio.es) and Maria de la Iglesia-Vayá (delaiglesia_mar@gva.es)

This work was supported in part by the TARTAGLIA Project from Banco de Imagen Médica de la Comunidad Valenciana (BIMCV), part of the Research and Development Missions in Artificial Intelligence Program of the Spanish Digital Agenda 2025 and the National Artificial Intelligence Strategy, funded by the European Union through the Next Generation EU Funds, Regional Ministry of Health of the Valencian Region, corresponding to the funds of the Recovery, Transformation, and Resilience Plan, under Grant MIA.2021.M02.0005; in part by the Bioinformatics and Biostatistics Unit from Principe Felipe Research Center (CIPF), co-funded by European Regional Development Funds (FEDER) in the Valencian Community 2014–2020; and in part by the Biomedical Imaging Mixed Unit from Fundació per al Foment de la Investigació Sanitaria i Biomedica (UMIB-FISABIO), co-funded by FEDER in 2014–2020.

ABSTRACT Prostate cancer (PCa) is one of the leading causes of cancer-related mortality among men worldwide. Accurate and efficient segmentation of clinically significant prostate cancer (csPCa) regions from magnetic resonance imaging (MRI) plays a crucial role in diagnosis, treatment planning, and monitoring of the disease, however, this is a challenging task even for the specialized clinicians. This study presents SAM-UNETR, a novel model for segmenting csPCa regions from MRI images. SAM-UNETR combines a transformer-encoder from the Segment Anything Model (SAM), a versatile segmentation model trained on 11 million images, with a residual-convolution decoder inspired by UNETR. The model uses multiple image modalities and applies prostate zone segmentation, normalization, and data augmentation as preprocessing steps. The performance of SAM-UNETR is compared with three other models using the same strategy and preprocessing. The results show that SAM-UNETR achieves superior reliability and accuracy in csPCa segmentation, especially when using transfer learning for the image encoder. This demonstrates the adaptability of large-scale models for different tasks. SAM-UNETR attains a Dice Score of 0.467 and an AUROC of 0.77 for csPCa prediction.

INDEX TERMS Artificial intelligence, deep learning, prostate cancer, semantic segmentation.

I. INTRODUCTION

PCa presents a significant global health challenge, impacting millions of men worldwide, with 1.4 million new cases and 375,000 deaths recorded in 2020 alone [1]. The early detection and prompt treatment of PCa are critical for improving patient outcomes, given the poor prognosis associated with advanced-stage PCa. However, the complex nature of PCa detection poses a significant challenge that requires a high level of expertise and skill from radiologists.

MRI is a powerful imaging modality that can provide detailed images of the prostate gland and surrounding tissues, facilitating the detection and diagnosis of PCa by clinicians [2]. Nonetheless, the interpretation of MRI images can be a daunting task, even for experienced radiologists, given the intricate nature of prostate anatomy and the variability of the

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno M. Garcia.

disease. Artificial intelligence (AI) models are computational techniques that use existing data to learn patterns and make predictions on new, unseen data [3]. In the past, traditional machine learning approaches were used, which involved two steps: domain experts designing features to extract quantitative variables from the data, and then feeding these features into computational models to learn how to combine them to maximize accuracy in classifying data into categories [4]. Recent advances in graphics processing unit (GPU) computing power have enabled the development of deep learning models, which automate the process of identifying features and using them for downstream tasks [5]. Deep learning (DL) models have revolutionized the field of AI by achieving unprecedented performance that often exceeds human performance, particularly in image analysis tasks. Medical imaging is one of the areas where AI has shown immense potential to improve the accuracy and efficiency of cancer detection and diagnosis. AI models can analyze large amounts of imaging data quickly and accurately, potentially improving the speed and accuracy of PCa detection and diagnosis [6]. In particular, deep learning models have shown promise in analyzing MRI images for PCa detection, with several studies demonstrating improved sensitivity and specificity [7], [8], [9], [10].

For example, the work by Haozhe Jia et al. [11] uses a 3D Convolutional Neural Network (CNN) to segment the prostate, following an encoder-decoder architecture and involving novel elements such as anisotropic convolutional decoder with pyramid convolutional skip-connections. Furthermore, this segmentation CNN is trained in an adversarial-style, where an additional CNN differentiates between a segmentation result and its corresponding ground truth, achieving model regularization. Another approach by Zhu et al. [12] proposes a boundary-weighted loss function that makes the CNN segmentation model more sensitive to object borders, aiming to tackle the lack of clear edges between the prostate and other anatomical structures, as well as, large annotated datasets. On the other hand, Ushinsky et al. [13] proposed a 3D-2D hybrid CNN to leverage information from multiple axial slices simultaneously, imitating how radiologists interpret multiple axial images before making decisions about one 2D slice, this approach also favors rapid prostate organ segmentation, taking only 0.363 seconds per image as reported by the authors. In a different line of work, specific to csPCa detection and segmentation, the work by Singla et al. [14] uses a transformer-based U-Net architecture for the detection and segmentation of PCa using MRI scans, achieving a 0.80 Dice Score on the PROMISE-12 dataset outperforming other conventional DL models. Similarly, Dai et al. [15] aimed to detect and delineate intraprostatic lesions for PCa radiation therapy, they used T2-weighted (T2w) images and a 2D Mask R-CNN architecture. They report high Dice Scores, 0.88 and 0.86, for two different cohorts of patients.

As such, the integration of AI into MRI-based PCa detection and diagnosis has the potential to revolutionize clinical practice and lead to better patient outcomes. However, the successful implementation of AI in clinical practice requires robust validation and ongoing evaluation of the AI algorithms to ensure their reliability and effectiveness. For this, PI-RADS (Prostate Imaging - Reporting and Data System) was developed by an international collaboration of the American College of Radiology (ACR), the European Society of Urology (ESUR) and the AdMetech Foundation to promote global standardization and reduce variation in the acquisition, interpretation and reporting of prostate multiparametric Magnetic Resonance Imaging (mpMRI) exams. PI-RADS v2 uses a 5-point scale based on the likelihood that a combination of mpMRI findings correlates with the presence of clinically significant cancer for each lesion in the prostate [16].

- **PI-RADS 1:** Very low (csPCa is highly unlikely to be present).
- **PI-RADS 2:** Low (csPCa is unlikely to be present).
- **PI-RADS 3:** Intermediate (the presence of csPCa is equivocal.
- **PI-RADS 4:** High (csPCa is likely to be present).
- **PI-RADS 5:** Very high (csPCa is highly likely to be present).

This grading is underpinned by careful evaluation of MRI scans. The primary indicators are the detection and localization of nodules or shadows, often characterized as T2-weighted morphology. At the same time, Diffusion-Weighted Imaging (DWI) and Apparent Diffusion Coefficient (ADC) maps evaluate the density of these abnormalities and measures the diffusion rate of water molecules, as diffusion differs between healthy and cancerous prostate tissue. In addition, Dynamic Contrast Enhanced (DCE) images provide insight into the uptake of contrast agents, revealing blood flow patterns characteristic of tumour cells. These parameters are complemented by MRI spectroscopy, which compares the secretions or metabolites of suspicious regions with those found in normal prostate tissue. Together, these nuanced assessments provide the PI-RADS score, which serves as an indicative measure of the likelihood of PCa [16].

Acknowledging the potential of automated computer-aided diagnosis (CAD) systems in medical imaging, csPCa represents a key area where early and accurate detection can make a significant difference in patient outcomes and treatment trajectories. A critical challenge is the design and training of highly complex and novel DL architectures, which often require massive computational resources and specialized infrastructure.

Against this backdrop, our research is centered in two main goals:

- **Exploiting transfer learning in high-end architectures**: Large companies have invested heavily in developing and training complex DL models that, while incredibly powerful, are often beyond the reach of many researchers due to their computational requirements. This study seeks to leverage these efforts through the use of transfer learning. By leveraging the pre-trained

weights of these sophisticated models, we aim to initiate and adapt them for PCa imaging tasks, bypassing the enormous computational overhead typically associated with training such models from scratch.

- **A novel architecture for csPCa detection:** The promise of automated CAD systems to improve medical diagnostic workflows is undeniable. Building on this potential, we present a DL-based methodology for prostate lesion segmentation in MRI scans using our novel architecture, SAM-UNETR. This architecture is a harmonious blend of the SAM image encoder and a UNET-style transformer decoder. By leveraging the pre-trained weights of complex models, SAM-UNETR provides a powerful yet efficient approach to csPCa detection, combining the richness of transfer learning with the precision of custom segmentation.

The rest of the paper is organized as follows: Section II Materials and Methods describes the dataset, DL architectures, experiments and training set-up; Section III Results goes over the results at lesion and patient level and Section IV Discussion analyze the presented results; finally Section V presents the conclusion of this work and some future work directions.

II. MATERIAL AND METHODS

The proposed methodology entails a systematic approach involving preprocessing of MRI data obtained from two distinct datasets. These datasets encompass T2w and DWI sequences accompanied by ADC maps. The preprocessing procedure comprises multiple stages aimed at extracting relevant information from the MRI scans.

Initially, a 3D-UNET architecture is employed to perform segmentation of the entire prostate volume. Subsequently, the same network is utilized to discriminate and segment the Central Zone (CZ) and Peripheral Zone (PZ) of the prostate. Following this step, the MRI data is cropped to isolate the region of interest, specifically the prostate area, thereby facilitating subsequent analysis. In the final stage of the methodology, four distinct DL networks are trained for the purpose of detecting and segmenting prostate lesions. Three of these networks are well-established models widely employed in the field, while the fourth network represents a novel approach proposed in this research.

A. DATASETS

We utilized two datasets for our study: Prostate158 and the dataset from the PI-CAI Challenge. Each dataset provides unique insights and values to our research, and their details are elaborated below:

1) PROSTATE158 DATASET

The Prostate158 dataset [17] includes 158 carefully curated biparametric 3T prostate MRI scans. These scans encompass sequences such as T2w, DWI, and ADC maps. Expert radiologists provided annotations to ensure accurate segmentations at the pixel level for different regions of interest: CZ, PZ, and PCa lesions. PCa lesions were demarcated as areas with a PI-RADS score of 4 or higher. For zonal segmentations, axial T2w sequences were primarily used, with segmentations being pixel-specific.

TABLE 1. Dataset distribution summary.

Dataset	Training	Validation	Test	Total
PI-CAI Challenge	895	298	298	1,491
Prostate158	119	20	19	158
Merged	**1,014**	**318**	**317**	**1,649**

2) PI-CAI CHALLENGE DATASET

Introduced as a significant grand challenge, the PI-CAI Challenge acts as a benchmarking platform for advanced AI algorithms and also evaluates radiologist performance in diagnosing csPCa [18]. The challenge has over 10,000 prostate MRI scans. However, for research purposes, only 1,500 cases are made public. Of these, 328 cases are sourced from the ProstateX dataset [19], a dataset with the same data distribution but collected under different conditions and scanners, which ensures a diverse representation of clinical scenarios. Each MRI case in the dataset includes T2w, DWI, and ADC sequences.

Out of the 1,500 cases:

- 1,075 cases present benign or indolent PCa.
- 425 cases are associated with csPCa, but only 220 of these have expert annotations.

For consistency, all annotations are rescaled to the T2w sequence's dimension and resolution.

3) UTILIZATION OF DATASETS

In our study, the Prostate158 dataset was used exclusively for prostate segmentation tasks, primarily due to its comprehensive zonal segmentation. However, for the purpose of lesion detection, the Prostate158 and PI-CAI Challenge datasets were merged to increase the variability of the data. During the pre-processing phase, nine corrupted images from the PI-CAI Challenge dataset were identified and subsequently discarded. This left 1,491 images from the original data. These were distributed in a 3:1:1 ratio for training (60%), validation (20%) and testing (20%) - a distribution that protects the integrity of the patient data and ensures a sufficient number of cases to accurately assess model performance. This resulted in a configuration of 895 training images, 298 validation images and 298 test images for the PI-CAI Challenge dataset. The Prostate158 dataset had predefined partitions: 139 for training and 19 for validation. From the initial 139 training images, 20 were reallocated to the validation set. Cumulatively, the inclusion of the Prostate158 data resulted in a final distribution of 1,014 images for training, 318 for validation and 317 for testing. Table 1 shows a summary of the full dataset partitions.

B. PREPROCESSING

Prior to model training, it is essential to preprocess the data in order to enhance results and promote data homogeneity. Specifically, for prostate zonal segmentation, T2w images were utilized. To ensure uniformity, these images underwent a series of preprocessing steps. Firstly, the images were resampled to a spacing of $0.5 \times 0.5 \times 0.5$ mm and oriented in the radiological anatomical system (RAS) orientation. Subsequently, a two-step normalization process was implemented, encompassing both min-max normalization and z-score normalization. These measures contribute to aligning the intensity values across the dataset, facilitating consistent and reliable segmentation outcomes.

Regarding lesion segmentation, a similar preprocessing pipeline was employed. The modalities utilized for this task included T2w, ADC maps, and DWI images. In order to establish compatibility between the modalities, the ADC and DWI images were resampled to match the shape and spacing of the T2w images, then images were cropped using the dimensions of the complete prostate segmentation. Subsequently, all images were resized to dimensions of $128 \times 128 \times n_c$, where n_c corresponds to the number of channels required for 2D segmentation. Then, in order to ensure the preservation of crucial image information while mitigating any potential bias, each image was individually normalized using z-score normalization as shown in Equation 1 where μ represents the mean and σ the standard deviation of every image.

$$X_{\text{normalized}} = \frac{X - \mu}{\sigma} \tag{1}$$

To enhance the robustness and generalizability of the trained models, multiple random data augmentation techniques were employed. These augmentations included spatial transformations such as rotations, the addition of Gaussian noise, and slight intensity variations. These augmentation strategies promote the creation of a more diverse training set, effectively increasing the model's ability to handle variations and generalize well to unseen data.

C. PROSTATE SEGMENTATION AND CROPPING

Building on the successful results of Adams et al. on the Prostate158 dataset [17], our study applies the same methodology and network architecture to prostate segmentation. For training, we extract random patches of size $96 \times 96 \times 96$ pixels from each image and use them during each training iteration. From the Prostate158 dataset, 119 images are specifically selected for training.

The chosen architecture for our prostate segmentation task is a 3D-UNET, which will be explained in Section II-D. The input to this network is a T2w image, and it outputs a segmentation mask with three channels: background, CZ, and PZ. When the model produces these masks, the CZ and PZ zones are merged into a single semantic segmentation representing the entire prostate region. This combined mask is then transformed into a bounding box, and its boundaries are extended by 20 pixels in all directions. The prostate region is then cropped based on these expanded dimensions, in T2w, ADC maps, and DWI images. Figure 1 shows the mentioned process.

D. DEEP LEARNING NETWORKS

As mentioned above, a 3D model architecture was chosen for the zonal segmentation of the prostate. However, only 2D networks were chosen for lesion segmentation, driven by two key considerations. Firstly, this decision was made to facilitate a direct and meaningful comparison with SAM-UNETR, a model purposefully designed for 2D images, since it utilizes the image encoder from the SAM, which was exclusively trained on 2D image modalities. Further details regarding the training of SAM will be expounded upon in subsequent sections of this paper. By employing 2D networks, the study aims to establish a fair and insightful evaluation of the selected models in relation to SAM-UNETR. Secondly, the utilization of 2D networks ensures the preservation of vital information during the lesion segmentation process. By conducting segmentation in the 2D space, the risk of losing valuable details on individual image slices is mitigated. This approach safeguards against the potential loss of crucial spatial context and ensures that the segmentation algorithms can effectively capture important features within the lesions.

1) UNET

UNET is a CNN architecture commonly used for image segmentation tasks. It was introduced by Ronneberger et al. in 2015 [20]. UNET has gained significant popularity in the field of medical image analysis due to its effectiveness in segmenting anatomical structures and abnormalities. This architecture consists of an encoder-decoder structure with skip connections. The encoder part of the network gradually reduces spatial dimensions while increasing the number of feature channels through successive convolutional and pooling layers. This process captures hierarchical feature representations at multiple scales. The decoder part, on the other hand, performs upsampling and convolution operations to progressively reconstruct the segmented output. Skip connections are established between corresponding encoder and decoder layers to enable the integration of both low-level and high-level features, aiding in the preservation of fine details during the segmentation process.

The 3D-UNET previously mentioned for prostate segmentation is a derivative of the standard UNET architecture. The main difference between the 3D-UNET and the conventional UNET is that the former uses 3D convolutional kernels, as opposed to the 2D counterparts used by the latter. This adaptation allows the 3D-UNET to exploit enhanced spatial information from the input volumes. As described in [17] and Figure 2 shows, the proposed architecture consists of five residual blocks accompanied by five successive downsampling stages. These downsampling processes result in a systematic reduction in spatial dimensions while

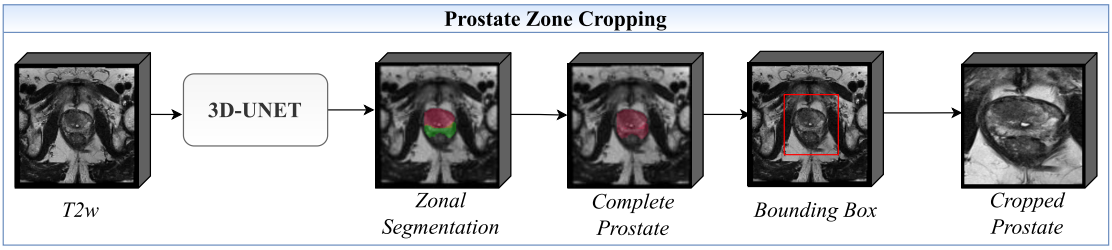

FIGURE 1. Prostate cropping process. The network takes a T2w image as input and generates a segmentation mask with three channels: background, CZ, and PZ. The CZ and PZ channels are then combined to form a single mask that represents the whole prostate region. We then compute a bounding box around this mask and expands it by 20 pixels in each direction. The expanded bounding box is used to crop the prostate region from the T2w image, as well as from the ADC maps and DWI images that correspond to the same slice.

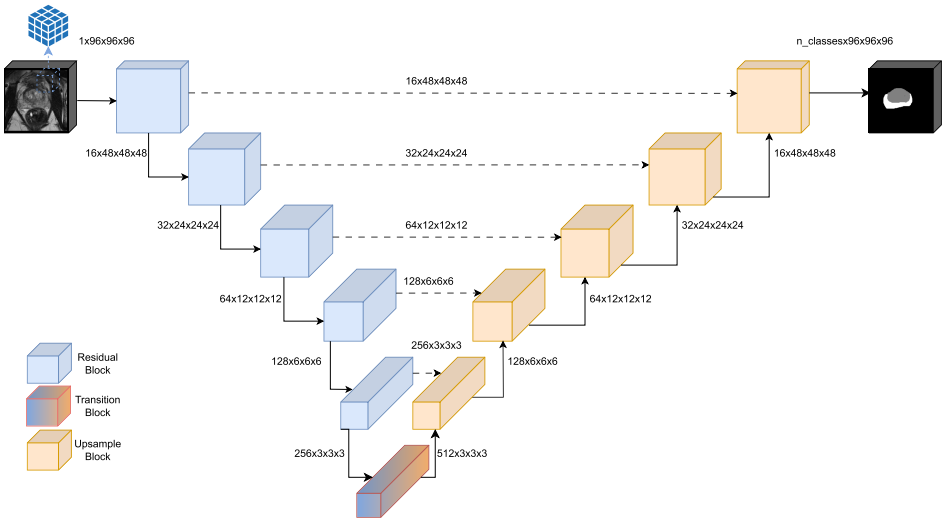

FIGURE 2. 3D-UNET for prostate zonal segmentation. Multiple random patches of size 96 × 96x96 are extracted from the images, then these patches are sent to the 3D-UNET model which employs five residual blocks and five downsampling steps. The downsampling progressively reduces spatial dimensions and increases channel count, starting from 16 and doubling to 512. The input is a T2w image, and the output is a segmentation mask with three channels: background, CZ, and PZ. Adapted from [17].

simultaneously increasing the number of channels, starting from an initial 16 channels and successively doubling at each stage, culminating in 512 channels and batch normalization. Such a design configuration is instrumental in the extraction of hierarchical features, thereby enhancing segmentation efficiency. For lesion segmentation, a 2D version of the network, originally designed for zonal segmentation of the prostate, was implemented.

2) UNETR

UNETR, derived from the renowned UNET architecture, represents a novel variation that incorporates Transformers. This network was introduced by Hatamizadeh et al. in 2021 [21], specifically designed to address the challenges encountered in 3D Medical Image Segmentation. By integrating a transformer encoder, UNETR excels at capturing and comprehending global multi-scale information. Furthermore, it employs a UNET decoder, enriched with multiple layers of convolutions and skip connections, to facilitate precise segmentation. The architectural design proposed for this task is a 2D version of the model with a feature size of 16, a hidden size of 768, 24 attention heads and batch normalization.

3) SwinUNETR

SwinUNETR is an advanced model derived from UNETR, a state-of-the-art U-shaped network renowned for its effectiveness in 3D medical image segmentation. Building upon the foundation of UNETR, SwinUNETR introduces a significant enhancement by replacing the transformer encoder with a Swin transformer encoder. The Swin transformer is a hierarchical transformer architecture that employs shifted windows to compute self-attention in an efficient and highly effective manner [22]. In SwinUNETR, the transformer encoder operates by extracting features at five distinct resolutions, capitalizing on the utilization of shifted windows for self-attention computation. This approach facilitates the capture of intricate spatial dependencies across different scales, enabling the model to discern fine details and

contextual information crucial for accurate segmentation. Through the integration of skip connections, the transformer encoder at each resolution is seamlessly connected to a Fully CNN-based decoder [23]. This connectivity ensures the fusion of both local and global information, enabling the network to generate precise segmentation outputs. For the specific task at hand, a 2D version of the SwinUNETR model is employed. This variant exhibits a feature size of 48, incorporates instance normalization to enhance stability during training, and integrates a dropout rate of 0.15 for each layer.

4) SAM-UNETR

The proposed method combines the image encoder from the SAM, a pioneering segmentation model introduced by Kirillov et al. [24], with the decoder architecture inspired by UNETR. SAM, as implied by its name, serves as a foundational segmentation model that exhibits remarkable versatility in accommodating various types of input prompts, including points, boxes, or text. It proficiently generates masks for all objects present in an image.

SAM's training procedure involved an extensive dataset comprising 11 million images and 1.1 billion corresponding masks, called SA-1B. The creation of the dataset involved a unique ''data engine'' approach that included three stages: assisted-manual, semi-automatic, and fully automatic annotation. In the first phase, the Segment Anything Model (SAM) assists annotators in annotating masks, similar to traditional interactive segmentation. In the subsequent semi-automatic phase, SAM autonomously generates masks for certain objects while human annotators focus on the rest. In the final stage, SAM produces an average of about 100 high-quality masks per image when prompted with a grid of foreground points. Impressively, SA-1B has 400 times more masks than any existing segmentation dataset, ensuring both quality and diversity. It should be noted that all this procedure was developed by Kirillov et al. [24].

SAM architecture encompasses an image encoder responsible for extracting comprehensive features from the input image. Additionally, a prompt encoder is employed to encode the input prompts into a spatial representation. The mask decoder harmoniously integrates the image features with the prompt representation, effectively generating masks for each prompt in a coherent and precise manner.

To take advantage of the knowledge gained from training on the large SA-1B dataset, our method emphasizes the extraction of pre-trained weights from the image encoder. This encoder, a Masked Autoencoder Vision Transformer, is optimized for high-resolution images. Given our transfer learning goals, it's important to be able to adapt to different image resolutions and channel counts, ensuring the adaptability of the weights.

The encoder has 32 attention blocks, where blocks 8, 16, 24, and 32 are special global attention blocks (GABs). These GABs are modified to handle any image size, ensuring the model's versatility across different resolutions. For a smooth integration into our transfer learning framework, we also adapted the weights shape and input channels as mentioned before. Additionally, for potential skip connections, the encoder returns the hidden states from each transform block. However, as in [24], most of the structure of the encoder is equal to the original SAM encoder, maintaining an embedding size of 1,280.

Similar to the Swin-UNETR and UNETR architectures, the image encoder employed in this study adopts a vision transformer approach. The U-shaped architecture incorporates skip connections to facilitate the transmission of information from the encoder to the decoder, being GABs used for this purpose. The output of each GAB assumes a shape of $\frac{H}{16} \times \frac{W}{16} \times \frac{E_s}{16}$, where E_s corresponds to the embedding size. These outputs are processed by a deconvolution block, which involves a 3×3 convolution followed by normalization layers, as highlighted in [21]. At the bottleneck of the SAM encoder, a deconvolution layer increases the resolution of the feature map by a factor of two. The enhanced feature map is then merged with the feature map from the previous transformer output. This combined feature map undergoes 3×3 convolutional layers and is upscaled using a deconvolutional layer. This cycle is repeated until the original resolution is restored. Finally, the output is processed through a 1×1 convolutional layer with softmax activation, allowing for pixel-wise semantic prediction. The architecture is shown in Figure 3.

E. EXPERIMENTS

After preprocessing the T2w, ADC, DWI, and zonal masks, we concatenate the images. The zonal masks (CZ and PZ) are encoded using a one-hot representation, yielding a two-channel image. The final concatenated image has five channels. We transpose the three-dimensional images from $H \times W \times D \times 5$ to $D \times H \times W \times 5$, with D indicating the number of slices. This format supports batch processing with dimensions $B \times H \times W \times 5$ for 2D models.

It's worth noting that for SAM-UNETR training we slightly adjusted the preprocessing steps described in section II-B to suit our transfer learning methodology, especially since the images from SA-1B have different intensity range. First, we applied min-max normalization as in equation 2 to adjust image pixel values to fall within the [0, 255] range of the SA-1B dataset. Then, we used the z-score method (Equation 1) by subtracting the mean and dividing by the standard deviation derived from the SA-1B dataset. It's important to note that while the SA-1B dataset contains RGB images, our dataset consists entirely of concatenated gray images. The SA-1B dataset has mean pixel values of 123.675, 116.28, and 103.53 for each of the RGB channels, and standard deviations of 58.395, 57.12, and 57.375, respectively. From this, we derived an overall mean (μ) of 114.495 and a standard deviation (σ) of 57.63. This calculated mean and standard deviation were then used for z-score normalization on each individual channel (T2w,

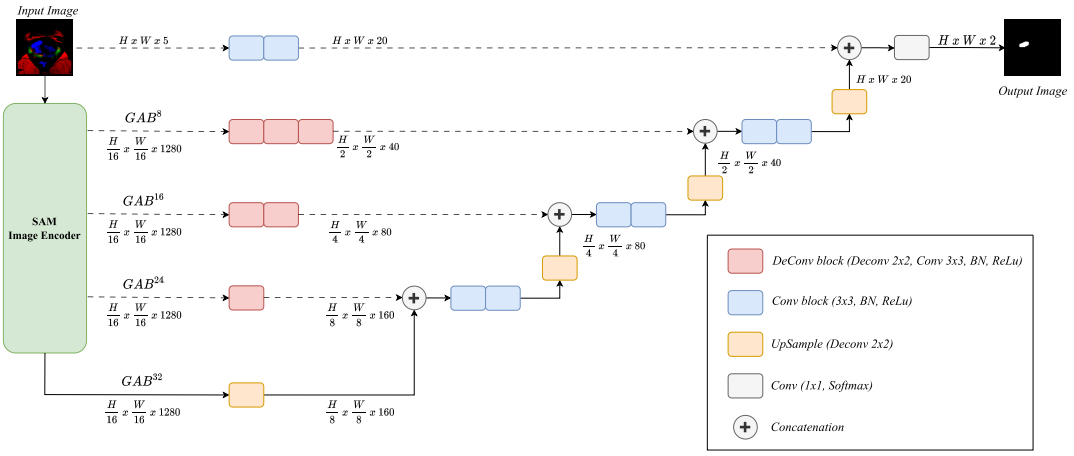

FIGURE 3. SAM-UNETR Architecture. This model uses a SAM transformer-based image encoder with 32 attention blocks. It has a U-shaped architecture with skip connections using Global Attention Blocks (GAB). The resolution of feature maps is enhanced using deconvolutional blocks. The feature maps are then combined with preceding transformer output and processed through convolutional layers and upsampling. This process continues until the original input resolution is reached. Finally, the output is passed through a 1 × 1 convolutional layer with softmax activation. Based on [21].

ADC, DWI). This preprocessing adjustment ensures that our dataset remains compatible and maintains consistent performance when merged with the transferred model. For reference, SAM-UNETR is trained in two different ways: either without pretraining (via random initialization) or with pretraining that uses weights from the SAM model.

$$X_{0\text{-}255} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{2}$$

Our methodology is outlined in Figure 4 and involves:

1) Segmenting the entire prostate using only the T2w sequence and cropping the segmented region.
2) Resampling the ADC and DWI sequences to match the T2w sequence, followed by cropping to the prostate region.
3) Applying the preprocessing techniques detailed in II-B.
4) Concatenating all images, including the CZ and PZ.
5) Transposing the images for compatibility with 2D models, resulting in a format of Batch size x channels x H x W.
6) Training each model with a Dice-Focal Loss Function. We use the Novograd optimizer [25] and an initial learning rate of 0.001, running for 200 epochs.

Following this methodology, we obtain the lesion segmentation map for each model, facilitating accurate prostate lesion identification.

Training was conducted on an Nvidia Tesla V100 with 32GB of memory using PyTorch 2.0 and MONAI Core library version 1.1.0. These tools were chosen for their robustness in handling medical imaging analysis. Table 2 provides a breakdown of each model's complexity, capturing number of parameters, Multiply-Accumulate Operations (MACs), which means multiply and add two numbers, basic operations for many linear algebra operations, such as matrix multiplications, convolutions, and dot products; and an approximate of epoch training time. While SAM-UNETR stands out with higher values, it's justified given its versatile, large-scale encoder designed for broad segmentation tasks.

All code used for this project is available at https://github.com/BIMCV-CSUSP/SAM-UNETR

III. RESULTS

After the training process is completed, all models are evaluated on the test partition of the dataset, which contains 317 images from different patients. The evaluation procedure can be divided into two main phases: lesion level and patient level. The lesion level evaluation focuses on the ability of the model to detect and segment the entire lesion in the prostate region, regardless of its malignancy. To measure this ability, two common metrics are used: Dice Score and Intersection over Union (IoU) Score. These metrics compare the overlap between the predicted lesion mask and the ground truth lesion mask. The higher the Dice Score and IoU, the better the model's performance. These metrics are suitable for evaluating segmentation tasks, as they account for both the size and shape of the lesions. The patient level evaluation, on the other hand, concentrates on the ability of the model to detect csPCa. To do this, the model's predictions are compared with the biopsy results of each patient. If the patient has lesions that are confirmed to be csPCa by biopsy, then the model should also detect lesions in that patient. If the patient has no lesions, then the model should not detect any lesions in that patient. The Area Under Receiver Operating Curve (AUROC) is used as a metric for this phase. The AUROC measures how well the model can distinguish between patients with csPCa and patients with without csPCa (no-csPCa), based on their predicted lesion scores. The higher the AUROC, the better the model's performance. This metric is suitable for evaluating classification tasks, as it accounts for both true positives and false positives, Equation 3 shows how AUROC is calculated, where TPR(f) is the True Positive Rate

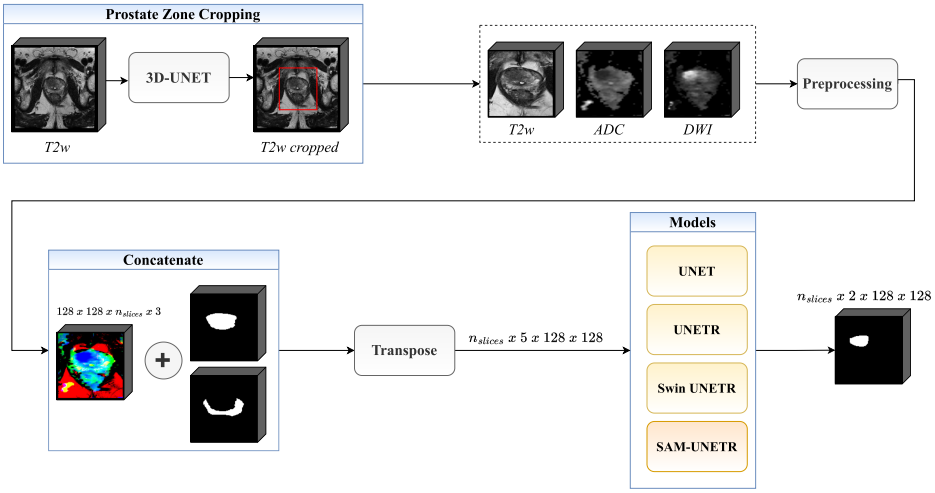

FIGURE 4. Methodology overview: Prostate segmentation on the T2w sequence followed by cropping. ADC and DWI sequences are then resampled, cropped, and preprocessed. Post concatenation, the images are transposed for 2D models. Training is done using Dice-Focal Loss and the Novograd optimizer over 200 epochs.

TABLE 2. Model complexity and epoch training time.

Model	Architecture Type	Parameters	MACs	Epoch Time
UNET	Convolutional Encoder and Decoder with Residual Connections	12.82 M	540.51 MMac	~ 10 min
UNETR	Transformer encoder and convolutional decoder	87.96 M	6.63 GMac	~ 13 min
UNETR	Swin Transformer encoder and convolutional decoder	25.14 M	4.94 GMac	~ 14 min
SAM-UNETR	Transformer encoder and convolutional decoder	635.97 M	66.57 GMac	~ 24 min

TABLE 3. General lesion level results.

Model	Mean Dice Score	Mean IoU
UNET	0.417	0.314
UNETR	0.376	0.275
SwinUNETR	**0.479**	**0.361**
SAM-UNETR pretrained	0.467	0.346
SAM-UNETR no-pretrained	0.447	0.33

(Sensitivity) at a given threshold f, d[FPR(f)] represents the change in False Positive Rate (1 - Specificity) as the threshold f changes. The integral sums up the TPR values over all possible thresholds, effectively calculating the area under the ROC curve.

$$\text{AUROC} = \int_{-\infty}^{\infty} \text{TPR}(f)\, \text{d}[\text{FPR}(f)] \tag{3}$$

A. LESION LEVEL

As previously stated, the metrics used for lesion level detection are Dice Score and IoU. Table 3 presents the results of these two metrics for each model. The best performing model is SwinUNETR, followed by SAM-UNETR pretrained, which outperforms the non-pretrained variant. These metrics reflect the performance of the model in detecting the entire lesion, regardless of its PIRADS score.

TABLE 4. Results based on each PIRADS score.

PIRADS	Model	Dice Score	IoU
PIRADS 2	UNET	0.428	0.312
	UNETR	0.410	0.297
	SwinUNETR	**0.478**	**0.355**
	SAM-UNETR pretrained	0.470	0.344
	SAM-UNETR no-pretrained	0.458	0.338
PIRADS 3	UNET	0.455	0.344
	UNETR	0.430	0.313
	SwinUNETR	0.496	0.376
	SAM-UNETR pretrained	**0.533**	**0.404**
	SAM-UNETR no-pretrained	0.490	0.366
PIRADS 4	UNET	0.498	0.377
	UNETR	0.490	0.360
	SwinUNETR	0.589	0.451
	SAM-UNETR pretrained	**0.624**	**0.491**
	SAM-UNETR no-pretrained	0.610	0.480
PIRADS 5	UNET	0.490	0.357
	UNETR	0.341	0.240
	SwinUNETR	0.469	0.341
	SAM-UNETR pretrained	**0.520**	**0.380**
	SAM-UNETR no-pretrained	0.438	0.317

In addition, the models are assessed on their performance in predicting lesions in each PIRADS category. For this purpose, only images with an assigned PIRADS score are considered, which implies that only human annotated images are utilized for this task. Table 4 presents the

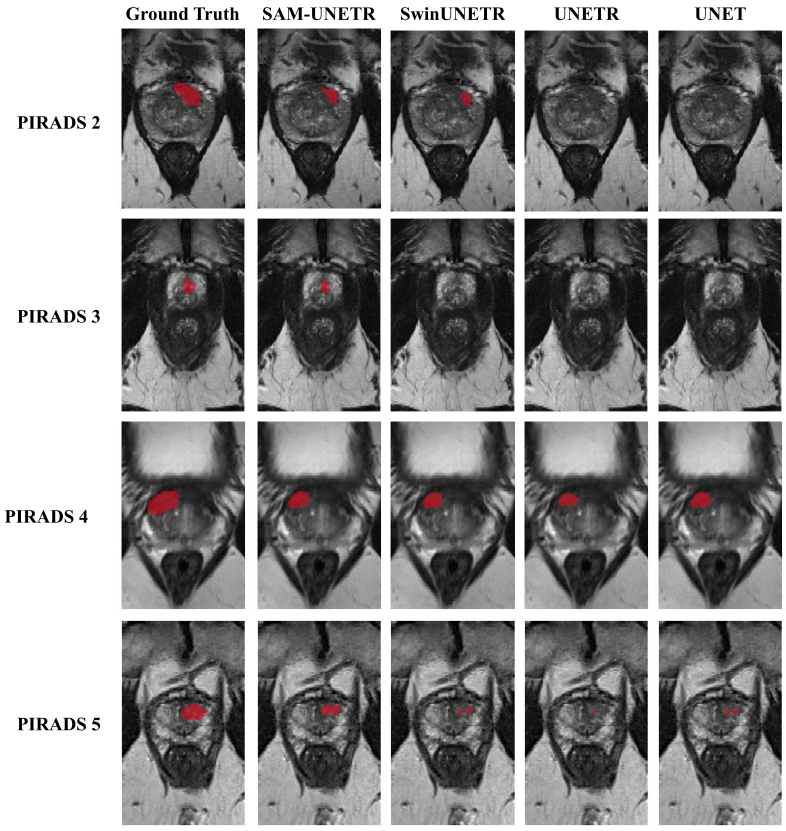

FIGURE 5. Lesion detection based on PIRADS score.When the PIRADS score is low, the detection of lesions becomes more challenging. However, the proposed SAM-UNETR algorithm exhibits a commendable detection rate for both difficult-to-detect lesions, characterized by low PIRADS scores, as well as relatively easier lesions such as those categorized as PIRADS 4 and 5.

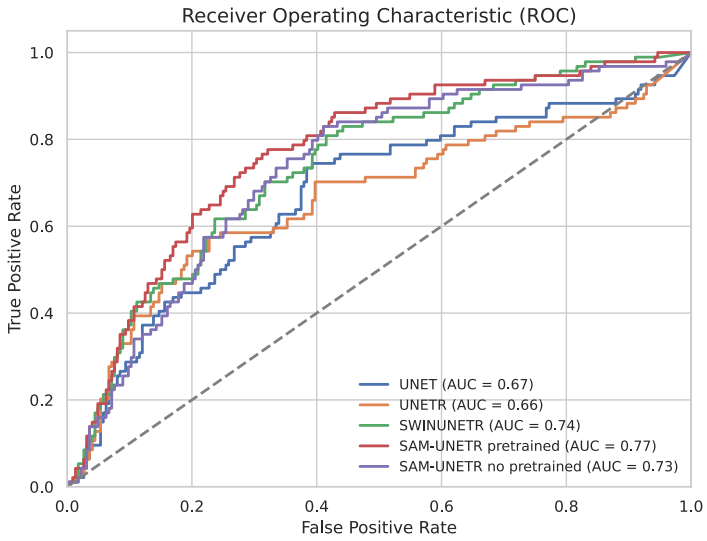

FIGURE 6. AUROC curve for each model. This curve accurately assesses the ability of a model to detect csPCa in images that genuinely exhibit csPCa. Remarkably, the proposed SAM-UNETR model surpasses all other models.

outcomes of each model on every PIRADS classification. SAM-UNETR was the most accurate model in almost all the PIRADS categories except PIRADS 2, where SwinUNETR achieved the highest score, albeit with a marginal difference. Furthermore, the pretrained version of the model consistently outperformed its non-pretrained counterpart.

Figure 5 illustrates the performance of the models on images with different levels of lesion severity, as indicated by the PIRADS score. The models are able to detect lesions even in images where the detection is challenging due to the low severity of the lesions, such as those corresponding to PIRADS 2 and 3. In PIRADS 2, only SAM-UNETR and SwinUNETR were able to detect the lesion, while in PIRADS 3, only SAM-UNETR detected the lesion. Moreover, SAM-UNETR can detect a larger area of the lesion compared to the other models in images where the lesion is more prominent and easier to detect, such as those corresponding to PIRADS 4 and 5.

B. PATIENT LEVEL

As previously stated, the models were also trained with no-csPCa images, which implies that no lesions were segmented in those images. Therefore, it is essential to evaluate the performance of the models in discriminating between csPCa and no-csPCa cases. Figure 6 illustrates the results of this evaluation, based on the AUROC metric. The higher the AUROC, the better the model's performance. Among the models, SAM-UNETR pretrained achieved the highest AUROC of 0.77, followed by SwinUNETR and its non-pretrained variant. These results demonstrate the superior ability of SAM-UNETR pretrained to detect csPCa lesions accurately and reliably.

IV. DISCUSSION

The presented results demonstrate that SAM-UNETR pretrained generally surpasses all other models in this work in the challenging task of csPCa lesion segmentation. Even though overall results show that SwinUNETR performs better in lession level segmentation, as evidenced by Table 3, when evaluating images labeled for human experts at each PI-RADS score, SAM-UNETR shows more robustness and better results in most categories compared to all other models as reported in Table 4. A possible explanation for this difference could be that SAM-UNETR is better tuned to understand and detect the subtle features associated with specific PI-RADS scores. The benefit of pre-training is evident here, suggesting that such a model may benefit from an inductive bias that aids its ability to handle complex patterns in the dataset.

Moreover, it is important to note that while segmentation metrics provide a detailed view of model performance, their translation into clinical utility remains to be seen. A patient-level evaluation provides a clearer perspective on the potential real-world performance of these models, particularly in differentiating csPCa from no-csPCa scenarios. At this point, SAM-UNETR achieves the highest AUROC score (Figure 6) in discriminating between csPCa and no-csPCa cases, indicating its reliability and effectiveness in detecting csPCa lesions. It is noteworthy that SAM-UNETR pretrained consistently outperforms its non-pretrained version, which validates the value of using a transfer learning strategy for this task, even when the pretrained weights originate from a different task and the decoder architecture differs from the original SAM model.

Regarding the presented results, comparing them directly to other studies is complex. Many factors can cause a model to perform differently, such as the dataset size, the metrics used, and the complexity of the methods. In particular, the work by Adams et al. on the Prostate158 dataset [17] used the same type of images and a simple 3D UNET model. They got a dice score of 0.453 but only used images with a PI-RADS score higher than 3 while our model methodology included a comprehensive range of PI-RADS scores. On the other hand, Bosma et al. [26] had access to the entire PICAI challenge dataset since they were the organizers of the challenge [18]. Their study included a large dataset of 6,578 MRI scans with a PI-RADS score of 4 or higher. They reported a high AUROC of 0.91 but used an ensemble of 15 models and a semi-supervised learning method, which makes their approach more costly and less practical for places with limited resources, such as some developing countries.

V. CONCLUSION

This study demonstrates the adaptability of existing model architectures through a transfer learning approach to address multiple tasks. Specifically, the SAM-UNETR architecture is proposed, leveraging the spatial representations of a transformer-encoder derived from a large pretrained model like SAM, in combination with a decoder utilizing residual convolutions of the UNETR. Multiple 2D models were trained using a consistent methodology for the purpose of clear comparison with SAM-UNETR.

Results show that the proposed SAM-UNETR architecture achieves a general dice score of 0.467 for lesion detection, demonstrating its effectiveness despite not being the best performing model. Further analysis shows that the proposed methodology outperforms alternative models in accurately identifying lesions labeled by human experts, as well as demonstrating superior performance across different PIRADS levels. In particular, SAM-UNETR exhibits an AUROC of 0.77 in the detection of csPCa, providing greater confidence in its lesion detection capabilities compared to other models. This shows that while our model may not delineate the entire lesion with perfect accuracy, its predictions can be invaluable in suggesting potential lesion locations. This may then serve as an additional tool for radiologists, allowing them to make more informed decisions based on their expertise and the model's indications.

Additionally, a comparison between the pretrained and non-pretrained versions of the model reveals the beneficial impact of the learned weights from the SAM encoder, trained on a large dataset. This highlights the value of employing a transfer learning approach, even when the task at hand and the original architecture significantly differ, particularly in challenging tasks such as those addressed in this study.

As previously discussed, direct comparison of detection percentages is inherently complex due to the variety of metrics used and the unique characteristics of the dataset.

While a larger image dataset may improve model performance, it is important to recognize that the adoption of complex methodologies may negatively impact clinical feasibility in terms of both resource allocation and time commitment. While the proposed methodology may not be the most superior approach documented in the literature, this work elucidates the efficacy of the transfer learning approach in conjunction with contemporary large-scale models. SAM-UNETR can potentially be applied in tandem with other methodologies, as it accommodates images of varying shapes due to the modifications made to the original encoder.

As future work, efforts will be directed towards training SAM-UNETR to enhance further the results pertaining to the detection and segmentation of csPCa lesions. Additionally, future studies on diverse datasets can further cement these findings and guide the development of clinically robust algorithms. An alternative decoder approach based on a transformer-decoder, akin to the original SAM model, will be explored.

ACKNOWLEDGMENT

The views and opinions expressed are those of the author(s) and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission is responsible for them.

REFERENCES

[1] (2020). *Global Cancer Observatory: Cancer Today*. [Online]. Available: https://gco.iarc.fr/today/home

[2] A. B. Rosenkrantz, S. Verma, P. Choyke, S. C. Eberhardt, S. E. Eggener, K. Gaitonde, M. A. Haider, D. J. Margolis, L. S. Marks, P. Pinto, G. A. Sonn, and S. S. Taneja, "Prostate magnetic resonance imaging and magnetic resonance imaging targeted biopsy in patients with a prior negative biopsy: A consensus statement by AUA and SAR," *J. Urology*, vol. 196, no. 6, pp. 1613–1618, Dec. 2016, doi: 10.1016/J.JURO.2016.06.079.

[3] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[4] P. P. Shinde and S. Shah, "A review of machine learning and deep learning applications," in *Proc. 4th Int. Conf. Comput. Commun. Control Autom. (ICCUBEA)*, Aug. 2018, pp. 1–6, doi: 10.1109/ICCUBEA.2018.8697857.

[5] A. Reuther, P. Michaleas, M. Jones, V. Gadepally, S. Samsi, and J. Kepner, "AI accelerator survey and trends," in *Proc. IEEE High Perform. Extreme Comput. Conf.*, Sep. 2021, pp. 1–9, doi: 10.1109/HPEC49654.2021.9622867.

[6] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/J.MEDIA.2017.07.005.

[7] M. Arif, I. G. Schoots, J. M. Castillo, M. J. Roobol, W. Niessen, and J. F. Veenland, "Computer aided diagnosis of clinically significant prostate cancer in low-risk patients on multi-parametric MR images using deep learning," in *Proc. Int. Symp. Biomed. Imag.*, Apr. 2020, pp. 1482–1485, doi: 10.1109/ISBI45749.2020.9098577.

[8] P. Mehta, M. Antonelli, H. U. Ahmed, M. Emberton, S. Punwani, and S. Ourselin, "Computer-aided diagnosis of prostate cancer using multiparametric MRI and clinical features: A patient-level classification framework," *Med. Image Anal.*, vol. 73, Oct. 2021, Art. no. 102153, doi: 10.1016/J.MEDIA.2021.102153.

[9] G. Lu and L. Zhou, "Localization of prostatic tumor's infection based on normalized mutual information MRI image segmentation," *J. Infection Public Health*, vol. 14, no. 3, pp. 432–436, Mar. 2021, doi: 10.1016/j.jiph.2019.08.011.

[10] M. Gibbons, O. Starobinets, J. P. Simko, J. Kurhanewicz, P. R. Carroll, and S. M. Noworolski, "Identification of prostate cancer using multiparametric MR imaging characteristics of prostate tissues referenced to whole mount histopathology," *Magn. Reson. Imag.*, vol. 85, pp. 251–261, Jan. 2022, doi: 10.1016/J.MRI.2021.10.008.

[11] H. Jia, Y. Xia, Y. Song, D. Zhang, H. Huang, Y. Zhang, and W. Cai, "3D APA-net: 3D adversarial pyramid anisotropic convolutional network for prostate segmentation in MR images," *IEEE Trans. Med. Imag.*, vol. 39, no. 2, pp. 447–457, Feb. 2020, doi: 10.1109/tmi.2019.2928056.

[12] Q. Zhu, B. Du, and P. Yan, "Boundary-weighted domain adaptive neural network for prostate MR image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 3, pp. 753–763, Mar. 2020, doi: 10.1109/TMI.2019.2935018.

[13] A. Ushinsky, M. Bardis, J. Glavis-Bloom, E. Uchio, C. Chantaduly, M. Nguyentat, D. Chow, P. D. Chang, and R. Houshyar, "A 3D-2D hybrid U-Net convolutional neural network approach to prostate organ segmentation of multiparametric MRI," *Amer. J. Roentgenology*, vol. 216, no. 1, pp. 111–116, Jan. 2021, doi: 10.2214/ajr.19.22168.

[14] D. Singla, F. Cimen, and C. A. Narasimhulu, "Novel artificial intelligent transformer U-NET for better identification and management of prostate cancer," *Mol. Cellular Biochemistry*, vol. 478, no. 7, pp. 1439–1445, Jul. 2023, doi: 10.1007/s11010-022-04600-3.

[15] Z. Dai, E. Carver, C. Liu, J. Lee, A. Feldman, W. Zong, M. Pantelic, M. Elshaikh, and N. Wen, "Segmentation of the prostatic gland and the intraprostatic lesions on multiparametic magnetic resonance imaging using mask region-based convolutional neural networks," *Adv. Radiat. Oncol.*, vol. 5, no. 3, pp. 473–481, May 2020, doi: 10.1016/J.ADRO.2020.01.005.

[16] J. C. Weinreb, J. O. Barentsz, P. L. Choyke, F. Cornud, M. A. Haider, K. J. Macura, D. Margolis, M. D. Schnall, F. Shtern, C. M. Tempany, H. C. Thoeny, and S. Verma, "PI-RADS prostate imaging–reporting and data system: 2015, version 2," *Eur. Urology*, vol. 69, no. 1, pp. 16–40, Jan. 2016, doi: 10.1016/J.EURURO.2015.08.052.

[17] L. C. Adams, M. R. Makowski, G. Engel, M. Rattunde, F. Busch, P. Asbach, S. M. Niehues, S. Vinayahalingam, B. van Ginneken, G. Litjens, and K. K. Bressem, "Prostate158—An expert-annotated 3T MRI dataset and algorithm for prostate cancer detection," *Comput. Biol. Med.*, vol. 148, Sep. 2022, Art. no. 105817, doi: 10.1016/J.COMPBIOMED.2022.105817.

[18] A. Saha, J. Bosma, J. Twilt, B. van Ginneken, D. Yakar, M. Elschot, J. Veltman, J. Fütterer, M. de Rooij, and H. Huisman, "Artificial intelligence and radiologists at prostate cancer detection in MRI—The PI-CAI challenge," in *Proc. Med. Imag. Deep Learn.*, 2023, pp. 1–5. [Online]. Available: https://openreview.net/forum?id=XfXcA9-0XxR

[19] L. Geert, D. Oscar, B. Jelle, K. Nico, and H. Henkjan, "Prostatex challenge data," Cancer Imag. Arch., 2017, doi: 10.7937/K9TCIA.2017.MURS5CL.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[21] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "UNETR: Transformers for 3D medical image segmentation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 1748–1758, doi: 10.1109/WACV51458.2022.00181.

[22] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002, doi: 10.1109/ICCV48922.2021.00986.

[23] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12962. Cham, Switzerland: Springer, 2022, pp. 272–284, doi: 10.1007/978-3-031-08999-2_22.

[24] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment anything," 2023, *arXiv:2304.02643v1*.

[25] B. Ginsburg, P. Castonguay, O. Hrinchuk, O. Kuchaiev, V. Lavrukhin, R. Leary, J. Li, H. Nguyen, Y. Zhang, and J. M. Cohen, "Training deep networks with stochastic gradient normalized by layerwise adaptive second moments," in *Proc. ICLR*, 2020, pp. 1–13. [Online]. Available: https://openreview.net/forum?id=BJepq2VtDB
[26] J. S. Bosma, A. Saha, M. Hosseinzadeh, I. Slootweg, M. de Rooij, and H. Huisman, "Semisupervised learning with report-guided pseudo labels for deep learning–based prostate cancer detection using biparametric MRI," *Radiol., Artif. Intell.*, vol. 5, no. 5, Sep. 2023, Art. no. 230031, doi: 10.1148/ryai.230031.

JESUS ALEJANDRO ALZATE-GRISALES received the B.Sc. degree in biomedical and electronic engineering from Universidad Autónoma de Manizales. Since 2018, he has been contributing as a Dedicated Member of the esteemed Research Group on Bioinformatics and Artificial Intelligence. Throughout his tenure, he has demonstrated exceptional competence in various projects, such as Applying Deep Learning Techniques in Digital Media Steganalysis, Detecting Alzheimer's From 3-D Magnetic Resonance, and Detecting Respiratory System Diseases From Chest X-Ray Imaging. Currently, he is a Data Scientist with the Biomedical Imaging Unit (UMIB), FISABIO, Valencia, Spain, where he plays a pivotal role in developing and executing projects centered around medical imaging. His profound expertise and relentless pursuit of innovation continue to contribute significantly to the advancement of medical science.

ALEJANDRO MORA-RUBIO received the B.Sc. degree (Hons.) in biomedical engineering and electronics engineering from Universidad Autónoma de Manizales (UAM), Colombia, in 2022. He has been a member of the Research Team on Bioinformatics and Artificial Intelligence, UAM, since 2018. Currently, he is a Data Scientist with the Biomedical Imaging Unit (UMIB), FISABIO, Valencia, Spain, where he plays a pivotal role in developing and executing projects centered around medical imaging. He has undertaken several research projects in digital image and signal processing, applying machine learning techniques. His current research interest includes healthcare applications of computer vision.

FRANCISCO GARCÍA-GARCÍA received the degree in statistical sciences and techniques and the Ph.D. degree in biomedicine and biotechnology from the University of Valencia, Spain. He has developed his technical and research career in different health institutions. During the last 20 years, he has been part of the team of the National Institute of Bioinformatics and the Bioinformatics Platform for Rare Diseases, generating new methods of functional enrichment analysis, designing and implementing tools for the analysis of omics, and clinical and biomedical image data. He has participated in numerous training activities in bioinformatics and computational biology, aimed at professionals and university students. He currently heads the Bioinformatics and Biostatistics Unit, Principe Felipe Research Centre, Valencia, Spain.

REINEL TABARES-SOTO received the first B.Sc. degree in electronic engineer from Universidad Nacional de Colombia, in 2009, the second B.Sc. degree in systems and computer engineer from Universidad de Caldas, Colombia, in 2016, the M.Sc. degree in electronic engineer from Universidad Nacional de Colombia, in 2017, and the Ph.D. degree in engineering from Universidad Autónoma de Manizales, Colombia, in 2021. He is currently a Professor with the Engineering Faculty, Universidad de Caldas and Universidad Autónoma de Manizales, Colombia. His main research interests include steganalysis, machine learning, deep learning, bioinformatics, and high-performance computing.

MARIA DE LA IGLESIA-VAYÁ received the Ph.D. degree from the Polytechnic University of Valencia. She is currently pursuing the Postdoctoral degree with the Max Planck Institute for Human Cognitive and Brain Sciences, Germany. She is leading the Brain Connectivity Laboratory, FISABIO-CIPF Centre. She is also the Head of the Biomedical Imaging Unit (UMIB), FISABIO, Valencia, Spain. Her expertise lies in high-performance computing for functional and structural imaging analysis. Her research focuses on investigating various neurological disorders, including Alzheimer's, schizophrenia, minimal hepatic encephalopathy, and utilizing imaging technologies. Additionally, she serves as the Spanish Delegate to Euro-BioImaging, contributing to population studies based on medical imaging in the Valencia Region.