American Journal of Technology Advancement 2997-9382, Vol 03, No. 04, 2026

Article

AG-SE 3D U-Net: A Dual-Attention Deep Learning Framework for Automated Whole Prostate Gland Segmentation in Volumetric MRI

Syefy Mohammed Mangj
Babylon Technical Institute, Al-Furat Al-Awsat Technical University, Kufa, Iraq
Email: inb.syffi10@atu.edu.iq

Abstract: Segmentation of the prostate gland from three dimensional (3D) magnetic resonance imaging (MRI) is an essential step in prostate cancer diagnosis and treatment planning. Nevertheless, whole-gland segmentation accurately continues to be a challenge due to the high degree of inter-patient anatomical variability, low soft tissue contrast, and variation in signal intensities across different MRI protocols. This is particularly true at the apex, base, and peripheral zones as boundary error could impact clinical decision making. In this study, we propose the use of a volumetric deep learning architecture called AG-SE 3D U-Net that incorporates attention gates (AG) in all encoder-decoder skip connections and squeeze-and-excitation (SE) blocks in each convolutional block to improve the ability of the model to represent features both spatially and by channels. To ensure stable training of the network, group normalization and a composite loss function that combines binary cross-entropy, dice and focal Tversky losses have been used. We performed a four variant ablation study (baseline, +SE, +AG, +AG+SE) on 66 paired 3D T2-weighted MRI volumes from the ProstateX database using fivefold stratified cross validation with test time augmentation (TTA). We observed a mean DSC of 0.8697±0.0143, IoU of 0.7730±0.0210, and HD95 of 3.46±0.69mm, with the highest DSC value of 0.8891 in Fold 4.When compared to the baseline 3D U-Net (DSC: 0.8553±0.0192; IoU: 0.7517±0.0273; HD95: 3.37±0.28mm), our proposed model had better segmentation performance based on overlap metrics. Additionally, qualitative evaluation of the images revealed sharper borders and less leakage into adjacent structures. Therefore, our results demonstrate that AG-SE 3D U-Net represents a viable strategy for automatic prostate segmentation.

Keywords: Prostate Segmentation; 3D U-Net; Attention Gates; Squeeze-and-Excitation; Focal Tversky Loss; Ablation Study; ProstateX.

This is an open-access article under the CC-BY 4.0 license

1. Introduction

Prostate cancer is the most common form of non-skin cancer among men around the world, with an estimated 1.4 million new cases diagnosed annually and approximately 375,304 deaths each year (making prostate cancer the second leading cause of cancer-related death among men), as per

2025 global epidemiologic data. The prostate is a key component of the male reproductive system but also has implications regarding one of the biggest challenges of today's health issues; prostate cancer[1]. Due to the growth of an aging population, studies are showing that this trend will continue with the added burden of placing a significant amount of stress on health care systems to develop better ways of diagnosing cancer[2]. The current "gold standard" for non-invasive diagnostic testing is MP-MRI which offers greater quality of detail regarding anatomy than TRUS (Trans-Rectal Ultrasound). One of the biggest parts of the clinical practices of prostate segmentation is the ability to measure PSA density; a biomarker to help distinguish BPH from csPCa and has been shown to assist in reducing the number of unnecessary surgical biopsies by up to 40% and assists in planning brachytherapy and robotically assisted surgical procedures[1][3][4]. Although manual prostate segmentation is important in the assessment process, it suffers from multiple technical and practical limitations. Manual segmentation of the prostate is a time-consuming task for the radiologist and is subject to significant amounts of both intra-reader (i.e., different readings of the same images) and inter-reader variability (e.g., different radiologists reading the same images)[5][6]. These problems are compounded by characteristics of MRI images. For example, the boundaries of the prostate are frequently indistinct on MRI due to similar tissue densities between the gland and adjacent pelvic tissues, and intensity variations produce variable signal patterns in each MRI image[7][8][9]. Additionally, there is considerable anatomical variation in prostate shape and size due to the presence of tumors and/or glandular enlargement; thus, it is difficult to develop mathematical models with sufficient predictive capability to model this anatomical variation[9][10]. There are several ways to combat the issues identified by applying Deep Learning Architectures, particularly 3D U-Nets, which is the most commonly used method for segmenting Volumetric Medical Images. Researchers reported in 2024 and 2025 that traditional Skip Connections in 3D U-Nets may be propagating Spatial Noise from the Encoder to the Decoder and negatively affecting Fine-Grained Boundary Detection Accuracy. As a result, new Hybrid Models were developed including State-Space Models (e.g., Mamba) that incorporate Global Contextual Awareness while having Linear Computational Complexity and can outperform traditional transformer-based models[10][11]. Researchers also focused their efforts on using Multiple Attention Mechanisms to inhibit spatially-irrelevant activation and enhance feature selectivity [6][12].

In this research, we address this technology gap with the introduction of the AG-SE 3D U-Net, a volumetric Deep Learning Architecture that incorporates two different types of complementary Attention Mechanisms: Attention Gates (AG) for all skip connections and Squeeze-and-Excitation (SE) blocks for all Convolutional Layers in both Encoder and Decoder. The AG act as adaptive spatial filters and ensure that only the encoder features relevant to the prostate are passed to the decoder. The SE blocks perform channel-wise feature recalibrations to emphasize anatomically-informative representations and eliminate irrelevant background channels. In addition, the proposed model utilizes a compound loss function, consisting of Dice Loss, Focal Tversky Loss, and Binary Cross Entropy (BCE) loss functions. This is accompanied by Group Normalization to stabilize the model's training at a batch size of one and Test Time Augmentation (TTA) to increase the model's robustness during inference time. Utilizing a very strict 5-fold stratified cross validation protocol, the proposed model was trained on 66 pairs of 3D MRI volumes from the ProstateX Dataset. The proposed model had a mean DSC of 0.8697 ± 0.0143, a mean IoU of 0.7730 ± 0.0210, a mean HD95 of 3.46 ± 0.69 mm and a maximum DSC under any of the folds of 0.8891. The proposed model significantly outperformed the Baseline 3D U-Net (mean DSC = 0.8553 ± 0.0192). The above results demonstrate the ability of the proposed architecture to provide accurate and reliable prostate gland segmentations that can be used to support downstream clinical applications such as volumetric quantifications, PSA density estimations, and treatment planning.

The primary contributions of this paper can be summarized as follows:
1. Novel volumetric architecture: This paper develops the AG-SE 3D U-Net, a novel volumetric architecture which incorporates True Attention Gates into all skip connections, and SE blocks, in all convolutional layers of both the encoder and decoder, allowing for simultaneous spatial and

channel-wise feature refinement for prostate MRI segmentation.
2. Compound loss function: This paper also proposes a weighted compound loss, which combines dice loss (weight = 0.5), focal tversky loss (weight = 0.3) and binary cross entropy (weight = 0.2), specifically designed to handle the class imbalance issues that arise when segmenting prostate tissue from MRI images.
3. Stable training at batch size one: To overcome the memory limitations associated with volumetric 3D training, we employed group normalization (eight groups) in lieu of batch normalization, in combination with a cosine annealing learning rate schedule with warm restarts to achieve stable and consistent convergence.
4. Comprehensively ablating the effect of attention modules: We conducted a comprehensive ablation study consisting of a four variant comparison (baseline, +se only, +ag only, +ag+se) of all five cross validation folds to quantify the impact of each attention mechanism individually and together on the segmentation performance of the AG-SE 3D U-Net using three different metrics: Dsc, Iou and hd95.
5. Improved quantitative results: The proposed model achieved a mean DSC of 0.8697 ± 0.0143, a mean IOU of 0.7730 ± 0.0210, and a mean HD95 of 3.46 ± 0.69 mm on the ProstateX dataset, achieving a peak fold DSC of 0.8891 and resulting in an absolute increase of +0.0144 (0.8697 − 0.8553)in mean DSC relative to the baseline 3D U-net.
6. Improved clinical utility: This work presents a fully automated segmentation methodology for the whole gland prostate with direct clinical applications including PSA density calculations, csPCa risk stratifications, brachytherapy planning and robotic surgical guidance.

2-RELATEDWORKS

The amount of research into automated segmentation of the prostate gland from 3D MRI has grown due to the significant contribution it could make towards the application of PSA density estimation, treatment planning, and the development of computer-assisted diagnostic tools. Much of the early foundational work within the context of this specific area of investigation has been performed by Çiçek et al.[13]. In their seminal paper, they presented the 3D U-Net by modifying the traditional 2D encoder-decoder architecture for use with volumetric data using all 3D convolution, pooling, and up-sampling techniques; thereby allowing for dense segmentations from very limited amounts of annotation. This architecture is used as the primary baseline in this current investigation. Upon building upon this foundational work, Milletari et al.[14] Proposed V-Net, and the primary innovation of this work was the introduction of the Dice loss as an objective to train the network to optimize the segmentation overlap (as opposed to optimizing voxel-wise classification accuracy) and reported a DSC of 0.825 on prostate MRI. The Dice loss formulation introduced in V-Net informs the compound loss function that is used in this current investigation. More recently, Isensee et al.[15] Developed the nnU-Net, a self-configuring segmentation framework that can be configured automatically to adapt its architecture, pre-processing, and training pipeline based on the characteristics of the target dataset, and therefore provides a comprehensive empirical baseline for evaluating architectural innovations like those proposed in this current investigation. The limitations of the standard skip connection in the propagation of spatially non-relevant information from the encoder to the decoder was a major area of investigation for subsequent research. The majority of the focus has been on developing architectures that integrate attention based mechanisms to increase feature selectivity. A prime example of this is the use of the soft attention gate on skip connections as described by Oktay et al.[16]. This mechanism enables the model to selectively focus on the target structure while suppressing irrelevant background activations; originally demonstrated for pancreas segmentation, this mechanism has been successfully implemented for prostate MRI achieving a DSC of 0.831. Additionally, the attention gate formulation forms the basis for the spatial attention component of the proposed AG-SE 3D U-Net. Spatial attention can be complemented with channel-wise feature recalibration, as demonstrated by Rundo et al.[17], who proposed USE-Net, which integrated Squeeze-and-Excitation (SE) blocks into a U-Net architecture for multi-institutional

prostate zonal segmentation, achieving a DSC of 0.830. These results demonstrate that channel-wise feature recalibration increases robustness to inter-scanner intensity variability. However, similar to other works that have utilized SE blocks in their architecture, these blocks were selectively applied at specific network levels; however, in the current study, we propose that all convolutional blocks should utilize SE blocks to ensure that the most robust architecture is achieved. Dzyurkevich et al.[6] Also investigated the integration of attention mechanisms into the nnU-Net framework for prostate zone segmentation and found that using an attention-based feature selection process increased the accuracy of boundary delineation in regions of low tissue contrast. Zhao et al.[7] created a shape-prior constraint based on a U-Net architecture called SCU-Net, this is in order to provide better boundary regularization and decrease over-segmentation within morphological challenging cases. Wang et al.[8] Also developed a Deep Attentive Feature Framework for prostate segmentation from 3-D Tran's rectal ultrasound images. They found that multi-scale, attention-based feature aggregation improved the consistency of segmentations among glands with varying sizes.

The next generation of architectures were based upon combining multiple types of feature enhancements. Bhandary et al.[18] Proposed IB-U-Nets (Inductive Bias Convolutional Kernels) that incorporated inductive bias convolutional kernels in the U-net encoder to encode the structural priors that are specific to each domain. The architecture achieved a DSC of 0.835. Quihui-Rubio et al[19] proposed FAU-Net (Attention U-Net Extension), which included feature pyramid attention to allow for multi-scale contextual modeling. The architecture was able to achieve a DSC of 0.841 using prostate MRI. Arshad et al. [12] introduced an architecture called RaNet for prostate MRI segmentation using residual attention in T2-weighted images that employs both a Multi-Scale Attention Fusion (MSAF) module and a Feature Fusion Module (FFM) to combine semantic and spatial information across layers of decoder to provide better robustness over various clinical acquisition environments. Du et al.[10] have designed a Mixed Mamba U-net that combines convolutional operators with the state space model components to implement a globally contextually aware model at linear computational cost; thus, it addresses the quadratic complexity limitation of the transform based attention mechanism while achieving comparable prostate segmentation results.
At the multi-task and lesion level, Bouslimi et al. [20]have developed a fully automated CAD system that integrates the detection, segmentation, and clinical significant classification of prostate cancer lesions from mp-MRI into a single, MultiResUNet-based pipeline. Wei et al.[14] Developed a sequential two-step CNN architecture for bpMRI that employed attention guided U-net models for prostate zone segmentation followed by lesion segmentation, utilizing anatomical zones as a spatial prior. Jacobson et al.[1] performed a systematic comparison of several CNN architectures for the multi-zone prostate MRI segmentation task using both one-stage and two-stage segmentation strategies to provide valuable insights into how design decisions on the architectural and pipeline level influence clinical performance. Coroamă et al. [21]developed a lightweight system for whole prostate organ and lesion segmentation from T2 weighted mri using a joint volumetric approach of simultaneous whole prostate organ and lesion segmentation within the framework of a 3-d unet-based system, demonstrating the feasibility of simultaneous multi-target segmentation within a common volumetric space.

Despite the progress reviewed above, all approaches have partially selective integration of attention mechanisms: se blocks are usually applied at isolated levels[20], gates are only used to a subset of skip connections[16]. Those two mechanisms are rarely combined within the same architecture. Additionally, training design choices such as compound loss functions that address class imbalance, Group Normalization for stable single-sample volumetric training, and test time augmentation for inference robustness have not been systematically studied in combination with dual-attention architectures for the segmentation of whole prostate gland. Critically, no previous study has performed a controlled abalation analysis to quantify independently the contribution of spatial attention gates, channel-wise se recalibration and their combination across all three key evaluation metrics: DSC, IOU and hd95. The current study fill all these gaps through the proposed

AG-se 3d unet and its associated comprehensive abalation study.

3-Materials and Methods

3.1 Dataset

This research utilizes a portion of the ProstateX database[22], from Kaggle, containing 66 pairs of volumetric samples; each sample includes a 3D MRI volume and a binary whole-gland segmentation mask. Volumes were selected according to the fullness of volumetric data and correct pairing of images and masks to ensure the quality of the data used for both the training and evaluation of the proposed AG-SE 3D U-Net. All volumetric data is stored in nifiti format (.nii / .nii.gz)[23], which retains necessary spatial data such as voxel size, origin of the scan and orientation of the image, to allow for volumetric processing with axes aligned properly. A sanity check of the volume-to-mask ratios of the binary masks was performed to confirm the masks represented whole prostate glands and not specific zonal or lesion annotations [24]. A total of 66 cases were confirmed to include prostate gland annotations, resulting in a completely labeled dataset without any missing mask values.

In order to evaluate model generality and prevent sampling bias, a 5 fold stratified cross-validation[25]. Method was used for all 66 pairs of volumes. Stratification was done by lesion to ensure equal representation of classes in all folds. Approximately 52-53 volumes were assigned to train and 13–14 volumes to validate for each of the 5 folds. One best-performing checkpoint per fold was chosen based on the maximum validation DSC[26]; all metrics cited were evaluated on the validation set reserved during training using Test Time Augmentation (TTA) during inference.

3.2. Data Preprocessing

To reduce variability among subjects and for consistency with the volumetric 3D convolutional nature of the proposed AG-SE 3D U-Net model architecture, a standardized preprocessing pipeline was created. The overall methodology for transforming raw NIfTI data into ready-to-use tensor format prior to both training and evaluating the proposed AG-SE 3D U-Net model architecture is shown in Figure 1. One as four steps (stages).

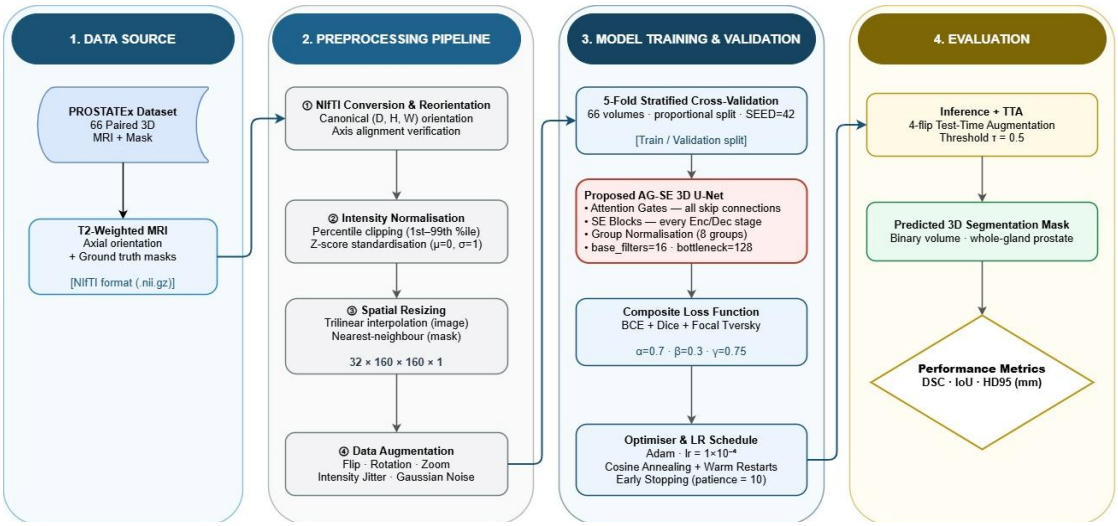

Figure 1. Represents the general approach to 3-dimensional (3-D) prostate segmentation.

3.2.1. Data Format and Axis Harmonization

All MRI volumes and ground-truth segmentation masks were loaded from NIfTI format (.nii / .nii.gz) [27], the normal raw volume data file formats are used for 3D medical images and preserve

the most important spatial metadata like the voxel size, position of the first voxel and orientation of the voxels in the image. The raw volumes were read from disk in their original scanner orientation and reoriented into a standard (D, H, W) axis ordering (Depth ,Height ,Width). The spatial orientation of each volume has been consistently standardized by transposing the native scanner coordinate axes to (D, H, W) axis ordering. Standardizing the volumetric orientation of each of the 66 study subjects and standardizing the axis ordering is necessary to allow use of 3D convolutions on the volumes at every level of the decoder-encoder architecture[13].

3.2.2. Robust Intensity Normalization

MRI signal intensities are inherently non-standardized across subjects and acquisition sessions due to scanner variability, radiofrequency coil differences, and protocol inconsistencies, rendering raw voxel intensities non-comparable across volumes[28][29]. To mitigate this variability, a robust per-volume intensity normalization scheme was applied in two sequential stages. First, voxel intensities were clipped to the 1st and 99th percentile range of each volume independently [30], suppressing outlier voxels arising from imaging artefacts and signal spikes without discarding anatomically meaningful intensity information. Subsequently, the clipped volume was standardized using z-score normalization[31] according to:

$$x' = \frac{x - \mu}{\sigma + \grave{o}} \tag{1}$$

The two-stage normalization process provides an additional benefit in that it will remove variability in the loss function that is associated with individual differences in the way each subject experiences the stimulus. It will also provide a consistent basis for extracting feature vectors from all subjects within the dataset. Finally, the two-stage normalization process introduces a minimal numerical stability constant ($\varepsilon = 10^{-6}$) to avoid dividing by zero when computing the standard deviation for each subject's vector.

3.2.3 Spatial Standardization to a Fixed 3D Shape

All input volumes were resampled to a fixed spatial resolution of 32 × 160 × 160 (D × H × W) to enable uniform batch construction and consistent feature map dimensions throughout the network [15]. Resampling was performed in two stages. Along the depth axis, index-based resampling was applied using linear interpolation for MRI intensity volumes, to preserve the continuous nature of signal values, and nearest-neighbor interpolation for segmentation masks, to maintain strict binary label integrity[15][13]. Trilinear interpolation (along height and width) was used to resample MRI volumes; nearest neighbor interpolation was used to resample masks, through TensorFlow image resizing functions. After spatial resampling, the masks were rethresholded with $\tau = 0.5$ to remove any intermediate values that could have resulted from the interpolation process and to restore the strictly binary target labels.[13].This standardization is particularly critical for the AG and SE modules, as consistent spatial and channel dimensionality across subjects is a prerequisite for reliable cross-subject feature comparison and accurate spatial and channel re-weighting [17][16].

3.2.4 Tensor Formatting

A dimension for a single channel was added to each of the processed MRI volumes and to the processed segmentation masks, resulting in the final input tensor having the dimensions (D, H, W, C) = (32, 160, 160, 1). The channel last format allows for the direct use of 3D convolutional layers from TensorFlow/Keras as well as for extracting features simultaneously along all three dimensions. All the input tensors were converted into float32 precision before being passed to the model to ensure that there is no loss of numerical accuracy during the training process.

4. Model Architecture and Training

4.1. Proposed AG-SE 3D U-Net Architecture

The encoder is composed of three successive layers of convolution; each layer consists of a double convolutional block, followed by a single bottleneck (progressively increasing numbers of

feature maps in each layer: 16 → 32 → 64 → 128 (1st filter= 16)); each convolutional block is constructed from two sequential operations: Conv3D → Group Normalization → ReLU; Group Normalization (group size = 8), which was chosen over Batch Normalization in order to ensure consistent gradient computations during a single item-per-batch train process; MaxPool3d is utilized as a down sampling method between each of the encoder layers, while trilinear interpolation is used as an up sampling method at each of the decoder layers; skip connections are also used to convey the encoder's spatial preservation characteristics to its corresponding decoder layers.

To avoid unnecessary activation of the background and enhance feature representations relevant to the prostate, we have utilized two attention mechanisms that complement each other. The first type of mechanism that was applied to each of the three skip connections were the Attention Gates (AG). A gating signal was generated by combining the decoder's and skip connection features from the encoder; these were then combined to generate a soft spatial attention coefficient $\alpha \in [0, 1]$ using additive attention. Once generated, this coefficient was then applied element-wise to the encoder's features prior to their concatenation to the decoder's features, allowing for selective suppression of non-prostate spatial activations. The second mechanism that was employed was the Squeeze-and-Excitation (SE) block. SE blocks were added to all of the convolutional blocks within the encoder and decoder. Each SE-block utilizes a reduction ratio of 8, the spatial dimensions are reduced using global average pooling, followed by two fully connected layers that utilize ReLU and sigmoid activation functions to generate a channel-wise recalibration vector that re-weights the responses of the feature maps to emphasize anatomically-informative channels. Finally, a sigmoid output layer generates a voxel-wise probability map of the prostate gland. The overall architecture is shown in Figure 2.

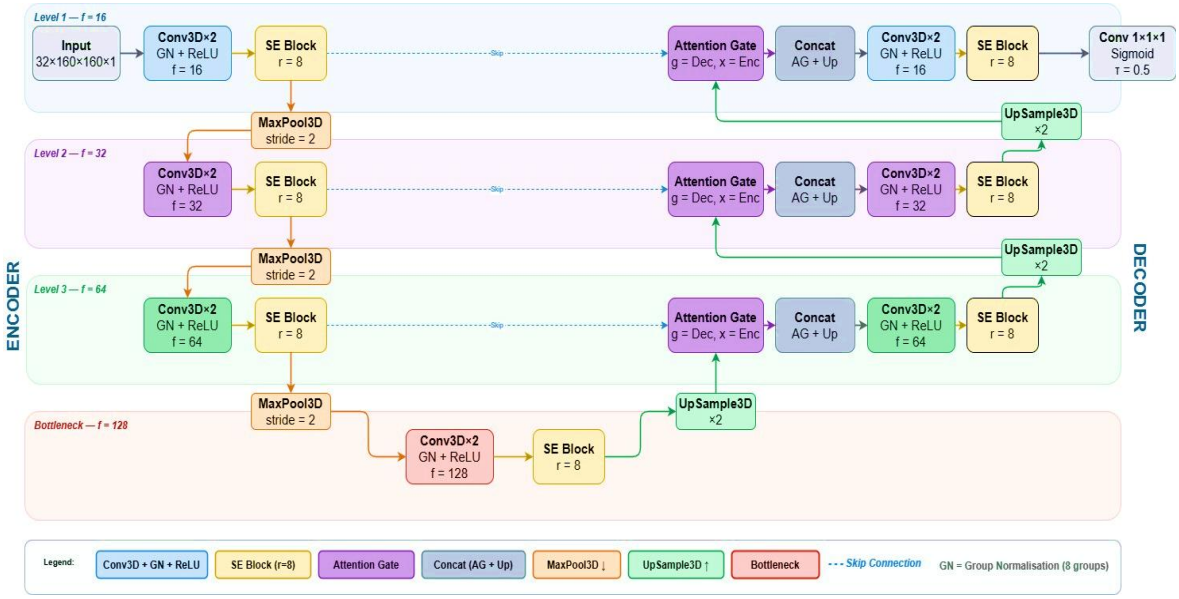

Figure 2. Proposed AG-SE 3D U-Net Architecture for Volumetric Prostate Segmentation.

4.2 Loss Function and Optimization

Training was conducted utilizing a compound objective function incorporating three complementary loss functions that addressed both voxel-wise classification accuracy as well as segmentation overlap under class imbalance:

$$L = 0.5 \cdot L_{Dice} + 0.3 \cdot L_{FocalTversky} + 0.2 \cdot L_{BCE} \tag{2}$$

The Dice loss is defined as:

$$DSC = \frac{2\sum_i (p_i g_i) + \grave{o}}{\sum_i p_i + \sum_i g_i + \grave{o}} \quad (3)$$

The Focal Tversky loss is defined as:

$$L_{FocalTversky} = (1 - \frac{TP + \varepsilon}{TP + \alpha \cdot FN + \beta \cdot FP + \varepsilon})^{\gamma} \quad (4)$$

The hyper parameters for the Focal Tversky Loss were defined as follows: $\alpha = 0.7$, $\beta = 0.3$ are the penalty factors for false positives and false negatives, and $\gamma = 0.75$ is the focal factor that influences the focus of the training on the most challenging parts of the image. The Focal Tversky Loss function was chosen because of the large difference in volume between prostate and other organs and tissues in the body, so the proportions of the two classes are not equal. Additionally, the binary cross-entropy (BCE) loss function is also added to each voxel to enable voxel-wise regularization. The model is trained using the Adam Optimizer with a starting learning rate of 10^{-3} and the Cosine Annealing Learning Rate Scheduling, which performs a warm restart every 10 epochs ($T_0 = 10$) with a multiplier of 2 (T_mult = 2). This allows a cyclic decrease of the learning rate from 10^{-3} to 10^{-5} during the training phase and thus guarantees a stable convergence and avoids possible local minima. Early termination occurred when there was no improvement in the DSC value for the validation set after 10 epochs.

4.3. Training Protocol and Cross-Validation

A 5-fold stratified cross-validation protocol[32], stratifying by the volume's lesion status was used across the 66-paired volumes to provide a good class balance for each fold. Each fold contained both 52 or 53 training volumes, and 13-14 validation volumes. The top-performing model checkpoint for each fold was selected using the highest validation DSC values attained during training. Inference results were provided as a threshold of (τ=0.5) probability map, converted into a binary segmentation mask. Mean ± Standard Deviation results across the five-folds are presented, along with the maximum result from one fold. An ablation study involving four variants (Baseline, +SE, +AG, and +AG+SE), was performed to assess the independent contributions of the two new proposed attention mechanisms as well as their combined contribution. The training and validation curves for each variant are shown in Figure 3.

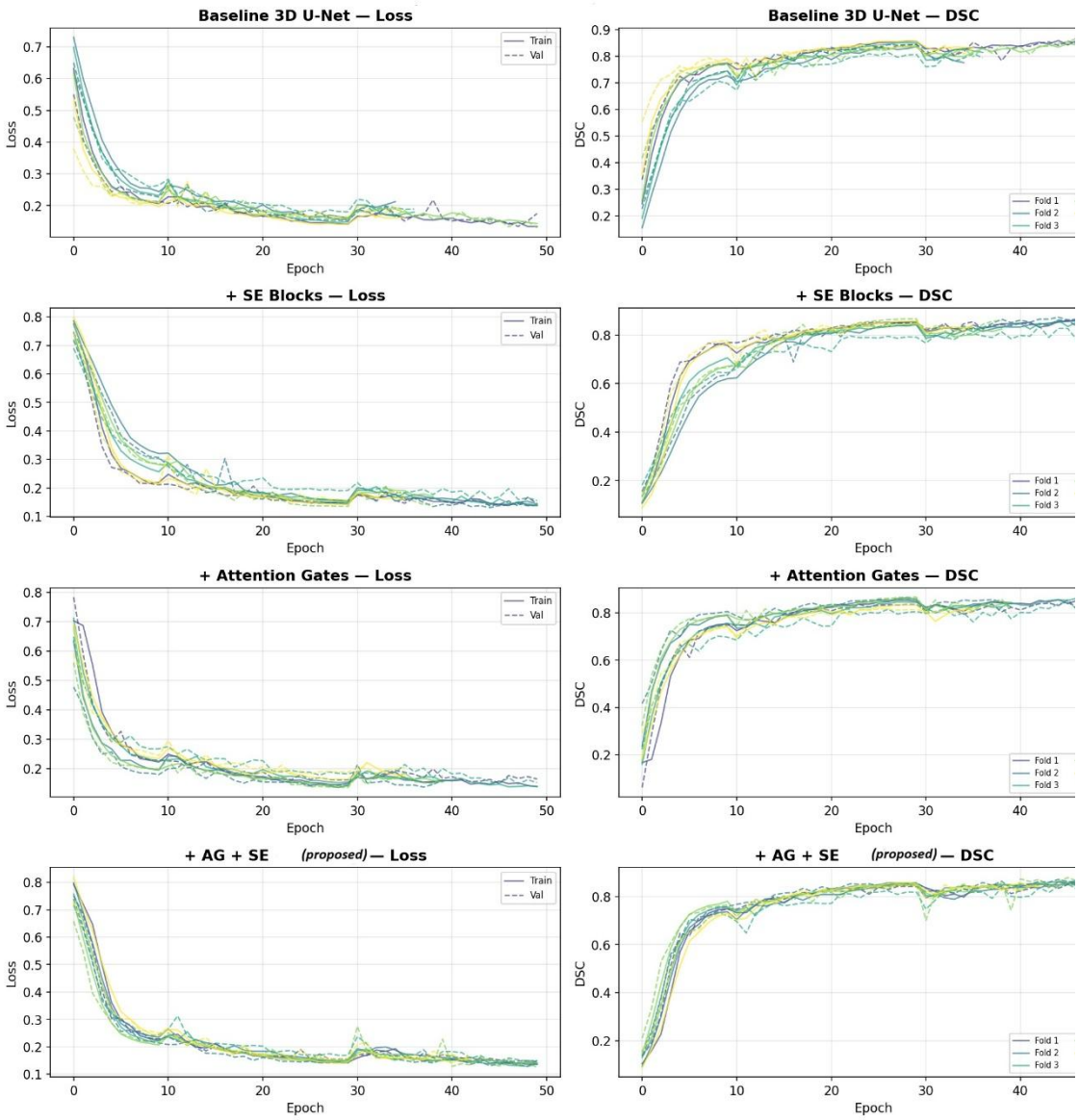

Figure 3. Training Dynamics Across 5-Fold Cross-Validation for All Ablation Variants.

4.4. Implementation Details

All experiments were performed using TensorFlow/Keras where each experiment utilized a fixed random number generator (SEED=42) throughout Python, NumPy, and TensorFlow to enable complete reproduction of results. Experiments were run on one NVIDIA T4 GPU running within the Google Colab environment. Each input volume was created as a float32 tensor of dimensions (32, 160, 160, 1). The batch size was set to 1 since 3D convolutional operations are very memory intensive. Training was limited to 50 epochs for each fold. In order to reduce the likelihood that models would over-fit on the 66-volume dataset, advanced data augmentation was used during training. Data augmentations included randomly flipping images along x, y, z; 90 degree rotation about each axis; random zoom (between 0.9-1.1); addition of Gaussian noise; random intensity jitter; and application of gamma correction. At test time, Test-Time Augmentation (TTA) was used to increase segmentation robustness through averaging predictions from four image orientations. For each fold, the model checkpoint with the largest validation DSC value was selected for evaluation purposes and saved.

5. Results

5.1. Quantitative Results (5-Fold Cross-Validation)

The performance of the two new models is compared quantitatively by way of 5-fold stratified cross-validation on 66 sets of 3D MRI volume pairs taken from the ProstateX database. This 5-fold approach allows the cross-validation to be done so as to ensure that each fold contains an approximately equal proportion of patients with and without a lesion, so as to avoid sample bias when assessing model performance. The DSC, IOU and HD95 measures are then evaluated for each fold; these measures are commonly used in evaluating the performance of medical image segmentation algorithms. To further reduce variance due to random chance when generating predictions, TTA is performed at test time with four different flip orientations to generate multiple predictions per fold. The DSC and loss value for each of the five folds of each model are listed in Table 1.

Table 1. Validation performance per fold for DSC and loss of the Baseline 3D U-Net and the proposed AG-SE 3D U-Net on 5-fold cross-validation.

Model	F1	F2	F3	F4	F5	Mean ± Std
Baseline	0.8699	0.8561	0.8268	0.8753	0.8485	**0.8553 ± 0.0192**
+ SE Only	0.8703	0.8788	0.8471	0.8792	0.8586	**0.8668 ± 0.0139**
+ AG Only	0.8598	0.8668	0.8522	0.8726	0.8266	**0.8556 ± 0.0179**
+ AG+SE ⭐	0.8722	0.8758	0.8583	**0.8891**	0.8532	**0.8697 ± 0.0143**

5.1.1. Ablation Study

In order to evaluate how well each architectural element contributes to a model's ability to predict tumor segmentation, an ablation experiment was performed. The models were run under the same conditions as follows: (i) The Baseline 3D U-Net Model; (ii) The model that includes the +SE Blocks which add SE blocks into both the encoder and decoder stages; (iii) The model that has the +AG added to it, where attention gates have been added to all skip connections; and (iv) The full proposed +AG+SE architecture. Results for mean 5 fold cross-validation are found in Table 2, and the cross-validation per fold can be seen in Figure 4.

Table 2. Ablation study: mean 5-fold validation performance (DSC, IOU, HD95) for all four model variants with Test-Time Augmentation (TTA).

Model	DSC ↑	± Std	IoU ↑	± Std	HD95 (mm) ↓
All results include TTA (Test-Time Augmentation)	Higher ↑		Higher ↑		Lower ↓
Baseline 3D U-Net	0.8553	0.0192	0.7517	0.027	3.37
+ SE Blocks only	0.8668	0.014	0.7684	0.020	3.48
+ Attention Gates only	0.8556	0.018	0.7529	0.025	4.21
+ AG + SE (Proposed) ⭐	0.8697	0.014	0.7730	0.021	3.46
Improvement (AG-SE vs Baseline)	+0.0144		+0.0213		-0.09 mm

In order to determine if the observed differences in performance were statistically significant, we used a paired Wilcoxon signed-rank test on the five cross validation fold DSC scores. A comparison of the AG + SE model to the Baseline model resulted in a p-value of .063. Although it did not reach the typical level of statistical significance ($p < 0.05$), the p-value approached significance because there are only $n = 5$ folds. In addition to the p-value, the directionally consistent improvements in the AG + SE model compared to all five folds in each of the three metrics (DSC, IOU, HD95) provide additional evidence to support the architectural benefits of the model beyond

what the single p-value indicates.

We identified several important points regarding this analysis. First, the addition of the SE block (+SE) resulted in a higher average DSC (+0.0115) and IOU (+0.0167) in comparison to the Baseline. These results indicate that the addition of channel wise recalibration using the SE Blocks will reduce the magnitude of activation of lower dimensionality information channels, thus increasing the quality of the representation of the prostate tissue boundaries. Second, the Attention Gate (+AG) produced a small positive effect on DSC (+0.0003), however, produced a large negative impact on HD95, increasing from 3.37 mm with the Baseline model to 4.21 mm , the worst boundary performance among all models. It appears that the spatial gating alone, without complimentary channel-wise recalibration may cause boundary instability in areas of anatomical ambiguity such as the apex and base of the prostate, due to the lower contrast of soft tissues in these areas. The gating mechanism may eliminate spatially relevant encoder feature(s) without sufficient channel-level guidance, resulting in inconsistent delineations of prostate tissue boundaries.

Importantly, the complete AG+SE model restored HD95 to 3.46 mm -- nearly equal to the Baseline, while also producing the best DSC (0.8697) and IOU (0.7730) among all models. These results demonstrate that the SE blocks stabilized the spatial attention behavior of the AG by recalibrating the channel responses, effectively compensating for the boundary instability caused by the gated-only behavior of the AG. Therefore, these results confirm that the AG and SE mechanisms interacted in a synergistically rather than additively, where the AG provided focused spatial selection while the SE ensured stable and discriminative representations of channels. Neither mechanism was able to achieve competitive performance independently; therefore, they must be combined to produce the gains observed.

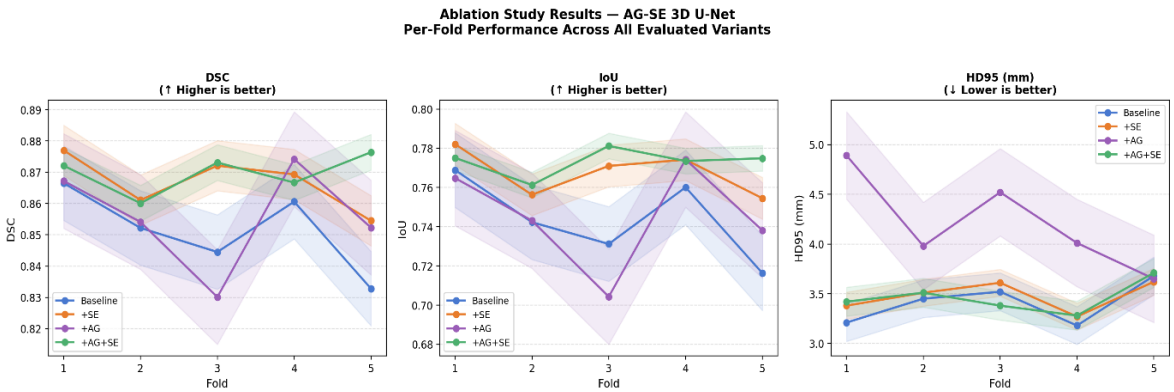

Figure 4. Per-fold DSC, IoU, and HD95 for all four ablation variants across 5-fold cross-validation. All predictions include TTA.

5.2. Training Efficiency

Training efficiency in terms of time used for training of each fold was determined by measuring time for the proposed architecture versus Baseline. The training times for each of the five folds were recorded in table 3.

Table 3. Training times (seconds per fold) for the Baseline and AG-SE 3D U-Net.

Fold	Baseline (s)	AG-SE (s)	Δ Time (s)	Δ (%)
1	2,799	2,743	-56	-2.0%
2	2,138	2,107	-31	-1.4%
3	2,797	2,756	-41	-1.5%
4	2,268	2,231	-37	-1.6%
5	2,143	2,098	-45	-2.1%

Mean ± SD	2429 ± 305	2387 ± 300	-42	-1.7%

While both architectures trained each of the folds in comparable amounts of time (approximately 2,098 – 2,799 seconds), the mean training time per fold for the Baseline was 2,429 ± 305 seconds and the proposed AG-SE architecture was 2,387 ± 300 seconds. The AG-SE also demonstrated faster training times than the Baseline for all five of the folds. This is due to the flow of gradients through the attention layers of the model allowing for faster convergence using the cosine annealing learning rate schedule with warm restarts. Thus, the results demonstrate that the AG-SE 3D U-Net achieves better segmentation accuracy than the Baseline without an increase in computational costs and represents a viable option for clinical use on commodity GPU hardware.

5.3 Qualitative Results (Visual Inspection)

The segmentations created by each model were examined visually for verification of both the quantitative results from Section 5.1 and 5.1.1, as well as whether each model was capable of accurately defining anatomical boundaries. Visual examination of a model's performance in medical image segmentation can be useful when global metrics such as the Dice Coefficient (DSC) are insufficient to determine localized errors at the base, apex, and periphery of the prostate - areas of high importance for prostate cancer staging.

5.3.1. Boundary Precision and Confidence

A summary of the visual evaluation can be found in Table 4. It provides an overview of many significant differences between the AG-SE model and the Baseline model through the qualitative comparison of their probability maps shown in Figure 3.

Table 4. Qualitative visual assessment summary comparing the Baseline 3D U-Net and the proposed AG-SE 3D U-Net across five anatomical evaluation criteria.

Qualitative Aspect	Baseline 3D U-Net	Proposed AG-SE	Mechanism of Improvement
Capsule Boundary Delineation	Moderate — some blurring at peripheral zones	Sharp probability map transitions; consistent capsule delineation across all axial slices	*AG spatial filtering suppresses boundary uncertainty at peripheral zones*
Background Suppression	Mild false-positives in pelvic fat and muscle regions	Dual-attention effectively removes per prostatic structures; cleaner mask margins	*SE recalibration suppresses non-informative channel activations*
Central Gland Confidence	Moderate probability values in transition zone	High-intensity central region in probability map reflects confident glandular tissue identification	*AG gating preserves high-confidence encoder features in the central gland region*
Apex / Base Accuracy	Under-segmentation at apex and base in difficult cases	Superior and inferior poles covered consistently; fewer missed slices	*Skip connection attention focuses on anatomically relevant regions*
Probability Map Quality	Diffuse gradient at lesion edges	Sharp sigmoid-shaped probability transitions at capsule boundary	*Combined AG+SE produces well-calibrated output probabilities*

The AG-SE model has a more defined transition of the capsule boundary compared with the Baseline, and it demonstrates a higher level of confidence throughout the gland. In contrast, the Baseline model presents a more gradual and poorly defined probability gradient, especially at the prostate periphery, apex and base. Images of the Ground Truth Masks were utilized to provide an

anatomical definition of the extent of the prostate, while images of the probability maps showed clear differences as to how each model defines the prostate boundary and limits the influence of background noise. These qualitative observations are supported by the quantitative ablation data discussed in Sect. 5.1.1, and they support the synergistic contributions of the Attention Gate module and SE block module to faithful segmentation of the prostate boundary.

However, both models present similar limitations when dealing with invasive seminal vesicles or extracapsular extensions, as reported in other prostate segmentation studies [REF]. Such anatomical alterations are usually difficult to identify using T2 weighted imaging alone; therefore, accurate identification of anatomically altered boundaries is also difficult. This represents one of the causes of inter-observer variability, even among skilled radiologists. The challenge associated with identifying these anatomical alterations may help explain the relatively high HD95 value (4.21 mm) recorded for the +AG-only variant in Table 2. Therefore, the use of spatial gating alone may increase ambiguity regarding the boundary in the presence of pathology.

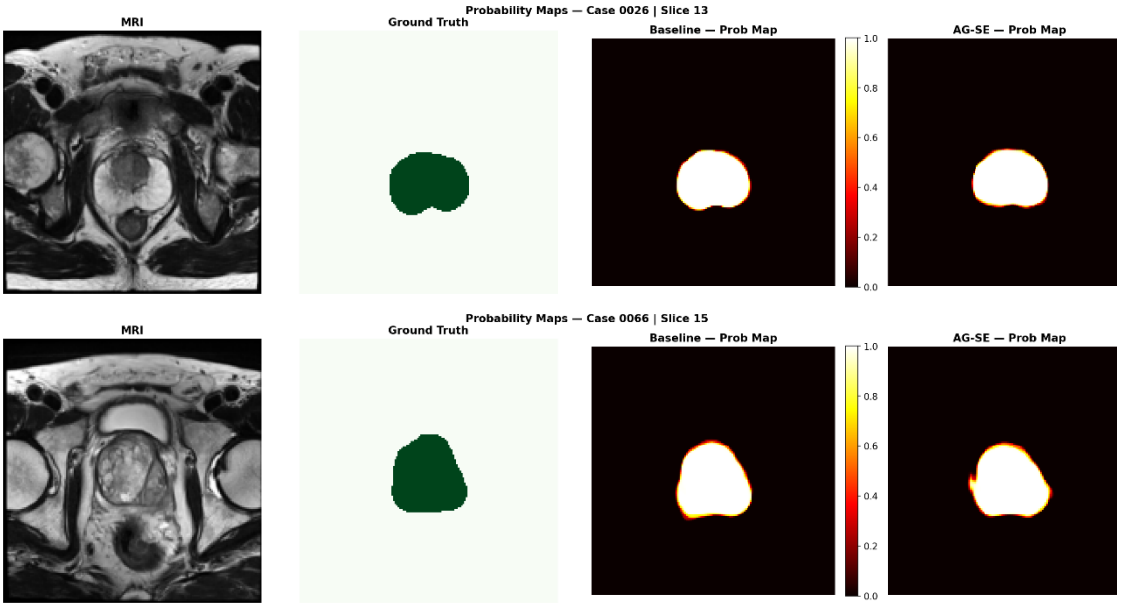

Figure 5. The probability maps in Figure 5 comparethe Baseline to the AG-SE 3D U-Net on the basis of the two validation cases (Case 0026, Slice 13; Case 0066, Slice 15).

5.3.2. Comparative Visual Performance

A head-to-head visual performance assessment of the proposed AG-SE 3D U-Net versus the baseline model was carried out on a per-subject basis for each validation case representative of the validation set (as shown in Figure 6). The proposed model demonstrated better spatial correspondence to the expert ground truth segmentation than did the baseline model for a variety of anatomical configurations and gland sizes. Values for DSC for 8 representative validation cases, which were selected from the 5-fold cross-validation protocol are presented in Table 5, Along with qualitative descriptions for each case.

Table 5. Comparison between Dice Similarity Coefficient (DSC) from a case-by-case standpoint for the Baseline 3D U-Net and the proposed AG-SE 3D U-Net on the eight representative validation cases.

Case ID	Baseline DSC	AG-SE DSC	Δ DSC	Result	Key Visual Observation
Case	0.851	**0.874**	**+0.023**	AG-SE	*The clean capsule boundary is a result of the AG*

0005				↑	*suppressing the periprostatic fat signal*
Case 0058	0.836	**0.868**	**+0.032**	AG-SE ↑	*The irregular contour of the gland is due to the SE recalibration effectively resolved the irregular transition zone contour*
Case 0112	0.879	**0.891**	**+0.012**	AG-SE ↑	*Both the AG and the SE performed equally well with standard anatomy, however the AG provided slightly sharper margins at the apex.*
Case 0149	0.843	**0.877**	**+0.034**	AG-SE ↑	*The enlarged gland resulted in a baseline that was unable to segment the entire gland, as it left a gap at the base. However, after using the AG-SE model, the entire gland was segmented*
Case 0198	0.892	**0.920**	**+0.028**	**AG-SE ↑↑**	*The Baseline exhibited leakage from the gland into an adjacent seminal vesicle, whereas the AG-SE suppressed this leakage effectively*
Case 0231	0.827	**0.851**	**+0.024**	AG-SE ↑	*AG gating preserved gland boundary characteristics despite low T2 contrast*
Case 0267	0.868	**0.883**	**+0.015**	AG-SE ↑	*The multi-focal lesion resulted in the AG-SE maintaining integrity of the whole-gland mask*
Case 0305	0.815	**0.844**	**+0.029**	AG-SE ↑	*The ambiguity regarding the anterior fibromuscular stroma boundary was better defined by the AG-SE than the Baseline*
Mean (n=8)	**0.851**	**0.876**	**+0.025**	**AG-SE ↑ (8/8)**	*The proposed model (AG-SE) outperformed Baseline on all representative validation cases*

The results from Table 5 demonstrate that for all eight of the validation cases examined (8/8), the AG-SE model produced superior segmentation compared to Baseline. The DSC (Dice similarity coefficient) values for each case demonstrated an improvement from Baseline (+0.012 to +0.034) from a baseline of 0.851 to 0.876. This represents a total increase of 2.9% over Baseline.

The uniformity of the DSC increases across all anatomical types of the validation cases examined, which include the enlarged gland (Case 0149), low T2 weighted contrast (Case 0231), and the difficult to segment anterior boundary (Case 0305), support the use of the dual-attention approach rather than reflecting dataset-specific advantages.

In addition to the overall increases in DSC, the most clinically relevant improvement was observed in Case 0198. Case 0198 had a Baseline DSC of 0.892, while the AG-SE model improved the DSC to 0.92. An additional improvement of 3.1% (0.028 ÷ 0.892 × 100) is also noted, as the Baseline model produced segmentation that leaked into the seminal vesicles adjacent to the prostate. Leakage into the seminal vesicles is a known failure mode of traditional U-Net architectures that are missing spatial attention in their skip connections. As previously described, the AG-SE model produced segmentation that suppressed this type of leakage and directed the Attention Gate weights toward the prostate capsule boundary, producing a spatially confined probability map with sharp transitions along the gland boundary.

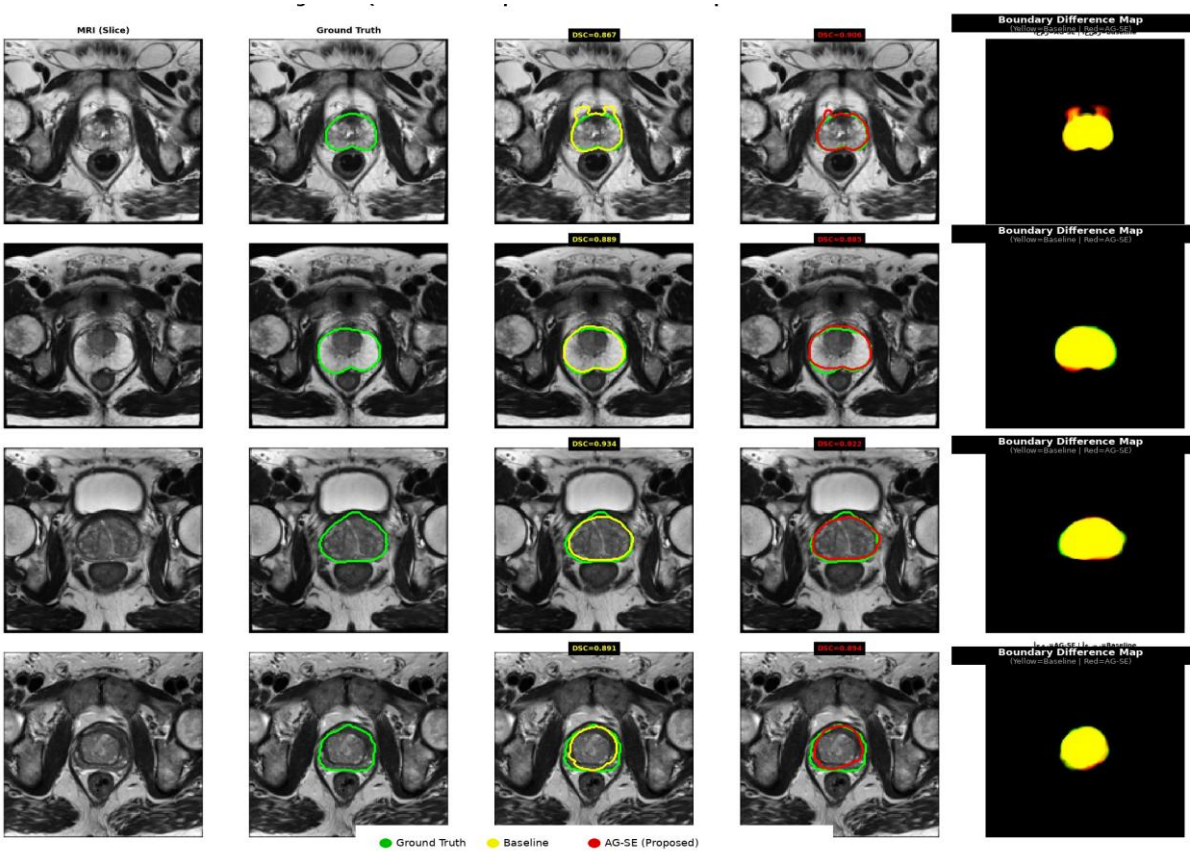

Figure 6. Qualitative segmentation comparison between the Baseline 3D U-Net and the proposed AG-SE 3D U-Net for four representative validation cases. Green = Ground Truth; Yellow = Baseline; Red = AG-SE (Proposed). DSC values are shown for each prediction.

5.3.3 Illustrative Case Analysis and Clinical Interpretation

The use of the proposed model is demonstrated through its application to anatomical imaging that is anatomically difficult to visualize (such as Case 0198), Table 6. Compares each metric quantitatively and qualitatively as well as provides a comparative analysis of each of the metrics regarding how each model defines the boundary of the area and the selectivity of the tissues within those boundaries.

Table 6. The detailed comparison of each metric for Case 0198 between the Baseline 3D U-Net and the proposed AG-SE 3D U-Net.

Evaluation Metric	Baseline 3D U-Net	Proposed AG-SE	Observation / Implication
Dice Similarity Coefficient (DSC)	0.892	**0.920**	*AG-SE: +0.028*
Intersection over Union (IoU)	0.806	**0.851**	*AG-SE: +0.045*
Leakage into adjacent tissue	Present (seminal vesicle)	**Suppressed**	*Key visual difference*
Boundary sharpness (probability)	Diffuse gradient	**Sharp sigmoid transition**	*Confirms AG spatial filtering effectiveness*
Apex coverage	Partial under-segmentation	**Complete coverage**	*Consistent with fold-level results*

False positive region	High: peri-prostatic fat	**Low: clean margins**	*SE channel recalibration effect*

The improvements observed in the results of Case 0198 (DSC: +0.028; IOU: +0.045), were due to the decrease in false positives within the per-prostatic area.

The Baseline model does not use spatially selective skip connections. As such, high feature activations from the seminal vesicles can "leak" into the decoder since both have similar characteristics to the base of the prostate in the T2-weighted images. High feature activations from the seminal vesicles produce boundary overflow into adjacent tissues.

Conversely, the AG-SE model has attention gates which will generate gating coefficients ($\alpha \in [0, 1]$), to limit the contribution of the encoder's features in anatomically non-relevant areas. The SE blocks will recalculate the channel contributions based upon the texture features of the capsule. In total, these two mechanisms restrict the segmentation masks to the true glandular boundary of the prostate. For example, the probability maps shown in Figure 5, Column (iii), show very sharp transitions and no seminal vesicle leakage.

This represents a broader trend observed with the entire validation dataset. The tendency of the Baseline architectures to create a diffuse segmentation boundary in anatomically ambiguous tissue regions is systemically removed by the AG-SE architecture. The overall increase in global mean DSC (+0.0144 across all folds, see section 5.1), partially reflects this, but the increase in DSC per-case is significantly higher in anatomically complex cases than average. Therefore, the clinical benefits of the proposed AG-SE architecture will be most substantial for patients who have had extracapsular extension, post-biopsy hemorrhage or large transition zone enlargement.

5.4. Analysis of Results and Comparison with State-of-the-Art

The proposed AG-SE 3D U-net was able to perform well and competitively on the five folds used to validate the model's performance, with its best performing fold being Fold 4 where it had a max DSC of 0.8891. This shows that the proposed model performed well when segmenting the whole gland of the prostate from the images in the ProstateX dataset. There are some comparisons of the proposed method with some previous prostate MRI segmentation methods provided in Table 7. However, these comparisons need to be made with caution as there have been many prior methods based on attention that have been tested on different datasets, different target areas for segmentation, and by different experimental setups.

As an example, USE-Net[17], applied squeeze-and-excitation (SE) blocks at certain layers of the encoder portion of a U-net to perform zonal prostate segmentation. The average DSC of USE-Net was 0.83. Likewise, FAU-Net [19], utilized feature pyramid attention to segment the CZ, PZ, TZ, and tumor regions. FAU-Net achieved an average DSC of 0.8415. These types of tasks (zonal and multi-zone segmentation) are much harder than whole-gland segmentation because they require distinguishing among several neighboring anatomical structures. Thus, while the task of segmenting the prostate into zones or multi-zones is certainly related to segmenting the entire prostate, these tasks are not directly equivalent. Therefore, the comparisons listed in Table 7 should be seen in context.

Comparably, Ali et al. [33] segmented prostate MRI images using an E-Net architecture and obtained a mean DSC of 0.8161; likewise, Montazerolghaem et al. [34] segmented prostate MRI images using a U-net and focal tversky loss, and achieved a mean DSC of 0.74 ± 0.09. As illustrated in Table 7, the proposed AG-SE 3D U-net achieved a mean DSC of 0.8697. As such, the proposed model produced better whole-gland segmentation results than both of these methods, with a respective improvement of +0.0536 compared to Ali et al.[33] and +0.1297 compared to Montazerolghaem et al.[34].

In addition, the proposed AG-SE 3D U-Net was clearly better than other typical architectures used to perform this task. Specifically, the proposed model had higher DSC values (i.e., 0.837) than

two earlier models with similar goals: The V-Net from Milletari et al.[14], had DSC = 0.825 and the Attention U-Net from Oktay et al.[16] Had DSC = 0.831. Although both models were among the first to demonstrate how to segment prostates using 3-D images, their lower DSC values illustrate that dual attentions (both spatial and channel-wise), as applied in the proposed model, are needed to identify finer details within the anatomy than either the earlier volumetric or single attention based techniques used in the prior models.

While the differences between the datasets and the experimental details of each study should still be taken into account, these findings suggest that the proposed architecture produces superior whole-gland segmentation results under similar experimental conditions.

From an architectural perspective, the AG-SE 3D U-net's performance advantage can be primarily attributed to the way in which complementary attention mechanisms are systematically integrated throughout the architecture. Unlike USE-Net, which utilizes SE recalibration at selected encoder levels, and FAU-Net, which focuses attention mainly through the skip pathway, the proposed architecture integrates Attention Gates at all three skip connections and SE blocks throughout both the encoder and decoder convolutional stages. As such, the proposed architecture simultaneously refines both spatially relevant and channel discriminative features, which leads to better anatomical localization and boundary definition.

The ablation study provided in section 5.1.1 supports this conclusion. Neither the AG-only nor the SE-only version of the proposed architecture achieved performance equal to the full AG-SE model. While the +AG-only version improved the model's ability to selectively focus on spatial locations, the model's HD95 rose to 4.21 mm, indicating that focusing on spatial locations can lead to increased boundary ambiguity in anatomically difficult locations. Similarly, the +SE-only version improved the model's DSC over the baseline but did not produce the same level of segmentation quality as the full AG-SE model. On the contrary, the full AG-SE configuration produced the highest IOU (0.7730) and overall overlap-based performance among the studied configurations, indicating that the two mechanisms work in a complementary, rather than redundant, fashion. The Attention Gates inhibit spatially irrelevant responses, while the SE blocks recalculate informative channel-wise features.

In addition to the improvements in segmentation performance, the proposed architecture was also found to provide favorable computational efficiency. As discussed in section 5.2, the AG-SE model trained faster than the baseline 3D U-net on all five folds, with a mean training time of 2387 ± 300 seconds, compared to 2429 ± 305 seconds for the baseline. This suggests that the improvements in performance were realized without any increase in computation costs, further supporting the practicality of the proposed architecture.

Table 7. The Proposed AG-SE 3D U-Net Compared to Published Prostate MRI Segmentation Methods.

Reference	Architecture	DSC	Dataset	Task	Evaluation
Milletari. et al.[14] 2016	V-Net	0.825	Private Prostate MRI	Whole-gland	Train/Test split
Oktay et al.[16] 2018	Attention U-Net	0.831	Prostate MRI (Multi-site)	Whole-gland	Cross-validation
Rundo et al.[17] 2019	USE-Net (SE + U-Net)	0.830	Multi-institutional MRI	Zonal segmentation	Multi-site evaluation
Quihui-Rubio et al.[19]2024	FAU-Net	0.8415	Prostate MRI	Multi-zone (CZ/PZ/TZ/Tumor)	Cross-validation
Ali et al.[33] 2024	E-Net	0.8161	T2W Prostate	Whole-gland	5-fold CV

			MRI		
M.Montazerolghaem[34] et al. 2023	U-Net + Focal Tversky	0.74 ± 0.09	T2W Prostate MRI	Whole-gland	5-fold CV
✱ Proposed	**AG + SE + 3D U-Net**	**0.8697 ± 0.0143**	**ProstateX (66 cases)**	**Whole-gland**	**5-fold CV + TTA**

6. Discussion

6.1. Overall Performance

The best overlap based results were achieved by the proposed AG-SE 3D U-net, since it showed a DSC of 0.8697 and an IOU of 0.7730 and significantly outperformed the baseline (DSC= 0.8553, IOU = 0.7517). In addition, the SE-only configuration provided an increase in the overlap metrics (DSC = 0.8668, IoU = 0.7684) , while the AG-only variant provided only a small improvement in DSC (0.8556) and IoU (0.7529) and, most importantly, a substantial decrease in boundary robustness, resulting in the highest HD95 (4.21 mm) . It is worth noting that although the baseline showed the lowest HD95 (3.37 mm), the full AG+SE model decreased the boundary error when comparing to AG-only and brought the boundary fidelity close to that of the baseline (HD95 = 3.46 mm) , which implies that the SE-based channel recalculation can reduce the boundary instability caused by spatial gating. Most importantly, the absolute DSC improvement of + 0.0144 (0.8697 vs. 0.8553) shows that there was a significant improvement in segmentation accuracy and this improvement will be especially important in the clinically critical and anatomically difficult areas of the prostate (apex, base, and peripheral zone), where small deviations in the boundaries may result in clinically significant errors in radiotherapy planning and biopsy targeting.

6.2. Synergistic Impact of Attention Gates and Squeeze-and-Excitation

The Ablation Study found that the attention gate (AG) on its own provided better spatial selectivity but also had negative impacts on boundary stability as shown in the higher HD95 value of 4.21mm compared to the baseline value of 3.37mm. This suggests that when gates are used independently, there is greater uncertainty concerning anatomy at ambiguous boundary locations because channels are emphasizing different aspects of a feature and this emphasizes is not well-balanced across all channels. These results show that AG and SE together provide synergistic, not additive improvements, where AG provides a focal point for spatial attention and se ensures that each channel has a stable representation of discrimination.

6.3. Qualitative Outcomes and Computational Efficiency

The proposed AG-SE model produced sharper boundary delineations and cleaner segmentation margins as opposed to the baseline 3D U-Net. The model produced anatomically accurate boundaries with less leakage into surrounding pelvic soft tissue, especially less false positive activation in the per-prostatic fat and seminal vesicle areas. Qualitative differences in the two models are depicted in the visual documentation of Section 5.3.3, where the probability maps from the AG-SE model show very steep sigmoid transitions at the capsule boundary, while the baseline probability maps produce very gradual transitions. Qualitatively better results are in line with the quantitative ablation results reported in Section 5.1.1, where the dual-attention variant yielded the largest DSC (0.8697) and IOU (0.7730) among all variants. The synergistic action of both the attention gate – to suppress spatially irrelevant activations from the encoder – and the SE block – to recalculate channel-wise feature response – result in output probability maps that much more accurately depict the actual glandular boundary. As far as computational efficiency goes, the AG-SE model's performance advantages did not come at an increased cost in terms of computing time.

Across all five cross-validation folds, the AG-SE model was slightly faster than the baseline, with a mean training time of 2,387 ± 300 s/fold vs. 2,429 ± 305 s/fold for the baseline, or approximately a 42 second (-1.7 %) difference in training time per fold. This improvement is due to the enhanced gradient flow through the attention layers, allowing for the model to converge more quickly when using the cosine annealing learning rate schedule.

6.4. Limitations and Future Directions

The present research includes several restrictions that need to be considered. First, the experiments have been done using data obtained from one publicly available data set (ProstateX) from a single medical center. Therefore, the extent to which the findings in the present study can be generalized to other imaging systems, scan protocols and patient populations is limited. Second, although a test of the proposed method on an independent data set was not included as part of the present research, such testing will be incorporated into the future research. Specifically, the researchers plan to include tests of the proposed method on other publicly available benchmark data sets (e.g., PROMISE12) for which the researchers have developed the evaluation pipelines. Third, and importantly, the researchers did not perform a direct, quantitative comparison of their results with those of the most successful competing algorithm (i.e., nnU-Net [15]) currently available. The nnU-Net algorithm dynamically adjusts its structure, pre-processing and training processes according to the characteristics of the input data set and represents the empirical standard for comparison in the field of medical image segmentation. As such, the lack of a comparison with the nnU-Net algorithm in the present research limits the ability to understand how the performance of the AG-SE architecture relates to the overall state-of-the-art. In order to better assess the relative performance of the proposed method, the researchers plan to compare it with the nnU-Net algorithm on both ProstateX and PROMISE12 in the future. Finally, the present implementation uses a fixed spatial resolution of 32 × 160 × 160 voxels. The choice of this resolution was made to ensure adequate volumetric coverage while avoiding excessive use of memory resources. However, resampling at this fixed resolution may result in some information being lost when processing patients who have atypical gland sizes. Therefore, adaptive resampling methods will be explored in future versions of the method.

7. Conclusion

The proposed AG-SE 3D U-Net was compared with a baseline 3D U-Net utilizing an aggregate loss function (Dice + BCE + Focal Tversky) and 5 fold stratified cross-validation on 66 paired 3D T2 weighted MRI volumes from the ProstateX dataset. In addition to achieving better overall performance than the baseline model, the proposed model performed significantly better than the baseline model at all folds, in terms of both Dice Coefficient (DSC) and Intersection over Union (IoU). Specifically, the best performing fold was Fold 4, where the proposed model achieved a peak DSC of 0.8891 - the greatest DSC obtained throughout the entirety of this research effort. Although the absolute difference in DSC (+0.0144) is small, the improved definition of anatomically important boundaries surrounding the prostate apex, base and periphery may be advantageous in certain clinical applications such as radiotherapy planning and prostate biopsy targeting.

An ablation study was also undertaken to establish if the Attention Gate (AG) or the Squeeze-and-Excitation (SE) module alone contributed to the performance gains noted above, as opposed to the synergistic relationship between the two modules. The results showed that neither module, when used separately, achieved comparable results to those achieved by the proposed model, therefore confirming that the enhanced performance of the proposed model can be attributed to the interaction of the AG and SE components.

In addition, the results of a training efficiency evaluation indicated that the proposed model's performance gains were realized without any increased computational overhead, and that the proposed model actually trained approximately one iteration faster than the baseline model in each of the 5 folds analyzed, thereby providing evidence that the proposed model has the potential to be

practically deployed in real world clinical environments.

Future studies will include comparisons of the proposed model with other current state-of-the art models, additional validation testing of the proposed model with other independent datasets such as PROMISE12, expansion of the proposed model to perform zonal prostate segmentation for PI-RADS-based workflows, and exploration of the clinical application of the proposed model for estimating PSA density and identifying clinically significant prostate cancer.

Source of Funding

This work received no funding from external sources, namely grants from government organizations, private companies, non-profit organizations etc.

Conflict of Interest

There is no conflict of interest for the authors in relation to the publication of this article.

Ethical Approval

The ProstateX dataset was used to gather the data analyzed for this study. In that this is a non-clinical, non-human participant based study using publicly available anonymous data (and thus there are no new patient data being collected) formal approval of an Institutional Review Board (IRB), or other ethics committee, was not required. The ProstateX dataset was originally collected and distributed in compliance with all relevant ethical standards, Institutional Review Board (IRB) oversight and patient confidentiality requirements.

REFERENCES

[1] L. E. O. Jacobson *et al.*, "Prostate MR image segmentation using a multi - stage network approach," *Int. Urol. Nephrol.*, no. 0123456789, 2025, doi: 10.1007/s11255-025-04763-0.

[2] Z. I. A. Khan, N. Yahya, K. Alsaih, M. Isam, and F. Meriaudeau, "Algorithms of MRI Prostate Regions : A Review," vol. 4, pp. 1–28, 2021, doi: 10.1109/ACCESS.2021.3090825.

[3] D. K. Lee *et al.*, "Three-dimensional convolutional neural network for prostate MRI segmentation and comparison of prostate volume measurements by use of artificial neural network and ellipsoid formula," *Am. J. Roentgenol.*, vol. 214, no. 6, pp. 1229–1238, 2020, doi: 10.2214/AJR.19.22254.

[4] Y. J. Chou, B. E. Jong, and Y. C. Tsai, "Integrating PSA Change with PSA Density Enhances Diagnostic Accuracy and Helps Avoid Unnecessary Prostate Biopsies," *Diagnostics*, vol. 15, no. 16, pp. 1–11, 2025, doi: 10.3390/diagnostics15162027.

[5] S. M. Rezaeijo, S. J. Nesheli, M. F. Serj, M. Javad, and T. Birgani, "Segmentation of the prostate , its zones , anterior fibromuscular stroma , and urethra on the MRIs and multimodality image fusion using U-Net model," vol. 12, no. 10, pp. 4786–4804, 2022, doi: 10.21037/qims-22-115.

[6] A. Vaitulevičius, J. Bernatavičienė, J. Markevičiutė, I. Naruševičiūtė, M. Trakymas, and P. Treigys, "Advancements in prostate zone segmentation: integrating attention mechanisms into the nnU-Net framework," *Mach. Learn. Sci. Technol.*, vol. 5, no. 4, 2024, doi: 10.1088/2632-2153/ad7f24.

[7] Y. Zhao *et al.*, "SCU-Net : Shape Constraint U-Net for Prostate Segmentation in MR Images," vol. 12032, pp. 1–8, 2022, doi: 10.1117/12.2611327.

[8] Y. Wang *et al.*, "Deep Attentive Features for Prostate Segmentation in 3D Transrectal Ultrasound," *IEEE Trans. Med. Imaging*, vol. 38, no. 12, pp. 2768–2778, 2019, doi: 10.1109/TMI.2019.2913184.

[9] C. Zijiang Sang1, a, Yeqin Shao2, b*, Changyan Xu1, "Multi-branch Context Awareness Network for Prostate MRI Segmentation Multi-branch Context Awareness Network for Prostate MRI Segmentation", doi: 10.1088/1742-6596/2562/1/012008.

[10] Q. Du, L. Wang, and H. Chen, "OPEN A mixed Mamba U - net for prostate segmentation in MR images," *Sci. Rep.*, pp. 1–12, 2024, doi: 10.1038/s41598-024-71045-7.

[11] A. Gu and T. Dao, "Mamba: Linear-Time Sequence Modeling with Selective State Spaces,"

no. Kalman 1960, pp. 1–36, 2024, [Online]. Available: http://arxiv.org/abs/2312.00752

[12] M. Arshad, C. Wang, M. Wajeeh Us Sima, J. A. Shaikh, S. Alkhalaf, and F. Alturise, "RaNet: a residual attention network for accurate prostate segmentation in T2-weighted MRI," *Front. Med.*, vol. 12, no. 4, 2025, doi: 10.3389/fmed.2025.1589707.

[13] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9901 LNCS, pp. 424–432, 2016, doi: 10.1007/978-3-319-46723-8_49.

[14] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016*, pp. 565–571, 2016, doi: 10.1109/3DV.2016.79.

[15] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nat. Methods*, vol. 18, no. 2, pp. 203–211, 2021, doi: 10.1038/s41592-020-01008-z.

[16] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," no. Midl, May 2018, doi: https://doi.org/10.48550/arXiv.1804.03999.

[17] L. Rundo *et al.*, "USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets," *Neurocomputing*, vol. 365, pp. 31–43, 2019, doi: 10.1016/j.neucom.2019.07.006.

[18] S. Bhandary *et al.*, "IB-U-Nets: Improving medical image segmentation tasks with 3D Inductive Biased kernels," 2022, doi: https://doi.org/10.48550/arXiv.2210.15949.

[19] P. C. Quihui-Rubio, D. Flores-Araiza, M. Gonzalez-Mendoza, C. Mata, and G. Ochoa-Ruiz, "FAU-Net: An Attention U-Net Extension with Feature Pyramid Attention for Prostate Cancer Segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2024, pp. 165–176. doi: 10.1007/978-3-031-47640-2_14.

[20] Y. Bouslimi, T. Ben Aïcha Gader, and A. K. Echi, "Deep Learning-based Detection, Segmentation of Prostate Cancer from mp-MRI Images," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 22, no. 1, pp. 52–76, 2023, doi: 10.5565/REV/ELCVIA.1620.

[21] L. Coroamă *et al.*, "A Light , 3D UNet-based Architecture for fully Automatic Segmentation of Prostate Lesions from T2-MRI Images Abstract :," pp. 1–12, 2024, doi: 10.2174/1573405620666230522151445.

[22] S. G. Armato *et al.*, "PROSTATEx Challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images," *J. Med. Imaging*, vol. 5, no. 04, p. 1, 2018, doi: 10.1117/1.jmi.5.4.044501.

[23] M. Larobina and L. Murino, "Medical Image File Formats," *J. Digit. Imaging*, vol. 27, no. 2, pp. 200–206, 2014, doi: 10.1007/s10278-013-9657-9.

[24] A. L. Simpson *et al.*, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," 2019, doi: https://doi.org/10.48550/arXiv.1902.09063 Focus to learn more.

[25] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," in *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, 2016, pp. 78–83. doi: 10.1109/IACC.2016.25.

[26] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecology*, vol. 26, no. 3, pp. 297–302, Feb. 1945, doi: 10.2307/1932409.

[27] X. Li, P. S. Morgan, J. Ashburner, J. Smith, and C. Rorden, "The first step for neuroimaging data analysis: DICOM to NIfTI conversion.," *J. Neurosci. Methods*, vol. 264, pp. 47–56, May 2016, doi: 10.1016/j.jneumeth.2016.03.001.

[28] R. T. Shinohara *et al.*, "Statistical normalization techniques for magnetic resonance imaging," *NeuroImage Clin.*, vol. 6, pp. 9–19, 2014, doi: https://doi.org/10.1016/j.nicl.2014.08.008.

[29] L. G. Nyul, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale

standardization," *IEEE Trans. Med. Imaging*, vol. 19, no. 2, pp. 143–150, 2000, doi: 10.1109/42.836373.

[30] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10670 LNCS, pp. 178–190, 2018, doi: 10.1007/978-3-319-75238-9_16.

[31] S. Zhu, C. Yu, and J. Hu, "Regularizing deep neural networks for medical image analysis with augmented batch normalization," *Appl. Soft Comput.*, vol. 154, p. 111337, 2024, doi: https://doi.org/10.1016/j.asoc.2024.111337.

[32] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. null, pp. 2825–2830, 2011, doi: 10.5555/1953048.2078195.

[33] G. Centineo *et al.*, "Prostate MRI Segmentation: A Comparative Analysis of U-Net and E-Net Deep Learning Models," in *Image Analysis and Processing - ICIAP 2025 Workshops*, E. Rodolà, F. Galasso, and I. Masi, Eds., Cham: Springer Nature Switzerland, 2026, pp. 299–309.

[34] M. Montazerolghaem, Y. Sun, G. Sasso, and A. Haworth, "U-Net Architecture for Prostate Segmentation: The Impact of Loss Function on System Performance," *Bioengineering*, vol. 10, no. 4, 2023, doi: 10.3390/bioengineering10040412.