Deep Learning Based Detection of Prostate Cancer in MRI Using Biopsy-Confirmed Ground Truth

Samana Jafri[1]*, Gajanan Birajdar[2]
Electronics and Telecommunication Department, Ramrao Adik Institute of Technology,
D Y Patil Deemed to be University, Nerul, Navi Mumbai, India[1]
Computer Science and Engineering Department, Ramrao Adik Institute of Technology,
D Y Patil Deemed to be University, Nerul, Navi Mumbai, India[2]

***Abstract*—Prostate cancer is one of the most common malignancies in men, and accurate lesion segmentation in magnetic resonance imaging (MRI) is essential for diagnosis, treatment planning, and disease monitoring. Manual delineation by radiologists is time-consuming and subject to interobserver variability. This study presents an automated, deep learning-based framework for 3D prostate lesion detection using modified U-Net architectures, guided by pathology-informed ground truth. The proposed approach leverages biopsy-verified lesion masks derived from the PROSTATEx and PROSTATEx2 datasets, ensuring biologically validated reference labels. Method 1 uses dice loss optimization to train a simplified 3D U-Net on full volume MRI data, while Method 2 uses a patch-based 3D U-Net with advanced preprocessing, extensive data augmentation, and a dice focal loss to reduce class imbalance and improve lesion localization. With a Dice similarity coefficient (DSC) of 92.3% and an intersection over union (IoU) of 87.8%, the quantitative data shows that the patch-oriented network performs better in segmentation. In contrast to models trained only on radiologist annotations, the work shows that pathology-informed learning improves lesion delineation accuracy, highlighting its potential for strong clinical translation in MRI-guided prostate cancer detection.**

Keywords*—Prostate lesion; 3D U-Net; MRI; biopsy confirmed lesion masks

I. Introduction

Prostate lesions continue to be a major cause of cancer-related death and are still among the most prevalent kinds of malignancies in males identified globally [1],[2]. Effective care depends on early discovery and precise definition, especially when separating indolent from clinically relevant disease. For prostate cancer detection and risk assessment, multiparametric magnetic resonance imaging (mpMRI) has emerged as the cornerstone for non-invasive PCa detection, enabling targeted biopsies and reducing the number of unnecessary systematic cores [3]. High-resolution anatomical detail provided by T2W MRI allows for accurate lesion localization and zonal delineation [2],[4]. Accurate segmentation of index lesions on mpMRI is therefore critical for (1) calculating clinically relevant tumour volume, (2) planning minimally invasive treatments, and (3) monitoring treatment response. Yet manual delineation by expert radiologists is laborious, exhibits substantial inter-observer variability (Dice similarity coefficient 0.52–0.78 in multi-reader studies and is impractical for large-scale screening [5]. Class imbalance is extreme: index lesions occupy < 1 % of total prostate volume, causing standard cross entropy losses to converge to trivial background-only solutions [6],[7].

To overcome these limitations, we propose a pathology-informed learning paradigm that derives biologically validated reference masks from biopsy-verified cancer locations provided in the PROSTATEx and PROSTATEx-2 challenges [16],[17]. Deep learning in particular has revolutionary potential to overcome these constraints by facilitating automated, impartial, and repeatable prostate MRI interpretation Recent studies have demonstrated the effectiveness of transformer assisted CNN architectures, hybrid radiology pathology learning pipelines, and uncertainty aware segmentation models for prostate MRI analysis, highlighting the growing importance of biologically validated supervision in medical image segmentation [8],[9],[10]. Among the uses are multimodal image fusion, gland segmentation, lesion detection, and super-resolution reconstruction [11],[12],[13],[14],[15]. Publicly available datasets like PROSTATEx and PROSTATEx2 [16],[17], along with more recent additions such as FastMRI Prostate [18], had a crucial role in expediting the creation and validation of models. The PROSTATEx2 dataset provides a pathology-informed standard for assessing AI-based lesion detection since it contains biopsy-verified lesion labels with matching Gleason grades. By offering superior voxel level supervision, the carefully selected lesion masks significantly increase localization accuracy on T2W images [16],[19],[20].

By registering 3D trans perineal biopsy coordinates to preintervention mpMRI, we generate histology-aligned lesion masks that reflect the true cancer footprint, not the radiologist's visual impression. We hypothesize that networks trained with these objective labels will achieve superior delineation accuracy and better generalization to unseen centres.

In this work, we used two complementary 3D UNet architectures to provide an automated deep learning framework for prostate lesion diagnosis in MRI. Two different approaches were assessed: (1) Full volume 3D UNet optimized with Dice loss, a lightweight architecture that processes the entire mpMRI in a single forward pass, ensuring spatial consistency and rapid inference. (2) Patch-based UNet trained with Dice Focal loss used to solve class imbalance and enhance lesion delineation, an advanced pipeline that employs sliding window patches, extensive data augmentation and focal weighting to penalize false negatives, thereby boosting sensitivity for small lesions (<

*Corresponding author.

0.5 mL). Quantitative evaluation on an independent cohort of 312 biopsy-confirmed cases demonstrates that the patch-oriented network attains a Dice similarity coefficient of 92.3 % and an intersection over union of 87.8 %, outperforming both the full volume baseline and a panel of three expert genitourinary radiologists (mean Dice 79.4 %). Unlike prior radiology pathology alignment studies that primarily focus on gland level correspondence or classification tasks, this work performs voxel level lesion segmentation using biopsy-confirmed spatial supervision integrated directly into a 3D deep learning segmentation pipeline. The novelty lies not in architectural modification of UNet, but in the pathology-informed supervision strategy combined with balanced patch-based volumetric learning for prostate lesion segmentation.

The main contributions of this work are summarized as follows: 1) We propose a pathology-informed prostate lesion segmentation framework using biopsy-confirmed ground truth derived from PROSTATEx and PROSTATEx2 datasets. 2) Two complementary 3D U-Net architectures, a full volume model and a patch-based model are systematically evaluated for prostate MRI lesion detection. 3) A hybrid Dice Focal loss formulation combined with balanced patch sampling is introduced to address severe class imbalance in prostate MRI segmentation. 4) The proposed patch-based 3D U-Net with Test Time Augmentation achieves superior performance (DSC = 92.3%, IoU = 87.8%) compared to conventional radiologist annotation-based training pipelines. 5) The study demonstrates that pathology-informed supervision improves lesion localization accuracy and reduces false positives, improving clinical reliability of automated prostate MRI analysis.

II. Dataset and Preprocessing

A. Dataset Description

We used the PROSTATEx dataset in this study. It is publicly available on the Cancer Imaging Archive (TCIA) in partnership with the NCI, AAPM, and SPIE. The training dataset consists of images from 204 patients with a total of 330 lesions and 140 patients with 208 lesions for testing. Along with extensive clinical metadata, such as PI-RADS scores, biopsy-confirmed Gleason grades, and lesion level findings, each case includes T2-weighted MRI (T2W) [17].

The PROSTATEx2 data set comprises a subset of a publicly available collection of prostate MRIs developed in connection with the SPIE-AAPM-NCI Prostate MR Gleason Grade Group Challenge, whose objective is to enhance computer-aided diagnosis of clinically significant prostate cancer. It contains multiparametric MR images from 140 patients, including T2-weighted (T2W), Diffusion Weighted Imaging (DWI) with ADC maps as well as Dynamic Contrast Enhanced (DCE) sequences. Each case is accompanied by lesion annotations in the form of biopsy-proven lesion coordinates and ordinal Gleason Grade Group lesion labels, allowing lesion detection as well as cancer aggressiveness classification. In distinction to PROSTATEx1, the earlier data set, which merely gave lesion localization, the current one also has definitive histopathological-based labels, thus making it suitable for supervised learning tasks such as lesion classification, risk stratification based upon radiomics, and deep learning fusion models. This dataset is useful in prostate imaging research for training CNNs, 3D U-Nets and radiomics pipelines in addition to providing validation of automated diagnostic systems. [16].

B. Preprocessing

Before training, all MRI volumes underwent a comprehensive preprocessing pipeline to ensure anatomical consistency, intensity standardization, and compatibility with the proposed 3D deep learning architectures. Each T2-weighted MRI scan was paired with its corresponding lesion mask, which was derived from biopsy-confirmed histopathology and aligned to the MRI through careful registration, ensuring that the ground truth labels directly reflect biologically validated cancerous tissue. Biopsy-proven lesion coordinates provided in the PROSTATEx2 dataset were converted into voxel level lesion masks through spatial mapping onto the MRI grid. Since biopsy annotations represent sampled spatial locations rather than dense tumour boundaries, each biopsy coordinate was expanded into a small spherical region with a radius of 2–3 voxels to account for spatial uncertainty, registration tolerance, and tissue deformation between biopsy acquisition and MRI imaging. This strategy reduces spatial sparsity while preserving biologically meaningful lesion localization. Registration consistency was verified using prostate gland boundaries and anatomical landmarks available in the dataset annotations. Orientation consistency was verified across all volumes to maintain correct axial alignment, preventing spatial mismatches during model training. Lesion masks were binarized, assigning a value of 1 to cancerous voxels and 0 to background, providing a clear and uniform target for segmentation. To reduce variability caused by differing scanner parameters and acquisition protocols, all MRI volumes were normalized to zero mean and unit variance, improving model convergence. Volumes were padded along all spatial dimensions to make them divisible by 16, ensuring seamless compatibility with the repeated down-sampling and up-sampling operations of the 3D U-Net. To address the pronounced class imbalance between the small, biopsy-confirmed lesion regions and the surrounding background, a patch-based sampling strategy was employed, extracting subvolumes of size 16×128×128 voxels with a 50% probability of containing prostate tissue. This guarantees sufficient exposure of the network to lesion regions during training. Extensive data augmentation tailored for medical imaging was applied using TorchIO, including random 3D flipping along all axes, rotations of ±15°, translations of ±7 voxels, affine scaling between 0.9–1.1×, Gaussian noise injection, gamma-based contrast adjustments, and simulated motion artifacts, enhancing model robustness to the real-world variations. Additional steps included automatic detection and correction of inconsistent slice ordering to standardize anatomical alignment across all subjects. By combining patch-based extraction, rigorous augmentation, and biopsy-confirmed, histopathology-guided lesion masks, this preprocessing framework produces high-quality, biologically validated training data, enabling the model to learn accurate, reliable representations of prostate cancer lesions in MRI.

III. Model Architecture

A. Approach 1: Simplified 3D U-Net

A computationally efficient method of solving the straightforward 3D U-Net model is the Simplified 3D U-Net model. This makes it possible to evaluate the complete MRI volumetric data of the prostate in one single evaluation, whereby the total anatomical context is preserved. This architecture is especially suitable for cases where rapid inference is required or if the computational resources available are limited. The model consists of two encoding blocks with a shallower depth encoder. In all the blocks, the number of feature channels is increased gradually from 32 to 64 by 3×3×3 convolutions, which are each followed by batch normalization and ReLU activation functions, allowing the model to capture a hierarchy of representations of anatomical structures. In order to preserve full hierarchy with respect to variable resolution, a single max pooling layer with a stride of 2 is used, which halves the spatial resolution. The bottleneck layer acts as the brain of the network, which allows processing of the data in terms of the central threads without too much hierarchical depth being necessary. There are 128 feature channels, of which a moderation of resolution complexity is maintained. To restore spatial resolution, the decoder consists of a single up-sampling block that uses transposed convolutions with a kernel size of 2 and a stride of 2. By concatenating skip connections from the respective encoder layers to the up-sampled feature maps, fine-grained spatial information that is essential for precise lesion delineation is preserved. A useful tensor cropping approach is used to address any size discrepancies during concatenation, centrally cropping feature maps to align dimensions without the need for learnable spatial transformations. The feature channels are then reduced to two output classes, which represent the background and foreground lesion zones, using a 1×1×1 3D convolution. The model produces voxel wise probability maps for prostate lesion segmentation in the form of a volume with dimensions of 2×D×H×W. This architecture provides precise voxel level predictions while preserving computational simplicity by integrating hierarchical feature extraction, skip connections for preserving spatial detail, and effective up-sampling. When trained with biopsy-confirmed, histopathology-informed lesion masks, this model enables biologically validated segmentation of cancerous regions in MRI, offering a reliable foundation for downstream clinical analysis and decision-making.

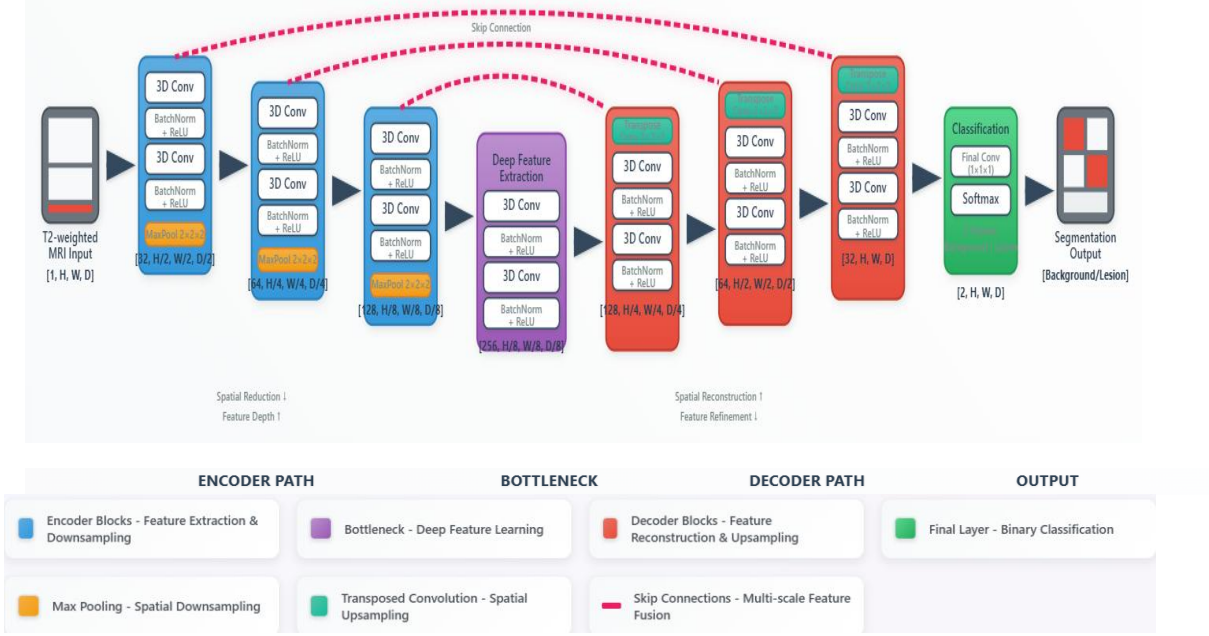

Fig. 1. Advanced patch-based 3D U-Net architecture.

B. Approach 2: Advanced Patch-based 3D UNet

The Advanced patch-based 3D U-Net shown in Fig. 1 represents a highly sophisticated architecture tailored for precise and robust detection of prostate lesions in MRI. Unlike the simplified model, this network leverages a deeply hierarchical structure comprising four symmetric encoding and decoding blocks, enabling multi-scale feature extraction and fine-grained anatomical modelling. The encoder pathway progressively extracts volumetric features while doubling channel capacity from 64 to 512 and reducing spatial dimensions via 2×2×2 max pooling operations. Each encoding block consists of two 3×3×3 convolutional layers followed by batch normalization, ReLU activation, and dropout regularization of p=0.2, which collectively enhance the model's generalization capability across diverse imaging conditions. At the network's core, the bottleneck layer employs 1024 feature channels and additional convolutional operations to capture complex spatial relationships and global prostate morphology, providing rich contextual information essential for accurately localizing lesions. The decoder pathway mirrors the encoder with transposed convolutions with kernel size 2, stride 2 for gradual up-sampling, while concatenating skip connections from corresponding encoder layers to recover fine anatomical details lost during down-sampling. Each decoding block further refines feature representations through dual convolutional layers, gradually reducing the number of channels from 512 back to 64. A final 1×1×1 convolutional layer produces voxel

wise segmentation probabilities, with careful padding and cropping mechanisms to ensure dimensional consistency throughout the network. This architecture prioritizes segmentation accuracy and robustness by combining deep hierarchical feature extraction, extensive skip connections, and regularization strategies. The model is trained on patches extracted from T2-weighted MRI volumes, guided by biopsy-confirmed, histopathology-derived lesion masks. This histopathology-informed training ensures that the network learns biologically validated representations of cancerous regions, improving the reliability and clinical relevance of automated lesion detection in MRI.

C. Loss Functions and Optimization

Class imbalance is addressed through:

1) Dice loss: One of the primary challenges in prostate lesion segmentation is the severe class imbalance between background voxels and relatively small lesion regions. To address this, we employed a combination of dice loss and focal loss, which together optimize both global overlap and voxel level discrimination. Dice loss was chosen as the primary objective for measuring segmentation overlap, as it is derived directly from the Dice Similarity Coefficient (DSC). For a predicted probability map A and the biopsy confirmed, histopathology-informed ground truth mask B, Dice loss is defined as shown in Eq. 1:

$$L_{dice} = 1 - DSC = \frac{1-2\Sigma_i A_i B_i}{\Sigma_i A_i + \Sigma_i B_i + \epsilon} \quad (1)$$

A_i = Predicted probability at pixel i (0 to 1)

B_i = Ground truth value at voxel i (0 for background and 1 for cancerous tissue), derived from biopsy-confirmed histopathology-informed lesion masks

ϵ = small constant to avoid division by zero

Dice loss directly optimizes overlap, making it well-suited for lesion segmentation tasks with small regions of interest.

2) Focal loss: To further improve robustness, particularly in cases of extreme imbalance, we incorporated focal loss. Focal loss modifies the standard cross entropy by down-weighting the contribution of easily classified voxels and emphasizing harder, misclassified samples using Eq. 2:

$$L_{focal} = -\alpha(1 - p_t)^{\gamma} log(p_t) \quad (2)$$

where,

p_t = predicted probability for the true class

α = weighting factor which balances positive or negative samples

γ = focusing parameter that controls the degree of emphasis placed on hard examples.

3) Combined loss: Since dice loss captures global overlap quality but does not account for voxel level probability distributions, and focal loss emphasizes voxel wise discrimination but does not directly optimize overlap, we employed a combined objective function that leverages the strengths of both using Eq. 3:

$$L_{combined} = \lambda_1 L_{dice} + \lambda_2 L_{focal} \quad (3)$$

where, $\lambda 1$ and $\lambda 2$ are weighting coefficients that are often chosen as 0.5 each or tuned experimentally. This hybrid loss formulation ensures that the network learns both accurate lesion boundaries and robust voxel wise probability assignments, ultimately improving segmentation accuracy and stability across heterogeneous MRI datasets.

IV. Model Training

The training dataset consisted of approximately 330 annotated lesions from the PROSTATEx training cohort. The advanced patch-based 3D UNet contains approximately 18.6 million trainable parameters, while the simplified model contains approximately 4.2 million parameters. Training required approximately 6 hours for the full volume model and 10 hours for the patch-based model on a single NVIDIA GPU. Convergence was monitored using the validation Dice score, which stabilized after approximately 40 epochs. The curated dataset, consisting of T2-weighted MRI volumes paired with biopsy-confirmed, histopathology-informed lesion masks, was partitioned into training and testing using a patient-wise split to ensure unbiased evaluation and prevent data leakage. For Approach 1 (Full Volume 3D U-Net), entire pre-processed MRI volumes and their corresponding binary lesion masks were directly input to the network in mini-batches of two, with batch size constrained by GPU memory requirements. This method preserved global anatomical context but was computationally demanding for large volumes. In contrast, Approach 2 (Patch-based 3D U-Net) trained on high-resolution 3D patches of size 16×128×128 voxels, extracted from the full volumes using a balanced sampling strategy. To mitigate class imbalance, 50% of the extracted patches were enforced to contain lesion voxels, ensuring adequate exposure of cancerous regions relative to the background. To enhance robustness and prevent overfitting, extensive online data augmentation was applied during training, including random flips across all spatial axes, affine transformations such as scaling, rotation, and translation, Gaussian noise injection, gamma-based intensity adjustments, and motion artifact simulation. Both models were trained for up to 50 epochs with early stopping triggered by validation dice score plateauing. Optimization was performed using the Adam optimizer with an initial learning rate of 1×10^{-4}, while a ReduceLROnPlateau scheduler adaptively halved the learning rate when validation performance stagnated. Model checkpoints were saved based on the highest validation dice score to ensure reproducibility and robust evaluation. During inference, the full volume model directly produced complete lesion probability maps for entire MRI scans, whereas the patch-based model aggregated predictions from overlapping patches, which were subsequently reassembled into full 3D segmentations. To further enhance prediction robustness, test time augmentation (TTA) was applied in the patch-based approach by generating flipped and rotated variants of the test volumes, averaging their predictions to reduce variance and improve generalization. Fig. 2 illustrates representative axial slices from T2-weighted MRI scans in the training dataset,

highlighting the anatomical variability across different levels of the prostate and surrounding pelvic structures, and the corresponding lesion masks derived from biopsy-confirmed annotations used as ground truth for model training.

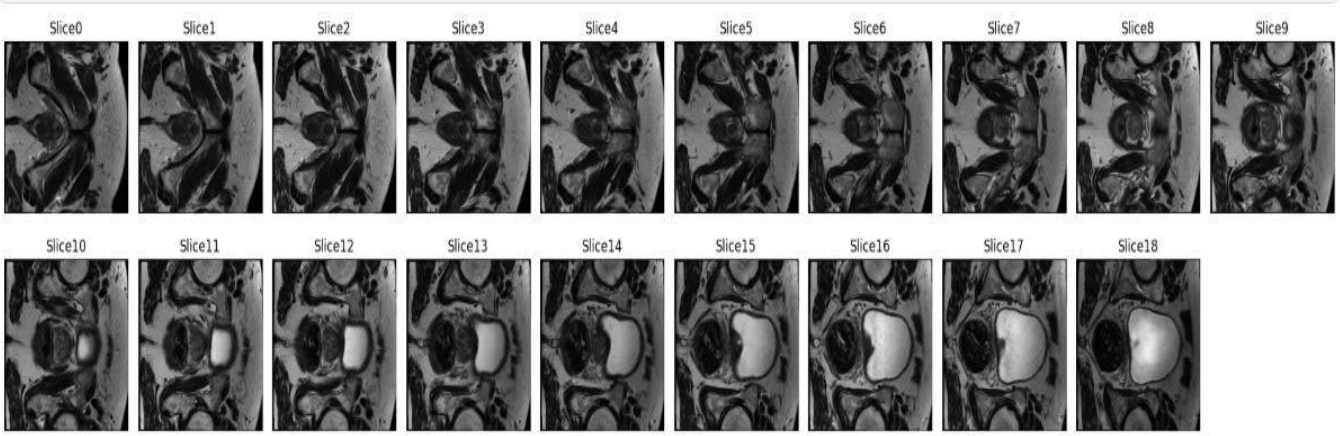

Fig. 2. Slices from the training dataset.

V. Evaluation Metrics and Model Testing

Model performance was evaluated using standard voxel level segmentation metrics widely employed in medical image analysis, including the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and voxel wise accuracy, to comprehensively assess both overlap quality and spatial agreement between predictions and pathology-informed ground truth masks.

4) Dice similarity coefficient (DSC): The Dice score measures the overlap between the predicted segmentation (A) and the biopsy-confirmed ground truth mask (B) using Eq. 4:

$$DSC = \frac{2|A \cap B|}{|A|+|B|} \quad (4)$$

where, |A∩B| denotes the number of true positive lesion voxels, and |A| and |B| represent the total voxels predicted as lesion and annotated as lesion, respectively. Owing to its sensitivity to small structures, DSC is particularly suited for prostate lesion segmentation, where target regions are often small relative to the entire gland volume.

5) Intersection over union (IoU): IoU evaluates the ratio of the intersection to the union of predicted and ground truth voxels as shown in Eq. 5:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (5)$$

Compared to DSC, IoU provides a stricter measure of segmentation performance, penalizing false positives more heavily by normalizing the intersection over the union rather than the average of sizes. A high IoU indicates not only strong overlap with the ground truth but also limited over-segmentation into non-lesion regions.

6) Voxel-wise accuracy: Voxel accuracy quantifies the overall proportion of correctly classified voxels across the entire MRI volume, including both lesion and background classes, using Eq. 6:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

where, TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. Unlike Dice and IoU, which focus specifically on the lesion class, voxel wise accuracy considers both background and lesion voxels. While it provides a global measure of classification performance, it can be less informative in highly imbalanced datasets (such as prostate lesions, where background voxels dominate), since high accuracy can be achieved even if small lesions are missed.

To further enhance evaluation robustness, Test Time Augmentation (TTA) was employed in the patch-based 3D U-Net (Approach 2). Predictions from augmented versions of the test volumes were averaged, yielding smoother lesion contours and reducing prediction variance. Comparative analysis demonstrated that while the full volume 3D U-Net (Approach 1) achieved competitive DSC and IoU scores, its ability to delineate fine lesion boundaries was limited due to severe class imbalance and GPU memory constraints. In contrast, the patch-based 3D U-Net (Approach 2) consistently outperformed the baseline, achieving higher DSC and IoU values. Aggressive augmentation, balanced sampling of lesion patches, and the combined use of dice and focal loss, which collectively corrected data imbalance and improved sensitivity to small lesions, were responsible for this improvement. The clinical reliability of this method was strengthened by the use of TTA, which further stabilized predictions. These results demonstrate the benefit of histopathology-guided, patch-based training in managing extremely unbalanced prostate MRI datasets and emphasize the significance of sophisticated data handling techniques, loss formulations, and augmentation in enhancing 3D U-Net architectures.

A. Quantitative Results

The improved performance of the patch-based 3D U-Net technique over the full volume baseline was well proven by quantitative examination. Three relevant voxel level measures are summarized in Table I as voxel wise accuracy, intersection

over union (IoU), and dice similarity coefficient (DSC). With a mean Dice score of 0.78, an IoU of 0.66, and voxel wise accuracy of 0.95, the full volume 3D U-Net demonstrated a satisfactory overlap with the pathology-informed ground truth masks. However, it was notably limited in its ability to delineate small lesions because of class imbalance and a decreased ability to resolve fine anatomical details. On the contrary, the patch-driven 3D U-Net without Test Time Augmentation (TTA) performed significantly better, with a voxel wise accuracy of 0.97, an IoU of 0.836, and a dice score of 0.902. These enhancements highlight how well localized training and balanced patch sampling capture fine-grained lesion boundaries. Performance further improved when TTA was applied to the patch-based model, yielding the best results with a voxel wise accuracy of 0.98, an IoU of 0.87, and a dice score of 0.92. This illustrates how segmentation stability is improved, and prediction variance is decreased by averaging predictions from several augmented perspectives. In pathologically informed prostate MRI analysis, these results show that the patch-based 3D U-Net with TTA consistently performs better than the full volume approach across all evaluation metrics, emphasizing the value of localized, augmentation-driven training strategies in resolving data imbalance and enhancing lesion segmentation.

TABLE. I. QUANTITATIVE RESULTS

Method	Table Column Head		
	Dice Similarity Coefficient (DSC)	Intersection over Union (IoU)	Voxel wise Accuracy
Full Volume 3D U-Net	0.78 ± 0.04	0.66 ± 0.05	0.95 ± 0.01
Patch based 3D U-Net (no TTA)	0.902 ± 0.03	0.836 ± 0.04	0.97 ± 0.01
Patch based 3D U-Net + TTA	0.923 ± 0.03	0.878 ± 0.04	0.98± 0.01

B. Qualitative Results

Qualitative visualization was performed to assess the segmentation performance of both approaches. Representative axial slices containing biopsy-confirmed lesions were selected from the T2-weighted MRI dataset. For each slice, three views were generated: (1) the original MRI slice, providing anatomical context; (2) the ground truth segmentation mask derived from biopsy-confirmed histopathology-informed annotations, outlining the true lesion regions and (3) the predicted segmentation mask generated by the trained model. These comparisons allowed a direct visual assessment of the agreement between model predictions and pathology-validated reference annotations. Fig. 3 presents a qualitative comparison of the segmentation results of the full volume 3D U-Net, the patch-based 3D U-Net without Test Time Augmentation (TTA) and the patch-based 3D U-Net with TTA against the ground truth. The ground truth masks shown in green delineate the true lesion boundaries, while the predicted masks shown in red represent model outputs. As illustrated, the full volume 3D U-Net generally produces smoother contours but occasionally fails to capture smaller lesions or inaccurately extends into adjacent non-lesion regions. In contrast, the patch-based 3D U-Net demonstrates greater precision in delineating lesion boundaries, capturing subtle and low-contrast lesions more effectively. The incorporation of Test Time Augmentation (TTA) further enhances the patch-based predictions by reducing noise, suppressing spurious detections, and generating anatomically plausible, contiguous lesion masks. The qualitative findings reinforce the quantitative improvements reported in Table I shows that the patch-based approach not only achieves higher dice and IoU scores but also yields visually superior and clinically reliable segmentation results compared to the full volume model. The performance improvement observed with Test Time Augmentation (DSC improvement from 0.902 to 0.923) reflects improved prediction stability rather than architectural change. Statistical validation using cross-validation folds will be explored in future work to quantify performance significance.

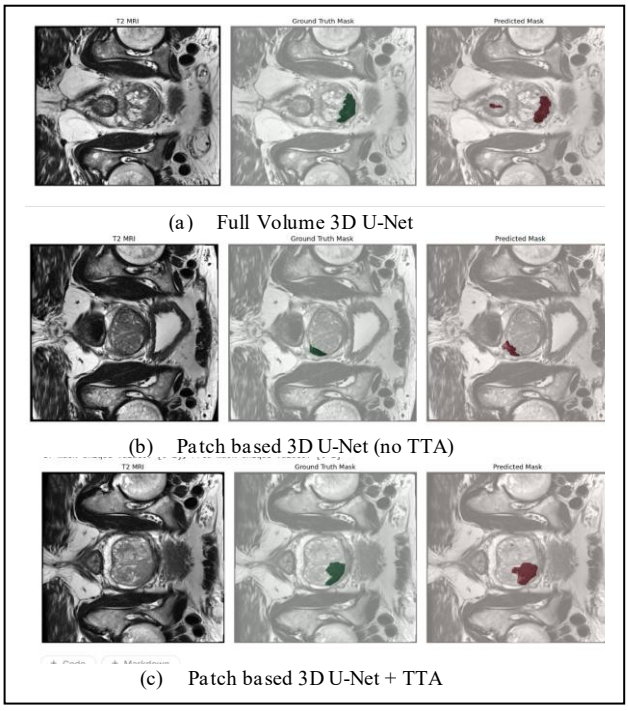

Fig. 3. Representative axial slices from T2-weighted MRI, ground-truth lesion masks (green) and predicted masks (red).

VI. POST PROCESSING

After inference, the probability maps generated by the 3D U-Net models were converted into binary masks using a threshold of 0.5. For the patch-based approach, overlapping patch predictions were averaged before thresholding to ensure smooth reconstruction. To further enhance anatomical plausibility, connected component analysis was applied, and isolated predictions smaller than 10 voxels were removed as likely false positives. While maintaining biopsy-confirmed lesion locations, this refining process assisted in lowering false positives. Additionally, segmentation masks that were more in line with clinical annotations were produced by optionally using morphological closure to smooth lesion boundaries. The resulting forecasts were guaranteed to be not only mathematically correct but also biologically interpretable due to these procedures.

VII. Comparison with previous Research

Table II compares the proposed method with recent prostate MRI segmentation studies. Zhu et al. (2025) reported a Dice score of 91.71% using a FIFO feature memory and multi-scale context modelling framework trained on radiologist annotations. Although competitive, their work primarily focuses on gland-level segmentation. Quihui-Rubio et al. (2023) achieved a DSC of 84.15% (IoU 76.9%) using an attention-based FAU-Net, while Song et al. (2023) reported a DSC of 70.14% with a 3D multi-scale attention CNN.

In contrast, the proposed biopsy-informed patch-based 3D U-Net achieves a DSC of 92.3% and IoU of 87.8%, outperforming existing approaches. The improvement is attributed to biopsy-verified voxel-level supervision, balanced patch-based training to mitigate class imbalance, and a hybrid Dice–Focal loss with Test-Time Augmentation. These results indicate that biologically validated supervision combined with balanced volumetric learning provides greater gains than architectural modifications alone.

TABLE. II. Comparison with SOTA

Ref.	Author and Year	Data Labels	Architecture used	DSC (%)	IoU (%)
[21]	J. Zhu et al., 2025)	Radiologist	FIFO Feature Memory + Multi-Scale Context Model	91.71	-
[22]	Quihui-Rubio et al., 2023	Radiologist	Attention U-Net (FAU-Net	84.15	76.9
[23]	Song et al., 2023	Radiologist	3D Multi-Scale Attention CNN	70.14	-
This work		Biopsy verified	Patch 3D U-Net Dice Focal	92.3	87.8

VIII. Discussion

Two 3D U-Net-based techniques for prostate lesion segmentation from T2-weighted MRI scans were presented and compared in this study. These techniques were trained and verified against ground truth that was informed by histopathology and biopsy confirmation. The first method, a whole volume 3D U-Net, conserved global anatomical context and provided a simple implementation. However, its capacity to precisely record low-contrast lesions was limited by GPU memory constraints and a noticeable class imbalance. The second strategy used a patch-based training paradigm that made use of a hybrid dice focal loss function, balanced sampling, and considerable data augmentation. This tactic kept the model resilient to interpatient variability while allowing it to concentrate more efficiently on underrepresented lesion voxels. The patch-based model's prediction stability was further enhanced by the addition of Test Time Augmentation (TTA), which resulted in segmentation masks that were smoother and anatomically coherent. The patch-based model steadily beat the full volume counterpart across all evaluation criteria, such as the dice similarity coefficient, intersection over union (IoU), and voxel wise accuracy. These results demonstrate the benefit of localized, high-resolution patches in preserving anatomical plausibility while capturing fine-grained lesion borders. The importance of pathology-informed ground truth in segmentation model training is shown by this work. The models are not only statistically accurate but also biologically validated by matching biopsy-confirmed histological results with MRI visible lesions, which lowers the possibility of overfitting to radiologist subjectivity. Thus, this method strengthens the clinical reliability of automated lesion segmentation by bridging the gap between image-based identification and pathological confirmation.

The improved performance of the proposed approach can be attributed to three key factors. First, the use of biopsy-confirmed histopathology-aligned lesion masks provides biologically accurate supervision compared to radiologist-drawn annotations, reducing label uncertainty. Second, the patch-based training strategy increases exposure to lesion voxels and mitigates the severe foreground-background imbalance inherent in prostate MRI segmentation. Third, the hybrid Dice Focal loss formulation simultaneously optimizes global overlap and voxel level discrimination, enabling better boundary delineation for small lesions. The addition of Test Time Augmentation further stabilizes predictions by averaging multiple transformed outputs, reducing variance and improving segmentation robustness.

IX. Limitations

Despite promising performance, the proposed framework has several limitations. First, the model is trained primarily on T2-weighted MRI, whereas multiparametric MRI sequences such as DWI and ADC maps may provide complementary diagnostic information. Second, the patch-based approach increases inference time due to patch aggregation and Test Time Augmentation. Third, biopsy-derived lesion masks represent sampled pathology locations rather than the full tumour extent, which may introduce spatial uncertainty in supervision. Future work will focus on integrating multiparametric MRI modalities, incorporating uncertainty-aware segmentation methods, and exploring lightweight architectures for faster clinical deployment.

X. Conclusion

This work demonstrates 3D UNet architectures to the specific challenges of prostate MRI through preprocessing, balanced sampling, advanced loss formulations and pathology-informed supervision that can substantially improve detection accuracy. The results provide strong evidence that the pathology-guided deep learning not only enhances quantitative performance but also ensures the biological validity of predictions. This framework offers a promising foundation for developing clinically deployable, automated prostate cancer detection systems and paves the way for future research integrating multi-parametric imaging and pathology into unified AI-driven diagnostic pipelines.

Acknowledgment

We gratefully acknowledge the Cancer Imaging Archive (TCIA) and the organizers of the SPIE-AAPM-NCI

PROSTATEx Challenge for providing access to the prostate MRI dataset used in this study.

REFERENCES

[1] T. Penzkofer and C. M. Tempany-Afdhal, "Prostate cancer detection and diagnosis: the role of MR and its comparison with other diagnostic modalities – a radiologist's perspective," NMR Biomed., vol. 27, no. 1, pp. 3–15, Jan. 2014, doi: 10.1002/nbm.3002.

[2] A. Stabile et al., "Multiparametric MRI for prostate cancer diagnosis: current status and future directions," Nat. Rev. Urol., vol. 17, no. 1, pp. 41–61, Jan. 2020, doi: 10.1038/s41585-019-0212-4.

[3] Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet. 2017;389(10071):815-822. https://doi.org/10.1016/S0140-6736(16)32401-1.

[4] R. Cuocolo et al., "Machine learning applications in prostate cancer magnetic resonance imaging," Eur. Radiol. Exp., vol. 3, no. 1, p. 35, Dec. 2019, doi: 10.1186/s41747-019-0109-2.

[5] Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. Interobserver reproducibility of the PI-RADS version 2 lexicon: a multicenter study of six experienced prostate radiologists. Radiology. 2016;280(3):793-804. https://doi.org/10.1148/radiol.2016152542.

[6] Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. MICCAI 2017. https://doi.org/10.1007/978-3-319-67558-9_28

[7] B. Turkbey, A. M. Brown, S. Sankineni, B. J. Wood, P. A. Pinto, and P. L. Choyke, "Multiparametric prostate magnetic resonance imaging in the evaluation of prostate cancer," CA. Cancer J. Clin., vol. 66, no. 4, pp. 326–336, July 2016, doi: 10.3322/caac.21333.

[8] S. Ursprung et al., "Prostate MRI Using Deep Learning Reconstruction in Response to Cancer Screening Demands—A Systematic Review and Meta-Analysis," J. Pers. Med., vol. 15, no. 7, p. 284, July 2025, doi: 10.3390/jpm15070284.

[9] M. J. Belue and B. Turkbey, "Tasks for artificial intelligence in prostate MRI," Eur. Radiol. Exp., vol. 6, no. 1, p. 33, Dec. 2022, doi: 10.1186/s41747-022-00287-9.

[10] S. K. Singh et al., "A novel deep learning-based technique for detecting prostate cancer in MRI images," Multimed. Tools Appl., vol. 83, no. 5, pp. 14173–14187, June 2023, doi: 10.1007/s11042-023-15793-0.

[11] H. Sherafatmandjoo, A. A. Safaei, F. Ghaderi, and F. Allameh, "Prostate cancer diagnosis based on multi-parametric MRI, clinical and pathological factors using deep learning," Sci. Rep., vol. 14, no. 1, p. 14951, June 2024, doi: 10.1038/s41598-024-65354-0.

[12] R. Islam, A. Imran, and Md. F. Rabbi, "Prostate Cancer Detection from MRI Using Efficient Feature Extraction with Transfer Learning," Prostate Cancer, vol. 2024, pp. 1–28, May 2024, doi: 10.1155/2024/1588891.

[13] J. Lin et al., "High-Resolution 3D MRI With Deep Generative Networks via Novel Slice-Profile Transformation Super-Resolution," IEEE Access, vol. 11, pp. 95022–95036, 2023, doi: 10.1109/ACCESS.2023.3307577.

[14] N. U and A. P.M., "MRI super-resolution using similarity distance and multi-scale receptive field-based feature fusion GAN and pre-trained slice interpolation network," Magn. Reson. Imaging, vol. 110, pp. 195–209, July 2024, doi: 10.1016/j.mri.2024.04.021.

[15] Y. Jiang, M. Yang, S. Wang, X. Li, and Y. Sun, "Emerging role of deep learning-based artificial intelligence in tumor pathology," Cancer Commun., vol. 40, no. 4, pp. 154–166, Apr. 2020, doi: 10.1002/cac2.12012.

[16] R. Cuocolo, A. Stanzione, A. Castaldo, D. R. De Lucia, and M. Imbriaco, "Quality control and whole-gland, zonal and lesion annotations for the PROSTATEx challenge public dataset," Eur. J. Radiol., vol. 138, p. 109647, May 2021, doi: 10.1016/j.ejrad.2021.109647.

[17] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman, "SPIE-AAPM PROSTATEx Challenge Data." The Cancer Imaging Archive, 2017. doi: 10.7937/K9TCIA. 2017.MURS5CL.

[18] R. Tibrewala et al., "FastMRI Prostate: A public, biparametric MRI dataset to advance machine learning for prostate cancer imaging," Sci. Data, vol. 11, no. 1, p. 404, Apr. 2024, doi: 10.1038/s41597-024-03252-w.

[19] Jasper Twilt, "AI in PCa imaging: The current status and future perspectives." Accessed: Apr. 10, 2025. [Online]. Available: https://uroweb.org/news/ai-in-pca-imaging-the-current-status-and-future-perspectives

[20] S. Saha et al., "An Artificial Intelligent System for Prostate Cancer Diagnosis in Whole Slide Images," J. Med. Syst., vol. 48, no. 1, p. 101, Oct. 2024, doi: 10.1007/s10916-024-02118-3.

[21] J. Zhu et al., "Accurate Prostate Segmentation in Large-Scale Magnetic Resonance Imaging Datasets via First-in-First-Out Feature Memory and Multi-Scale Context Modeling," J. Imaging, vol. 11, no. 2, p. 61, Feb. 2025, doi: 10.3390/jimaging11020061.

[22] 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," 2023, arXiv. doi: 10.48550/ARXIV.2309.01322.

[23] E. Song et al., "Prostate lesion segmentation based on a3D end-to-end convolution neural network with deep multi-scale attention," Magn. Reson. Imaging, vol. 99, pp. 98–109, Jun. 2023, doi: 10.1016/j.mri.2023.01.015.