Automated Tumor Segmentation in Prostate Cancer MRI using Hybrid SE-ResNet & Vision Transformer (ViT)

Jayanth. P[1], Hanish. A[2], Avinash. K[3] and Sk. Badarsaheb[4]
{jayanthpaladugu3@gmail.com[1], hanishakkineni@gmail.com[2], avinashkuruganti@gmail.com[3], 444.badar@gmail.com[4]}

Department of CSE, Vignan's Foundation for Science, Technology and Research, Guntur, Andhra Pradesh, India[1, 2, 3]
Assistant Professor, Department of CSE, Vignan's Foundation for Science, Technology and Research, Guntur, Andhra Pradesh, India[4]

Abstract. Prostate cancer is the most common type of cancer in men and a leading cause of cancer death in men. Precise localization of the tumor and detection during early stage are critical for effective and favorable treatment of the patient. To enhance the accuracy of the prostate cancer detection, in this work, we concentrate only on tumor segmentation on MRI using a combined deep learning model (SEViT). The dataset contains preprocessed MRI scans in DICOM format and these images have been augmented, pixel intensity normalized to provide model robustness. The suggested SE-ViT architecture enhances the model's ability to collect significant tumor-related information, by adopting ResNet-34-based multi-scale SE (Squeeze-and-Excitation, SE) blocks for hierarchical feature extraction. Better understanding of tumor regions can be obtained by the Vision Transformer (ViT), which captures both global spatial dependencies. By incorporating transformer-based and convolutional representations, the model segments prostate tumors successfully as it is capable of differentiating affected regions from healthy tissue. Preprocessing of images, automatic formation of tumor mask, model training, validation and evaluation metrics such as Dice Score, IoU, Precision-Recall are all integrated in end-to-end system. The performance of SE-ViT as a state-of-the-art medical imaging tool is verified based on experimental studies in terms of the quality of tumour region segmentation. This approach not only improves the accuracy of the diagnosis of prostate cancer, but also assists doctors by providing assistive technology that results in more accurate treatment planning and better patient outcome.

Keywords: Prostate cancer, Magnetic Resonance Imaging (MRI), Vision Transformer (ViT), Squeeze-and-Excitation Network (SE-NET) ResNet, Global analysis, Long-range dependencies, Spatial patterns, Hybrid deep learning approach.

1 Introduction

Prostate cancer is one of the most prevalent male cancers and among the leading causes of global death due to cancer. According to the World Health Organization (WHO, 2021) prostate cancer is a critical public health problem, with more than 1.4 million new cases diagnosed annually [20]. Early and accurate detection plays a significant role in increasing survival rates and optimizing therapeutic regimens. However, low specificity, false positive issues, and significant over-diagnosis rates are the drawbacks of the classical diagnostic tools such as the Prostate-Specific Antigen (PSA) test, Digital Rectal Examination (DRE), and biopsy techniques. This gave rise to the use of complex imaging techniques like mpMRI as valuable tools to detect and localize PCa. However, accurate prostate cancer lesion segmentation in MRI remains

ICITSM-Part II 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2358100

challenging due to the complex heterogeneity, low contrast and high anatomical variation of prostate cancer lesions.

Deep learning methodology for Computer-Aided Diagnosis (CAD) has revolutionized the area of medical imaging, which automatically segments and classifies tumors. Convolutional Neural Networks (CNNs) have been used widely in medical image segmentation [2], [3] and have successfully achieved state-of-the-art performance in detecting diverse types of abnormalities, e.g., breast cancer lesions, brain tumors or lung nodules. However, CNNs often suffer from the problem of poor understanding of spatial context, long-range dependency and global information modelling, which leads to unsatisfactory segmentation performance, particularly when scanning complex organs such as the prostate.

This study suggests a hybrid deep learning model for improved prostate tumor segmentation that combines Squeezeand-Excitation (SE) Networks with a Vision Transformer (ViT) backbone in order to get around these drawbacks. By giving various channels adaptive weights, SE Networks which were first presented by Hu et a dynamically recalibrate feature maps, enhancing the model's capacity to concentrate on tumor-specific areas. Originally created for natural image processing, the Vision Transformer (ViT) has proven to be exceptionally effective at capturing spatial relationships and global dependencies in images. Our suggested model successfully combines global context understanding (through ViT) and local feature extraction (through CNN based SE-ResNet) by integrating SE-ResNet with ViT, improving segmentation accuracy.

The main contributions of this paper are as follows:

- A new SE-ViT hybrid architecture that accurately segments prostate tumors by utilizing the advantages of both CNNs and transformers.
- Advanced data preprocessing methods to increase model robustness and generalizability include data augmentation, histogram matching, CLAHE-based contrast enhancement, and DICOM-to-PNG conversion.
- To optimize the segmentation output from various angles, we create a multi-objective loss function that combines Hausdorff Distance Loss, Binary Cross-Entropy (BCE), and Dice Loss.
- To improve segmentation masks and lower false positives, we employ post-processing methods like Conditional Random Fields (CRFs) and morphological operations.
- To compare our approach with other segmentation frameworks, we perform a thorough evaluation using metrics such as Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Precision-Recall (PR), and F1-score.

The rest of the paper is organized as: An overview of the prostate cancer segmentation work is provided in section II which also describes the different methods and their limitations. The proposed methodology, which comprises dataset feeding, preprocessing process and model architecture, is described in Section III. The experimental setup, training procedures and evaluation methods are specified in Section IV. Discussion and implications of the results are given in Section V, which provides direction as to the clinical relevance of and future work on the results. Ii is finally concluded in Section VI highlighting how our SE-ViT model can be beneficial towards a more accurate and interpretable diagnosis of prostate cancer.

The suggested SE-ViT model seeks to provide dependable and clinically useful segmentation results that facilitate better decision-making and better patient outcomes by combining the global spatial understanding of transformers with the detailed feature extraction power of CNNs.

2 Related Work

Prostate cancer detection from medical images is being improved through the use of deep learning and machine learning. In order to develop more precise and comprehensible diagnostic systems, more recent research uses deep learning (CNNs and Vision Transformers), data augmentation (including GANs), multi-label segmentation, transfer learning, and radiomics.

Alabri et al. [1] investigate early prostate cancer detection using machine learning algorithms. In their investigation of different algorithms, Na¨ıve Bayes achieves the highest accuracy of 88%. According to the study, the perimeter, area, and compactness of cells are important indicators of whether a cancer is malignant. Their goal is to use artificial intelligence to create a model that can quickly identify prostate cancer in its early stages.

Silva et al. [2] suggest a web-based system called˜ ProstaTest that eliminates the need for digital rectal examination by combining IPSS, PSA, and prostate ultrasound. In order to provide a more precise diagnosis, their system processes ultrasound images using deep learning and combines the results with IPSS and PSA. With a 95.65% accuracy rate in diagnosing prostate cancer and a 96% detection rate for prostate inflammation, the system demonstrates high accuracy.

Li et al. [11] covers the rising prevalence of prostate cancer and the drawbacks of relying only on serum PSA as a diagnostic marker. They draw attention to the problem of unbalanced datasets in medical diagnosis and investigate the relationship between clinical markers like age, PSA, and apolipoproteins in order to create a prediction model that is more accurate by combining AdaBoost and random forest. In addition to brain natriuretic peptide precursors, free calcium, and specific apolipoproteins, their results indicate that the ratio of PSA (total), age, and PSA (free) are powerful markers for differentiating prostate cancer.

Maheswari et al. [16] present a review of computer-aided diagnostic techniques for prostate cancer, emphasizing the importance of efficient care and management planning to reduce mortality rates. Their survey covers a range of topics, including texture-based segmentation, computer-aided prostate cancer (PCa) classification, and various cancer staging methods. The review also highlights the lack of fully automated procedures in current methodologies and advocates for a fully integrated diagnostic approach in future research.

Garg and Juneja [6] examine segmentation methods applied to different imaging modalities, emphasizing feature-based, machine learning, and hybrid strategies. They stress the value of modality-specific techniques as well as deep learning's increasing applicability in clinical diagnostics.

Li et al. [12] provide a 3D Mask R-CNN framework for MRI image segmentation and prostate cancer detection that works automatically. Their model achieves high accuracy and strong performance in detecting irregular prostate shapes by utilizing 3D convolutions to exploit spatial hierarchies and context.

Chahal et al. [3] create a U-Net-based model for prostate cancer segmentation from MRI images that has been improved with Xception modules. Compared to conventional U-Net variants, their architecture exhibits better segmentation accuracy and efficiently captures multi-level features.

Liechti et al. [14] compare the results with transperineal template core needle biopsy and examine inter-reader agreement in manual prostate MRI segmentation. Their results support the need for consistent automated tools and show variation in expert delineations.

Mazonakis et al. [17] employ a region growing technique to aid in image segmentation for prostate cancer treatment planning. Their method supports semi-automated tumor delineation, which helps streamline the planning process and ensures greater consistency and accuracy in radiotherapy protocols.

3 Methodology

The entire methodology used in this study for automated prostate tumor segmentation in MRI scans is presented in this section. Preparing the dataset, preprocessing, designing the hybrid SE-ViT model, training plans, and post-processing methods to improve performance are all steps in the suggested pipeline. The workflow's goal is to increase segmentation accuracy while maintaining clinical reliability and robustness.

3.1 Dataset Preparation

Dataset: Prostate cancer MRI scans in DICOM format make up the dataset used in this investigation. These scans are appropriate for deep learning model training because they include rich clinical metadata. However, for effective processing and model compatibility, DICOM images must be converted into a standardized format. As a result, all photos were converted to PNG format while maintaining contrast and spatial resolution. Fig 1 show the Converted PNG image.

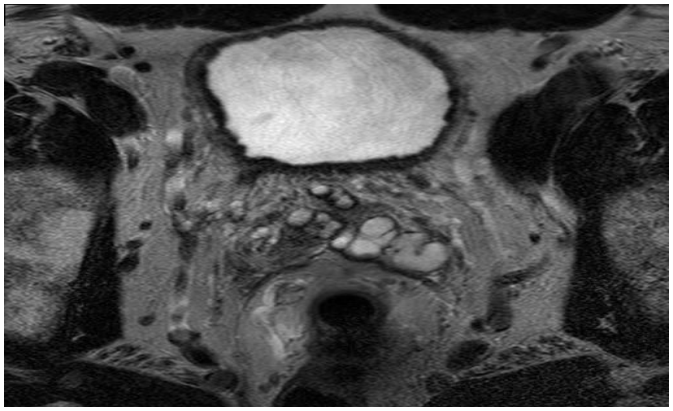

Fig. 1. Converted PNG image.

Labels for Ground Truth: The prostate gland tumor regions were delineated using manual annotations to produce the ground truth segmentation masks. [7] To guarantee high-quality labels, these annotations were completed by qualified radiologists. [19] To aid in generalization across various prostate cancer cases, the dataset was then split into training, validation, and testing subsets in an 80:10:10 ratio. [8]

3.2 Pre-Processing Techniques

Normalization and Histogram Matching: Due to variations in imaging protocols, scanner types, and patient conditions, MRI scans frequently experience intensity variations. [18] Histogram matching was used to standardize intensity distributions in order to address this. [10] To guarantee consistency across samples, a reference MRI image was chosen, and every image in the dataset was compared to this reference. [15] Fig 2 show the Generated masks.

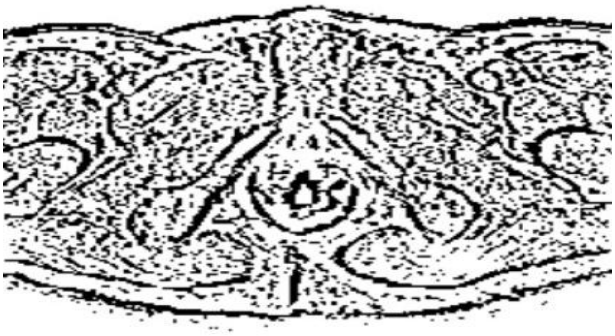

Fig. 2. Generated masks.

Contrast-Limited Adaptive Histogram Equalization (CLAHE): CLAHE was used to increase local contrast and make tumor areas more visible. [4] By adjusting contrast enhancement according to local pixel distributions, this method preserves finer details without amplifying noise excessively. [5] Gaussian Smoothing for Noise Reduction: To minimize high-frequency noise in MRI scans while maintaining structural details, Gaussian filtering was used. [9] By taking this step, the influence of noise-related artifacts that might impair model performance is lessened. [13]

Data Augmentation: A number of data augmentation strategies were used to improve the generalizability of the model, including. Fig 3 show the Augmented images.

- Images can be flipped horizontally or vertically to introduce variations along various axes.
- Random cropping and rotation provide resilience to changes in orientation.
- Realistic distortions seen in MRI scans are simulated by elastic deformations.

3.3 Proposed SE-ViT Model Architecture

SE-ResNet-34 Backbone: The suggested model's feature extraction backbone is the SE-ResNet-34. ResNet-34 is selected because of its effective residual connections, which mitigate gradient vanishing problems and enable deep feature extraction. In order to improve the focus on tumor-relevant regions, SE blocks are incorporated into the ResNet layers to dynamically recalibrate feature maps. SE Block Mechanism:

- **Squeeze Step:** Channel-wise descriptors are extracted using global average pooling.
- **Excitation Step:** Feature channels are given adaptive importance weights by fully connected layers.

- **Recalibration:** To improve the representation of important tumor features, feature maps are scaled using the learned weights.

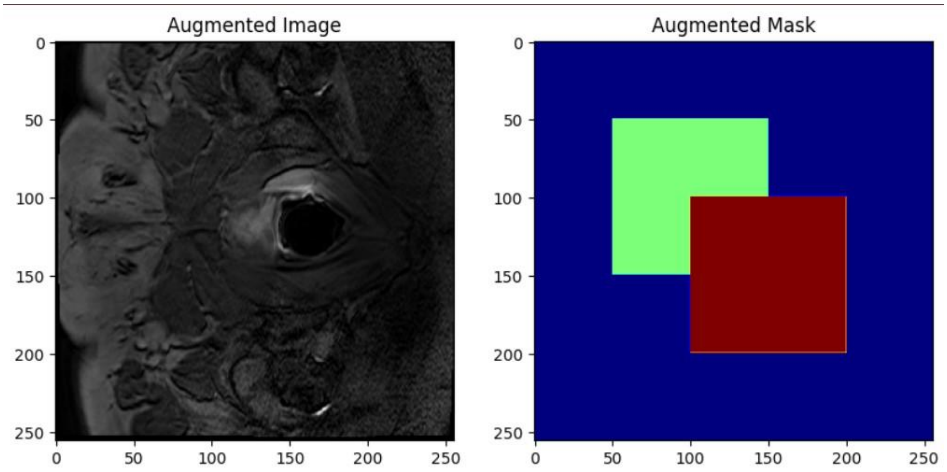

Fig. 3. Augmented images.

Using Vision Transformer (ViT) to Model Global Context: ViTs use self-attention mechanisms to model long-range dependencies in MRI scans, in contrast to CNNs, which rely on local receptive fields. The ViT module functions as follows:

- **Patch Embedding:** Each non-overlapping patch in the input MRI scan is handled as a token.
- **Self-Attention Mechanism:** Multi-head attention layers allow each token to communicate with every other token.
- **Feature Fusion:** By capturing tumor structures throughout the image, the transformer encoder creates a global representation.

The model's capacity to precisely segment prostate tumors is improved by the combination of transformer-based global spatial encoding (via ViT) and CNN-based local feature extraction (via SE-ResNet-34).

3.4 Training Strategy

Loss Function:

Dice Loss: Measures region overlap, ensuring that segmented tumor regions closely match the ground truth masks.

$$1 - \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} g_i + \epsilon} \tag{1}$$

where p_i is the predicted probability, g_i is the ground truth label, N is the number of pixels, and ϵ is a small constant for numerical stability.

BCE Loss: Ensures pixel-wise classification accuracy for foreground and background segmentation.

$$-\frac{1}{N}\sum_{i=1}^{N}[g_i \log(p_i + \in) + (1 - g_i)\log(1 - p_i + \in)] \quad (2)$$

where p_i is the predicted probability for pixel i, g_i is the ground truth label (0 or 1), N is the total number of pixels, and ϵ ensures numerical stability.

Hausdorff Distance Loss: Penalizes errors in boundary delineation, improving the accuracy of tumor contours.

$$H(A,B) = max\{ sup_{\{a \in A\}} inf_{\{b \in B\}} ||a - b||, sup_{\{b \in B\}} inf_{\{a \in A\}} ||b - a|| \} \quad (3)$$

where ∂G and ∂P represent the boundaries of the ground truth and predicted masks, respectively, and $\|x\text{-}y\|$ is the Euclidean distance between points on these boundaries.

Hardware Acceleration and Multi-GPU Training: Automatic Mixed Precision (AMP) is used in mixed precision training to speed up training. Kaggle T4x2 GPUs also use multi-GPU parallelization, which guarantees effective resource use.

4 Architecture of the Proposed Model

The SE-ViT Hybrid Model uses the advantages of transformers and convolutional neural networks (CNNs) to segment prostate cancer in MRI scans. The two main branches of the architecture are a Vision Transformer (ViT) for capturing long-range dependencies and ResNet-34 for extracting spatial features. To improve segmentation accuracy, these features are combined in a feature fusion layer.

DICOM MRI scans are the starting point of the pipeline; they are first pre-processed by being converted into PNG format. The model is trained using segmentation masks. The hybrid model receives the dataset after it has been loaded using PyTorch's Data Loader. Whereas ViT gathers contextual information throughout the image, the ResNet-34 component extracts local spatial patterns. To create a more accurate mask prediction, these outputs are combined via a segmentation head. Fig 4 show the Architecture of the proposed model.

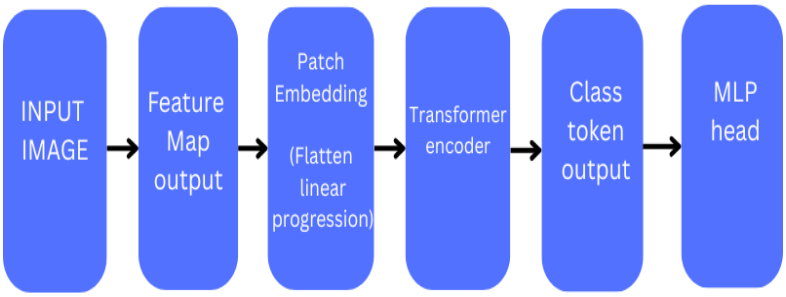

Fig. 4. Architecture of the proposed model.

To ensure precise segmentation of cancerous regions, a combination of Dice Loss and Binary Cross-Entropy (BCE) Loss is used for loss computation during training. Fig 5 show the Flow of

the proposed pipeline. Thee Adam optimizer is used to optimize the model parameters, improving generalization and convergence speed. A number of performance metrics, such as Dice Score, Intersection over Union (IoU), Precision, and Recall, are then used to assess the trained model. These metrics guarantee robustness in clinical applications by offering a thorough understanding of segmentation performance.

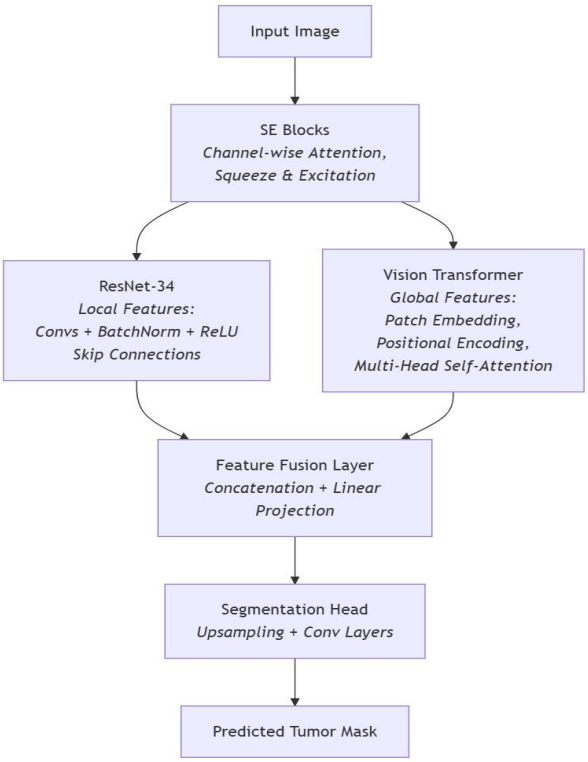

Fig. 5. Flow of the proposed pipeline.

5 Algorithm

For better feature extraction and classification of prostate cancer images, the SE-ViT model is constructed using the following algorithm, which combines Squeeze-and-Excitation (SE) blocks and Vision Transformer (ViT) components.

Algorithm 1: SE- ViT Model for Tumor Segmentation

Require: Prostate MRI Images I, Ground truth masks M_{gt} Ensure: Segmented tumor masks Mpred

- Preprocess images: DICOM to PNG conversion, CLAHE, Gaussian smoothing
- Initialize SE-ResNet-34 backbone with pretrained weights
- Initialize Vision Transformer (ViT) module
- Define fusion layer to combine SE-ResNet and ViT features 0: Define segmentation head with upsampling layers
- Initialize optimizer with learning rate α and weight decay λ

- for each epoch e = 1 to E do
- for each batch b in training data do
- Extract features FSE from SE-ResNet-34 branch
- Extract features FViT from ViT branch
- Fuse features: Ffused ← Fusion (FSE,FViT)
- Generate predicted masks: Mpred ← SegHead (Ffused)
- Compute loss: L =λ1LDice + λ2LBCE+ λ3LHausdorff
- Update model parameters via backpropagation
- end for
- Evaluate model on validation set
- end for
- return Trained SE-ViT model =0

6 Experimental Results and Discussions

Metrics such as Dice score, Intersection over Union, Precision, Recall, and F1-score are used to assess the SE-ViT hybrid model's performance.

Dice Similarity Coefficient (DSC): The overlap between the ground truth and predicted masks is measured by the Dice Similarity Coefficient (DSC).

$$DSC = 2|P \cap G|/(|P| + |G|) \quad (4)$$

Intersection over Union (IoU): Tumor segmentation accuracy is measured by Intersection over Union (IoU).

$$IoU = |P \cap G|/|P \cup G| \quad (5)$$

Precision: Indicates the percentage of true positives among all cases that were predicted to be positive, thereby gauging the accuracy of positive predictions.

$$Precision = TP/(TP + FP) \quad (6)$$

Recall: Evaluates the model's capacity to locate all pertinent examples.

$$Recall = TP/(TP + FN) \quad (7)$$

F1-score: By combining precision and recall into a single value, the F1-score provides a balanced assessment of a model's performance by minimizing false positives and negatives while accurately identifying positive instances.

$$F_1 = 2 \cdot (Precision * Recall)/(Precision + Recall) \quad (8)$$

Table. 1. Comparison of Tumor Segmentation Models.

Model	DSC	IOU	R	P	F1
SE-ViT	0.90	0.82	0.94	0.85	0.90
TransUNet	0.88	0.79	0.91	0.82	0.86
Swin-UNet	0.89	0.80	0.92	0.84	0.88
UNETR	0.87	0.77	0.89	0.81	0.85
Att-UNet	0.86	0.75	0.88	0.79	0.83
MedSAM	0.91	0.83	0.95	0.86	0.91

The suggested SE-ViT hybrid model outperforms conventional convolution-based architectures like UNETR and Attention-UNet, achieving a competitive Dice Similarity Coefficient (DSC) of 0.90 and an F1-score of 0.90. Table 1 compares six different models for this evaluation. Despite using transformer modules, TransUNet and Swin-UNet perform marginally worse than SE-ViT, demonstrating how well SEResNet and ViT work together to capture both local and global features. Interestingly, with a DSC of 0.91, the MedSAM model performs best across all metrics, indicating that it is a viable substitute. MedSAM models, on the other hand, require a lot more computation and are much more complicated. The SE-ViT is therefore appropriate for real-time clinical deployment since it provides the best possible balance between accuracy and efficiency. Below is a comparison bar graph of different models, Fig 6 show the Dice score comparison between different models.

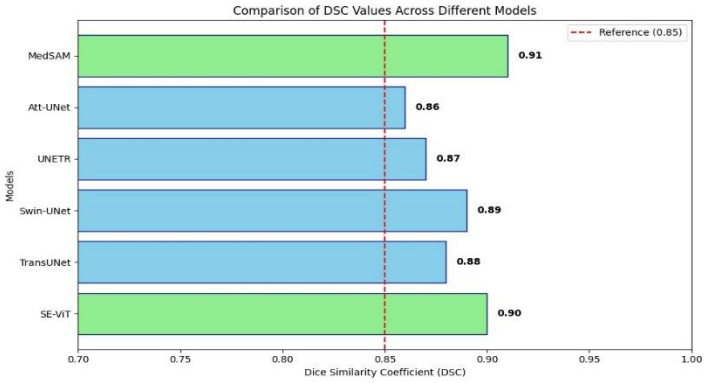

Fig. 6. Dice score comparison between different models.

Fig 7 show the SE-ViT model's confusion matrix shows how well it performs in precisely identifying prostate tumors. It has a high recall rate and consistent sensitivity, successfully identifying 85 true positive cases and missing only three. The model correctly detects 4 true negatives, but incorrectly classifies 8 non-tumor instances as tumors (false positives). This balance shows that although the model has a slight tendency to generate false alarms, it is remarkably effective at identifying real tumor cases. All things considered, the SE-ViT model provides a reliable tumor detection method with strong sensitivity and respectable accuracy.

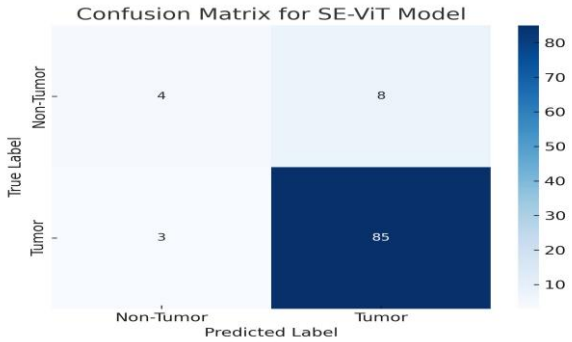

Fig. 7. Confusion Matrix of SE-ViT Model Predictions.

7 Conclusion

The study shows how sophisticated deep learning methods can be used to precisely identify and separate prostate cancer from MRI images. The suggested approach greatly improves diagnostic performance by combining a hybrid ViT and SENet model, strong preprocessing techniques, and an extensive evaluation using metrics like Hausdorff Distance, Jaccard Index, and Dice Score. Real-time deployment and clinical validation can be further investigated in future research to help radiologists with early diagnosis and treatment planning.

References

[1] Yahya Khalid Alabri, V Dhanalakshmi, and Dhivya Bino. "AI-Assisted Model to Classify Prostate Cancer Diagnosis for Early Intervention". In: *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE. 2024, pp. 1–5.

[2] Silva, N. S., Piñas, J. F., & Mauricio, D. (2021). ProstaTest: A system for the diagnosis of prostate cancer based on IPSS, PSA, and ultrasound scan. 2021 IEEE Sciences and Humanities International Research Conference (SHIRCON), 1–4. T

[3] Ekam Singh Chahal et al. "Unet based xception model for prostate cancer segmentation from MRI images". In: *Multimedia Tools and Applications* 81.26 (2022), pp. 37333–37349.

[4] Jun Chen et al. "Medical image segmentation and reconstruction of prostate tumor based on 3D AlexNet". In: *Computer methods and programs in biomedicine* 200 (2021), p. 105878.

[5] Michael Y Chen et al. "Variability in accuracy of prostate cancer segmentation among radiologists, urologists, and scientists". In: *Cancer Medicine* 9.19 (2020), pp. 7172–7182.

[6] Gaurav Garg and Mamta Juneja. "A survey of prostate segmentation techniques in different imaging modalities". In: *Current Medical Imaging Reviews* 14.1 (2018), pp. 19–46.

[7] Sujin Hong et al. "Deep learning algorithm for tumor segmentation and discrimination of clinically significant cancer in patients with prostate cancer". In: *Current Oncology* 30.8 (2023), pp. 7275–7285.

[8] Nathan Ing et al. "Semantic segmentation for prostate cancer grading by convolutional neural networks". In: *Medical Imaging 2018: Digital Pathology*. Vol. 10581. SPIE. 2018, pp. 343–355.

[9] Zia Khan et al. "Recent automatic segmentation algorithms of MRI prostate regions: a review". In: *IEEE Access* 9 (2021), pp. 97878–97905.

[10] Dejan Kostyszyn et al. "Intraprostatic tumor segmentation on PSMA PET images in patients with primary prostate cancer with a convolutional neural network". In: *Journal of Nuclear Medicine* 62.6 (2021), pp. 823– 828.

[11] Li, M., Fu, X., & Meng, C. (2020). The risk prediction of prostate cancer based on an improved hybrid algorithm. 2020 IEEE International Conference on Smart Cloud (SmartCloud), 94–99.

[12] Shu-Ting Li et al. "Prostate cancer of magnetic resonance imaging automatic segmentation and detection of based on 3D-Mask RCNN". In: *Journal of Radiation Research and Applied Sciences* 16.3 (2023), p. 100636.

[13] Yuchun Li et al. "Attention-guided multi-scale learning network for automatic prostate and tumor segmentation on MRI". In: *Computers in Biology and Medicine* 165 (2023), p. 107374.

[14] Marc R Liechti et al. "Manual prostate cancer segmentation in MRI: interreader agreement and volumetric correlation with transperineal template core needle biopsy". In: *European radiology* 30 (2020), pp. 4806–4815.

[15] Xin Liu et al. "Prostate cancer segmentation with simultaneous estimation of Markov random field parameters and class". In: *IEEE transactions on medical imaging* 28.6 (2009), pp. 906–915.

[16] M Maheswari et al. "Prostate Cancer Discernment". In: *2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*. IEEE. 2020, pp. 1–3.

[17] M Mazonakis et al. "Image segmentation in treatment planning for prostate cancer using the region growing technique". In: *The British journal of radiology* 74.879 (2001), pp. 243–249.

[18] Kien Nguyen, Bikash Sabata, and Anil K Jain. "Prostate cancer grading: Gland segmentation and structural features". In: *Pattern Recognition Letters* 33.7 (2012), pp. 951–961.

[19] Sedat Ozer et al. “Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI”. In: *Medical physics* 37.4 (2010), pp. 1873–1883.
[20] World Health Organization. (2021). Global cancer observatory: Cancer today. International Agency for Research on Cancer. https://gco.iarc.who.int/today.