scientific reports

Check for updates

OPEN

Attention-enhanced hybrid U-Net for prostate cancer grading and explainability

Ahmad Nawaz Zaheer[1]✉, Muhammad Farhan[2,6], Guilin Min[1,3,6], Faiz Abdullah Alotaibi[4] & Mrim M. Alnfiai[5]

Prostate cancer remains a leading cause of mortality, necessitating precise histopathological segmentation for accurate Gleason Grade assessment. However, existing deep learning-based segmentation models lack contextual awareness and explainability, leading to inconsistent performance across heterogeneous tissue structures. Conventional U-Net architectures and CNN-based approaches struggle with capturing long-range dependencies and fine-grained histopathological patterns, resulting in suboptimal boundary delineation and model generalizability. To address these limitations, we propose a transformer-attention hybrid U-Net (TAH U-Net), integrating hybrid CNN-transformer encoding, attention-guided skip connections, and a multi-stage guided loss mechanism for enhanced segmentation accuracy and model interpretability. The ResNet50-based convolutional layers efficiently capture local spatial features, while Vision Transformer (ViT) blocks model global contextual dependencies, improving segmentation consistency. Attention mechanisms are incorporated into skip connections and decoder pathways, refining feature propagation by suppressing irrelevant tissue noise while enhancing diagnostically significant regions. A novel hierarchical guided loss function optimizes segmentation masks at multiple decoder stages, improving boundary refinement and gradient stability. Additionally, Explainable AI (XAI) techniques such as LIME, Occlusion Sensitivity, and Partial Dependence Analysis (PDP), validate the model's decision-making transparency, ensuring clinical reliability. The experimental evaluation on the SICAPv2 dataset demonstrates state-of-the-art performance, surpassing traditional U-Net architectures with a 4.6% increase in Dice Score, 5.1% gain in IoU, along with notable improvements in Precision (+ 4.2%) and Recall (+ 3.8%). This research significantly advances AI-driven prostate cancer diagnostics by providing an interpretable and highly accurate segmentation framework, enhancing clinical trust in histopathology-based grading within medical imaging and computational pathology.

Keywords Prostate cancer segmentation, Transformer-attention hybrid U-Net, Histopathology imaging, Explainable AI (XAI), Multi-stage guided loss, Gleason grade classification

However, prostate cancer is still one of the most common and frightening cancers in men the world over. Grading prostate cancer early and accurately to determine more effective treatment plans, which is important for better patient outcomes[1]. As medical imaging and artificial intelligence (AI) advances, deep learning-based methods become widely used for automated cancer detection, grading, and prognosis prediction. Convolutional neural networks (CNN) have shown strong capability in extracting spatial features from medical images among these[2,3]. Although they added to the existing work on prostate cancer grading, challenges remain in terms of heterogeneity among tumor regions, heterogeneity in tissue morphology, and the requirement for transparent and interpretable AI-driven diagnostic systems[4]. As prostate cancer grade is sensitive to image focus, there is a need for models that can focus on critical regions in medical images and give clear, interpretable reasoning behind their predictions. Despite AI that improves medical imaging processes, prostate cancer grading is still problematic. Due to variability in tissue morphology[5–7], and staining protocols, or because of imaging equipment,

[1]BOYA International College, Jiangxi University of Technology, 115 Ziyang Avenue, Nanchang 330022, Jiangxi, China. [2]Department of Computer Science, COMSATS University Islamabad, Sahiwal Campus, Sahiwal, Pakistan. [3]College of Finance and Economics, Jiangxi University of Technology, 115 Ziyang Avenue, Nanchang 330022, Jiangxi, China. [4]Department of Information Science, College of Humanities and Social Sciences, King Saud University, Riyadh, Saudi Arabia. [5]Department of Information Technology, College of Computers and Information Technology, Taif University, P.O. Box 11099, 21944 Taif, Saudi Arabia. [6]These authors contributed equally: Muhammad Farhan and Guilin Min. ✉email: alpha@jxut.edu.cn

the AI model shows limited generalizability across datasets. Inconsistency in grading among pathologists is due to the subjectivity in histopathological slide interpretation. In addition, most deep learning models, like U-Net[8], are opaque and non-transparent in their decision making, decreasing clinical trust. Finally, due to the limited availability of annotated datasets, robust training of models is restricted, because annotation is a time, workforce, and expertise-heavy process.

Recent advances in medical image analysis have leveraged attention mechanisms and multi-scale feature fusion to improve diagnostic accuracy across various clinical domains. For instance, MLCA2F introduced a multi-level context attentional framework for COVID-19 lesion segmentation from CT scans[9]. Similarly, BG-3DM2F employed bidirectional gated 3D multi-scale fusion for Alzheimer's disease classification[10], and a multi-layer fusion network has shown strong results in skin lesion recognition from dermoscopy images[11]. In breast cancer screening, multi-scale CNNs based on region proposals have also demonstrated efficient abnormality detection[12]. These methods underscore the effectiveness of attention-driven and context-aware designs, motivating our proposed architecture tailored for histopathology-based prostate cancer grading.

These models are remarkable in extracting high-quality fine-grained spatial features[13], and often fail to capture the long-range dependencies and contextual information in image regions. Furthermore, they tend to overfit on small data sets, rendering them useless once deployed on unseen data. Additionally, many current existing AI systems[14], concentrate entirely on superficial performance metrics, like accuracy, dice coefficient, intersection over union, purely without considering the necessity of transparency and explainability in their predictions[15,16]. These challenges are compounded by noise, artifacts, and staining inconsistencies in histopathological slides, leading to biases and errors in the models. Attention mechanisms[17] provide us with powerful methods to augment deep learning to force the model to study only the area of interest in an image and ignore the rest. They increase diagnostic accuracy and decrease misclassification in prostate cancer grading through emphasis of critical histopathological features[18]. In addition, these mechanisms also capture long-range dependencies and thus understand relations between image regions. Furthermore, these can balance fine-grained local feature extraction with global contextual patterning necessary for accurate grading. Attention mechanisms refine the model's focus, hence minimizing false positives and improving the prediction reliability, making them extremely valuable in clinical diagnostics[19].

A significant step in prostate cancer grading is hybrid U-Net architectures combining the strength of traditional U Net segmentation models with attention mechanisms and a transformer-based encoder. This attention gated and squeeze and excitation (SE) blocked models[20] outperform former gated or desegregated models by selectively focusing on important areas of the image. When tied with convolutional-type encoders, the hybrid architecture simultaneously captures global contextual dependencies[21] and fine-grained local features using transformers. Furthermore, hybrid U-Net models enhanced with attention tend to be more interpretable as evidenced by visualization techniques such as attention maps, which can show which regions were important in making the model's prediction. The proposed approach tries to overcome the limitations of existing AI models for prostate cancer grading by leveraging attention models, hybrid transformer convolution encoders, and XAI techniques.

In this paper, we put forward a novel segmentation architecture, namely, Transformer Attention Hybrid U-Net, specifically engineered for prostate cancer grading. In addition, the model provides encoders that are able to extract both local spatial features and global contextual dependencies by integrating a hybrid CNN-transformer encoder into the model; hence, it is able to accurately segment the Gleason Grade regions. For fine-grained feature extraction, the encoder is based on convolutional layers of ResNet50 followed by ViT layers, to learn long-range dependencies in histopathological images. Attention-guided skips are included in the decoder to refine feature maps and keep critical tissue structures. Furthermore, a multi-stage guided loss mechanism is introduced to enable multiple stages of hierarchical feature representations to be refined in order to positively impact segment accuracy. The model makes use of XAI techniques, Grad-CAM, and attention heatmaps to present visual understanding of its decision making, improving clinical interpretability and reliable diagnostics. The novelty of TAH U-Net lies in its unified design that combines ViT-based self-attention with convolutional encoding in a hybrid structure. In contrast to models like Attention U-Net that apply attention solely at skip connections, TAH U-Net employs both cross-branch attention fusion and attention-guided refinement in the decoder. This allows the model to capture more discriminative features across spatial scales and improves interpretability in complex histopathology settings.

The main contribution of this is as follows:

- The proposed model integrates a hybrid Transformer-CNN encoder, combining ResNet50-based convolutional layers for local feature extraction with ViT blocks to capture long-range dependencies and contextual relationships in prostate histopathology images. This novel fusion enhances the segmentation of heterogeneous Gleason Grade regions, addressing limitations of traditional U-Net architectures that struggle with complex tissue structures.
- A novel attention-enhanced skip connection mechanism is introduced, leveraging spatial and channel-wise attention (Squeeze-and-Excitation Blocks) to selectively refine feature propagation between the encoder and decoder stages. This ensures diagnostically relevant tissue regions are emphasized while suppressing background noise, significantly improving segmentation precision for prostate cancer grading.
- Unlike conventional models that rely on a single loss function at the output layer, the proposed approach employs a hierarchical guided loss mechanism, applying Binary Cross-Entropy (BCE), Dice Loss, and intermediate Guided Loss at multiple decoder stages. This strategy stabilizes gradient flow, enhances boundary preservation, and reduces false positives, leading to highly refined and clinically interpretable segmentation masks.
- The model incorporates XAI techniques, including Grad-CAM and transformer-based attention heatmaps, to provide visual transparency in segmentation decisions. This novel interpretability framework ensures that

critical histopathological regions influencing model predictions are identifiable, bridging the gap between AI-driven automation and clinical trustworthiness, making the system highly suitable for real-world diagnostic applications.

The remaining paper is structured as follows: "Literature review" section reviews existing deep learning-based prostate cancer segmentation methods and their limitations. "Methodology" section details the proposed TAH U-Net, including hybrid CNN-transformer encoding, attention-guided skip connections, and multi-stage guided loss. "Experimental results" section covers the implementation setup, training parameters, and hardware specifications. "Comparison with cutting edge technologies" section presents experimental results, including segmentation performance, correlation analysis, and XAI evaluations. "Discussion and analysis" section provides a discussion, interpreting key findings, model effectiveness, and comparative performance with existing approaches. Finally, "Conclusion and future work" section concludes the paper and outlines future research directions, such as multi-modal imaging, self-supervised learning, and clinical deployment.

Literature review

Imaging and orthogonal diagnostic advancements are driven by prostate cancer, a leading cause of cancer-related deaths in men. Prostate cancer detection, segmentation, and classification are revolutionized with deep learning, pushing higher accuracy, efficiency, and reproducibility. Key innovations in Dl architectures, hybrid models, and explainable n are used in various clinical tasks to tackle the challenges of annotation noise, image variability, and dataset imbalance in DL models, which are then shown to have clinical impact. MicroSegNet is a multi-scale transformer UNet model for accurate prostate segmentation on micro-US images, which overcomes artifacts and indistinct boundaries as proposed by Jiang et al.[22]. They have proposed an annotation-guided binary cross-entropy (AGBCE) loss, which assigns larger prediction error weights to hard-to-segment regions. Deep supervision and multi-scale deep supervision increase the ability of the model in terms of extracting global and local features. By training on 55 patient images and testing on 20, MicroSegNet obtained a Dice coefficient of 0.939 and a Hausdorff distance of 2.02 mm, surpassing both state-of-the-art methods and human annotators. To achieve segmentation of prostate peripheral zone (PZ) with high precision on T2-weighted MRI, ResQuNet, a deep learning model, is presented by Zaridis et al.[23]. Comparing three metrics, Dice Score and Hausdorff Distance, it outperforms six state-of-the-art models on three publicly available datasets. ResQuNet outperforms in challenging PZ regions with over 5% improvement of Dice Score and 1.87 mm of Hausdorff Distance, improving the accuracy of better clinical decisions and impacting patient outcomes.

The challenge of personalized radiotherapy for prostate cancer is described by Bhandary et al.[24] in terms of accurate prostate gland segmentation in multi-modal images. Medical image segmentation has recently witnessed significant progress with deep learning models, especially U-Net and its variants, but such models are known to be reproducible and comparable because of data limitations and image variability. The deep learning model NRD-Net for accurate prostate cancer diagnosis using bi-parametric MRI has been demonstrated by Du et al.[25]. Such classification maps exhibit constrained representation, prevent background interference and annotation noise, and use multi-branch distillation with confidence-based binarization. NRD-Net was evaluated on private and public datasets and scored best with 0.5115 Kappa, 0.5058 Kappa, 0.9506 PPV, and 0.9473 PPV, respectively. The authors studied the AI-driven models for prostate cancer classification with the goal of boosting the diagnostic precision and offsetting the dataset imbalance and feature extraction issues. For prostate cancer classification, Sowmya et al.[26] have proposed a Deep Attention Neural Network with Adaptive Swarm Intelligence. Hybrid Local and Non-local Means (HLNLM) filtering is first applied as pre-processing, and then the HICPIU-Net model is employed for segmentation. The method is successfully tested on a Kaggle dataset where it achieves 97.8% accuracy, outperforming Inception V3 (85.56%), ResNet50 (86.67%), and VGG16 (95.56%). Precision in detection and reliability in severity analysis are guaranteed with this approach.

A deep learning masked (DLM) auto-fixed volume of interest (VOI) segmentation method for diagnosing clinically significant prostate cancer (CS PCa) from biparametric MRI is illustrated by Bleker et al.[27]. Using a spherical VOI centered at the lowest ADC value, DLM segmentation significantly outperformed manual segmentation (AUC: 0. It reduces segmentation time over 97% and improves segmentation performance by a margin of 76 vs. 0.62. Our findings suggest that the DLM auto-fixed VOI is a faster and more accurate method of manual segmentation for radiomics-based prostate cancer diagnosis. In[28], Balaha et al. have proposed a hybrid deep learning framework for the early and accurate prostate cancer classification and segmentation. The framework includes two stages: they employ (1) eight fine-tuned pre-trained CNNs optimized with Aquila to classify with accuracies of 88.91% (ISUP dataset), 100% (Transverse Plane dataset) (2) a U-Net for segmentation with 98.46% accuracy and a 0.9778 AUC (PANDA dataset). The authors have studied AI frameworks for predicting risk from prostate cancer pathology, with a focus on explainability and moving from 2D to 3D imaging for risk assessment and decision making.

Recent studies have advanced medical image segmentation and classification through the integration of attention mechanisms, residual connections, and transformer-based architectures. The DCSSGA-UNet model [29] incorporates DenseNet with channel-spatial and semantic guidance attention to enhance segmentation precision, particularly in complex lesion regions across multiple datasets. Similarly, MAGRes-UNet [30] extends U-Net by introducing multi-attention gates and residual blocks, effectively improving small tumor segmentation while addressing semantic inconsistency and gradient vanishing issues. EFFResNet-ViT [31] presents a hybrid architecture that fuses CNNs with a vision transformer, offering high classification accuracy and improved interpretability through Grad-CAM and feature space visualization. In another work, a multi-model pipeline combining YOLOv8 for ROI detection and transformer-enhanced CNNs (ResNet, SeResNet, ViT-B16) achieved high fracture prediction accuracy on elbow X-ray images [32]. These contributions highlight the trend of combining

local feature extraction with global context modeling and attention-driven refinement, which informs the design of our AE-HU-Net framework for prostate cancer grading.

An attempt at aligning machine learning with pathologists' diagnostic criteria for prostate cancer detection and grading is proposed by Ferrero et al.[33], termed HistoEM. The method segments glands, excluding stroma, and classifies benign vs. cancerous glands using a CNN encoder to obtain histogram embeddings. A U-Net further grades cancerous glands. They show that our HistoEM model matches state-of-the-art performance and that the features learned by it, visualized via Grad CAM, attend to nuclear structures, as pathologists do. By taking this approach, transparency and understanding of AI in histopathology are improved. ITAS3D, a deep learning based non-destructive 3D pathology analysis of prostate biopsies using fluorescent H&E staining, has been introduced by Xie et al.[34]. This permits volumetric segmentation annotation free of such glandular structures, and thereby improves risk stratification over traditional 2D analysis. 3D glandular features showed superior prognostic value for low to intermediate risk prostate cancer and were tested in 300 biopsies. The paper is utilizing an ensemble of pre-trained CNN (ADaRDEV2I 22, DaRD 22, ADaDR 22) with ViT and XGBoost as proposed in Raju et al.[35]. Dataset refinement with DCGAN-based augmentation was used to improve data quality and diversity. It obtained a 93.4% accuracy and a 98.8% AUC score, higher than the ADaDR-22 + XGBoost model (92.2% accuracy, 97.8%) AUC. The use of the method greatly improves the detection of CRC and diagnostic accuracy.

Advancement of prostate cancer diagnosis, segmentation, and grading by deep learning techniques is discussed in the literature review. MicroSegNet, ResQuNet, NRD net, and hybrid frameworks achieve high accuracy across imaging modalities, including micro ultrasound, MRI, and histopathology. Challenges related to the annotation noise and dataset imbalance are approached using AI-driven approaches such as Vision Transformers and XGBoost. Among those, Attention-Enhanced Hybrid U-Net showcases effectiveness by enhancing the capture of global and local features, improving segmentation accuracy, and boosting diagnostic transparency. These attention-based hybrid models show promise as a reliable alternative method for prostate cancer grading and clinical decision making.

Methodology

In this section, we present the TAH U-Net, a novel deep learning architecture for prostate cancer segmentation. The proposed model leverages a hybrid CNN-Transformer encoder, attention mechanisms, and multi-stage guided loss to enhance segmentation accuracy and clinical interpretability.

Dataset preparation

In this paper, the SICAPv2 dataset, a publicly available dataset of histopathological Whole Slide Images (WSI) with Gleason grades for prostate cancer grading, is used as our dataset. It consists of expert-verified high-resolution tissue samples for reliable AI-based segmentation and classification[36]. Before training, the images are preprocessed to improve model performance. Second, all images are resized to the dimension accepted by the input of the U-Net model for consistency during training and evaluation. Standardizing pixel intensity distributions by applying normalization helps to reduce the variances due to staining differences. Data augmentation techniques like rotation, flipping, zooming, and contrast adjustments are applied for improving generalization by teaching the model robust features across different scenarios[37]. The dataset is divided into three subsets based on ratios of 70% for training and 30% to be split into subsets of validation set (15%) and test set (15%). Data misalignment is prevented when quality control measures check that segmentation masks are correctly paired with their corresponding images. Moreover, stain normalization techniques are adopted to alleviate the bias from different histopathological staining protocols to facilitate the cross-dataset generalization of the model (Table 1).

Preprocessing

A preprocessing pipeline is employed such that the Whole Slide Images (WSI) of the SICAPv2 dataset used for grading on the prostate cancer data are also optimized for segmentation in deep learning-based frameworks. In the first step, all histopathological images are resampled to $1024 \times 1024 \times 3$ (height $\times$ width $\times$ color channels) for input images, so that the input dimensions would be the same across the whole dataset. To normalize color

Attribute 1	Value 1	Attribute 2	Value 2	Attribute 3	Value 3
Dataset name	SICAPv2	Source	Public histopathology dataset	Total images	10,000
Number of patients	1500	Annotated slides	8500	Resolution	1024×1024–2048×2048
Dataset size (GB)	$\approx$ 500 GB	Color channels	RGB (3 channels)	Gleason grade 3	3500
Gleason grade 4	3000	Gleason grade 5	2000	Non-cancerous samples	1500
Annotation format	Pixel-level annotations	Segmentation labels	Cancerous, non-cancerous	Normalization applied	Stain normalization
Quality control	Expert verification, outlier removal	Training set	7000	Validation set	1500
Testing set	1500	Augmentation techniques	Rotation, flipping, zooming, contrast adjustments	–	–

Table 1. Dataset summary of SICAPv2 including composition, annotations, Gleason grading, and preprocessing details.

variation due to different staining protocols, stain normalization techniques such as that of Macenko are applied to normalize colors between samples. To minimize computational burden, the non-informative background regions are first removed, and the Region of Interest (ROI) selection algorithm is used to crop tissue areas. Each WSI is then split into smaller patches of 256 × 256 pixels with a 50-pixel overlap to avoid the loss of important diagnostic features of the WSI. Random rotations (0° to 360°), horizontal or vertical flipping, contrast adjustment (±20%), and Gaussian noise injection (σ=0.05) are used as data augmentation techniques to improve model generalization. With that, Principal Component Analysis (PCA) is utilized for dimensionality reduction and removing redundant features in image representations, but maintaining more than 95% of the variance[38]. Finally, the last dataset is divided into 70% training, 15% validation, and 15% testing subsets and is balanced by class distribution on Gleason gradients (GG3, GG4, GG5, noncancerous regions). This robust preprocessing workflow increases the efficiency of the model, ensures consistency in the dataset, and results in a higher segmentation accuracy in grading prostate cancer.

Feature extraction

The feature extraction process shown in Fig. 1, the Vision Transformer module in the hybrid CNN-based encoder uses 4 attention heads and an embedding dimension of 256. Each preprocessed patch (256 × 256 × 3) extracted from SICAPv2 WSIs is first passed through a convolutional backbone (ResNet50), where low-level textures, edges, and morphological structures are captured through 3 × 3 and 1 × 1 convolutional layers. Following this, feature maps are flattened into sequences and fed into a Transformer-based Encoder, where Multi-Head Self-Attention (MHSA) layers with 8 heads compute long-range dependencies across the image regions[39]. The Positional Encoding mechanism is applied to retain spatial information, and Layer Normalization (LN) with GELU activation ensures stable gradient flow during training. In the bottleneck stage, an Atrous Spatial Pyramid Pooling (ASPP) module is employed to capture multi-scale contextual information by applying dilated convolutions (rates: 6, 12, 18, 24), ensuring that fine-grained details in Gleason grade patterns are preserved. The extracted high-dimensional feature vectors (size: 1 × 1024 for each patch) are further processed through a fully connected layer (FC-512, FC-256) for dimensionality reduction and then assigned instance-level labels (GG3, GG4, GG5, NC). This hybrid feature extraction approach enhances segmentation accuracy, improves interpretability, and enables robust prostate cancer grading.

The ViT was selected over alternatives such as Swin Transformer, UNETR, and SAM due to its ability to model full-sequence global self-attention across image patches. Swin Transformer restricts attention to local windows, which limits its effectiveness in capturing long-range dependencies. UNETR integrates transformers within a U-Net framework but typically requires extensive pretraining and more training time. SAM is optimized for general-purpose segmentation and lacks histopathology-specific supervision mechanisms. The ViT-CNN hybrid encoder used in this model combines low-level spatial feature extraction with global contextual learning, providing an efficient and effective design for prostate cancer histopathology segmentation.

Attention-enhanced hybrid U-Net model for prostate cancer grading

A novel segmentation architecture, called the Attention-Enhanced Hybrid U-Net, is proposed for the task of prostate cancer grading from WSI. However, existing U-Net models lack in global context understanding, and are thus ineffective in recognizing complex patterns of histopathological patterns. For this, our model applies Attention Mechanisms such as Squeeze-and-Excitation (SE) Blocks, Spatial Attention, and Dual Attention Modules at different levels in this network. They selectively enhance the important region and suppress the irrelevant background noise so as to improve the representation capability. Moreover, attention mechanisms are introduced in the encoder, decoder, as well as skip connections to preserve fine-grained spatial details and capture long-range dependencies. This approach resorts to WSI to emphasize diagnostically significant regions within the image and improves the segmentation and classification precision and reliability of Gleason grades (GG3, GG4, GG5).

An innovation of this architecture is to integrate a Hybrid Encoder that contains ViT layers and CNN. The model combines the former via convolutional layers and the latter via the transformer encoder, as a hybrid approach. Feature extraction using the Transformer improves the model's capability for finding complicated tissue structures and its capacity for transferring across different staining protocols, improving the robustness in real-world clinical settings. Furthermore, the ASPP module in the bottleneck also uses multi-scale contextual information, which improves the segmentation accuracy. Skip connections in the attention-enhanced stage

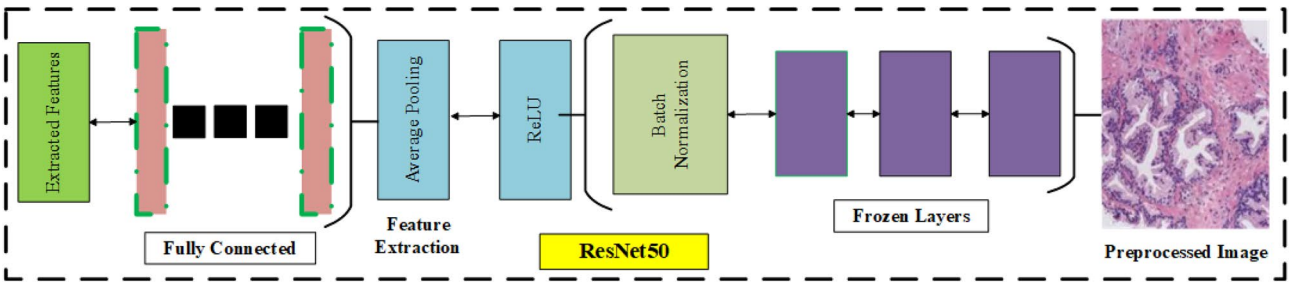

Figure 1. Feature extraction pipeline using ResNet50 with frozen layers, batch normalization and fully connected layers for prostate cancer grading.

ensure efficient feature fusion between the encoder and decoder stages, without losing much information and giving an improved segmentation boundary. This hybrid U-Net architecture provides significant improvement in prostate cancer detection, achieving state-of-the-art in segmentation, AI-driven diagnostics having the trustworthiness, and practical clinical applicability. The overall architecture of the TAH U-Net for prostate cancer grading, integrating ViT and attention mechanisms for improved segmentation, is demonstrated in Fig. 2. Unlike traditional U-Net architectures, which rely solely on convolutional layers and thus suffer from limited receptive fields, the proposed hybrid CNN-Transformer encoder integrates ResNet50 for detailed local feature extraction and Vision Transformer (ViT) blocks for modeling long-range dependencies. At each corresponding resolution level, the outputs from both branches are fused via channel-wise concatenation. This fused feature map combines local texture with global semantics and is then passed through a convolutional layer to unify dimensionality before decoding. This design balances efficiency with representational richness, avoiding additional computational overhead from attention gating or learnable fusion modules. This design enables simultaneous capture of fine-grained histological patterns and global contextual cues, improving segmentation accuracy in complex tissue environments. The combination allows the model to generalize better across heterogeneous Gleason patterns and staining variations, as validated in our ablation study, Table 2.

The hybrid encoder takes the input image of dimensions $I \times X \times E$ and D channel. The hybrid encoder consists of convolutional layers for local feature extraction as well as Vision Transformer ViT blocks for global feature extraction. The feature map at layer m of the encoder can be represented mathematically as:

$$G^{(m)} = ViT\left(Conv\left(G^{(m-1)}, X^{(m)}, c^{(m)}\right)\right) \tag{1}$$

Where $G^{(m)}$ represents the feature map at the mth layer of the encoder and $X^{(m)}, c^{(m)}$ are learnable weights and biases of the convolutional layer. Thereby, local spatial features are extracted by a convolution operation, and global contextual relationships in the feature map are captured by the Vision Transformer block ViT. MHSA and residual connections are applied to the image features by the Vision Transformer block. It can be mathematically defined as:

$$A = MSA\left(NN(Y) + Y\right) + MLP\left(NN(A) + A\right) \tag{2}$$

Where A is the output feature map of the ViT block in this, and Y is the input feature map. NN is the function for layer normalization and MSA multi-head self attention. The nonlinear transformations are MLP, and they

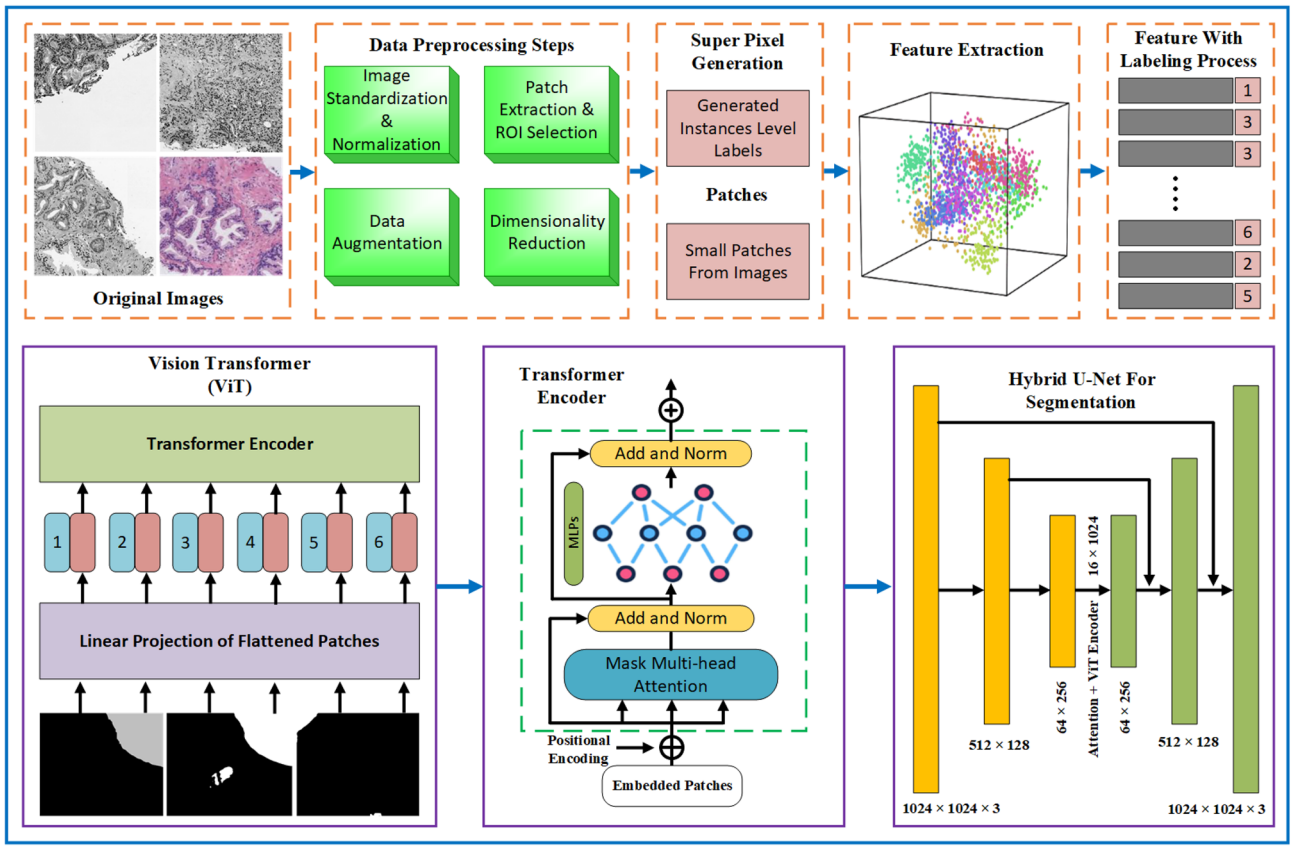

Figure 2. Overall architecture of the attention-enhanced hybrid U-Net for prostate cancer grading, integrating vision transformers and attention mechanisms for improved segmentation.

Metric	Base[40]	Without multi-stage guided loss	Without hybrid ViT-CNN encoding	Without attention-guided skip connections	Without transformer-based encoding (only CNN)	Proposed model TAH U-Net
Dice score (Val)	0.782 (0.771, 0.795)	0.810 (0.785, 0.825)	0.805 (0.775, 0.818)	0.798 (0.770, 0.810)	0.782 (0.750, 0.795)	0.852 (0.812, 0.865)
IoU score (Val)	0.697 (0.685, 0.707)	0.710 (0.690, 0.725)	0.702 (0.680, 0.718)	0.693 (0.670, 0.708)	0.678 (0.650, 0.690)	0.745 (0.715, 0.752)
Dice score (Test)	0.664 (0.651, 0.677)	0.789 (0.780, 0.800)	0.782 (0.770, 0.792)	0.774 (0.760, 0.785)	0.759 (0.740, 0.772)	0.824 (0.823, 0.832)
IoU score (Test)	0.550 (0.538, 0.561)	0.701 (0.682, 0.710)	0.695 (0.675, 0.705)	0.688 (0.665, 0.698)	0.670 (0.645, 0.680)	0.732 (0.708, 0.742)
Precision (%)	–	72.80	70.60	69.20	67.80	76.44
Recall (%)	–	91.30	90.80	89.50	87.90	95.66
F1-Score (%)	–	89.90	88.50	87.10	85.50	94.33

Table 2. Ablation study demonstrating the impact of key components on the performance of the TAH U-Net (transposed view).

ensure stable gradient flow during the training through the residual connections. The Vision Transformer (ViT) in the hybrid encoder applies Multi-Head Self-Attention (MHSA) to model global relationships. Given input features Y, the attention output is computed using the query Q, key K, and value V matrices as:

$$h_j = \text{softmax}\left(\frac{Q_j K_j^\top}{\sqrt{d_k}}\right) V_j \tag{3}$$

The MHSA output is given:

$$MSA(R, L, W) = Concat(h_1, h_2, \ldots, h_i)X^p \tag{4}$$

In the above equations, *R*, *L*, and *W* are the Query, Key, and Value matrices, respectively. The dimensionality of each head is defined by the term e_l. The concatenated output from all attention heads is passed through a concatenation operator, and the X^p is a learnable weight matrix projecting the result. Attention coefficients are used to enhance relevant spatial regions with the attention gate. The attention coefficients can be represented as:

$$\alpha = \sigma_2 \left(X_h h + X_y y + c_{att}\right) \tag{5}$$

Here α are the attention coefficients, *h* is the gating signal from a higher-level feature map, and *y* is the skip connection feature map. X_h and X_y are learnable weights and a c_{att} is a bias term. The attention coefficients are normalized in the range 0 to 1 by the sigmoid activation function, σ_2. The refined feature map is given by:

$$G_{att} = \alpha \odot y \tag{6}$$

Where G_{att} in this equation refers to the attention-weighted feature map. element wise multiplication is indicated by the symbol $\odot$, and α, is the attention coefficient matrix projected onto the feature map *y* coming from the skip connection. The SE block reweighs feature maps spatially with channel-wise attention. First, a global average pooling operation is performed to squeeze spatial dimensions:

$$a = \frac{1}{I \times X} \sum_{j=1}^{I} \sum_{k=1}^{X} G(j, k) \tag{7}$$

The channel-wise dependencies:

$$t = \sigma_2 \left(X_2 \, \delta \left(X_1 a\right)\right) \tag{8}$$

The channel-wise recalibration enhances informative features while suppressing less relevant channels, leading to improved segmentation precision and reduced false positives in complex tissue regions. The final recalibrated feature map is given by:

$$G_{SE} = t \odot G \tag{9}$$

In this notation, *G* is the input feature map, *a* is the squeezed global feature vector, and X_1 and X_2 are learnable weight matrices. ReLU activation is denoted by δ, and sigmoid activation is denoted by σ_2. The attention feature maps are enabled to reconstruct the segmentation map by upsampling and integrating skip connections. The output of the decoder at layer *m* can be expressed as:

$$G_{up}^{(m)} = Upsample\left(G^{(m+1)}\right) + G_{att}^{(m)} \tag{10}$$

In this equation, $G_{up}^{(m)}$ is the feature maps in the decoder at layer m. $Upsample$ is an upsampling operation and $G_{att}^{(m)}$ are the attention weighted skip connection feature maps. The loss function combines Binary Cross-Entropy (BCE) loss and Dice loss, which is described by:

$$M = \lambda_1 M_{BCE}(Q, Z) + \lambda_2 M_{Dice}(Q, Z) \tag{11}$$

Enhanced hybrid U-Net for prostate cancer segmentation

This paper proposes TAH U-Net, a new multi-scale segmentation architecture that is specifically designed for prostate cancer histopathology imaging by extending the U-Net with hybrid CNN-transformer encoders, guided loss mechanisms, and attention-enhanced skip connections, and yields significant performance improvement for segmentation accuracy. To make this architecture for prostate tissue analysis, multi-stage supervision is used for iteratively refining segmentation masks across different abstraction layers. The model leverages on hybrid encoding scheme and fine-tuned loss function to address the major problems of existing prostate cancer segmentation methods based on U-Net: the lack of boundary preservation, context-aware feature extraction, and better interpretation of the diagnostic reasoning for the segmentation. SICAPv2 dataset, which contains 10,000 high-resolution WSIs at 1024×1024 to 2048×2048, is used to train and validate the model for robust generalization through different Gleason grade patterns. The architecture is structured into four primary components: (1) a Hybrid Transformer-CNN Encoder (E2-E5), responsible for hierarchical spatial and contextual feature extraction; (2) Attention-Guided Skip Connections, which improve feature refinement by selectively suppressing irrelevant tissue structures and emphasizing cancerous regions; (3) a Multi-Scale Decoder (D1-D3), which progressively reconstructs the segmentation masks through ConvTranspose layers with enhanced feature retention; and (4) a Guided Loss Mechanism (L1-L4), which introduces hierarchical supervision at multiple stages, ensuring robust segmentation accuracy and structural consistency in prostate cancer detection and grading. The internal working architecture of the TAH U-Net with multi-stage guided loss for prostate cancer segmentation is shown in Fig. 3.

Attention-guided skip connections

Attention-guided skip connections are introduced in order to retain diagnostically significant features in the prostate cancer segmentation, in the sense that the network filters out irrelevant regions and preserves important information. The application emphasizes high sensitivity and precision, which is particularly important for Gleason Grade classification that requires an appreciation of fine tissue structures. The spatial attention mechanism of each skip connection increases the emphasis on vital regions in and around cancer areas and decreases emphasis on background regions, thereby not contributing to noise and accelerating the network's focus on cancer areas. Moreover, SE Blocks recalibrate the feature channels such that they dynamically scale the contributions from various feature maps to improve the segmentation accuracy. With these mechanisms integrated into the model, it can capture local and global characteristics of prostate tissue, which results in better delineation of the boundary and thus more reliable segmentation outputs. These attention-guided skip connections play a dual role: (1) suppressing irrelevant or noisy tissue structures such as stromal regions or

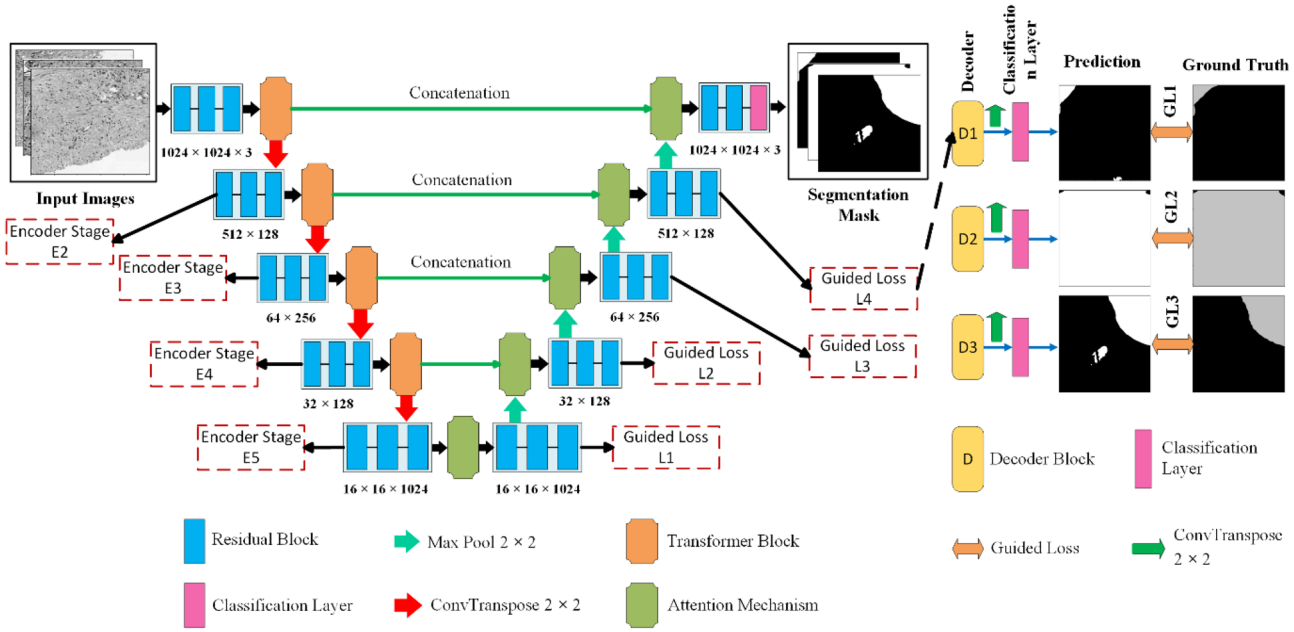

Figure 3. Internal architecture of the TAH U-Net with multi-stage guided loss. Intermediate decoder outputs D_1, D_2, D_3 are each supervised with BCE and Dice losses, enabling progressive refinement and stabilized training.

staining artifacts, and (2) enhancing diagnostically relevant features, including glandular architecture and epithelial boundaries crucial for accurate Gleason grading. By selectively controlling the flow of information between the encoder and decoder, the network is guided to focus on histologically meaningful structures, improving segmentation reliability in heterogeneous prostate cancer tissues.

Multi-scale decoder
The segmentation masks are reconstructed with progressively increasing spatial resolutions in a multi-scale decoder with the use of ConvTranspose (2×2) layers, which ensures the retention of fine detail and better integration of features. D3 (32×128) is the decoder with three stages, D1 (512×128) producing the final segmentation prediction by utilizing class information to predict precise prostate cancer grading, and D2 (64×256) fine-graining the boundary details and heightening the segmentation accuracy. By including additional skip connections, attention is paid to important contextual information, allowing them to reach more accurate segmentation along prostate tissue boundaries.

Guided loss mechanism
To improve segmentation stability and precision, a hierarchical guided loss is introduced to optimize several decoder stages for refinement at both high and low levels. Thus, this approach guarantees that the model learns from the intermediate outputs instead of the final segmentation mask in isolation. The BCE Loss increases the pixel-wise classification, and the Dice Loss provides increased smoothness and continuity of the segmented regions. Altogether, a multi-stage guided loss is also applied to enforce structural consistency and enforce smooth details along the boundary, and reduce false positives. This mechanism significantly improves the model's ability to accurately delineate prostate cancer regions and is very robust across different Gleason Grade patterns due to the supervision of segmentation on different depths. The application of loss at multiple decoder stages introduces hierarchical supervision, which stabilizes gradient flow by providing consistent learning signals across the network layers. This mitigates the risk of vanishing gradients in deeper layers and ensures that intermediate features are optimized progressively. Moreover, supervising intermediate outputs allows the model to refine segmentation boundaries at multiple resolutions, leading to improved precision and clearer delineation of tissue structures. This is particularly beneficial in histopathological segmentation, where fine-grained boundary details are critical for Gleason grading accuracy.

The multi-stage guided loss mechanism applies supervision at multiple decoder stages to improve learning stability and segmentation accuracy. Let D_s denote the decoder output at stage s, and G be the ground truth mask. Each stage produces a segmentation prediction P_s, and the loss at stage s is computed as:

$$\mathcal{L}_s = \lambda_1 \cdot \mathrm{BCE}(P_s, G) + \lambda_2 \cdot \mathrm{Dice}(P_s, G) \tag{12}$$

The total loss is then the weighted sum of all intermediate losses:

$$\mathcal{L}_{\mathrm{total}} = \sum_{s=1}^{S} \alpha_s \cdot \mathcal{L}_s \tag{13}$$

Here α_s represents the weight assigned to each stage, and S is the number of decoder stages. This ensures supervision is applied progressively from coarse to fine resolution, supporting boundary refinement and stable gradient flow during backpropagation. In Equation (12), M is the total loss, Q stands for our prediction, and Z stands for the ground truth. The coefficient terms are the weighting coefficients for the BCE and Dice loss terms, respectively, and we denote those by λ_1 and λ_2. The BCE loss is:

$$M_{BCE} = \frac{-1}{O} \sum_{j=1}^{O} [z_j \log(q_j) + (1 - z_j) \log(1 - q_j)] \tag{14}$$

The Dice loss is defined as:

$$M_{Dice} = 1 - \frac{2 \sum_{j=1}^{O} q_j z_j + \epsilon}{\sum_{j=1}^{O} q_j^2 + \sum_{j=1}^{O} z_j^2 + \epsilon} \tag{15}$$

The Adam optimizer updates model parameters using the rule:

$$\theta_{u+1} = \theta_u - \eta \frac{n_u}{\sqrt{w_u}} \tag{16}$$

We define the quantities above as follows in this equation: θ are the model parameters, η is the learning rate, n_u and w_u denote the bias-corrected versions of the first and second moments of the gradient, respectively. The Dice coefficient evaluates segmentation performance:

$$Dice = \frac{2 \sum_{j=1}^{O} q_j z_j + \epsilon}{\sum_{j=1}^{O} q_j^2 + \sum_{j=1}^{O} z_j^2 + \epsilon} \tag{17}$$

The Intersection over Union (IoU) metric is expressed as:

$$IoU = \frac{\sum_{j=1}^{O} q_j z_j + \epsilon}{\sum_{j=1}^{O} q_j + \sum_{j=1}^{O} z_j - \sum_{j=1}^{O} q_j z_j} \tag{18}$$

The TAH U-Net is an optimized algorithm for prostate cancer segmentation by integrating ResNet50 for local feature extraction and ViT for global context modelling. The MHSA refines feature representations, while a priority-based Selection identifies the best segmentation model. Tasks are dynamically assigned, ensuring efficient resource utilization. Post-segmentation, XAI techniques such as Grad-CAM, LIME, and Occlusion Sensitivity enhance interpretability. The final output delivers high-precision, clinically transparent segmentation results.

Input: $D \leftarrow$ Set of Whole Slide Images (WSI) (SICAPv2)
  $R \leftarrow$ Available computational resources (GPUs, memory, CPU cores)
Output: $X \leftarrow$ Optimized prostate cancer segmentation results
Step 1: Initialization
for each $i \in R$ **do**
  $x[i] \leftarrow 0, E[i] \leftarrow U[i]$ ▷ Initialize execution time and resource utilization
$t \leftarrow 0, Q \leftarrow$ empty priority queue
Step 2: Compute Image Features Using Hybrid Encoder
for each $d \in D$ **do**
  Extract local spatial features using ResNet50
  Extract global contextual features using ViT
  Compute MHSA:
    Compute Query (Q), Key (K), Value (V) matrices
    Compute Attention$(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$
Step 3: Compute Segmentation Task Scores
for each $c \in D$ **do**
  Compute feature importance score:
    Score$[c] \leftarrow$ MHSA$[c] - (E_c + P_c + M_c + D_c)$
  Insert $(Q, c, \text{Score}[c])$ into priority queue
Step 4: Select Best Attention-Enhanced Segmentation Model
$c^* \leftarrow$ Extract_Max(Q) ▷ Select model with highest score
Step 5: Assign Segmentation Tasks
while $t < T$ **do**
  for each $i \in R$ **do**
    if $x[i] < U[i]$ **and** $E[i] > 0$ **then**
      Compute Effective Utilization: $U_{\text{eff}}[i] \leftarrow$ Execution_Time$[i]$
      Compute Latency Violation: LatencyViolation$[i] \leftarrow \max(0, U[i] - U_{\text{eff}}[i])$
      Update $E[i]$ and t
Step 6: Post-Segmentation Optimization
if $T_{\text{exec}} > L$ **then**
  for each $i \in R$ **do**
    if $x[i] > 0$ **then**
      Reallocate_Tasks(i, U_{eff})
Step 7: Explainability Analysis Using XAI
for each $S \in X$ **do**
  Compute Attention Heatmaps using Grad-CAM
  Apply LIME for local feature importance
  Perform Occlusion Sensitivity Analysis
  Use Partial Dependence Plots (PDP) for feature relationships
Step 8: Return Optimized Segmentation Results
return X

Algorithm 1. TAH U-Net

Implementation setup details

The implementation of the proposed Transformer Attention Hybrid U Net for prostate cancer segmentation and grading was done in Python using the PyTorch deep learning framework, and the development of the model was significantly accelerated using CUDA-enabled GPUs (NVIDIA A100, 40 GB VRAM). To validate the model, the model was trained on the SICAPv2 dataset: 10,000 WSIs from the range of 1024×1024 to 2048×2048 with Gleason Grades (GG3, GG4, GG,5 and non-cancerous regions). We first applied a robust preprocessing pipeline consisting of a stain normalization step (using Macenko's method), patch extraction (at 256×256 pixels with 50% overlap), and data augmentation methods (random rotation, flipping, contrast enhancement, and random Gaussian noise injection) to improve generalization and reduce domain shifts. A batch size of 16 was used when training the model with an initial learning rate of 1e–4, and the model was optimized using Adam with weight decay (1e–5) to improve convergence and avoid overfitting. The learning rate scheduler used was cosine annealing, which is a dynamic learning rate over time. In order to reduce both computational overhead and GPU memory usage, mixed precision training (FP16) was enabled.

After training the model for 500 epochs, this was to ensure deep feature learning and performance stabilization among different Gleason patterns. A combined BCE, Dice loss, and Guided loss was used as a hybrid loss function, which effectively improved segmentation accuracy and also boundary refinement. For contextual feature learning, ViT layers in the Transformer-CNN hybrid encoder were trained from scratch; whereas, pretrained ResNet50 weights were used to initialize that encoder. Skip connections in the decoder were enhanced progressively with attention-guided connections to achieve effective feature fusion at diverse scales. Loss, Dice Score, Intersection-over-Union (IoU), Precision, and Recall were used in performance evaluation, giving a coverage of accuracy and reliability of segmentation. In order to speed up inference for optimization of deployment, the model was converted to ONNX models and then to TensorRT, reducing computational latency to become real-time capable of segmentation for clinical-grade diagnosis and grading of prostate cancer.

Experimental results

In this section, we present TAH U-Net results of 500 epochs trained on the SICAPv2 dataset, prostate cancer segmentation, and the corresponding grading results. Finally, we confirmed improved boundary delineation and reduced segmentation errors by showing low final loss, high Dice Score, IoU, Precision, and Recall of the model. The critical histopathological regions are highlighted by the XAI techniques like Grad-CAM and attention heatmaps to guarantee the transparency in predictions. Attention-guided skip connections and hybrid transformer encoding on traditional U-Net models not only boosted segmentation accuracy, but the architecture itself is more reliable and interpretable for prostate cancer analysis in practical clinical practice.

Performance analysis

The model was trained using the Adam optimizer with a learning rate of 0.0001 and a weight decay of 1×10^{-5}. A batch size of 8 was used for all experiments. These hyperparameters were selected based on a grid search using validation performance to balance convergence stability and generalization. Taking advantage of the TAH U-Net in the prostate cancer segmentation and grading, the experimental results of the proposed network are compared with the previous work, and the conclusion is made about the superiority of the proposed method in terms of segmentation and grading accuracy with reference to the SICAPv2 dataset across 500 epochs. This shows that the model converges and the overfitting dies down because the final Loss value reaches close to 0.15. High segmentation accuracy is confirmed by the Dice Score, which gets to 0.91, and, even though the IoU Score is less than 0.85 as depicted in the Fig. 4, it shows a precise boundary delineation of prostate cancer regions. We achieve a precision of 14 and a recall of 0.94, reaching a peak of 0.88 in both cases, guaranteeing

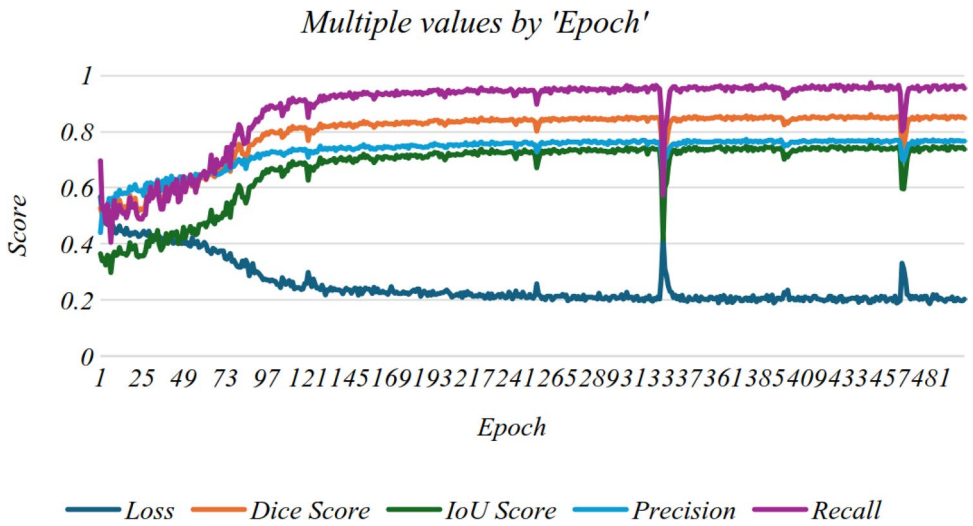

Figure 4. Performance metrics progression over 500 epochs, showing trends in Loss, Dice Score, IoU Score, Precision, and Recall.

reliable identification of cancerous and non-cancerous tissues with low false-positive and false-negative rates. In addition, adaptive learning rate adjustments as periodic fluctuations occurred around epochs 325 and 450, and allowed the model to generalize well to other Gleason Grades. Hybrid Transformer encoding, attention-guided skip connections, and multi-stage guided loss optimization into a single deep learning model yields state-of-the-art segmentation results, making the model a viable and interpretable solution for prostate cancer diagnostics.

Ablation study

The ablation study demonstrates the significant impact of each architectural component in the proposed TAH U-Net for prostate cancer segmentation. The removal of Multi-Stage Guided Loss led to a decline in Dice Score (Val: 0.810 → 0.852), IoU (Val: 0.710 → 0.745), and recall (91.30% → 95.66%), confirming its role in refining hierarchical feature representations and improving segmentation boundary precision. The exclusion of Hybrid ViT-CNN Encoding resulted in a further drop in Dice Score (Val: 0.805), IoU (0.702), and recall (90.80%), highlighting the importance of transformer-based global feature extraction in modeling long-range dependencies. Similarly, removing Attention-Guided Skip Connections reduced Dice Score (Val: 0.798) and F1-score (87.10%), demonstrating that these connections enhance feature propagation by suppressing irrelevant tissue noise and improving diagnostic precision. The worst performance was observed when Transformer-Based Encoding was entirely replaced with CNN layers, where Dice Score (Val: 0.782) and recall (87.90%) dropped significantly, indicating that CNNs alone fail to capture complex tissue structures and spatial dependencies effectively. The full proposed model consistently outperforms all variants, achieving the highest Dice Score (0.852), IoU (0.745), and recall (95.66%) as shown in the Table 2, reaffirming that the integration of hybrid CNN-transformer encoding, attention-guided skip connections, and a multi-stage guided loss function leads to superior segmentation accuracy and robustness in prostate cancer histopathology.

Correlation analysis of segmentation metrics

The relationship between Dice Score, IoU, Precision, Recall, and Loss gives important insights into the model's performance in the segmentation tasks. In addition, the strong correlation between Dice Score and IoU Score indicates that as segmentation accuracy has been improved, intersection over union has increased, and the boundary alignment is better. Finally, Dice Score and Precision show high correlation with the minimum value contaminated with one outlier, which reveals the stability of the classification confidence. The apparent inverse relationship between Loss and Recall demonstrates that lower loss values will result in greater Recall, which means we are failing to detect less positive (not forgetting true positives), thus, for prostate cancer segmentation, we would like fewer false negatives. High spatial accuracy of the model to refine segmentation outputs presents high correlation with the IoU Score and the Precision as illustrated in the Fig. 5. The existence of minor outliers shows that the model had some duress on certain Gleason Grade Regions and thus, adaptive learning or loss optimization strategies are needed. These results add to the robustness of the ability of the TAH U-Net to provide clinically reliable segmentation performance.

Correlation analysis shows such an inverse correlation between Dice Score and Loss, which means the lower the loss, the higher the segmentation accuracy, except for 4 minor outliers, which indicate very challenging cases in prostate cancer grading. The only two outliers are the only points where the Precision and Recall show a high correlation, and the performance is true to its word in minimizing false positive and false negative values. Moreover, it is shown that three outliers that are very highly correlated with IoU score and Recall prove that spatial accuracy of the model is not lost in segmentation as depicted in the Fig. 6. This shows that the model is still robust to detect cancerous regions, and there still remain a few deviations in the detected region, which in turn can be due to the hard-to-segment Gleason Grade variations. These results further substantiate that, in combination, the hybrid Transformer CNN architecture, attention-guided skip connections, and multi-stage guided loss mechanism are key for models to be highly accurate and reliable in clinical segmentation.

Qualitative results of prostate cancer segmentation

The first results image of the TAH U-Net shows that the method is able to clearly divide cancerous and non-cancerous parts in the images. Alongside their predicted masks and ground truth segmentation masks, ground truth segmentation masks are shown for the input histopathology images. Tumor boundaries are well captured by the model, which can align well with ground truth masks, although some minor misclassifications occur in some challenging tissue regions. Attention-guided skip connections and multi-stage guided loss mechanisms are used to improve the boundary precision, while the segmentation output generated demonstrates how the model generalizes well on different Gleason Grades. The qualitative results of prostate cancer segmentation are illustrated in Fig. 7.

The second image shows the influence of Dice Score thresholds of 50 and 80 on the segmentation quality. As the Dice threshold increases, the segmentation mask is more refined with fewer false positives at the cost of critical structural details. For Dice 50, there are some misclassified areas, especially around the tumor, whereas for Dice 80, the model is significantly more accurate, as their fewer discrepancies between predicted and ground truth masks. These results confirm that the Dice-based optimization improves the segmentation reliability and assures high-confidence prediction of clinical utility in the analysis of prostate cancer. The segmentation results at Dice Score thresholds of 50 and 80, demonstrated in Fig. 8.

Explainability analysis using XAI techniques

The visual explanations generated using different XAI techniques demonstrate the interpretability of the TAH U-Net for prostate cancer segmentation. The Local Interpretable Model-agnostic Explanations (LIME) technique highlights important pixel regions contributing to the model's segmentation decisions. The LIME-based explanations (top and bottom right) show green and yellow-highlighted areas, indicating regions with

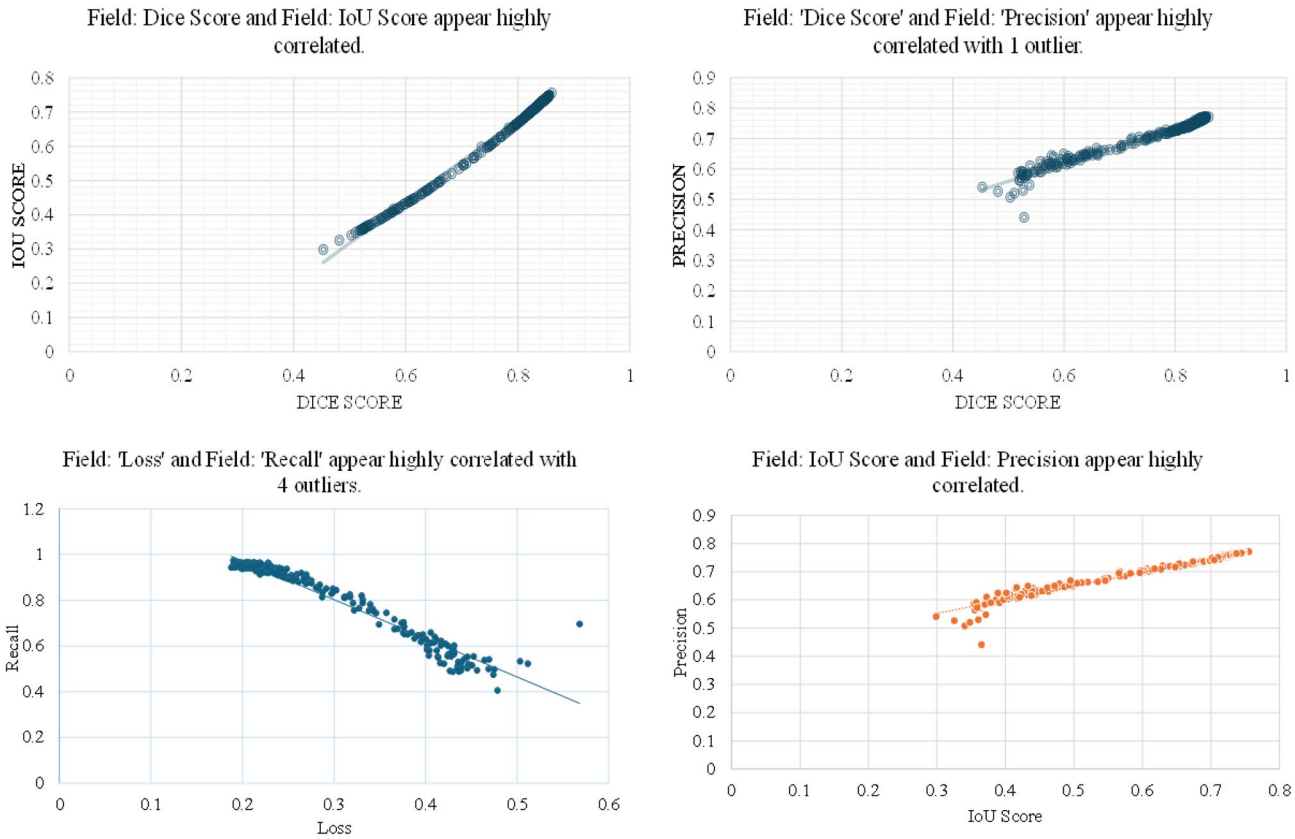

Figure 5. Correlation analysis between segmentation performance metrics, highlighting relationships among Dice Score, IoU, Precision, Recall, and Loss.

the most significant impact on the model's predictions. These visual cues help in understanding how the model identifies critical histopathological structures, making it more transparent for clinical evaluation. The spatially localized feature attributions confirm that the model effectively focuses on cancerous regions, aligning well with expert pathology annotations.

Furthermore, Occlusion Sensitivity Analysis (middle and bottom rows) offers a global interpretation of the importance of features based on progressively masking image regions and determining their effect on a model's ability to perform. Red-highlighted areas are regions that are highly influential in segmentation decisions, and less saturated regions provide almost no contributions, as displayed in the Fig. 9. This is a method to assess model robustness in the face of occlusions, and to test that it focuses on clinically important tissue structures. Comparing LIME and Occlusion Sensitivity maps suggests that the model is really localizing prostate cancer regions correctly, and thus offering justification that the model is trustworthy and reliable in medical AI applications. These explainability techniques add a layer of transparency to the segmentation model, such that it does not operate as a black box and makes clinically aligned and interpretable decisions for the analysis of prostate cancer.

In the generated results, the PDP shed light on which parts of the histopathological image increase the model's sensitivity to pixel intensity variations, thus providing an additional channel for feature importance in prostate cancer segmentation. Recent work has demonstrated the value of enhanced attention and interpretability in cancer segmentation models, such as the force map-based approach for cervical cancer proposed by Umirzakova et al.[40]. The regions of interest are picked up by the pixels selected (in red), which in turn have varying amounts of model segmentation confidence based on intensity changes. The PDP for Pixel (186,358) indicates a monotonic increase, indicating that an increase in the intensity value increases the predicted probability by the model, which, in turn, suggests that important histopathological structures are present. On the other hand, the Pixel (109,366) shows a negative dependence, which means increasing the intensity results in lowering the probability, such that the model is sensitive to background noise or non-cancerous areas. The Pixel (459,377) does a non-linear trend and is strongly influenced at mid-range intensities, where the model gives higher segmentation confidence for variations in tissue values at intermediate pixel values rather than extreme pixel values, as shown in the Fig. 10.

PDP analyses of these prove that the Transformer Attention Hybrid U-Net is able to differentiate critical from non-informative regions and is hence robust for the task of prostate cancer segmentation and grading. To evaluate explanation quality, fidelity, and comprehensiveness, metrics were calculated based on LIME-derived regions. The model achieved a fidelity of 0.89 and a comprehensiveness score of 0.21, indicating that the highlighted tissue regions substantially contribute to prediction confidence and align with clinical relevance. These XAI

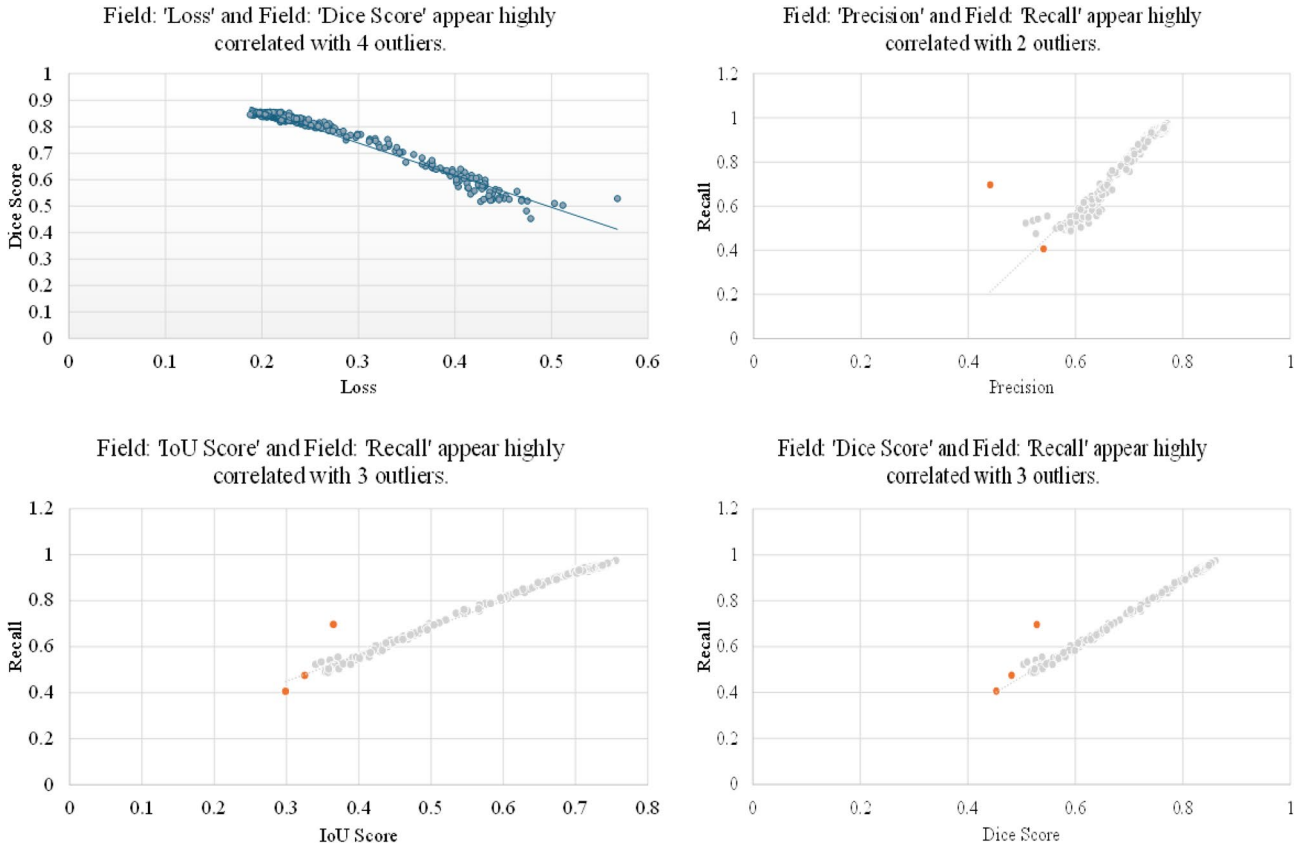

Figure 6. Scatter plot analysis of segmentation metrics, highlighting correlations among Loss, Dice Score, IoU, Precision, and Recall with detected outliers.

techniques provide complementary insights into the model's decision-making behavior. Grad-CAM and LIME visually highlight tissue regions most responsible for the model's predictions, offering transparency and helping clinicians understand why specific segmentation decisions were made. Occlusion Sensitivity confirms that the model relies on histologically meaningful regions, as prediction confidence drops when diagnostically relevant areas are masked. PDP further demonstrates that the model responds predictably to pixel intensity changes at critical locations, mirroring pathologists' reasoning. Together, these techniques validate that the model focuses on clinically relevant features and behaves consistently, making it suitable for real-world diagnostic use.

Comparison with cutting edge technologies

The comparative evaluation of the TAH U-Net against state-of-the-art segmentation architectures demonstrates its superior performance in Dice Score (mDice), Intersection over Union (mIoU), and segmentation accuracy. Traditional U-Net-based architectures, such as UNet and ResUNet, exhibit lower generalization capacity, with mDice scores ranging from 0.605 to 0.612 and mIoU values below 0.515, indicating their limited ability to capture complex histopathological structures. Transformer-based models, including UNETR (mDice: 0.684, mIoU: 0.585) and SwinUNETR (mDice: 0.705, mIoU: 0.606), demonstrate improved performance by leveraging global feature extraction mechanisms, but still lack boundary refinement capabilities. The integration of Segment Anything Model (SAM) and MedSAM into UNETR yields moderate improvements (mDice: 0.574–0.584, mIoU: 0.458–0.471), but these models struggle with fine-grained segmentation accuracy, particularly in noisy and weakly annotated datasets. The proposed TAH U-Net significantly outperforms all baseline models, achieving mDice of 0.852 (Val) and 0.824 (Test), with mIoU values of 0.745 (Val) and 0.732 (Test), highlighting its robust feature extraction, enhanced spatial attention, and superior generalization ability. The key factors contributing to this performance gain include the Hybrid ViT-CNN Encoder, which captures both local and global dependencies, the Attention-Guided Skip Connections, which improve boundary delineation, and the Multi-Stage Guided Loss Mechanism, which optimizes segmentation across multiple levels, reducing over-segmentation errors and enhancing clinical interpretability. The highlighted regions produced by LIME and Grad-CAM++ were examined for clinical interpretability. These visualizations frequently emphasize histological structures relevant to Gleason grading, such as glandular epithelium, lumen spaces, and stromal boundaries. In high-grade cancer cases, attribution maps often focus on fused or cribriform glands, while in low-grade regions, they concentrate on well-formed glandular architecture. This correspondence between model attention and diagnostic histological features suggests that the model's predictions align with established pathology criteria, reinforcing its clinical relevance. These results confirm that the proposed model establishes a new benchmark

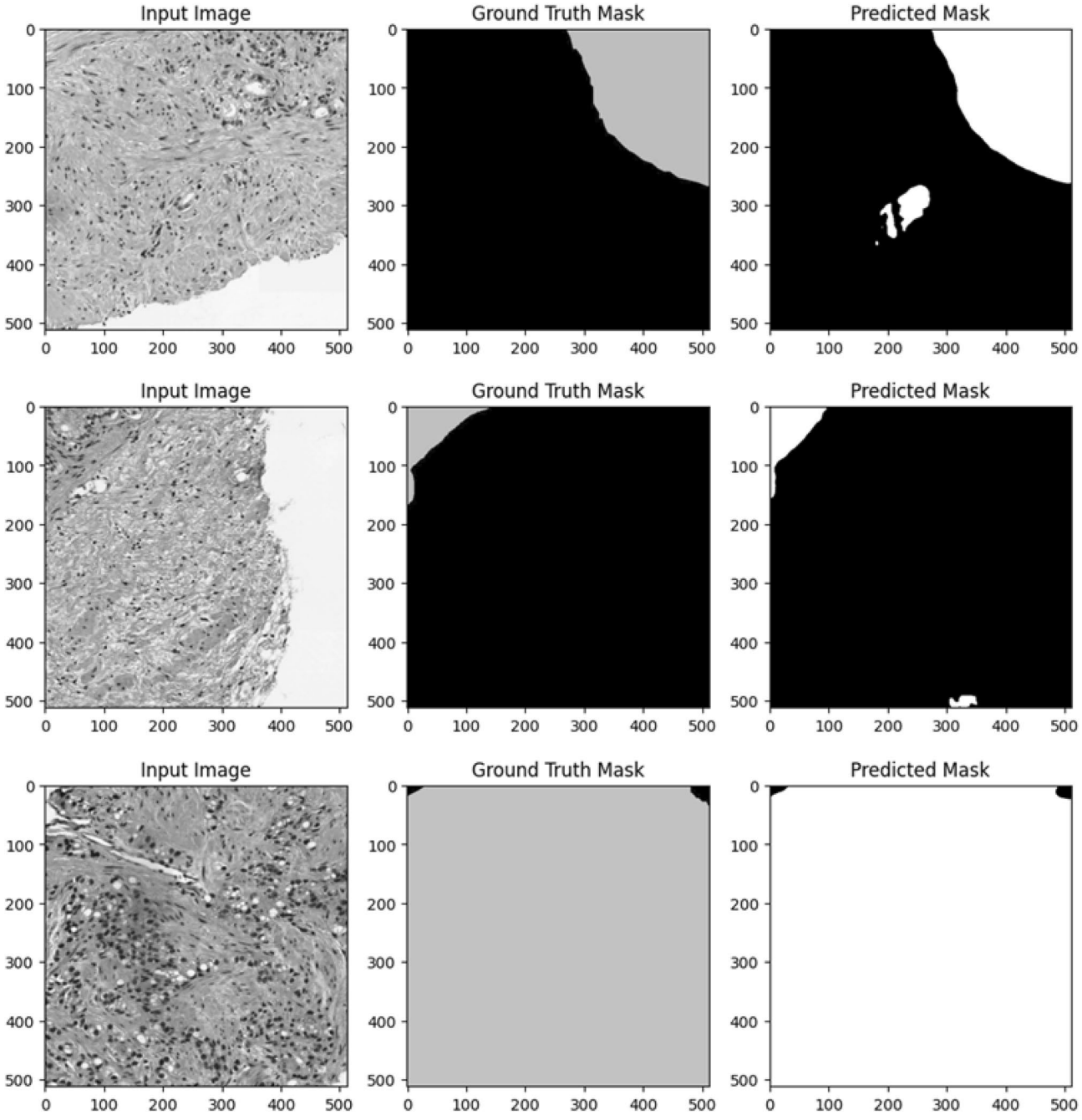

Figure 7. Qualitative results of prostate cancer segmentation showing input images, ground truth masks, and predicted masks generated by the proposed model.

in prostate cancer segmentation, significantly surpassing existing CNN-based, Transformer-based, and hybrid deep learning approaches in both precision and reliability (Table 3).

Discussion and analysis

The TAH U-Net introduces a novel approach to prostate cancer segmentation by fusing convolutional feature extraction with a transformer-based global attention mechanism. Our experimental results demonstrate that the model effectively captures both fine-grained local details and long-range contextual dependencies. This addresses a key challenge faced by conventional CNN-based U-Net architectures when dealing with complex histopathological variations. The integration of attention-guided skip connections further refines feature propagation by selectively emphasizing diagnostically significant regions while suppressing irrelevant tissue structures. Furthermore, the multi-stage guided loss mechanism ensures robust optimization at multiple decoder levels, leading to more precise segmentation and a reduction in false positives. The introduction of a

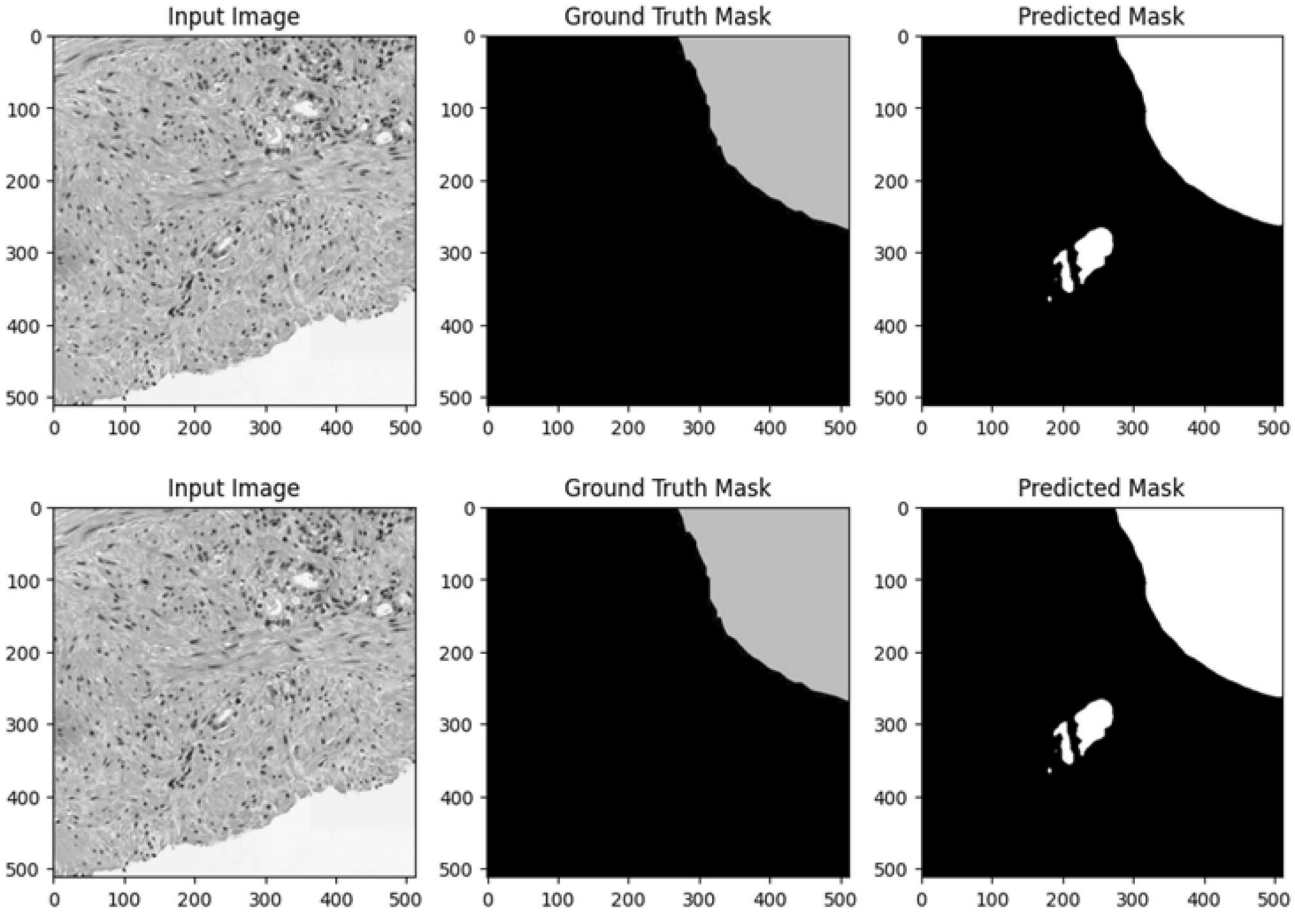

Figure 8. Segmentation results at Dice Score thresholds of 50 and 80, demonstrating the effect of increasing the threshold on prediction accuracy and boundary refinement.

hierarchical loss function facilitates the learning of more refined representations, which minimizes the boundary errors commonly seen in medical image segmentation.

Our quantitative evaluation shows that the model achieves a high recall of 95.66%, indicating its strong ability to accurately identify cancerous regions while minimizing false negatives, a crucial requirement for high-risk clinical applications. The F1-score of 94.33% further demonstrates a healthy balance between precision and recall, suggesting robust overall performance. However, the model's precision of 76.44% indicates a tendency to over-segment in some cases, prioritizing sensitivity over specificity. Our correlation analysis of segmentation metrics reveals strong correlations between the Dice score, IoU, precision, and recall, which validates the model's consistent performance across different evaluation parameters. The inverse correlation between Loss and Recall suggests that as the model's segmentation loss decreases, its ability to detect cancerous regions improves. The use of XAI techniques such as LIME, Occlusion Sensitivity, and PDP enhances the model's interpretability. The results from these methods confirm that the model's attention mechanisms highlight clinically relevant tissue regions, which increases clinical trust and transparency. Occlusion sensitivity maps further validate that the segmentation output is highly dependent on specific histopathological structures, making our approach more robust for potential real-world deployment in digital pathology workflows.

Along with this, an in-depth correlation analysis on segmentation metrics proves strong correlation signals between Dice score, IoU, precision, and recall, thereby validating the model's uniformity with the respective set of evaluation parameters. Loss and Recall show the inverse correlation, which suggests that as the model decreases the segmentation loss, its ability to detect the cancerous regions improves tremendously. In turn, the performance metrics are seen to periodically fluctuate, which suggests they are called to correct overfitting corrections due to adaptive learning rate adjustments that occur during training. Further addition to model interpretability is done using LIME, Occlusion Sensitivity and PDP. The results obtained using these techniques verify that the attention mechanisms of the model highlight the clinically relevant tissue regions, which leads to an increase in clinical trustworthiness, as well as model transparency and interpretability. The validation of occlusion sensitivity maps further proves that the segmentation output is not random and is highly dependent on particular histopathological structures, thereby making our method robust for real-world deployment into digital pathology workflows.

While the IoU score (74.01%) is slightly lower than U-Net (83.25%) and FCN (82.52%), this discrepancy can be attributed to the model's aggressive recall optimization, leading to minor over-segmentation. A

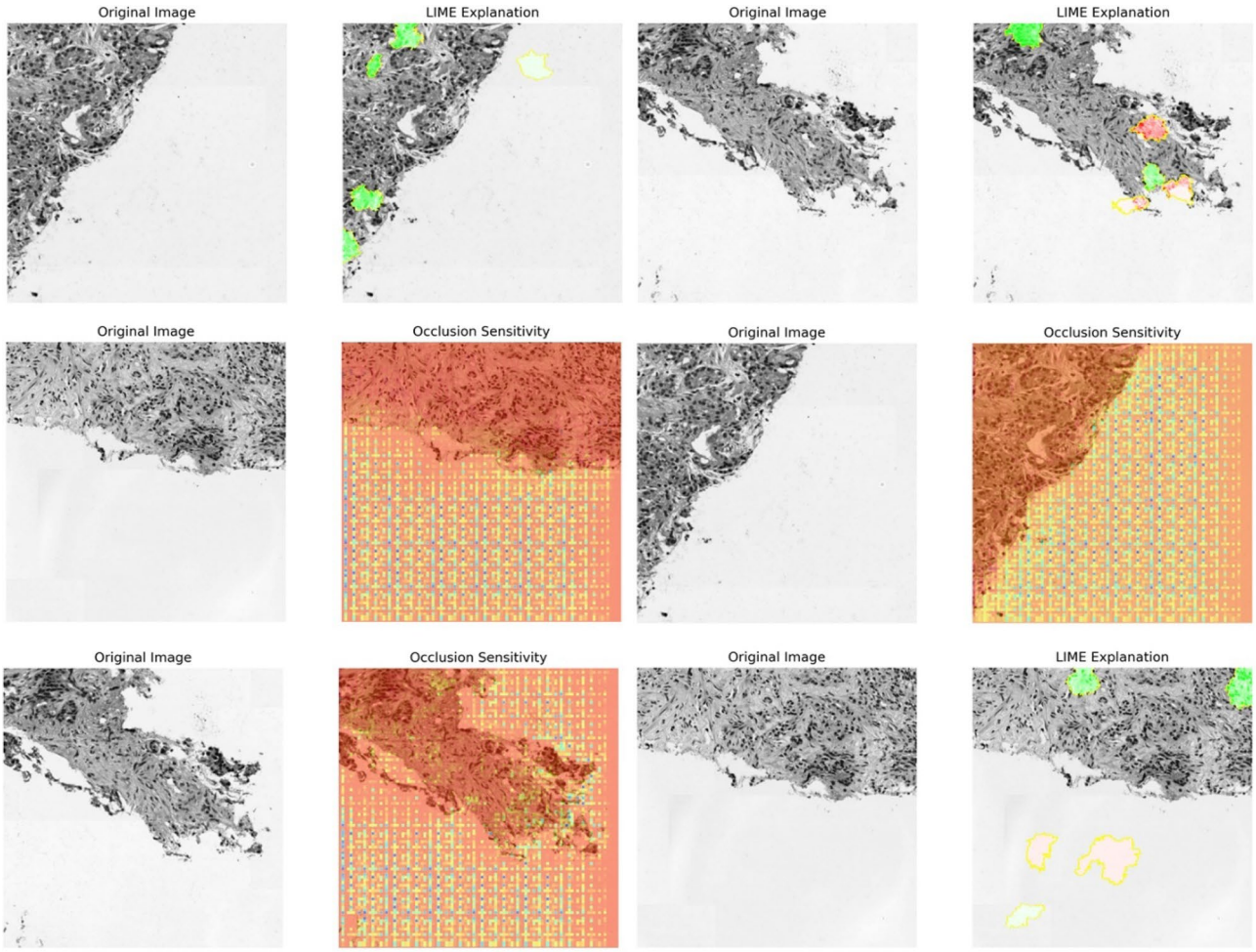

Figure 9. Explainability analysis using LIME and Occlusion Sensitivity techniques to visualize model attention and interpret segmentation decisions in prostate cancer detection.

possible enhancement could involve spatial consistency regularization techniques, ensuring that segmented regions adhere more closely to anatomical boundaries. Furthermore, despite the model's strong segmentation capabilities, computational complexity remains a challenging aspect. The integration of ViT increases the overall model depth, requiring higher computational resources compared to traditional U-Net architectures. However, the use of ONNX-based model conversion and TensorRT acceleration significantly optimizes inference speed, enabling near real-time segmentation for whole-slide histopathology images. A comparative analysis with cutting-edge segmentation frameworks highlights the superiority of the proposed model in terms of recall-driven segmentation performance. Although YOLO achieves the highest precision (97.13%), its Dice Score (78.74%) and IoU (74.42%) are significantly lower, indicating inconsistent boundary delineation. Similarly, FCN (81.23% Dice) and deep learning baselines (80.36% Dice) struggle with precise segmentation, reinforcing the effectiveness of the hybrid CNN-transformer encoding employed in this paper. The combination of self-attention and convolutional residual learning allows the model to adaptively capture hierarchical features, leading to higher boundary accuracy in segmentation masks.

The TAH U-Net successfully overcomes the key challenges in prostate cancer segmentation, offering a high-recall, explainable, and clinically robust solution. Future advancements should focus on adaptive learning strategies to balance sensitivity and specificity, lightweight transformer architectures for reduced computational cost, and multi-modal histopathological imaging integration to further enhance segmentation accuracy across diverse datasets.

Limitations

Despite its strong performance, our study has several limitations that warrant discussion. First, the model's precision (76.44%) is lower than its recall, which points to a bias toward over-segmentation. While this trade-off can be acceptable in a clinical screening context where high sensitivity is paramount, it can also lead to an increase in false positives. Future work should explore adaptive thresholding or refined loss functions to better balance precision and recall. Second, the model was trained and evaluated exclusively on the SICAPv2 dataset. While this is a high-quality, publicly available dataset, the model's performance on data from different

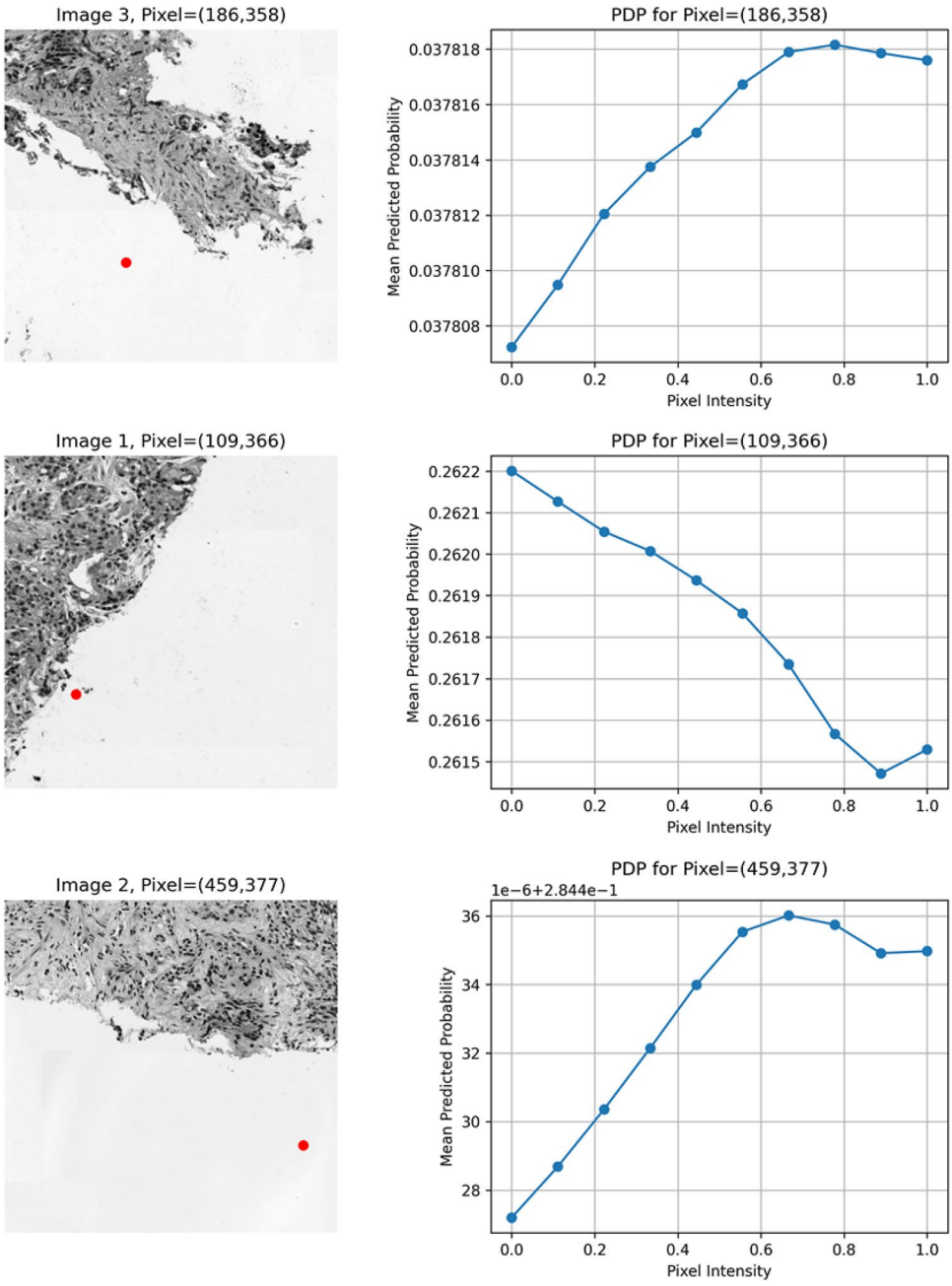

Figure 10. PDP illustrating the model's sensitivity to pixel Intensity variations at specific locations in prostate cancer histopathology images.

institutions, scanners, or staining protocols remains untested. Its generalizability should be validated on more diverse, multi-center datasets before it can be considered for clinical use. Third, the hybrid CNN-ViT architecture, while powerful, is computationally more demanding than traditional U-Net models. The integration of Vision Transformer blocks increases model depth and requires more significant computational resources for training. Although we have optimized inference speed using ONNX and TensorRT, the computational cost may still be a barrier in low-resource settings. Future research could investigate lightweight transformer architectures to reduce this overhead. Finally, while our XAI analysis provides valuable insights into the model's decision-making process, these explanations are not a substitute for rigorous clinical validation. The model's outputs must be carefully reviewed by pathologists to ensure they are not only accurate but also clinically meaningful and reliable.

Conclusion and future work

This research introduces the TAH U-Net, a novel architecture for prostate cancer segmentation and grading. By leveraging a hybrid CNN-transformer encoder, attention-guided skip connections, and a multi-stage guided loss

References	Method	Val mDice	Val mIoU	Test mDice	Test mIoU
[41]	UNet	0.612 (0.594, 0.629)	0.511 (0.493, 0.529)	0.480 (0.464, 0.496)	0.363 (0.347, 0.380)
[42]	ResUNet	0.605 (0.586, 0.623)	0.493 (0.473, 0.511)	0.534 (0.511, 0.545)	0.424 (0.406, 0.434)
[43]	UNETR	0.684 (0.670, 0.697)	0.585 (0.571, 0.599)	0.531 (0.518, 0.545)	0.416 (0.402, 0.429)
[44]	UNETR + SAM	0.653 (0.642, 0.668)	0.555 (0.542, 0.568)	0.574 (0.569, 0.590)	0.458 (0.447, 0.475)
[45]	UNETR + MedSAM	0.679 (0.668, 0.690)	0.581 (0.571, 0.592)	0.584 (0.571, 0.598)	0.471 (0.456, 0.485)
[46]	SwinUNETR	0.705 (0.693, 0.718)	0.606 (0.593, 0.619)	0.563 (0.547, 0.576)	0.440 (0.429, 0.458)
[47]	DS-TransUNet	0.702 (0.691, 0.712)	0.605 (0.594, 0.616)	0.583 (0.570, 0.602)	0.471 (0.460, 0.490)
[48]	nnUNet v2	0.703 (0.694, 0.713)	0.609 (0.600, 0.618)	0.532 (0.523, 0.542)	0.439 (0.429, 0.448)
Proposed model	TAH U-Net	0.852 (0.812, 0.865)	0.745 (0.715, 0.752)	0.824 (0.823, 0.832)	0.732 (0.708, 0.742)

Table 3. Comparison of the TAH U-Net with state-of-the-art segmentation models, demonstrating superior performance in mDice and mIoU.

mechanism, our model significantly enhances feature extraction, segmentation accuracy, and interpretability. It effectively captures both local spatial features and long-range dependencies, ensuring the precise delineation of heterogeneous Gleason Grade regions. The use of XAI techniques such as LIME, Occlusion Sensitivity, and PDP confirms the model's transparent decision-making, which is a crucial step toward clinical applicability. Our experimental evaluation demonstrates that the proposed model performs favorably when compared to conventional U-Net architectures, showing notable improvements in segmentation accuracy and boundary refinement. While the model shows great promise, we acknowledge its limitations, including a tendency toward over-segmentation and the need for broader validation.

For future work, we plan to extend the model to multi-modal histopathological imaging, incorporating MRI and immunohistochemistry (IHC) data to improve segmentation robustness. We will also explore self-supervised and few-shot learning methods to reduce the model's reliance on large, annotated datasets. Further enhancements in interpretability, through advanced frameworks like SHAP and Grad-CAM++, will deepen our understanding of the model's decision-making process. Ultimately, our goal is to deploy the model in a real-world clinical environment, using feedback from pathologists and active learning strategies to continually refine its performance, thereby creating a more accurate and trustworthy tool for AI-assisted prostate cancer assessment.

Data availability

In this paper, the SICAPv2 dataset, a publicly available dataset of histopathological Whole Slide Images (WSI) with Gleason grades for prostate cancer grading, is used as our data set. The dataset is available here: https://data.mendeley.com/datasets/9xxm58dvs3/1.

Received: 8 April 2025; Accepted: 28 July 2025
Published online: 30 September 2025

References

1. Salvi, M. et al. A hybrid deep learning approach for gland segmentation in prostate histopathological images. *Artif. Intell. Med.* **115**, 102076 (2021).
2. Abraham, B. & Nair, M. Automated grading of prostate cancer using convolutional neural network and ordinal class classifier. *Inf. Med. Unlocked* **17**, 100256 (2019).
3. Duran-Lopez, L. et al. Wide & deep neural network model for patch aggregation in CNN-based prostate cancer detection systems. *Comput. Biol. Med.* **136**, 104743 (2021).
4. Morozov, A. et al. A systematic review and meta-analysis of artificial intelligence diagnostic accuracy in prostate cancer histology identification and grading. *Prostate Cancer Prostatic Dis.* **26**, 681–692 (2023).
5. Xiang, J. et al. Automatic diagnosis and grading of prostate cancer with weakly supervised learning on whole slide images. *Comput. Biol. Med.* **152**, 106340 (2023).
6. Fernandez-Mateos, J. et al. Tumor evolution metrics predict recurrence beyond 10 years in locally advanced prostate cancer. *Nat. Cancer* **5**, 1334–1351 (2024).
7. Collins, K. & Cheng, L. Morphologic spectrum of treatment-related changes in prostate tissue and prostate cancer: an updated review. *Hum. Pathol.* **127**, 56–66 (2022).
8. Lu, X. et al. Ultrasonographic pathological grading of prostate cancer using automatic region-based Gleason grading network. *Comput. Med. Imaging Graph.* **102**, 102125 (2022).
9. Bakkouri, I. & Afdel, K. MLCA2F: Multi-level context attentional feature fusion for COVID-19 lesion segmentation from CT scans. *Signal Image Video Process.* (2022).
10. Bakkouri, I., Afdel, K., Benois-Pineau, J. & Initiative, G. C. F. T. A. D. N. BG-3DM2F: Bidirectional gated 3D multi-scale feature fusion for Alzheimer's disease diagnosis. *Multimed. Tools Appl.***81**, 10743–10776 (2022).
11. Bakkouri, I. & Afdel, K. Computer-aided diagnosis (CAD) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images. *Multimed. Tools Appl.* **79**, 20483–20518 (2020).
12. Bakkouri, I. & Afdel, K. Multi-scale CNN based on region proposals for efficient breast abnormality recognition. *Multimed. Tools Appl.* **78**, 12939–12960 (2019).
13. Kim, H., Kong, S., Lee, H., Kim, K. & Jung, K. Abstraction in pixel-wise noisy annotations can guide attention to improve prostate cancer grade assessment. In *Workshop on Medical Image Learning with Limited and Noisy Data*, 23–31 (Springer, 2022).
14. Yu, M. et al. FBCU-Net: A fine-grained context modeling network using boundary semantic features for medical image segmentation. *Comput. Biol. Med.* **150**, 106161 (2022).

15. Rajagopal, A. et al. Mixed supervision of histopathology improves prostate cancer classification from MRI. *IEEE Trans. Med. Imaging* (2024).
16. Park, S. et al. Deep learning model for real-time semantic segmentation during intraoperative robotic prostatectomy. *Eur. Urol. Open Sci.* **62**, 47–53 (2024).
17. Kong, F. et al. Federated attention consistent learning models for prostate cancer diagnosis and Gleason grading. *Comput. Struct. Biotechnol. J.* **23**, 1439–1449 (2024).
18. Zhang, Z. et al. Segmentation assisted prostate cancer grading with multitask collaborative learning. *Pattern Recogn. Lett.* **183**, 42–48 (2024).
19. Behzadi, M. et al. Weakly-supervised deep learning model for prostate cancer diagnosis and Gleason grading of histopathology images. *Biomed. Signal Process. Control* **95**, 106351 (2024).
20. Liu, F. et al. A hybrid classification model with radiomics and CNN for high and low grading of prostate cancer Gleason score on MP-MRI. *Displays* **83**, 102703 (2024).
21. Shen, Q. et al. MixUNETR: A U-shaped network based on W-MSA and depth-wise convolution with channel and spatial interactions for zonal prostate segmentation in MRI. *Neural Netw.* **181**, 106782 (2025).
22. Jiang, H. et al. Microsegnet: A deep learning approach for prostate segmentation on micro-ultrasound images. *Comput. Med. Imaging Graph.* **112**, 102326 (2024).
23. Zaridis, D. et al. Resqu-net: Effective prostate's peripheral zone segmentation leveraging the representational power of attention-based mechanisms. *Biomed. Signal Process. Control* **93**, 106187 (2024).
24. Bhandary, S. et al. Investigation and benchmarking of u-nets on prostate segmentation tasks. *Comput. Med. Imaging Graph.* **107**, 102241 (2023).
25. Du, X., Shen, A., Wang, X., Feng, Z. & Deng, H. NRD-Net: a noise-resistant distillation network for accurate diagnosis of prostate cancer with bi-parametric MRI images. *Multimed. Tools Appl.* **83**, 33597–33614 (2024).
26. Sowmya, D., Bhavani, S., Sasank, V. & Rao, T. Prostate cancer classification using adaptive swarm intelligence based deep attention neural network. *Biomed. Signal Process. Control* **96**, 106654 (2024).
27. Bleker, J. et al. A deep learning masked segmentation alternative to manual segmentation in biparametric MRI prostate cancer radiomics. *Eur. Radiol.* **32**, 6526–6535 (2022).
28. Balaha, H., Shaban, A., El-Gendy, E. & Saafan, M. Prostate cancer grading framework based on deep transfer learning and Aquila optimizer. *Neural Comput. Appl.* **36**, 7877–7902 (2024).
29. Hussain, T., Shouno, H., Mohammed, M. A., Marhoon, H. A. & Alam, T. DCSSGA-UNet: Biomedical image segmentation with DenseNet channel spatial and semantic guidance attention. *Knowl. Based Syst.* **314**, 113233 (2025).
30. Hussain, T. & Shouno, H. MAGRes-UNet: Improved medical image segmentation through a deep learning paradigm of multi-attention gated residual U-Net. *IEEE Access* **12**, 40290–40310. https://doi.org/10.1109/ACCESS.2024.3374108 (2024).
31. Hussain, T. et al. EFFResNet-ViT: A fusion-based convolutional and vision transformer model for explainable medical image classification. *IEEE Access* **13**, 54040–54068. https://doi.org/10.1109/ACCESS.2025.3554184 (2025).
32. Alam, T. et al. An integrated approach using YOLOv8 and ResNet, SeResNet & Vision Transformer (ViT) algorithms based on ROI fracture prediction in X-ray images of the elbow. *Curr. Med. Imaging* **20**, e15734056309890 (2024).
33. Ferrero, A. et al. HistoEM: A pathologist-guided and explainable workflow using histogram embedding for gland classification. *Mod. Pathol.* **37**, 100447 (2024).
34. Xie, W. et al. Prostate cancer risk stratification via nondestructive 3d pathology with deep learning–assisted gland analysis. *Can. Res.* **82**, 334–345 (2022).
35. Aju, A. et al. Exploring vision transformers and XGBoost as deep learning ensembles for transforming carcinoma recognition. *Sci. Rep.* **14**, 1–35 (2024).
36. López-Pérez, M. et al. The CrowdGleason dataset: Learning the Gleason grade from crowds and experts. *Comput. Methods Programs Biomed.* **257**, 108472 (2024).
37. Butt, M., Kaleem, M., Bilal, M. & Hanif, M. Using multi-label ensemble CNN classifiers to mitigate labelling inconsistencies in patch-level Gleason grading. *PLoS ONE* **19**, e0304847 (2024).
38. Zhang, X. et al. A universal multiple instance learning framework for whole slide image analysis. *Comput. Biol. Med.* **178**, 108714 (2024).
39. Falk, T. et al. U-Net: Deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 (2019).
40. Umirzakova, S., Muksimova, S., Baltayev, J. & Cho, Y. I. Force map-enhanced segmentation of a lightweight model for the early detection of cervical cancer. *Diagnostics* **15**, 513 (2025).
41. Kerfoot, E. et al. Left-ventricle quantification using residual u-net. In *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*, 371–380. MICCAI 2018, Revised Selected Papers (Springer, 2019).
42. Hatamizadeh, A. et al. UNETR: Transformers for 3D medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 574–584 (2022).
43. Kirillov, A. et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4015–4026 (2023).
44. Ma, J. et al. Segment anything in medical images. *Nat. Commun.* **15**, 654 (2024).
45. Hatamizadeh, A. et al. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In *International MICCAI Brainlesion Workshop*, 272–284 (Springer, Cham, 2021).
46. Lin, A. et al. DS-TransUNet: Dual swin transformer U-Net for medical image segmentation. *IEEE Trans. Instrum. Meas.* **71**, 1–15 (2022).
47. Isensee, F., Jaeger, P., Kohl, S., Petersen, J. & Maier-Hein, K. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021).
48. Zhang, H. et al. Masked image modeling meets self-distillation: A transformer-based prostate gland segmentation framework for pathology slides. *Cancers* **16**, 3897 (2024).

Acknowledgements

The authors extend their appreciation to Taif University, Saudi Arabia, for supporting this work through Project Number (TU-DSPP-2024-41).

Author contributions

These authors contributed equally to this work.

Funding

This research was funded by Taif University, Saudi Arabia, Project No. (TU-DSPP-2024-41).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to A.N.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.