

Deep learning approach for accurate prostate cancer identification and stratification using combined immunostaining of cytokeratin, p63, and racemase

Massimo Salvi ^{a,*}, Claudia Manini ^{b,c}, Jose I. López ^d, Dario Fenoglio ^a, Filippo Molinari ^a

^a Biolab, PoliToBIOMed Lab, Department of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy

^b Department of Pathology, San Giovanni Bosco Hospital, 10154 Turin, Italy

^c Department of Sciences of Public Health and Pediatrics, University of Turin, 10124 Turin, Italy

^d Biomarkers in Cancer Group, Biocruces-Bizkaia Health Research Institute, 48903 Barakaldo, Spain

ARTICLE INFO

Keywords:

Automated systems
Deep Learning
Digital Pathology
Prostate cancer
Supervised learning
Whole-Slide image

ABSTRACT

Background: Prostate cancer (PCa) is the most frequently diagnosed cancer in men worldwide, affecting around 1.4 million individuals. Current PCa diagnosis relies on histological analysis of prostate biopsy samples, an activity that is both time-consuming and prone to observer bias. Previous studies have demonstrated that immunostaining of cytokeratin, p63, and racemase can significantly improve the sensitivity and the specificity of PCa detection compared to traditional H&E staining.

Methods: This study introduces a novel approach that combines diagnosis-specific immunohistochemical (IHC) staining and deep learning techniques to provide reliable stratification of prostate glands. Our approach leverages a customized segmentation network, called K-PPM, that incorporates adaptive kernels and multiscale feature integration to enhance the functional information of IHC. To address the high class-imbalance problem in the dataset, we propose a weighted adaptive patch-extraction and specific-class kernel update.

Results: Our system achieved noteworthy results, with a mean Dice Score Coefficient of 90.36% and a mean absolute error of 1.64 % in specific-class gland quantification on whole slides. These findings demonstrate the potential of our system as a valuable support tool for pathologists, reducing workload and decreasing diagnostic inter-observer variability.

Conclusions: Our study presents innovative approaches that have broad applicability to other digital pathology areas beyond PCa diagnosis. As a fully automated system, this model can serve as a framework for improving the histological and IHC diagnosis of other types of cancer.

1. Introduction

Prostate cancer (PCa) is the most frequent cancer in the male population and the second-leading cause of male cancer death in Western countries (Rao et al., 2012). PCa is a complex androgen-dependent disease with different evolutionary patterns of growth and a wide spectrum of histologic appearances. Its prognosis is variable, ranging from indolent to very aggressive, with the latter showing accelerated transitions from the initial androgen-dependence to a castration-resistance status depending on specific genomic alterations (Karantanos et al., 2013). Such adverse genomic settings lead to uncontrolled growth, widespread dissemination, and patient death. The diagnosis of PCa usually relies on the histological analysis of trans-rectal

core biopsies. After the initial diagnosis, a subset of patients undergo a radical prostatectomy as the most common curative treatment. Pathologists identify the amount of cancer present in the core biopsies, assign the Gleason Score (GS), and evaluate the presence of specific histological features associated with prognosis such as perineural invasion and cribriform architecture. Recent studies have emphasized the importance of performing a correct classification of cribriform structures since their presence has a significant prognostic impact (van Leenders et al., 2020; Chan et al., 2022). In this context, the distinction between the cribriform type of high-grade intraepithelial neoplasia (HG-PIN), the intraductal adenocarcinoma, and the invasive adenocarcinoma with cribriform structures is highly recommended (Kench et al., 2022).

Although the routine diagnosis of PCa is made in most cases on H&E-

* Correspondence to: Biolab, Department of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy.

E-mail address: massimo.salvi@polito.it (M. Salvi).

<https://doi.org/10.1016/j.compmedimag.2023.102288>

Received 11 May 2023; Received in revised form 12 August 2023; Accepted 12 August 2023

Available online 19 August 2023

0895-6111/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

stained slides (Salvi et al., 2023), some particular situations advise performing additional immunostaining, for instance, when the tumor is scarcely represented in the sample -the so-called minute carcinoma- or when there is reasonable diagnostic doubt in distinguishing non-invasive versus invasive features. The detection of basal cells using p63 has been recommended to effectively reduce the diagnostic uncertainty in PCa (Chatrjian et al., 2021). However, an immunostaining that includes p63, basal cell-type cytokeratins (34 β E12), and racemase (p504s) antibodies appears to be a more reliable strategy to resolve any diagnostic dilemma (Paner et al., 2008). Several studies (Molini  et al., 2006; Dabir et al., 2012) have established that the combination of these markers can enhance the sensitivity and specificity of PCa detection. For example, Molini  et al. (2006), have achieved a sensitivity and specificity of 97 % and up to 100 %, respectively, using a cocktail combining p63 and p504s. Such a cocktail enabled the authors to correctly diagnose 93 % of ambiguous lesions on H&E slides. Indeed, the percentage of biopsies requiring this IHC evaluation varies from 25 % to 50 % depending on different institutions (Watson et al., 2013). The clinical practice demonstrates that the simultaneous combination of p63, 34 β E12, and p504s within the same slide is the best option and represents an opportunity to perform an *all-in-one* analysis of PCa based on artificial intelligence (AI). Although these markers allow for better differentiation between invasive and non-invasive lesions in the prostate, no AI-based approaches have been reported so far in the literature.

The last decades have witnessed an exponential growth in the application of AI algorithms in the biomedical field (Sardanelli et al., 2023), where deep learning (DL) systems have already achieved a state-of-the-art performance in almost all fields of medical imaging (Anaya-Isaza et al., 2021), including digital pathology (Deng, 2020). Traditionally, pathologists identify several quantitative parameters (e. g., number of cells, size of nuclei and cytoplasmic characteristics, tissue differentiation) used with diagnostic and/or prognostic purposes (Magi-Galluzzi, 2018; Khened et al., 2021). As a result, the diagnostic process becomes increasingly time-consuming (Khened et al., 2021) and subjected to inter- and intra-observer variability (Komura and Ishikawa, 2018; Ozkan et al., 2016). This fact directly influences the clinical management of patients. In this context, automated DL-based systems were introduced to emulate analytical protocols and provide objective quantification to support the pathologic diagnosis and prognosis.

The use of DL in oncology is increasingly being explored to integrate genomic, transcriptomic, and histopathological data with the goal of improving tumor detection, diagnosis, and treatments selection in the era of precision medicine (Tran et al., 2021). DL models have been developed not only for imaging-based diagnostic specialties like radiology or pathology but also for surgery and radiotherapy, contributing to the decision-making process. In the context of prostate cancer, DL is involved in various steps, from tumor detection to treatment. One prominent application is the analysis of magnetic resonance images, which aims to enhance tumor detection and provide more precise tumor evaluation (Michaely et al., 2022; Li et al., 2023; Netzer et al., 2023). DL facilitates decision-making processes related to surgical strategy selection, standardization of procedures, and training surgeons in robot-assisted surgery (T taru et al., 2021). It has also been applied to predict post-surgical adverse effects using magnetic resonance images (Nakamura et al., 2023). Additionally, radiomics-based information obtained from magnetic resonance and computed tomography images can assist radiation oncologists in the accurate application of brachytherapy and external beam radiation for prostate cancer patients who opt for these treatment modalities (T taru et al., 2021). Finally, DL models have gained worldwide traction in the pathological diagnosis of prostate cancer and have recently been reviewed by Busby et al (Busby et al., 2023). The paper highlights the utility of DL as an initial screening strategy and a second-read system to detect false negative diagnoses.

There has recently been a significant increase in the use of DL networks to assist pathologists in the PCa diagnosis. Although most applications have been performed on histological slides (Bulten et al., 1907;

Isaksson et al., 2017; Singhal et al., 2022; Salvi et al., 2021a), the development of specific systems for the segmentation of immunohistochemical samples in PCa diagnosis is also beginning to emerge. Bulten et al. (2019), for example, exploited the functional information of immunostaining (i.e., p63 and CK8/18) to build a more objective and accurate diagnostic basis. Subsequently, a CNN was trained on histological slides to perform the epithelial gland segmentation. However, this method has an inherent limitation as working on H&E staining makes it impossible to correctly classify certain cells, particularly in regions of active inflammation. In another study (Leng et al., 2019), hand-crafted features are extracted from both IHC (i.e., p63, 34 β E12, and p504s) and H&E-stained slides to perform the segmentation of adenocarcinoma regions with a regression model. As a result, the performance of the system is biased by color artifacts during the digitization process resulting in variations of the staining intensity and in the spatial correspondence of the two paired slides.

A possible limitation of the technique is that the extracted features correspond only to small regions of the gland without considering the entire marker distribution across different glands in different regions, which is essential for a correct classification of the transition phases (e. g., PIN). A recent paper by Zhdanovich et al. (2023), has shown that IHC (i.e., p63, 34 β E12, and p504s) is more suitable for prostate core classification compared to H&E staining and ERG expression. However, they employed hand-crafted features and machine learning algorithms (e.g., neural network, SVM, and random forest) were employed. To the best of our knowledge, only Blessin et al. have implemented so far a CNN-based system for PCa diagnosis on multiplex immunofluorescence. Nevertheless, their study focused only on the detection and segmentation of cells to assess the Ki-67 labeling index.

This paper introduces an innovative approach that combine a diagnosis-specific IHC staining for PCa detection (p63, 34 β E12, and p504s) with an advanced AI-based segmentation framework. The main contributions of this work can be summarized as follows:

- Successful identification of PCa in core biopsies and its distinction from non-invasive (intraductal) lesions. For this purpose, two expert independent uro-pathologists (C.M. and J.I.L.) performed precise multi-class annotations through the dataset to ensure the reliability of the results;
- Definition of a new segmentation network, called K-PPM, which combines adaptive kernels with multiscale features aggregation. This mechanism allows capturing relevant information across scales and effectively integrating it. The network identifies important features even in small image regions. We developed K-PPM specifically to address prostate cancer segmentation in histological images. The network's innovative approach resulted in a superior performance compared to all previous state-of-the-art methods checked during the validation process.
- Proposition of a novel method for patch extraction called "weighted-adaptive patch extraction," which addresses the challenge of working with unbalanced data. This approach facilitates dataset balancing during patch extraction and the rapid implementation in a wide range of other applications.
- Development of a novel inference method that generates smooth network predictions by effectively merging consecutive patches, without producing the common checkboard artifacts. This approach can be easily adapted to various other applications and has the potential to enhance the accuracy and visual quality of the predictions.

This paper is organized as follows: Section 2 presents a comprehensive description of the proposed method; experimental results are reported in Section 3; Sections 4 and 5 discuss the work as a whole.

2. Materials and methods

2.1. Dataset

The complete dataset consists of 32 annotated WSIs of prostate biopsies performed in the routine work-up at the San Giovanni Bosco hospital, Turin, Italy. We have obtained a total of 36,894 annotations for the following classes: benign glands, prostatic intraepithelial neoplasia (PIN), adenocarcinoma, and tissue of no interest (TNI) (stroma and background). Fig. 1 shows a tile of 2048×2048 pixels at $20\times$ magnification and its manual mask are shown in Fig. 1. The WSI dimensions range from $13,000 \times 24,000$ to $40,000 \times 44,000$ pixels at a magnification of $20\times$.

A routine double-staining technique using the basal cocktail 34 β E12 +p63 (Ventana-Roche, product numbers 790–4536 and 790–1010, ready-to-use) and p504s (Ventana-Roche, product number 7090–6011, ready-to-use) was performed on a Benchmark Ultra (Roche, Ventana) automated immunostainer at the Pathology Lab., San Giovanni Bosco Hospital, Turin, Italy. The cocktail 34 β E12 +p63 binds the basal cells of the prostatic glands and stains them with a dark brown color while p504s stains prostate adenocarcinoma cells in red color. As shown in Fig. 1, benign glands of the prostate show a continuous peripheral row of brown-stained basal cells, while adenocarcinoma displays only red-stained glands indicating malignancy. Finally, PIN glands show a mixed pattern of staining, with a discontinuous row of basal brown-stained (benign) cells at the periphery and central groups of red-stained (malignant) cells.

2.2. Semi-supervised data labeling

Manual labeling for semantic segmentation is time-consuming, involving the segmentation and classification of each gland in the WSI. Since WSIs may have up to 1200 glands on average, the whole process requires significant manual effort. To accelerate the labeling process, a semi-supervised approach was employed. Seventeen WSIs were manually annotated, and an initial segmentation was performed using a traditional U-Net over the entire dataset. Subsequently, manual inspection and correction of the automatic masks were carried out to change any erroneous segmentations. This semi-automatic data labeling procedure expedited the process approximately 5 times and resulted in a total of 36,894 glandular annotations.

2.3. Dataset partitioning and weighted-adaptive patch extraction

The dataset was partitioned into training set (22 WSI), validation set (5 WSI) and test set (5 WSI). Two parameters were considered for patch extraction: magnification and patch size. While higher magnification (e. g., $20\times$) provides clearer view of cellular details, it reduces the field of view (FOV) of the patch. Since a proper classification of a gland requires observing its entire staining, a larger FOV is preferred over cellular resolution. Several experiments were conducted, varying both FOV size (512×512 pixels, $1,024 \times 1,024$ pixels, 2048×2048 pixels) and magnification ($5\times$, $10\times$, $20\times$). The final configuration that yielded the best performance on the validation set was a FOV of $1,024 \times 1,024$ pixels at a magnification of $20\times$.

As shown in Fig. 2b, the dataset exhibits high-class, with the PIN class being the least represented, followed by the adenocarcinoma class. To mitigate this problem, a new weighted-adaptive patch extraction strategy was employed, following these criteria:

1. If the PIN class is present, the patch is extracted with a 75 % overlap with the next horizontal patch.
2. If the adenocarcinoma class is present, the patch is extracted with a 50 % overlap with the next horizontal patch.
3. Patches with the remaining classes are extracted without overlap, based on the percentage of white background area (i.e., regions without tissue). Below 30 % of the white area, the patch is extracted, while the extraction probability decreases as the percentage of white (i.e., background) within the patch increases.

Fig. 2a illustrates an example of weighted patch extraction with adaptive, where the path extraction becomes denser in regions with adenocarcinoma and PIN glands. In addition, an automatic patch selection process is used to reduce the high representation of the background class. The criteria include maintaining no more than 200 patches with over 90 % of background for each WSI. A comparison between traditional patch extraction with a constant 25 % overlap and the applied methods is shown in Fig. 2, demonstrating that the combination of the weighted path extraction with the adaptive overlap can reduce the imbalance.

Although PIN annotations comprise only 1.95 % of annotated pixels, representing less than one fifth of the benign or adenocarcinoma classes, the patch distribution in the training set is more balanced (Fig. 2c). At the end of this process, the final dataset consists of 9700 patches for the

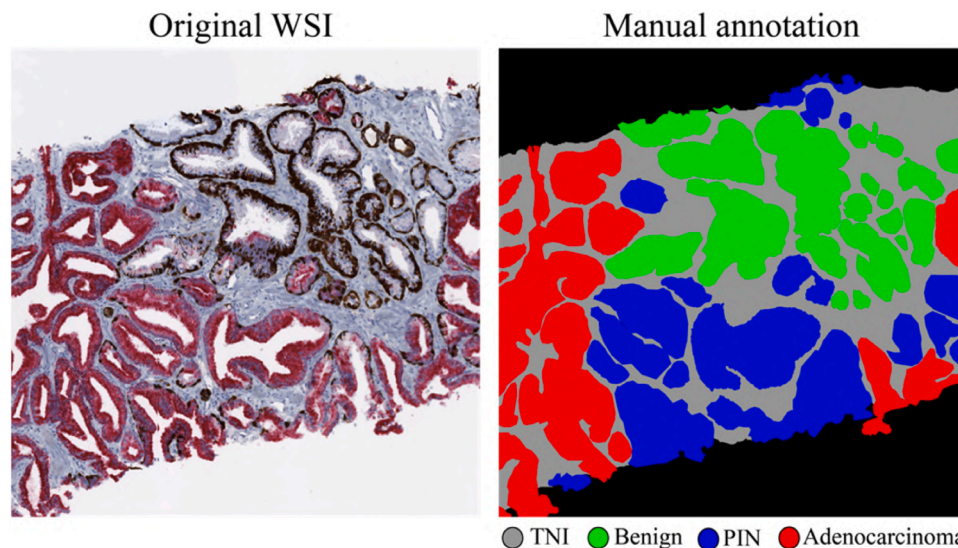


Fig. 1. An IHC tile of 2048×2048 pixels and its manual mask are shown. The legend indicates the color assigned to the respective classes. PIN: prostatic intra-epithelial neoplasia; Benign glands; Adenocarcinoma; TNI: tissue of no interest.

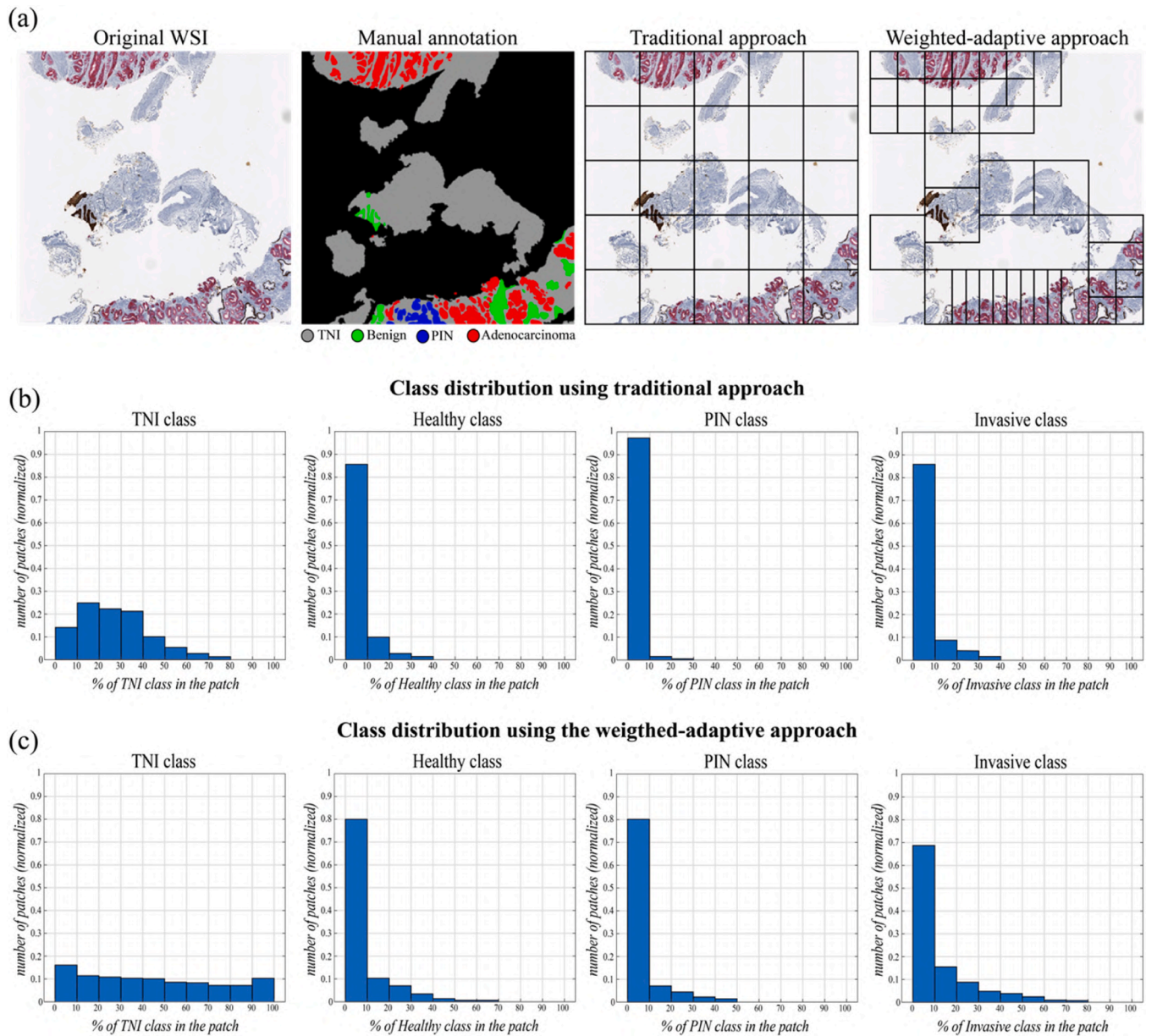


Fig. 2. Comparison between traditional patch extraction and the proposed weighted-adapted strategy to mitigate the class imbalance. (a) Patch extraction on a sample IHC tile. (b) Class distribution over the dataset using the traditional approach. (c) Class distribution over the dataset using the weighted-adaptive approach.

training set, and of 1807 and 1777 patches for the validation and test sets, respectively.

2.4. Proposed segmentation network (K-PPM)

Considering that the task is the semantic segmentation of three different types of glands (i.e., benign, PIN, and adenocarcinoma) and the tissue of no interest (i.e., stroma), the chosen models must exploit the information at different scales of the image. This is because the classification depends on the local and global disposition of the IHC markers on the whole gland.

This segmentation task has two main challenges: (i) discriminating features are encoded in a multiscale context; (ii) the dataset is not balanced, and some classes are underrepresented (i.e., PIN). To tackle these issues, we designed a new architecture called K-PPM, which extracts multiscale features thanks to a pyramid pooling module and perform precise segmentation by using different dynamic kernels for each class.

The K-PPM is based on the synergic combination between the PPM (Pyramid Pooling Module) (Zhao et al., 2017) and K-net architecture (Zhang et al., 2021). Our K-PPM employs a ResNet50 (He et al., 2016) as the features extractor. Next, a PPM is applied to the features map which exploits global context information by fusing features into four different pyramid scales. The featured map is divided into different subregions and forms a pooled representation for different locations. As illustrated in Fig. 3, the coarsest level in red represents the global information, whereas the higher levels contain features of the smaller regions of the image. After up-sampling, the different featured levels are concatenated with the entire feature map. This step is especially useful in gland classification, where both local and global information are relevant. In fact, the final prediction becomes more reliable when the fusion of local and global clusters is performed.

The ResNet50 followed by the PPM module represents the featured map generator of the K-PPM. Similar to the K-net, the K-PPM uses a kernel update head step to update the dynamic kernels with respect to their corresponding segmentation class and generates a more accurate

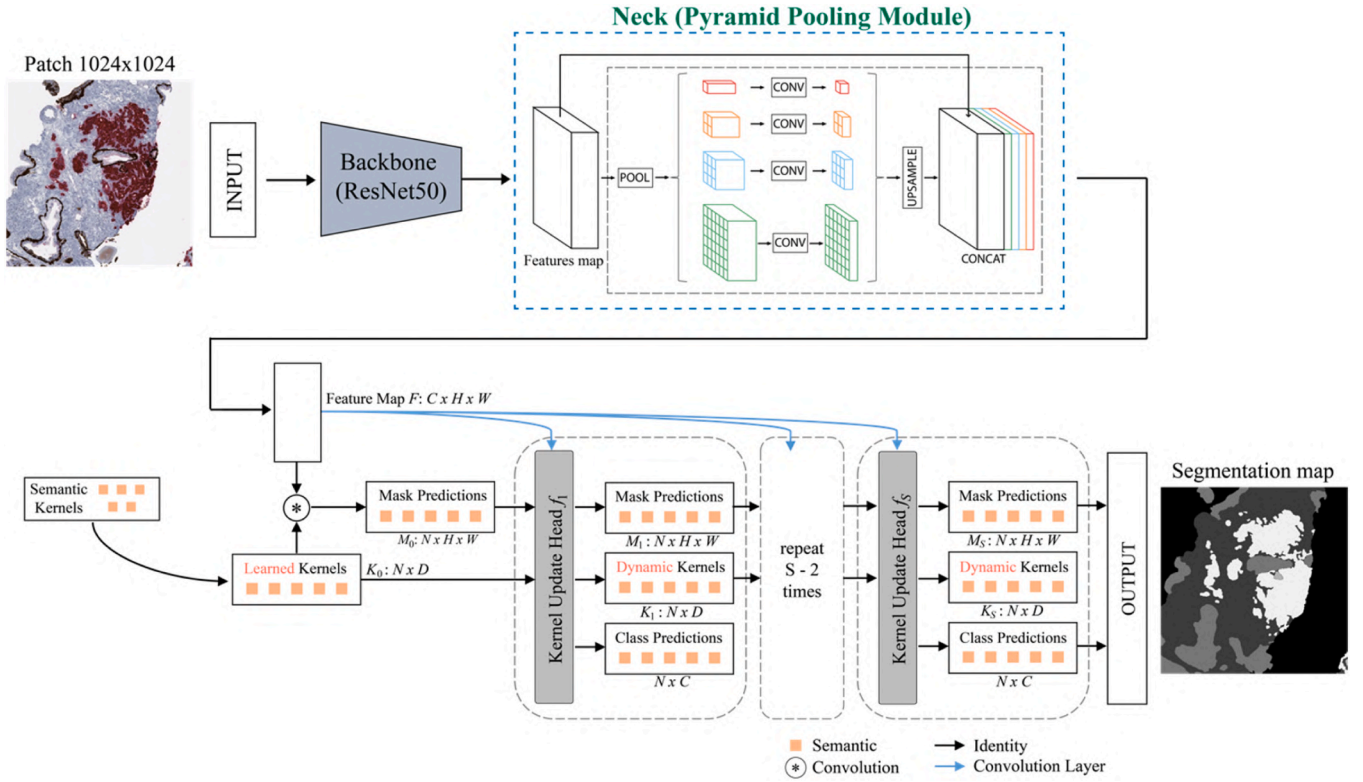


Fig. 3. Overall architecture of the K-PPM network. ResNet50 is employed as feature extractor (backbone). Then, the Pyramid Pooling Module perform multi-resolution feature aggregation (neck) and the K-net architecture generate the final prediction based on dynamic kernels for each class. Specifically, a convolution is performed between the feature map F and the learned kernels K_0 to predict the masks M_0 . After the kernel update, the network takes M_0 , F , and K_0 as input and produces M_1 mask predictions, K_1 dynamic kernels, and class prediction. The values of M_1 , K_1 , and F will be the inputs of the next kernel update. This step is repeated to progressively refine the mask predictions and dynamic kernels. N : number of input channels; C : number of classes; H : height of the image; W : width of the image.

mask prediction. Firstly, the group feature F^K , which is the content of each individual group, is produced by the multiplication between the previous mask prediction M_{i-1} and the features map F . After that step, the group feature F^K is used to update their corresponding kernel K_{i-1} , adaptively. Kernel refinement ends with a kernel interaction, where each kernel is updated based on the contextual information from other groups through a Multi-Head Attention block (Vaswani et al., 2017; Wang et al., 2021). Finally, the convolution between the obtained kernels K_i and the feature map F produces a more precise mask prediction M_i . In this way, our framework improves the discriminative ability of the kernels and boosts the final segmentation performance. Indeed, each kernel is specialized in the segmentation of the respective class, and during the training, its respective weights are improved according to its respective task.

2.4.1. Implementation details

Table 1 provides an overview of the architecture and layers of the K-PPM network. The backbone consists of a stem layer, max pool layer, and multiple ResNet layers with varying numbers of bottleneck layers. The decode head includes a kernel generate head, and three kernel update heads. The auxiliary head consists of segmentation convolutions and bottleneck convolutions. Overall, the network has approximately 79.6 million trainable parameters.

During training, the number of epochs was set to 80, with a batch size of 4 and an initial learning rate is 10^{-3} . Random flip was performed in real-time to avoid overfitting. Adam's optimization algorithm with a weight decay of 0.0005 and a weighted cross-entropy loss (Eq. 1) for both the decode and auxiliary head of the model are used. The weights W_c of the cross-entropy loss are inversely proportional to the number of pixels belonging to class c . This means that the least represented class (PIN) will have a greater contribution than the more represented one (e.

g., TNI) during the weight update. The loss used for the backbone and decode head is defined as:

$$L_{WCE} = -\frac{1}{N} \sum_{c=1}^C W_c \cdot \sum_{i=1}^N g_{ic} \cdot \log(p_{ic}) \quad (1)$$

Where p_{ic} and g_{ic} are the predicted segmentation probability and the ground truth label of class c at pixel i , respectively. N and C are the numbers of pixels and classes in the training dataset.

Regarding the auxiliary head, we have adopted a loss function that combines the Weighted Cross-Entropy Loss (Eq. 1) and the Weighted Dice Loss. Additionally, within the architecture of the auxiliary head, we have employed the OHEMPixelSampler (Online Hard Example Mining Sampler), which is a technique specifically designed for segmentation tasks. This method selects and samples the most challenging training examples during the training process of a segmentation model. By focusing on difficult instances or "hard examples" that are initially misclassified by the model, the OHEMPixelSampler improves the model's ability to handle complex scenarios and enhance its overall performance. This approach ensures that the model learns from the most challenging regions, reducing the risk of relying solely on easy examples that may not adequately represent the complexity of the segmentation problem.

2.5. Model inference

The proposed segmentation model runs on patches of dimension $1,024 \times 1,024$ pixels, but the final application is the segmentation mask over the entire WSI. A simple reconstruction approach is to apply the model on successive patches of the WSI without overlapping. It is the fastest method because each pixel of the WSI is classified only once by

Table 1
Overall architecture and layers of the proposed K-PPM network.

Section	Layer type	Settings	Activation functions	Trainable parameters
Backbone	Stem Layer	Conv3×3, BatchNorm (x3)	ReLU	28,768
	MaxPool Layer			
	Resnet Layer	Bottleneck (x3)	ReLU	215,808
	Resnet Layer	Bottleneck (x4)	ReLU	1,219,584
	Resnet Layer	Bottleneck (x6)	ReLU	7,098,368
Decode Head Kernel Generate Head Kernel Update Head (x3)	Resnet Layer	Bottleneck (x3)	ReLU	14,964,736
	Segm. Conv	Conv1×1, DropOut		2565
	PPM Module	AdapAvgPool, Conv1×1, BatchNorm (x4)	ReLU	4,198,400
	Bottleneck Conv	Conv3×3, BatchNorm	ReLU	18,875,392
	Lateral Conv	Conv1×1, BatchNorm (x3)	ReLU	92,0576
	FPN Conv	Conv3×3, BatchNorm (x3)	ReLU	7,080,960
	FPN	Conv3×3, BatchNorm	ReLU	9,438,208
	Bottleneck MultiHead	Attention layer, AttentionNorm		1,051,648
	Kernel Updator	Linear, LayerNorm (x5)	ReLU	463,104
	Feature Conv	Conv1×1		262,656
	Feed-Forward Network (FFN)	Linear, ReLU, Dropout, LayerNorm	ReLU	2,100,736
	FC	Linear, LayerNorm, ReLU, Linear	ReLU	525,824
	Auxiliary Head	Segm. Conv	Conv1×1, DropOut	1285
		Bottleneck Conv	Conv3×3, BatchNorm	2,359,808
Total				79.6 M

(*) Bottleneck is defined as a sequence of Conv1×1, BatchNorm, Conv3×3, BatchNorm, Conv1×1, BatchNorm, and ReLU.

the network, but it usually produces discontinuities in the final automatic mask on patch contours, as demonstrated by our previous work (Salvi et al., 2021b). This occurs because the predictions close to the contours are less reliable because the model has little information to make the classification. Furthermore, considering that a correct classification needs to look at the entire gland, the outer region has probably omitted some useful information from the input window.

In this work, we propose a novel reconstruction technique called SRM (smooth reconstruction method) to perform a spatially consistent segmentation map. First, an extended image is synthesized by padding the original image with mirror reflections of 512×512 pixels along each direction. This step is essential to give equal weight to regions near the edges. Then, a sliding window scrolls over the image with 50 % overlap, and for each patch, our segmentation model is applied to obtain the softmax. A pixel-wise multiplication is performed between the softmax of each patch and a second order spline window to gradually scale the value of each pixel with respect to the distance from the center of the image, indicating its reliability. The output of this operation is a weighted softmax. As shown in Fig. 4, the SRM repeats this process and each weighted softmax is placed with the same overlap through a sum operation. Considering that the sliding window size is 1024×1024 pixels and the horizontal translation is 512 pixels (50 %), two subsequent windows share half of the information. Since the same concept

also applies to the vertical translation of 512 pixels, each pixel of the WSI is viewed by the segmentation network from 4 different patches. Therefore, the classification of each pixel is a linear combination of 4 predictions, where the weights (i.e., a, b, c, e) are proportional to the distance (d) between the pixel and the center of the considered prediction:

$$P(x, y) = a(d_1) \bullet P_1(x, y) + b(d_2) \bullet P_2(x, y) + c(d_3) \bullet P_3(x, y) + e(d_4) \bullet P_4(x, y) \quad (2)$$

Here, P_i denotes the i^{th} prediction of the neural network on the considered pixel at position (x, y) ; $P(x, y)$ represent the final pixel prediction; d_i is the distance between the position of the considered pixel and the center of the i^{th} prediction. Finally, a crop is performed to obtain a segmentation mask with the same size of the original image. By using a smaller sliding window step, which means increasing the overlap between consecutive patches, the performance can be further increased. In that case pixel classification will depend on a linear combination of more than 4 predictions, and as a result the computation time will increase.

2.6. Performance metrics

Automatic masks were compared with manual ones to evaluate the performance of our strategy in the segmentation of PCa. The first validation metric was the Dice similarity coefficient (DSC), which is a statistic used to measure the similarity of two samples. The value ranges between 0 and 1, where 1 means that the automatic and manual mask are identical. This is one of the most used metrics for evaluating segmentation tasks (Crum et al., 2006; Gao et al., 2012). In addition, we computed the Balanced Accuracy (Bal_{ACC}), since it is a metric widely used in segmentation problems to deal with unbalanced datasets (Broderson et al., 2010; Grandini et al., 2008). Computational efficiency was also evaluated to identify the best trade-off between performance and processing time.

3. Results

3.1. Segmentation performance

The fully automated results provided by our framework are compared with manual masks drawn by an expert observer. We evaluated the performance of our method on both 1024×1024 patches and $5,120 \times 5,120$ tiles to assess the effectiveness of the smooth reconstruction technique. The processing was executed on a workstation equipped with a 4.1 GHz deca-core CPU with 64-GB of RAM and NVIDIA GeForce RTX 3070 GPU with 8-GB of VRAM. To compare the results, the Dice Similarity Coefficient (DSC), and the Balanced Accuracy (Bal_{ACC}) are used as evaluation metrics. The obtained metrics are presented in Table 2.

The evaluation of the network's generalization ability on the test and validation sets demonstrates that the implementation of the smooth overlap technique enhances the system, while allowing it to work on larger inputs. The DSC on the test set increases from 86.35 % on patches to 90.36 % on tiles, indicating a significant improvement. Notably, this improvement is not reflected in the training set, which was used to train the network and may have some overfitting. Fig. 5 shows three automatic segmentations extracted from the test set tiles and corresponding manual masks, highlighting a high consistency in gland classification. The few inaccuracies observed in Fig. 5 are primarily due to the lack of detail in the manual masks, such as the absence of small stromal regions within invasive glands and PINs. Hence, in most cases, the network achieves greater precision in identifying the contour of the glands than the manual observer.

Additionally, to evaluate the performance of our method across different classes, including TNI, composed mainly of stroma, benign and adenocarcinoma glands, and PINs, the evaluation metrics (i.e., DSC and

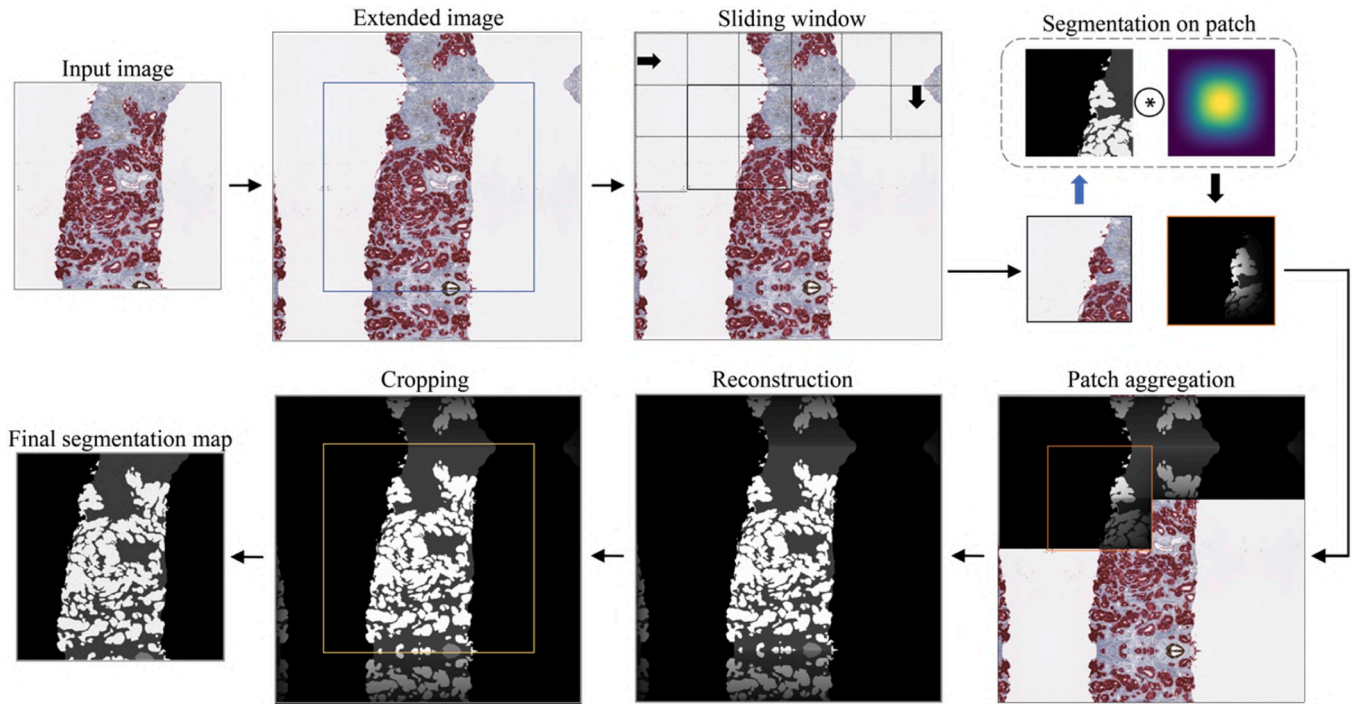


Fig. 4. Steps of the SRM (smooth reconstruction method): original image, padding, patch sliding window process, multiplication between the softmax of the patch and the second order spline window, patch aggregation and cropping to obtain the final softmax.

Table 2

Pixel-level performance of the proposed segmentation framework on patches 1024×1024 pixels and on tiles of 5120×5120 pixels. DSC: Dice similarity coefficient. BAL_{ACC}: balanced accuracy.

Subset	Patches 1024×1024			Tiles 5120×5120		
	DSC	BAL _{ACC}	Time (s)	DSC	BAL _{ACC}	Time (s)
Train set	94.95	95.75	0.33	90.43	95.11	26.1
Validation set	87.87	92.29		88.93	93.61	
Test set	86.35	90.95		90.36	94.24	

BAL_{ACC}) were computed on individual classes. Table 3 presents the individual metrics on tiles of 5120×5120 pixels from the train, validation, and test sets.

Based on our metrics, our network faces the greatest challenge in distinguishing the PIN class from benign and adenocarcinoma glands, as evidenced by a higher standard deviation in both DSC and BAL_{ACC}. Segmentation errors on the PIN class are mainly due to a partial identification of the entire gland as PIN. Although the PIN class is less represented in the training dataset (1.95 % of the annotated dataset), our method achieves a promising DSC of 79.8 % on the test set tiles. We also achieved high accuracy in segmenting TNI, benign, and adenocarcinoma glands, consistently achieving a DSC above 91.6 % and a BAL_{ACC} above 96.1 % on the test set. These results indicate strong agreement with manual annotation.

The proposed smooth reconstruction technique can handle images of any size, making the process well-suited to analyzing the entire WSI used in the diagnostic process. To demonstrate the effectiveness of our system, a representative example on a WSI is displayed in Fig. 6, where the original images are cropped to highlight the segmentation details. This result shows a high degree of spatial coherence with the segmentation performed by an experienced observer, obviating the need to analyze each individual gland. Although few inconsistencies in identifying the PIN class are observed, our method achieves high consistency in the identification of the different glandular classes, enabling a rapid and reliable biopsy interpretation.

Since the pathologist quantifies the extent of individual classes on

the slides, we assessed the average error between manual and automatic segmentations produced by our method during the evaluation of the area of each class. Specifically, we calculated the absolute error for each diagnostically relevant class (Benign, PIN, Adenocarcinoma) between the area identified by the network and the one identified by the pathologist on WSI. The results are presented in Table 4. Notably, the algorithm provides highly consistent results compared to manual annotation, with a maximum error of 1.64 % observed in the PIN class.

3.2. Comparison with state-of-the-art

We evaluated the performance of our proposed K-PPM against several state-of-the-art segmentation networks. These included a traditional UNet (Ronneberger et al., 2015) based on Fully Convolutional Network (FCN), as well as two models that incorporated receptive field enlargement and multi-scale context: PSPNet (Zhao et al., 2017) and DeepLabV3 (Chen et al., 2017). We also tested a Swin-Transformer to explore alternative architectures that are not based on convolutional approaches. Following the procedure described in Section 2.4, we trained each network on the training set, as detailed in Section 2.3, and evaluated them on the test set patches using the DSC and BAL_{ACC} metrics.

As shown in Table 5, our network achieved a mean DSC of 86.4 % and a mean BAL_{ACC} of 94.3 %, which were 1.1 % and 1.1 % higher than the second-best model, respectively. Our network outperformed all previous networks on most classes, except for the PIN class. These results demonstrate the importance of integrating local and global features for accurate prostate gland classification. It's worth noting that other networks, such as DeepLabV3 and PSPNet, which have achieved similar performance to ours (i.e., 85 % of DSC), have also used multiscale feature aggregation techniques like Atrous Spatial Pyramid Pooling (ASPP) and PPM. Our network outperformed state-of-the-art because we employ dynamic kernels that are specifically designed for the class being segmented. This approach resulted in a significant boost in performance compared to the previously mentioned networks. To demonstrate the effectiveness of the proposed K-PPM method in prostate cancer segmentation and stratification, a pairwise t-test was conducted between

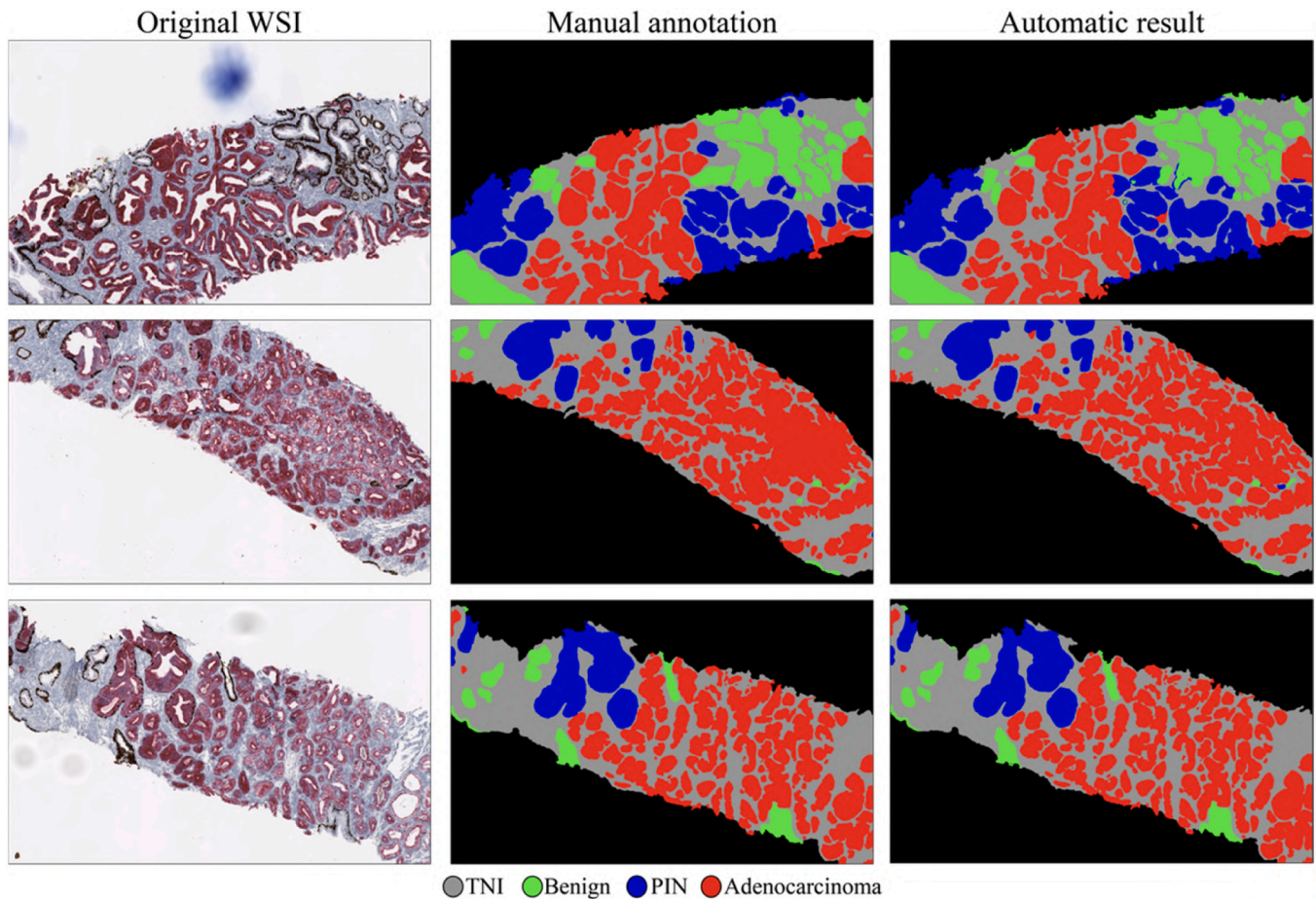


Fig. 5. Qualitative comparison between manual and automatic segmentation on three IHC tiles from test set.

Table 3

Performance metrics on the different classes on tiles of 5120×5120 pixels. TNI: tissue of no interest. PIN: prostatic intraepithelial neoplasia. DSC: Dice similarity coefficient. BAL_{ACC} : balanced accuracy.

Subset	Metric	Classes			
		TNI	Benign	PIN	Adenocarcinoma
Train set	DSC	97.9 ± 1.6	91.7 ± 16.1	80.3 ± 34.5	91.9 ± 20.2
	BAL_{ACC}	98.8 ± 1.1	95.7 ± 7.9	96.2 ± 9.7	96.9 ± 7.4
Validation set	DSC	96.8 ± 2.5	90.4 ± 8.7	80.7 ± 30.9	87.8 ± 28.6
	BAL_{ACC}	98.0 ± 2.0	94.9 ± 5.9	94.8 ± 10.0	96.0 ± 11.0
Test set	DSC	96.6 ± 3.0	93.4 ± 8.1	79.8 ± 31.1	91.6 ± 16.8
	BAL_{ACC}	98.3 ± 1.6	96.3 ± 5.2	94.6 ± 11.2	96.1 ± 7.3

the performance of our method and the compared techniques. All statistical tests were carried out with a significance level (p-value) of 0.05. The results of the paired t-test indicated a significant difference in balanced accuracy and dice score in the test set (Table 5).

In Fig. 7, we present three examples of cropped tiles obtained from the inference phase using the smooth reconstruction technique of the compared networks. Rows (a) and (b) demonstrate that networks without multiscale feature aggregation, such as UNET and Swin-Transformer, struggle to accurately distinguish large PIN nests from foci of adenocarcinoma with cribriform pattern. We discovered that correctly classifying large PIN nests requires global information about

the marker's spatial distribution (Fig. 7a), while cribriform patterns in adenocarcinoma also require local information visible at the cellular level (Fig. 7b). PSPNet and DeepLabV3 show similar segmentation results to our network, but they exhibit lower precision in the classification of challenging gland-like benign-mimickers, as shown in rows (c). This could be due to differences in the networks' ability to capture the fine details of class-specific glandular structures.

3.3. Faulty cases

Table 3 highlights that the PIN class was the most challenging to segment. However, it is important to consider three factors that contribute to this challenge. Firstly, the PIN class is underrepresented in our dataset, comprising only 1.95 % of the annotated pixels, which limits the amount of training data available for this class. Secondly, PIN annotations are subject to high inter-observer variability due to the difficulty in maintaining a consistent classification by pathologists during the labeling phase. As a precursor to PCa, PIN is defined as a neoplastic growth of epithelial cells within pre-existing benign prostatic ducts or acini, and its classification is not always clear-cut (Brawer, 2005). Therefore, this lack of clear delimitations in its classification contributes to the observed annotation variability. Fig. 8a provides a clear example of this situation, where our system correctly classified the glandular structure as PIN, which was previously misclassified by the manual operator, thus highlighting the system's capacity to reduce the inter-observer variability in gland classification. Thirdly, accurate PIN classification requires a complete understanding of the distribution of the marker over the entire surface of the glands, as only a small part of the gland may present the clues for correct classification (e.g., 34 β E12/p63-stained basal cells in an invasive pattern or vice versa, the

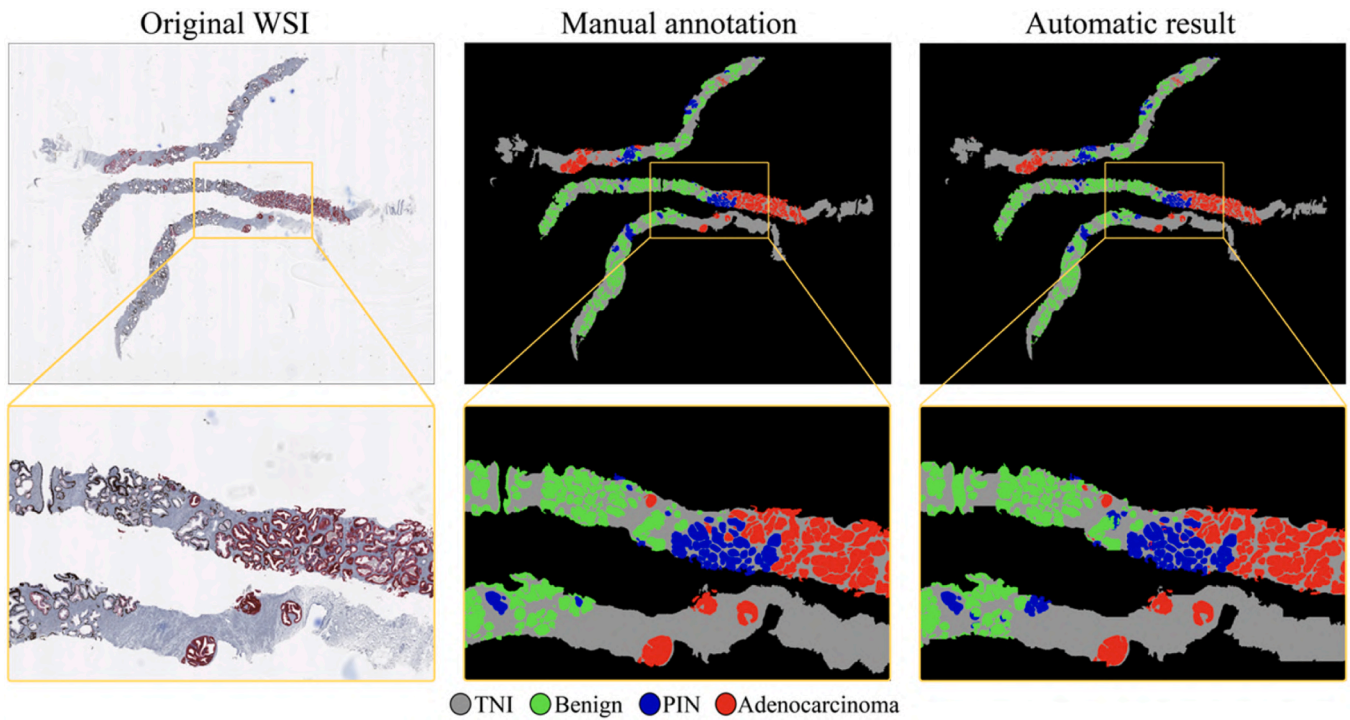


Fig. 6. Inference phase of the proposed method on an entire WSI. Whole slides were cropped images to visualize the segmentation details.

Table 4

Mean absolute error between manual and automatic segmentation during the evaluation of the area of each class in whole slide images. PIN: prostatic intraepithelial neoplasia.

Subset	Classes		
	Benign	PIN	Adenocarcinoma
Train set	1.39 ± 2.24	1.35 ± 2.29	0.32 ± 0.30
Validation set	1.31 ± 1.56	0.93 ± 0.78	0.86 ± 0.61
Test set	1.34 ± 2.10	1.64 ± 1.96	0.43 ± 0.45

presence of p504s in a benign-like gland). This is one of the primary reasons for incorporating a multi-scale features extraction mechanism. Nevertheless, since a PIN gland may not always appear entirely in the network's field of view, the neural network may not always have the necessary information to classify the gland accurately as a PIN. Fig. 8b shows a misclassification of a large PIN gland by our system, as only the left side of it presents the requirements to be classified as a PIN.

4. Discussion

The accurate identification and differentiation of benign glands from invasive or pre-invasive lesions in prostate core specimens is a critical issue for managing prostate cancer correctly in the era of precision oncology (Bulten et al., 2021; Arvaniti et al., 2018). The number of core biopsy specimens received in Pathology Departments to rule out prostate cancer continues to increase in Western Countries as a consequence of the generalization of PSA determinations in the blood analyses of the target male population, and also because of the implementation of screening programs (Hoffman, 2011). These reasons are the cause of the persistent leading position of prostate cancer incidence nowadays (Rao et al., 2012). A robust, reliable, and objective method of analysis would be very useful to overcome observer-dependent inconsistencies in a context of high clinical pressure. Traditionally, H&E-stained histological sections are the first step for these analyses. However, its accuracy may be eventually at risk in a still non determined number of cases due to several reasons, i.e., scarce material available, sub-optimal staining, diagnostic inexperience, and high-pressure working conditions, among others. A set of antibodies tested by IHC serve as ancillary tools in

Table 5

Comparison of the proposed network with current state-of-the-art methods on the test set. TNI: tissue of no interest. PIN: prostatic intraepithelial neoplasia. DSC: Dice similarity coefficient. BAL_{ACC}: balanced accuracy.

Method	Metric	Average values	Classes			
			TNI	Benign	PIN	Adenocarcinoma
SWIN-T Transformer	DSC	77.7 ± 33.0	82.8 ± 25.3	80.4 ± 28.8	75.3 ± 37.4	72.3 ± 38.7
	BAL _{ACC}	91.0 ± 14.0	93.1 ± 8.6	91.9 ± 12.2	90.9 ± 15.7	88.0 ± 17.8
UNet	DSC	79.1 ± 32.1	86.9 ± 23.3	83.6 ± 27.1	66.0 ± 42.2	79.8 ± 32.8
	BAL _{ACC}	92.9 ± 12.3	94.7 ± 8.5	92.5 ± 12.3	92.2 ± 13.5	92.1 ± 14.1
PSPnet	DSC	85.0 ± 28.4	88.9 ± 20.4	85.8 ± 26.6	81.7 ± 34.0	83.6 ± 30.9
	BAL _{ACC}	93.2 ± 12.8	94.7 ± 8.1	93.7 ± 12.4	92.3 ± 15.2	92.2 ± 14.4
DeepLabV3	DSC	85.3 ± 27.9	88.0 ± 21.3	86.5 ± 25.7	83.5 ± 32.4	83.4 ± 30.7
	BAL _{ACC}	92.7 ± 13.3	93.1 ± 10.0	93.7 ± 12.3	92.9 ± 14.9	91.3 ± 15.3
K-PPM (proposed)	DSC	$86.4 \pm 27.7(*)$	$91.3 \pm 18.6(*)$	$87.1 \pm 25.8(*)$	81.5 ± 34.0	$85.6 \pm 29.9(*)$
	BAL _{ACC}	$94.3 \pm 12.0(*)$	$95.4 \pm 9.3(*)$	$94.6 \pm 11.5(*)$	$94.4 \pm 12.4(*)$	$93.0 \pm 14.2(*)$

(*) Asterisks denote statistically significant difference ($p < 0.05$) compared to state-of-the-art methods.

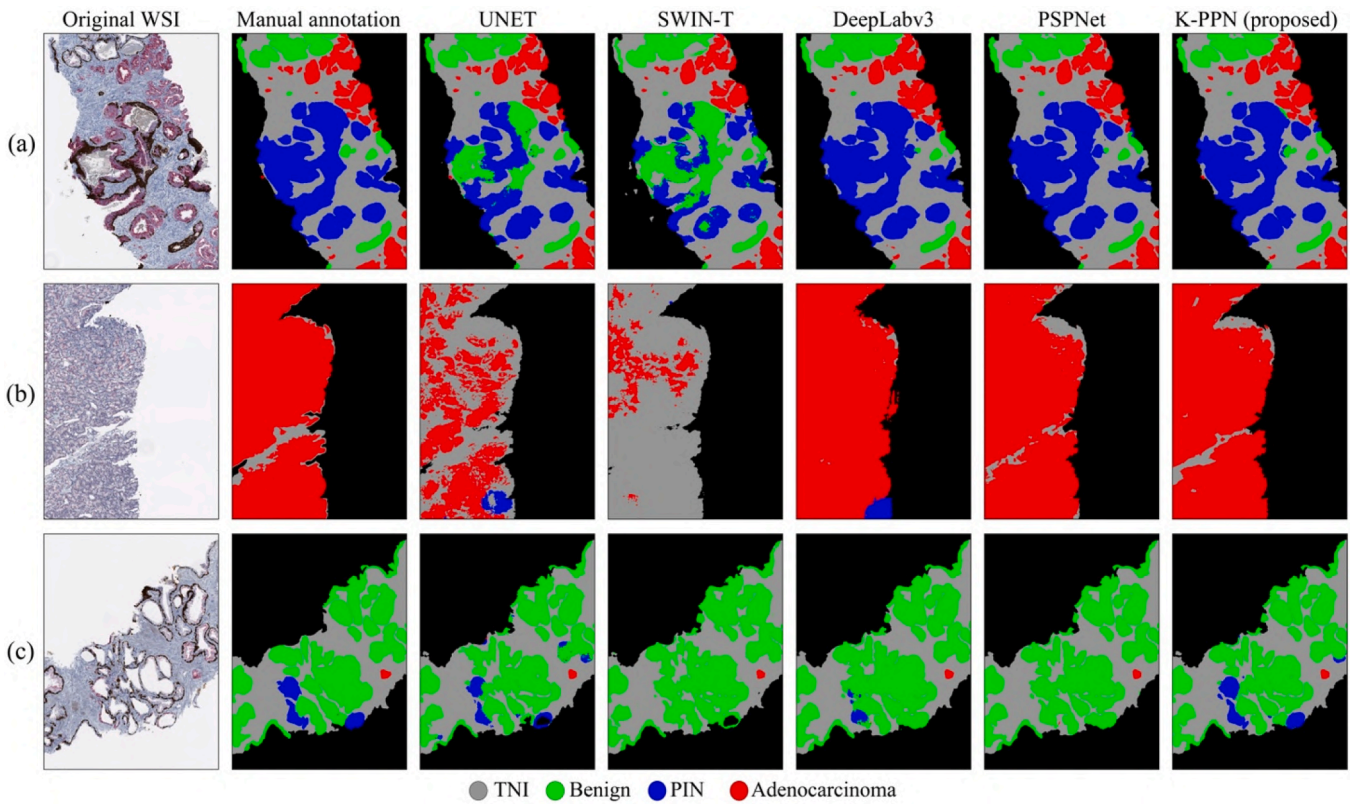


Fig. 7. Visual comparison between UNET, SWIN-T, DeepLabV3, PSPNet, and our proposed K-PPM models. Our network shows superior performance in accurately segmenting glandular structures compared to the other models.

selected cases.

In the present study, we propose a novel fully automatic approach for the accurate stratification and segmentation of prostate glands on IHC Whole-Slide-Images, which has the potential to improve the reliability of PCa diagnosis. Our new approach builds on a customized segmentation network, K-PPM, that effectively leverages the informative properties of IHC through adaptive kernels and multiscale feature integration. Additionally, to avoid common checkboard artifact, we incorporate a novel inference technique that enables smooth network prediction on any dimension of the input WSI, by weighted merging of consecutive patches.

Segmenting prostate cancer in IHC images poses two significant challenges, which we address through the following methodologies:

1. **Identification of Crucial Features:** One challenge is identifying crucial features that may be hidden in small regions while simultaneously classifying the entire gland. To tackle this issue, we propose a novel approach that integrates adaptive kernels and multiscale features. The adaptive kernels dynamically adjust their size and shape based on glandular structure characteristics, enabling them to focus on specific details critical for accurate classification. Meanwhile, the multiscale integration combines information from various scales, enhancing the model's robustness and reliability. This integration allows us to capture relevant features across the entire slide image, from large structures to minute details, significantly improving overall accuracy.
2. **Addressing Class Imbalance:** The second challenge arises from the infrequent occurrence of diagnostically important Prostatic Intraepithelial Neoplasia (PIN), leading to class imbalance. To overcome this issue, we employ two techniques: weighted adaptive patch extraction and specific-class kernel updates. The weighted patch extraction prioritizes underrepresented PIN regions during training, providing the classifier with adequate examples to learn from.

Additionally, the class-specific kernel updates refine decision boundaries, leading to better PIN identification despite limited cases.

In summary, our approach resolves key challenges in prostate cancer segmentation in IHC images by incorporating adaptive kernels for localized feature extraction, multiscale integration for enhanced robustness, weighted sampling to address class imbalance, and tailored kernel updates to improve minority class specificity.

The entire automatic method was validated using the Dice Score Coefficient (DSC) and Balanced Accuracy (BAL_{ACC}) to compare the segmentation predictions with manual masks drawn by an expert observer. Our system achieved an average DSC of 90.36 % and an average BAL_{ACC} of 94.24 % on the test set tiles, highlighting the potential of the proposed smooth reconstruction technique. Compared to the test set patch, the system's performance improved by 4.01 % and 3.29% for DSC and BAL_{ACC} , respectively. The results presented in Table 4 showcase the potential of our proposed method to accurately quantify the extent of individual classes on WSI, with highly consistent results compared to manual annotation. The network's ability to leverage information across scales and precisely segment small structures enables more accurate prostate cancer delineation. Thanks to a local and global feature extractor with adaptive kernel, our approach outperformed all other state-of-the-art methods tested (Zhao et al., 2017; Ronneberger et al., 2015; Szegedy et al., 2016; Liu et al., 2021). Our method accurately quantifies and identifies cribriform patterns, a critical histopathological feature of PCa that is linked to higher Gleason Scores and clinical aggressiveness (Chan and Nguyen, 2022). Furthermore, this model overcame the high-class imbalance present in the original dataset by integrating weighted adaptive patch extraction with specific-class kernel updates. This is a crucial point since the less-represented PIN class accounts for only 1.95 % of the entire annotations. In conclusion, our work introduces a reliable and precise multiclass segmentation system that reduces inter-observer variability and

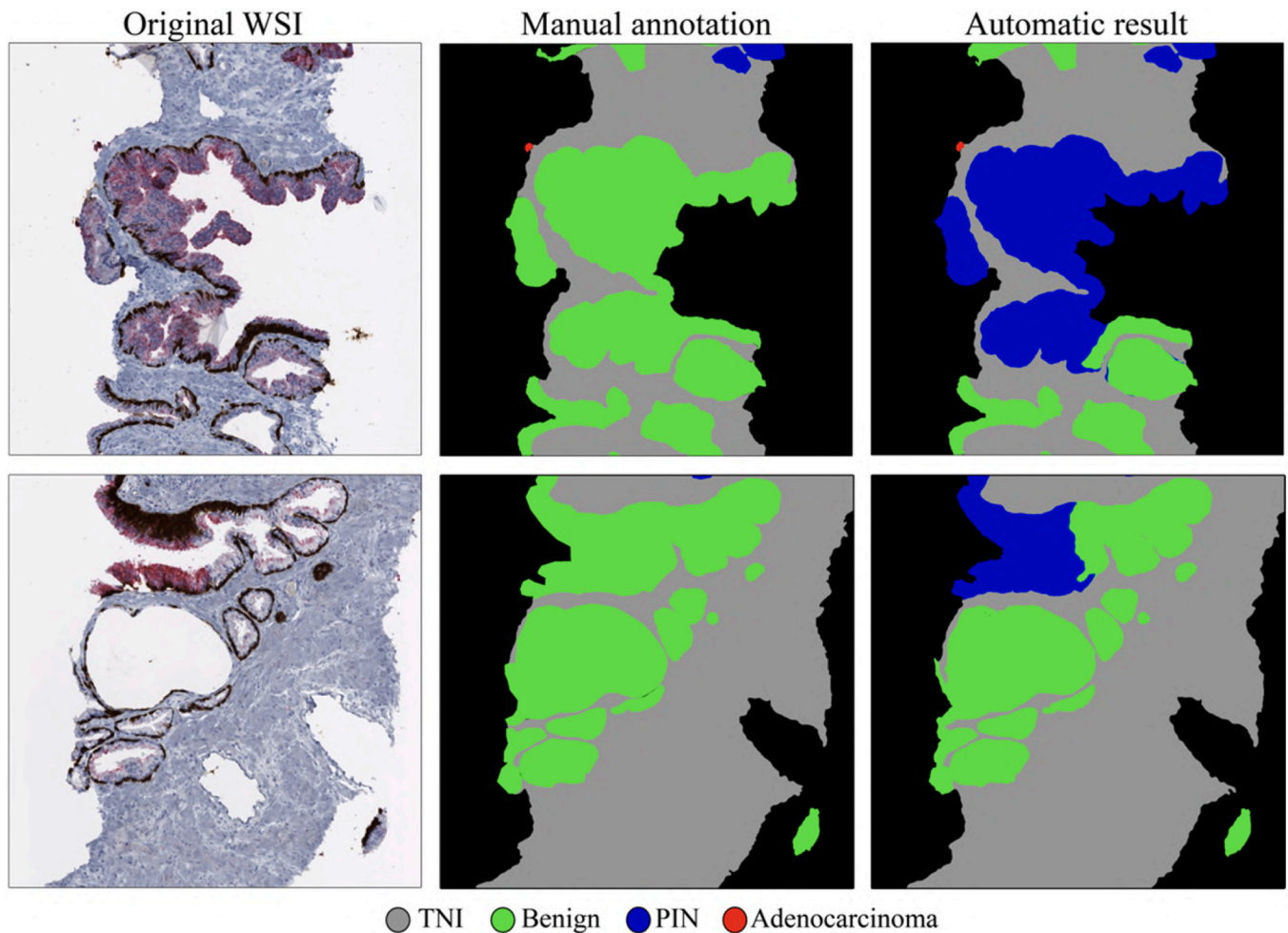


Fig. 8. Comparisons between manual and automatic mask in two cropped tiles. (a) our system accurately classified the glandular structure as PIN, which was previously misclassified by the manual operator; (b) our system misclassifies a large PIN gland, as only the left side presents the necessary requirements for classification as a PIN.

speeds up the PCA diagnostic process.

In this paper, we also introduce several technical novelties that can be applied to a range of cases beyond PCA diagnosis. Our first innovation is a new network, K-PPM, capable of enhancing the extraction of useful multiscale features for glandular stratification from IHC. Through adaptive kernels and class-specific enhancement of their weights, our method customizes the feature extraction according to the class to be segmented, increasing selectivity in classification. The potentiality of this network could be exploited for other segmentation task in digital pathology, such as for detection of lymph node metastases of breast cancer (Ehteshami Bejnordi et al., 2017) or lung cancer (Wang et al., 2018).

Secondly, we present a weighted-adaptive patch extraction method that has the potential to improve the performance of machine learning models trained on imbalanced datasets. This technique can be extended to a wide range of applications in digital pathology, particularly in the generation of datasets from large-scale images such as WSI (Ehteshami Bejnordi et al., 2017; Wang et al., 2018; Liu et al., 2020).

Finally, we introduce a smooth reconstruction technique that acts directly on softmax, which can be adapted to all segmentation or generative models. This method avoids the typical blocky effects associated with patch-based inference (Salvi et al., 2021b; Oskal et al., 2019; Tschuchnig et al., 2020), thus improving the interpretability and usability of machine learning models in several domains where the input image cannot be directly given to the models. Taken together, these innovations represent significant contributions to the field of digital

pathology in the automated diagnosis and have the potential to improve both the accuracy and efficiency of diagnostic processes for a wide range of diseases.

The limitations of our work are primarily related to the dataset at our disposal. Firstly, the dataset contains a limited number of biopsies from a small group of patients (i.e., 32 patients) and is sourced from only one center - the San Giovanni Bosco Hospital in Turin, Italy - which utilizes a Ventana DP 200 Slide Scanner (Roche Diagnostics) to digitize biopsy slides. Additionally, the dataset exhibits a poor representation of the PIN class and significant inter-observer variability, which adversely affects the performance of the network in stratification. These constraints limit the system's validation for reliability and generalizability in the clinical setting.

To overcome the limitations of the proposed method, future research should focus on increasing the sample size, particularly for the less represented classes. It would also be beneficial to employ a multi-center and multi-scanner dataset to enhance the model's generalization capabilities and to assess its robustness against various artifacts and color variations. Given the challenging nature of PIN classification, as reported in Leng et al. (2019) and Sabata et al. (2010), a multi-operator labeling phase could be explored to reduce inter-observer variability in the annotations of the PIN class, thus improving system performance. Moreover, we aim to develop a comprehensive framework that combines evaluations from the proposed system on IHC WSIs with observations from H&E staining. This integration of morphological and functional features could lead to more accurate diagnoses and highly

reliable tumor stratification.

5. Conclusions

This study presents a new AI-based segmentation framework that identifies prostate cancer in IHC WSIs. The proposed system utilizes the functional information of the IHC and introduces a customized network, K-PPM, with multiscale feature extraction and adaptive-dynamic kernels. Additionally, the system employs techniques such as smooth reconstruction of WSI, weighted-adaptive patch extraction for dataset creation, and class-specific kernel update during training to address class imbalance issues. The approach shows potential for improving the accuracy and efficiency of prostate cancer diagnosis, reducing inter-observer variability, thus aiding pathologists in their routine workload.

CRedit authorship contribution statement

Conceptualization: **M.S., and F.M.**; Data curation: **M.S., C.M., J.I.L. and D.F.**; Formal analysis: **M.S., C.M., J.I.L. and D.F.**; Methodology: **M.S. and D.F.**; Supervision: **F.M.**; Validation: **M.S.**; Visualization: **M.S. and D.F.**; Writing - original draft: **M.S.**; Writing - review & editing: **M.S., C.M., J.I.L., D.F. and F.M.**

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data and codes will be available under reasonable requests.

Acknowledgments

We would like to thank the laboratory technicians at San Giovanni Bosco Hospital for digitizing the slides.

References

- Anaya-Isaza, A., Mera-Jiménez, L., Zequera-Diaz, M., 2021. An overview of deep learning in medical imaging. *Inf. Med.* 26, 100723.
- Arvaniti, E., Fricker, K.S., Moret, M., Rupp, N., Hermanns, T., Fankhauser, C., Wey, N., Wild, P.J., Rueschoff, J.H., Claassen, M., 2018. Automated gleason grading of prostate cancer tissue microarrays via deep learning. *Sci. Rep.* 8, 1–11.
- N.C. Blessin, C. Yang, T. Mandelkoy, J.B. Raedler, W. Li, E. Bady, R. Simon, E. Vettorazzi, M. Lennartz, C. Bernreuther, Automated Ki-67 labeling index assessment in prostate cancer using artificial intelligence and multiplex fluorescence immunohistochemistry, *J Pathol.* (n.d.).
- Brawer, M.K., 2005. Prostatic intraepithelial neoplasia: an overview. *Rev. Urol.* 7, S11.
- Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M., 2010. The balanced accuracy and its posterior distribution. In: *Proceedings of the 2010 20th International Conference on Pattern Recognition, IEEE*. pp. 3121–3124.
- Bulten, W., Pinckaers, H., van Boven, H., Vink, R., de Bel, T., van Ginneken, B., van der Laak, J., de Kaa, C.H., Litjens, G., 1907. Automated gleason grading of prostate biopsies using deep learning. *ArXiv Preprint. ArXiv 07980*, 2019.
- Bulten, W., Bándi, P., Hoven, J., van de Loo, R., Lotz, J., Weiss, N., van der Laak, J., van Ginneken, B., Hulsbergen-van de Kaa, C., Litjens, G., 2019. Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard. *Sci. Rep.* 9, 864.
- Bulten, W., Balkenhol, M., Belinga, J.-J.A., Brilhante, A., Çakır, A., Egevad, L., Eklund, M., Farré, X., Geronatsiou, K., Molinié, V., 2021. Artificial intelligence assistance significantly improves gleason grading of prostate biopsies by pathologists. *Mod. Pathol.* 34, 660–671.
- Busby, D., Grauer, R., Pandav, K., Khosla, A., Jain, P., Menon, M., Haines III, G.K., Cordon-Cardo, C., Gorin, M.A., Tewari, A.K., 2023. Applications of artificial intelligence in prostate cancer histopathology. In: *Urologic Oncology: Seminars and Original Investigations*. Elsevier.
- Chan, E., Nguyen, J.K., 2022. Cribriform prostate cancer: an aggressive pattern where definition and size matter. *Diagn. Histopathol.*
- Chan, E., McKenney, J.K., Hawley, S., Corrigan, D., Auman, H., Newcomb, L.F., Boyer, H. D., Carroll, P.R., Cooperberg, M.R., Klein, E., 2022. Analysis of separate training and validation radical prostatectomy cohorts identifies 0.25 mm diameter as an optimal definition for “large” cribriform prostatic adenocarcinoma. *Mod. Pathol.* 35, 1092–1100.
- Chatrjian, A., Colling, R.T., Browning, L., Alham, N.K., Sirinukunwattana, K., Malacrino, S., Haghighat, M., Aberdeen, A., Monks, A., Moxley-Wyles, B., 2021. Artificial intelligence for advance requesting of immunohistochemistry in diagnostically uncertain prostate biopsies. *Mod. Pathol.* 34, 1780–1794.
- Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation, *ArXiv Preprint ArXiv:1706.05587*.
- Crum, W.R., Camara, O., Hill, D.L.G., 2006. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans. Med. Imaging* 25, 1451–1461.
- Dabir, P.D., Ottosen, P., Høyer, S., Hamilton-Dutoit, S., 2012. Comparative analysis of three-and two-antibody cocktails to AMACR and basal cell markers for the immunohistochemical diagnosis of prostate carcinoma. *Diagn. Pathol.* 7, 1–6.
- Deng, S., 2020. Deep learning in digital pathology image analysis: a survey. *Front. Med.* 14 (4), 470–487.
- Ehteshami Bejnordi, B., Veta, M., Diest, van, P.J., Ginneken, B., van, Karssemeijer, N., Litjens, G., et al., 2017. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 318, 2199–2210.
- Gao, Y., Kikinis, R., Bouix, S., Shenton, M., Tannenbaum, A., 2012. A 3D interactive multi-object segmentation tool using local robust statistics driven active contours. *Med. Image Anal.* 16, 1216–1227.
- Grandini, M., Bagli, E., Visani, G., 2008. Metrics for multi-class classification: an overview, *ArXiv Preprint. ArXiv 05756* (2020).
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: pp. 770–778.
- Hoffman, R.M., 2011. Screening for prostate cancer. *N. Engl. J. Med.* 365, 2013–2019.
- Isaksson, J., Arvidsson, I., Åström, K., Heyden, A., 2017. Semantic segmentation of microscopic images of H&E stained prostatic tissue using CNN. In: *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2017: pp. 1252–1256.
- Karantanos, T., Corn, P.G., Thompson, T.C., 2013. Prostate cancer progression after androgen deprivation therapy: mechanisms of castrate resistance and novel therapeutic approaches. *Oncogene* 32, 5501–5511.
- Kench, J.G., Amin, M.B., Berney, D.M., Compérat, E.M., Cree, I.A., Gill, A.J., Hartmann, A., Menon, S., Moch, H., Netto, G.J., 2022. WHO classification of tumours fifth edition: evolving issues in the classification, diagnosis, and prognostication of prostate cancer. *Histopathology* 81, 447–458.
- Khened, M., Kori, A., Rajkumar, H., Krishnamurthi, G., Srinivasan, B., 2021. A generalized deep learning framework for whole-slide image segmentation and analysis. *Sci. Rep.* 11, 1–14.
- Komura, D., Ishikawa, S., 2018. Machine learning methods for histopathological image analysis. *Comput. Struct. Biotechnol. J.* 16, 34–42.
- Leng, E., Henriksen, J.C., Rizzardi, A.E., Jin, J., Nam, J.W., Brassuer, B.M., Johnson, A. D., Reder, N.P., Koopmeiners, J.S., Schmechel, S.C., 2019a. Signature maps for automatic identification of prostate cancer from colorimetric analysis of H&E-and IHC-stained histopathological specimens. *Sci. Rep.* 9, 1–12.
- Li, C., Deng, M., Zhong, X., Ren, J., Chen, X., Chen, J., Xiao, F., Xu, H., 2023. Multi-view radiomics and deep learning modeling for prostate cancer detection based on multi-parametric MRI. *Front. Oncol.* 13, 1198899.
- Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G.E., Kohlberger, T., Boyko, A., Venugopalan, S., Timofeev, A., Nelson, P.Q., Corrado, G.S., 1703. Detecting cancer metastases on gigapixel pathology images, *ArXiv Preprint. ArXiv 02442* (2017).
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*: pp. 10012–10022.
- Magi-Galluzzi, C., 2018. Prostate cancer: diagnostic criteria and role of immunohistochemistry. *Mod. Pathol.* 31, 12–21.
- Michael, H.J., Aringhieri, G., Cioni, D., Neri, E., 2022. Current value of biparametric prostate MRI with machine-learning or deep-learning in the detection, grading, and characterization of prostate cancer: a systematic review. *Diagnostics* 12, 799.
- Molinié, V., Herve, J.M., Lugagne, P.M., Lebre, T., Botto, H., 2006. Diagnostic utility of a p63/α-methyl coenzyme A racemase (p504s) cocktail in ambiguous lesions of the prostate upon needle biopsy. *BJU Int.* 97, 1109–1115.
- Nakamura, W., Sumitomo, M., Zennami, K., Takenaka, M., Ichino, M., Takahara, K., Teramoto, A., Shiroki, R., 2023. Combination of deep learning and ensemble machine learning using intraoperative video images strongly predicts recovery of urinary continence after robot-assisted radical prostatectomy. *Cancer Rep.* e1861.
- Netzer, N., Eith, C., Bethge, O., Hielscher, T., Schwab, C., Stenzinger, A., Gnirs, R., Schlemmer, H.-P., Maier-Hein, K.H., Schimmöller, L., 2023. Application of a validated prostate MRI deep learning system to independent same-vendor multi-institutional data: demonstration of transferability. *Eur. Radiol.* 1–14.
- Oskal, K.R.J., Risdal, M., Janssen, E.A.M., Undersrud, E.S., Gulsrud, T.O., 2019. A U-net based approach to epidermal tissue segmentation in whole slide histopathological images. *SN. Appl. Sci.* 1, 1–12.
- Ozkan, T.A., Eruyar, A.T., Cebeci, O.O., Memik, O., Ozcan, L., Kusknmaz, I., 2016. Interobserver variability in Gleason histological grading of prostate cancer. *Scand. J. Urol.* 50, 420–424.
- Paner, G.P., Luthringer, D.J., Amin, M.B., 2008. Best practice in diagnostic immunohistochemistry: prostate carcinoma and its mimics in needle core biopsies. *Arch. Pathol. Lab Med.* 132, 1388–1396.
- Rao, H.-L., Chen, J.-W., Li, M., Xiao, Y.-B., Fu, J., Zeng, Y.-X., Cai, M.-Y., Xie, D., 2012. Increased intratumoral neutrophil in colorectal carcinomas correlates closely with malignant phenotype and predicts patients’ adverse prognosis. *PLOS One* 7, e30806.

- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 234–241.
- Sabata, B., Babenko, B., Monroe, R., Srinivas, C., 2010. Automated analysis of pin-4 stained prostate needle biopsies. In: *Proceedings of the International Workshop on Prostate Cancer Imaging. Computer-Aided Diagnosis, Prognosis, and Intervention, Held in Conjunction with MICCAI 2010, Beijing, China, September 24, 2010. Proceedings 1*, Springer, 2010: pp. 89–100.
- Salvi, M., Bosco, M., Molinaro, L., Gambella, A., Papotti, M., Acharya, U.R., Molinari, F., 2021a. A hybrid deep learning approach for gland segmentation in prostate histopathological images. *Artif. Intell. Med.* 115, 102076.
- Salvi, M., Mogetta, A., Gambella, A., Molinaro, L., Barreca, A., Papotti, M., Molinari, F., 2021b. Automated assessment of glomerulosclerosis and tubular atrophy using deep learning. *Comput. Med. Imaging Graph.* 90, 101930.
- Salvi, M., Caputo, A., Balmativola, D., Scotto, M., Pennisi, O., Michielli, N., Mogetta, A., Molinari, F., Fraggetta, F., 2023. Impact of stain normalization on pathologist assessment of prostate cancer: a comparative study. *Cancers* 15, 1503.
- Sardanelli, F., Castiglioni, I., Colarieti, A., Schiaffino, S., Di Leo, G., 2023. Artificial intelligence (AI) in biomedical research: discussion on authors' declaration of AI in their articles title. *Eur. Radiol. Exp.* 7, 2.
- Singhal, N., Soni, S., Bonthu, S., Chattopadhyay, N., Samanta, P., Joshi, U., Jojera, A., Chharchhodawala, T., Agarwal, A., Desai, M., 2022. A deep learning system for prostate cancer diagnosis and grading in whole slide images of core needle biopsies. *Sci. Rep.* 12, 3383.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826.
- Tătaru, O.S., Vartolomei, M.D., Rassweiler, J.J., Virgil, O., Lucarelli, G., Porpiglia, F., Amparore, D., Manfredi, M., Carrieri, G., Falagario, U., 2021. Artificial intelligence and machine learning in prostate cancer patient management—current trends and future perspectives. *Diagnostics* 11, 354.
- Tran, K.A., Kondrashova, O., Bradley, A., Williams, E.D., Pearson, J.V., Waddell, N., 2021. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med.* 13, 1–17.
- Tschuchnig, M.E., Oostingh, G.J., Gadermayr, M., 2020. Generative adversarial networks in digital pathology: a survey on trends and future potential. *Patterns* 1, 100089.
- van Leenders, G.J.L.H., Verhoef, E.I., Hollemans, E., 2020. Prostate cancer growth patterns beyond the Gleason score: entering a new era of comprehensive tumour grading. *Histopathology* 77, 850–861.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30.
- Wang, S., Chen, A., Yang, L., Cai, L., Xie, Y., Fujimoto, J., Gazdar, A., Xiao, G., 2018. Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome. *Sci. Rep.* 8, 10393.
- Wang, H., Zhu, Y., Adam, H., Yuille, A., Chen, L.-C., 2021. Max-deeplab: end-to-end panoptic segmentation with mask transformers. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5463–5474.
- Watson, K., Wang, C., Yilmaz, A., Bismar, T.A., Trpkov, K., 2013. Use of immunohistochemistry in routine workup of prostate needle biopsies: a tertiary academic institution experience. *Arch. Pathol. Lab. Med.* 137, 541–545.
- Zhang, W., Pang, J., Chen, K., Loy, C.C., 2021. K-net: Towards unified image segmentation. *Adv. Neural Inf. Process. Syst.* 34, 10326–10338.
- Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2881–2890.
- Zhdanovich, Y., Ackermann, J., Wild, P.J., Köllermann, J., Bankov, K., Döring, C., Flinner, N., Reis, H., Wenzel, M., Höb, B., 2023. Evaluation of automatic discrimination between benign and malignant prostate tissue in the era of high precision digital pathology. *BMC Bioinform.* 24, 1–14.