

## Article

# Automated Hybrid Model for Detecting Perineural Invasion in the Histology of Colorectal Cancer

Jiyoon Jung <sup>1</sup>, Eunsu Kim <sup>2</sup>, Hyeseong Lee <sup>2</sup>, Sung Hak Lee <sup>2,\*</sup> and Sangjeong Ahn <sup>3,\*</sup>

<sup>1</sup> Department of Pathology, Kangnam Sacred Heart Hospital, College of Medicine, Hallym University, 1, Singil-ro, Yeongdeungpo-gu, Seoul 07441, Korea

<sup>2</sup> Department of Hospital Pathology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 222 Banpodae-ro, Seocho-gu, Seoul 06591, Korea

<sup>3</sup> Department of Pathology, Korea University Anam Hospital, College of Medicine, Korea University, 73 Incheon-ro, Seonbuk-gu, Seoul 02841, Korea

\* Correspondence: hakjang@catholic.ac.kr (S.H.L.); vanitasahn@gmail.com (S.A.); Tel.: +82-2-2258-1617 (S.H.L.); +82-2-920-5590 (S.A.)

**Abstract:** Perineural invasion (PNI) is a well-established independent prognostic factor for poor outcomes in colorectal cancer (CRC). However, PNI detection in CRC is a cumbersome and time-consuming process, with low inter- and intra-rater agreement. In this study, a deep-learning-based approach was proposed for detecting PNI using histopathological images. We collected 530 regions of histology from 77 whole-slide images (PNI, 100 regions; non-PNI, 430 regions) for training. The proposed hybrid model consists of two components: a segmentation network for tumor and nerve tissues, and a PNI classifier. Unlike a “black-box” model that is unable to account for errors, the proposed approach enables false predictions to be explained and addressed. We presented a high performance, automated PNI detector, with the area under the curve (AUC) for the receiver operating characteristic (ROC) curve of 0.92. Thus, the potential for the use of deep neural networks in PNI screening was proved, and a possible alternative to conventional methods for the pathologic diagnosis of CRC was provided.

**Keywords:** colorectal cancer; perineural invasion; semantic segmentation; deep learning; computational pathology



**Citation:** Jung, J.; Kim, E.; Lee, H.; Lee, S.H.; Ahn, S. Automated Hybrid Model for Detecting Perineural Invasion in the Histology of Colorectal Cancer. *Appl. Sci.* **2022**, *12*, 9159. <https://doi.org/10.3390/app12189159>

Academic Editors: Qi-Huang Zheng and Mohib Ullah

Received: 7 June 2022

Accepted: 7 September 2022

Published: 13 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Perineural invasion (PNI) in colorectal cancer (CRC) is a well-established independent prognostic factor [1,2], with an incidence range from 9% to 30% [3,4]. PNI is defined as tumor invasion into, around, and through neural structures [5]; it is the distinct route through which cancer cells spread and metastasize to adjacent or distant organs [6,7]. PNI detection is associated with response to adjuvant chemotherapy [8]. Therefore, a meticulous evaluation of PNI and prognostication on the basis of the standardized pathology report are mandatory in routine pathology practice [9,10].

Despite its importance, the histologic evaluation of PNI is a cumbersome and time-consuming process, with a high risk of misdiagnosis [11]. Peng et al. [12] reported that only 7.5% of patients were PNI-positive in original pathologic reports; however, reviewing their PNI status revealed that 24.3% of patients were PNI positive. Thorough histological inspections are necessary for reducing the misdiagnosis rate. However, pathologic diagnosis of PNI is tedious and increases the workload of pathologists. Given the increasing workload for pathologists and their critical shortage nationally and globally [13,14], developing an automated screening tool for PNI is crucial.

Deep learning (DL) methods have achieved promising results in medical image analysis [15,16], surpassing human performance [17]. In computational pathology, histology-based DL approaches have facilitated computer-aided diagnostics, including tumor detection, classification, segmentation, and even quantification of established biomarkers,

such as tumor-infiltrating lymphocytes [18–20]. Until now, limited studies have focused on PNI detection in computational pathology [21,22]. However, as with DL-based approaches, these studies exhibited the drawback of failing to interpret unpredictable failures. The lack of interpretability in “black-box” modeling limits real-world application, which may lead to user distrust [23,24].

In this study, an interpretable DL-based PNI detector was developed through CRC histology, demonstrating the potential of computer-aided diagnosis for PNI screening. The proposed approach could become an alternative to the conventional methods of pathologic diagnosis of CRC.

## 2. Materials and Methods

### 2.1. Data Acquisition

A total of 77 whole-slide images (WSIs) of 63 patients with CRC who underwent surgical resection at International St. Mary’s Hospital, Catholic Kwandong University in Incheon Metropolitan City, Republic of Korea, were selected for the study. The specimens were formalin-fixed, paraffin-embedded, and stained with hematoxylin and eosin. An Aperio AT2 slide scanner (Leica Biosystems, Buffalo Grove, IL, USA) was used to scan the WSIs at 40× magnification. From the 77 WSIs, 530 regions were selected, including PNI with tumor and nerve tissue, non-PNI with tumor, non-PNI with nerve tissue, and normal tissue excluding neural tissue, respectively, denoted as “PNI”, “tumor”, “nerve”, and “normal”, (Table 1). All PNIs inside the region were annotated; non-PNI with tumor tissues, and non-PNI with nerve tissues were randomly extracted and annotated. Tumor, nerve, and normal tissues were annotated by board-certified pathologists (J.J. and S.A.) using the automated slide analysis platform (ASAP). All results in this study have undergone two rounds of reviews. Inconsistencies were discussed with another board-certified pathologist (S.H.L.). A total of 490 regions were used for training and validation, and the remaining 40 were used to test the model.

**Table 1.** Composition of regions and patches.

		No. of Regions	No. of Patches
	PNI	100	362
Non-PNI	Nerve	204	687
	Tumor	207	7547
	Normal	19	880
Total		530	9476

### 2.2. Patch Generation

We used a half-overlap sliding window algorithm for model input. This method can overcome the loss of information between adjacent areas. Moreover, the summation of the probability of adjacent regions, which comes from deep learning models, can increase overall accuracy. For the strategy of patch generation, we employed 1.0 mpp (micron per pixel) for resolution and  $512 \times 512 \times 3$  pixels for size. Patch prediction or extraction is skipped when the mean pixel value of the target patch is too high (>235) or too low (<50), because a high mean of the pixel value is mostly due to the background, and a low mean of the pixel value is mostly due to the low quality of the whole slide image.

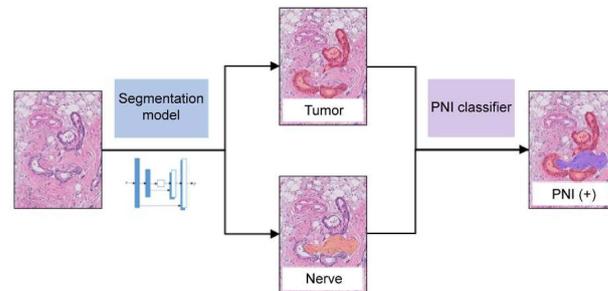
### 2.3. Image Preprocessing

Each patch generated from the same WSI follows a similar color distribution, but patches from different WSIs may not. To overcome the difference of color distribution among WSIs, we employ color augmentation, such as HSV shift and random brightness. To obtain a sufficient geometric pattern of patches, geometric augmentations, such as elastic transformation, shift scale rotation, random rotation, horizontal flipping, and vertical flipping, were used. All Augmentations were implemented using the Albumentation

open-source library <https://github.com/albumentations-team/albumentations> (accessed on 20 October 2021) [25]. Scaling (0–1) was employed for data normalization.

#### 2.4. Segmentation Network Development

A general scheme of the proposed model is displayed in Figure 1. A hybrid model was proposed to detect PNI using histology, comprising a semantic segmentation network and a rule-based PNI classifier.



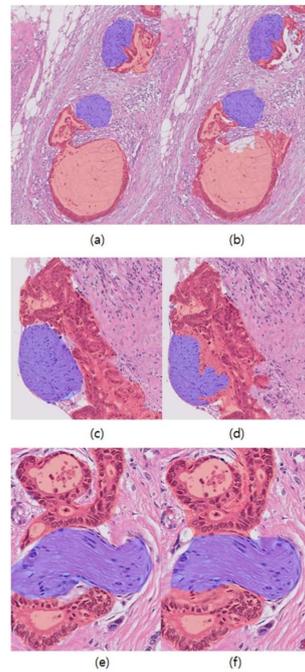
**Figure 1.** Proposed pipeline for PNI detection. The framework consists of a segmentation network and a PNI classifier. Tumor (red) and nerve masks (orange) were extracted according to the segmentation model. The extracted nerve areas were classified as PNI when they were close to the tumor.

A multiclass semantic segmentation network was trained to detect tumors and nerves. As an alternative approach, experiments were conducted using binary segmentation models for tumors and nerves. For the segmentation frameworks, we used the U-Net [26], Deeplabv3+ [27], and SegFormer [28] networks. U-Net, Deeplabv3+, and SegFormer were built on a pre-trained backbone. For an ablation study, SegFormer was trained from scratch, comparing the performance of a transformer-based segmentation network with transfer learning. To train the U-Net, three backbone models, namely Inception-Resnet-v2, EfficientNet-B0, and SE-ResNeXt-101, were used. For the training of the Deeplabv3+, two models—MobileNet and Xception—were used as backbones. A pre-trained model using the ImageNet database was used [29]. For the training of SegFormer, MiT-B0 (Mix Transformer Encoder) was used. An adaptive moment estimation optimizer was used, with an initial learning rate of  $10 \times 10^{-3}$ . The batch size for training was set to 32 and the maximum number of epochs was set to 200. In the multi-class segmentation network, a multi-loss function, calculated as a weighted sum of the dice loss and categorical focal loss ( $L_{\text{Multi}} = L_{\text{Dice}} + L_{\text{Focal}}$ ) was used. To train the binary segmentation network, a combination of dice loss and binary cross-entropy loss ( $L_{\text{Binary}} = L_{\text{Dice}} + L_{\text{CE}}$ ) was used.

#### 2.5. PNI Classifier

To generate tumor and nerve masks, which were input into the PNI classifier, we implemented six combinations of segmentation networks for tumors and nerves, denoted as Module1 (Md1) to Module6 (Md6). Md1 to Md4 employed binary segmentation networks for tumors and nerves, whereas Md5 and Md6 used multiclass segmentation networks. In the Md1 framework, U-Net was used for both nerve and tumor segmentation. For Md2, U-Net and DeepLabv3+ were used for nerve and tumor segmentation, respectively. In Md3, DeepLabv3+ and U-Net were utilized for nerve and tumor segmentation, respectively. In Md4, DeepLabv3+ was used for both nerve and tumor segmentation. In Md5 and Md6, U-Net and SegFormer were employed, respectively.

Based on the trained segmentation network, binary probability maps were inferred for tumors and nerves using a half-overlapped sliding window. The average probability was applied for the overlapped window. Tiny areas predicted as tumors or nerves (probability threshold = 0.5) were removed using morphological analysis. Nerves with PNI were extracted according to a rule-based approach, where the distances between the binary map of the tumor and the dilated nerve were calculated (Figure 2). PNI and non-PNI groups were defined as follows:



**Figure 2.** Three representative regions of extracted PNI with ground truths (a,c,e) and the corresponding pixel-wise predictions (b,d,f). Based on the spatial arrangement of the tumors (red) and the nerves (purple), a nerve close to a tumor was classified as PNI.

$$\text{PNI: } Area_{dilate\ nerve} \cap Area_{tumor} \neq \emptyset$$

$$\text{Non-PNI: } Area_{dilate\ nerve} \cap Area_{tumor} = \emptyset$$

$Area_{tumor}$  and  $Area_{dilate\ nerve}$  denote the binary map of the tumor and the dilated nerve, respectively.

## 2.6. Evaluation Metrics

The performance of the trained segmentation model using pixel-wise accuracy, intersection over union (IoU), sensitivity, precision, and the F1-score was compared. All metrics are defined as follows:

$$\text{Accuracy} = \sum p | p = g, \forall p \in P, g \in G$$

$$\text{IoU} = \frac{G \cap P}{G \cup P}$$

$$\text{Sensitivity} = \sum_i \frac{\sum(p | p=i, g=i, \forall p \in P, g \in G)}{\sum(p | p=i, g=i, \forall p \in P, g \in G) + \sum(p | p \neq i, g=i, \forall p \in P, g \in G)} \forall i=0,1,2$$

$$\text{Precision} = \sum_i \frac{\sum(p | p=i, g=i, \forall p \in P, g \in G)}{\sum(p | p=i, g=i, \forall p \in P, g \in G) + \sum(p | p=i, g \neq i, \forall p \in P, g \in G)} \forall i=0,1,2$$

$$\text{F1 - score : } F_1(\text{precision, recall}) = \frac{2 \text{ precision recall}}{\text{precision} + \text{recall}}$$

P and G denote predicted value and ground truth, respectively.

To compare the performance of the PNI classifier, we used region-wise Accuracy<sub>R</sub>, Sensitivity<sub>R</sub>, Specificity<sub>R</sub>, Precision<sub>R</sub>, Negative Predictive Value<sub>R</sub> (NPV<sub>R</sub>), F1-score<sub>R</sub>, and the area under the curve (AUC). The metrics used to evaluate the region-wise performance are defined as follows:

$$\text{Accuracy}_R = \frac{\sum(TP + TN)}{\sum(TP + FN + FP + TN)}$$

$$\text{Sensitivity}_R = \frac{\sum(\text{TP})}{\sum(\text{TP} + \text{FN})}$$

$$\text{Specificity}_R = \frac{\sum(\text{TN})}{\sum(\text{FP} + \text{TN})}$$

$$\text{Precision}_R = \frac{\sum(\text{TP})}{\sum(\text{TP} + \text{FP})}$$

$$\text{NPV}_R = \frac{\sum(\text{TN})}{\sum(\text{FN} + \text{TN})}$$

$$\text{F1 - score}_R : \text{F1}(\text{precision, recall}) = \frac{2 \text{ precision recall}}{\text{precision} + \text{recall}}$$

True positive, true negative, false positive, and false negative are denoted as TP, TN, FP, and FN, respectively.

Instead of a stochastic model, a rule-based model was designed. Thus, we used a simple receiver operating characteristic (ROC) curve, which indicates that the classifier predicts PNI as positive when the distance between the tumor and the nerve is zero, and the classifier predicts PNI as negative when the distance is infinite. The confidence interval was calculated by assuming that the distribution of the AUC is similar to that of the accuracy, which is a binomial distribution of sample length and probability.

### 2.7. Inference Timing

The inference times required in U-Net and SegFormer were measured by averaging the five execution times for a randomly selected patch.

## 3. Results

### 3.1. Results of Segmentation Networks

Table 2 details the performance of the algorithms for the trained segmentation models. For identifying nerves, the U-Net-based binary segmentation model exhibited excellent performance, with an IoU of 0.887. For identifying tumors, the DeepLabv3+ binary segmentation model outperformed other models, with an IoU of 0.769. The overall performance of the multiclass semantic segmentation models was lower than that of the binary semantic segmentation models.

**Table 2.** Performance of the segmentation models.

		Accuracy	IoU	Sensitivity	Precision	F1-Score
Nerve	U-Net <sup>a</sup>	0.987	0.887	0.943	0.937	0.940
	DeepLabv3+ <sup>a</sup>	0.985	0.837	0.892	0.931	0.911
	U-Net (m) <sup>b</sup>	0.893	0.801	0.867	0.924	0.891
	SegFormer (m) <sup>b</sup>	0.921	0.829	0.921	0.893	0.907
Tumor	U-Net <sup>a</sup>	0.900	0.676	0.887	0.740	0.805
	DeepLabv3+ <sup>a</sup>	0.922	0.769	0.903	0.839	0.869
	U-Net (m) <sup>b</sup>	0.893	0.611	0.856	0.681	0.757
	SegFormer (m) <sup>b</sup>	0.838	0.686	0.838	0.791	0.814

<sup>a</sup> Binary semantic segmentation; <sup>b</sup> Multi-class semantic segmentation.

### 3.2. Region-Wise Performance

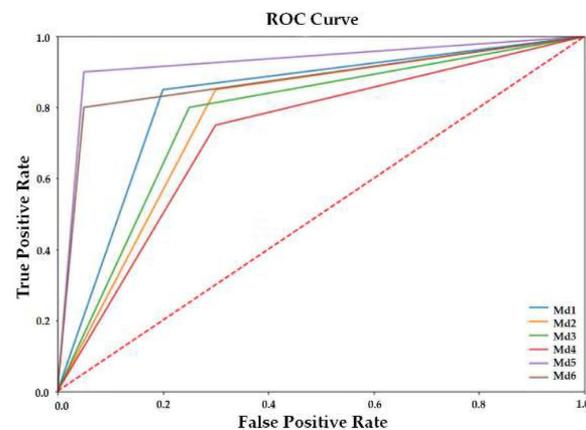
By inputting the tumor and nerve masks into the segmentation models, we extracted PNIs based on the distance between tumors and nerves (Figure 2). The pipeline using the multiple segmentation model, Md5, exhibited excellent performance, with an AUC of 0.92 (Table 3, Figure 3). Moreover, the standard deviation was the lowest among the models,

indicating that the multiple segmentation model was more stable than the combined binary segmentation model.

**Table 3.** Performance of PNI classifier according to the various combinations of segmentation models.

Module <sup>a</sup>	Accuracy <sub>R</sub>	Sensitivity <sub>R</sub>	Specificity <sub>R</sub>	NPV <sub>R</sub> <sup>b</sup>	Precision <sub>R</sub>	F1-Score <sub>R</sub>	AUC (95% CI)
Md1	0.85	0.85	0.85	0.85	0.85	0.85	0.85 ± 0.111
Md2	0.80	0.85	0.75	0.83	0.77	0.81	0.80 ± 0.124
Md3	0.80	0.75	0.85	0.77	0.83	0.79	0.80 ± 0.124
Md4	0.72	0.75	0.70	0.74	0.71	0.73	0.72 ± 0.138
Md5	0.92	0.90	0.95	0.90	0.95	0.92	0.92 ± 0.078
Md6	0.88	0.80	0.95	0.83	0.94	0.87	0.88 ± 0.102

<sup>a</sup> Architectures used for nerve and tumor segmentation in each sequential binary model are as follows: Md1: U-Net + U-Net; Md2: U-Net + DeepLabv3+; Md3: DeepLabv3+ + U-Net; Md4: DeepLabv3+ + DeepLabv3+. In Md5, U-Net is adopted for simple multiple segmentation; Md6: SegFormer is adopted for simple multiple segmentation. <sup>b</sup> Negative Predictive Value.



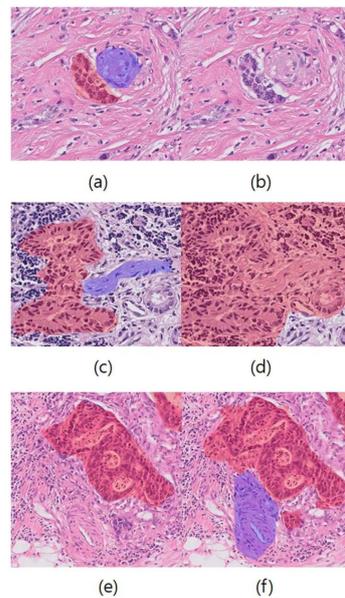
**Figure 3.** ROC curves of pipelines used for detecting PNI across various segmentation models. The results revealed that Md5, using a multi-class segmentation network, achieved the highest AUC of 0.92.

We found that the pipeline using the multiclass segmentation network showed better performance, although the binary segmentation network outperformed the simple multi-class segmentation model regarding pixel-wise performance for tumors and nerves. The combined networks are assumed to cause error accumulation and degrade performance.

### 3.3. Analysis of False Results

Falsely predicted images using to Md5 are presented in Figure 4. One of the FN regions exhibited relatively small tumor clusters and neural bundles (Figure 4b). In another region, the surrounding inflammatory cells and nerve cells were misclassified as tumor cells (Figure 4d). An FP region included thick blood vessels around a tumor, which were misclassified as PNI. The model falsely identified the smooth muscle cells of the wall of vessels as nerve bundle Schwann cells because of their similar spindle shapes. All the predicted results regarding Md1 to Md6 can be accessed through the web page (<http://pni.ssus.work/>, accessed on 13 September 2022).

In the current pipeline, falsely predicted results can be classified into six subgroups according to the types of tissue in error. FN prediction occurred in tumor, nerve, or both tissues (Figure S2a,c,e). FP originated from errors in tumor or nerve tissues (Figure S2b,d). The FP result was not observed in either the tumor or nerve tissues of our model. This classification allowed us to intuitively distinguish the tasks the model performed incorrectly, providing the interpretability of false results in the current model.



**Figure 4.** Examples of misclassification with ground truths (a,c,e) and the corresponding prediction (b,d,f). (b) In the false negative (FN) case, both tumor and nerve tissues were missed. (d) In the other FN case, nerve tissue surrounded by tumor tissue was predicted as tumor tissue. (f) In the FP case, nerve tissue was falsely predicted as a blood vessel. All the inference results from each pipeline are found on the webpage (<http://pni.ssus.work/>, accessed on 13 September 2022).

### 3.4. Effects of Pre-Training Tasks

We studied the impact of the pre-trained model on the performance of SegFormer (Table S2). The overall performance for identifying tumors increased when the pre-trained model weights were used. For identifying nerves, the F1 score was also improved. With transfer learning, improved performance was obtained.

### 3.5. Inference Time Comparison

Figure S3b reports the average inference time per patch for each architecture. SegFormer inference achieves the average of 126.754 ms, compared to the time of 1616.037 ms using the U-Net model. The number of parameters in SegFormer is about that in U-Net (3,714,915 vs. 6,251,759 for SegFormer and U-Net, respectively) (Figure S3a).

## 4. Discussion

In this study, a DL-based hybrid model was developed to detect PNI in CRC patients. The proposed framework exhibited excellent performance (accuracy of 0.92, sensitivity of 0.90, AUC of 0.92), with potential for computer-aided diagnosis in PNI screening. Considering the prognostic implications of PNI and the difficulty in detecting PNI in pathology slide images, the automated PNI detector exhibited a high potential for utility, and can potentially save medical resources.

In practice, PNI detection is a time-consuming and cumbersome task, with high intra- and inter-observer variation. A review of CRC slides in a study revealed that 46 of 55 PNI-positive cases (from a total of 249 cases) were reported as PNI-negative in the original pathology reports [2]. In another study, Peng et al. also revealed the differences in the PNI-positive rate (24.3%) after review compared to the PNI-positive rate (7.5%) recorded in the initial reports [12]. Furthermore, considerable variation is perceived between observers in defining PNI [11,30]. Some of the variations depend on the evaluation criteria among pathologists in terms of the distance between the cancer source and nerve cells [6]. Uncertainty in defining a nerve is another cause of poor inter-observer PNI detection reproducibility. These difficulties in detecting PNI can be solved by standardizing the evaluation criteria in the algorithms and refining the pixel-wise prediction

of neural bundles. Thus, expectedly, inter-observer reproducibility would increase, and underestimations could be reduced.

Some attempts have been made to use DL-based approaches to extract PNI histologically. Ström et al. used a convolutional neural network to classify PNI in prostate biopsies, and achieved a discrimination AUC of 0.98 [21]. Recently, an international Medical Image Computing and Computer Assisted Intervention Society-Pathology Artificial Intelligence Platform (MICCAI-PAIP) challenge was held to detect PNI in multiple organ cancers (<https://paip2021.grand-challenge.org>, accessed on 13 September 2022). The top-ranked team achieved the best F1 score of 41.55% using a feature pyramid network (FPN) [22]. This result revealed the capacity of the multi-resolution network, FPN, to detect PNI in histological images. However, these algorithms, where the representative images of PNI itself were learned, could not provide sufficient interpretation for false prediction because of the “black-box” nature of the DL methods.

An interpretable DL-based model was proposed to provide interpretability through semantic classification of the tissue type and the calculation of distances. This process sequence is similar to a diagnosis by a pathologist, enabling us to interpret the predicted results outputted by the proposed DL-based model, increasing model reliability. Thus, the method provided considerable advantages for medical image analysis and applying the models in practice [31,32].

Numerous challenges exist while using the DL algorithm for clinical applications, including its use in the current model to detect PNI in colon cancer. Therefore, these challenges should be resolved. First, an automated and efficient workflow using digitalized pathology images should be established to effectively use the DL model in practice. This high-cost digital pathology system does not exhibit any benefits, hindering most institutions from establishing this system [33]. Therefore, definite clinical values, such as reduced diagnostic time and improved quality and efficiency, should be achieved in DL implementation to incentivize DL adoption in clinical applications.

Second, safety should be guaranteed in the clinical application of the DL model. The robustness and generalizability of the trained DL model are critical for clinical application. Centralized digitalized data archives, which store large-scale biomedical images from various institutions, can be used to overcome this obstacle. The stored digital images can be used for model validation and bias optimization, allowing the algorithms to achieve generalized performance. Digital pathology guidelines for DL implementation and quality assessment have been established [34–37]. However, the digital pathology system is typically being established in large-scale university hospitals because the capital outlay for the system cannot be borne by every hospital. Schömig-Markiefka et al. investigated the accuracy of a deep-learning-based algorithm for datasets from various institutions digitized by different scanner systems. Although a model with high overall accuracy of >98% was used, substantial losses occurred in accuracy because of dependence on HE-staining quality, brightness, and contrast [38]. Therefore, national planning and systemic support for developing large-scale centralized biomedical image archives or databases is crucial for future clinical applications.

This study had some limitations. First, the dataset was limited in size. We used 77 WSIs to train the semantic segmentation network for extracting tumors and nerves. However, the proposed model exhibited performance comparable to that of a prior study, where 80k biopsy cores were used to achieve an AUC of 0.98 [21]. Considering the dataset size used in this study and the performance achieved, extremely large-scale data may not be required for the convergence of the proposed pipeline. Second, the results obtained were not externally validated. External validation with additional datasets can improve interpretability and generalizability. Finally, the classifier using distance calculation exhibits a drawback. Since the PNI classifier is a rule-based classifier, distinguishing whether tumor cells actually infiltrated a nerve sheath or are just adjacent to a nerve bundle is not possible. Despite these limitations, the PNI classifier can play a significant role as a screening tool.

Therefore, as a trial of PNI detection, this study provides researchers with new possibilities for the development and improvement of data-driven algorithms for PNI detection. Furthermore, by enabling the detection of accurate PNI status, this study can improve the clinical decisions made for individual patients, positively impacting their prognosis.

## 5. Conclusions

A novel DL-based approach was proposed to detect PNI in CRC using histopathological images. The hybrid model consists of two components: a segmentation network for tumor and nerve tissues, and a PNI classifier. The proposed framework exhibited high performance (with an accuracy of 0.92, a sensitivity of 0.90, and an AUC of 0.92), as well as the potential for computer-aided diagnosis in PNI screening. Considering the prognostic implication of PNI and the difficulty in detecting it in pathology slide images, the automated PNI detector exhibits significant potential for PNI diagnosis.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app12189159/s1>, Figure S1: Confusion matrix for the five proposed models; Figure S2: Representative FN and FP cases; Figure S3: Comparison of the number of parameters and inference time for SegFormer and U-Net; Table S1: Clinicopathological characteristics of colorectal cancer patients; Table S2: Performance for SegFormer trained on scratch and on pre-trained weights.

**Author Contributions:** Conceptualization, S.H.L. and S.A.; methodology, E.K. and H.L.; formal analysis, E.K.; resources, J.J. and S.A.; data curation, E.K.; writing—original draft preparation, J.J.; writing—review and editing, S.A. and S.H.L.; supervision, S.H.L. and S.A.; project administration, S.H.L. and S.A.; funding acquisition, S.H.L. and S.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded in part by research grants from the National Research Foundation (NRF) of Korea (grant number: NRF-2021R111A3043875) and the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea (grant number: HI21C0940).

**Institutional Review Board Statement:** The study was conducted in accordance with the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of International St. Mary's Hospital (IS21SIME0031).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support the findings of this study are available upon request from the corresponding author. The data are not publicly available because of privacy and ethical restrictions.

**Acknowledgments:** We participated in the international MICCAI-PAIP 2021 challenge and developed algorithms based on this competition.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Knijn, N.; Mogk, S.C.; Teerenstra, S.; Simmer, F.; Nagtegaal, I.D. Perineural Invasion is a Strong Prognostic Factor in Colorectal Cancer: A systematic review. *Am. J. Surg. Pathol.* **2016**, *40*, 103–112. [[CrossRef](#)] [[PubMed](#)]
2. Liebig, C.; Ayala, G.; Wilks, J.; Verstovsek, G.; Liu, H.; Agarwal, N.; Berger, D.H.; Albo, D. Perineural invasion is an independent predictor of outcome in colorectal cancer. *J. Clin. Oncol.* **2009**, *27*, 5131–5137. [[CrossRef](#)] [[PubMed](#)]
3. Tsai, H.L.; Cheng, K.I.; Lu, C.Y.; Kuo, C.H.; Ma, C.J.; Wu, J.Y.; Chai, C.Y.; Hsieh, J.S.; Wang, J.Y. Prognostic significance of depth of invasion, vascular invasion and numbers of lymph node retrievals in combination for patients with stage II colorectal cancer undergoing radical resection. *J. Surg. Oncol.* **2007**, *97*, 383–387. [[CrossRef](#)] [[PubMed](#)]
4. Hu, G.; Li, L.; Hu, K. Clinical implications of perineural invasion in patients with colorectal cancer. *Medicine* **2020**, *99*, e19860. [[CrossRef](#)]
5. Batsakis, J.G. Nerves and neurotropic carcinomas. *Ann. Otol. Rhinol. Laryngol.* **1985**, *94*, 426–427.
6. Liebig, C.; Ayala, G.; Wilks, J.A.; Berger, D.H.; Albo, D. Perineural invasion in cancer: A review of the literature. *Cancer* **2009**, *115*, 3379–3391. [[CrossRef](#)]

7. Marchesi, F.; Piemonti, L.; Mantovani, A.; Allavena, P. Molecular mechanisms of perineural invasion, a forgotten pathway of dissemination and metastasis. *Cytokine Growth Factor Rev.* **2010**, *21*, 77–82. [\[CrossRef\]](#)
8. Sun, Q.; Liu, T.; Liu, P.; Luo, J.; Zhang, N.; Lu, K.; Ju, H.; Zhu, Y.; Wu, W.; Zhang, L.; et al. Perineural and lymphovascular invasion predicts for poor prognosis in locally advanced rectal cancer after neoadjuvant chemoradiotherapy and surgery. *J. Cancer* **2019**, *10*, 2243–2249. [\[CrossRef\]](#)
9. Kim, B.H.; Kim, J.M.; Kang, G.H.; Chang, H.J.; Kang, D.W.; Kim, J.H.; Bae, J.M.; Seo, A.N.; Park, H.S.; Kang, Y.K.; et al. Standardized Pathology Report for Colorectal Cancer, 2nd Edition. *J. Pathol. Transl. Med.* **2020**, *54*, 1–19. [\[CrossRef\]](#)
10. Compton, C.; Fenoglio-Preiser, C.M.; Pettigrew, N.; Fielding, L.P. American Joint Committee on Cancer prognostic factors consensus conference: Colorectal Working Group. *Cancer* **2000**, *88*, 1739–1757. [\[CrossRef\]](#)
11. Chi, A.C.; Katabi, N.; Chen, H.S.; Cheng, Y.S.L. Interobserver Variation Among Pathologists in Evaluating Perineural Invasion for Oral Squamous Cell Carcinoma. *Head Neck Pathol.* **2016**, *10*, 451–464. [\[CrossRef\]](#)
12. Peng, J.; Sheng, W.; Huang, D.; Venook, A.P.; Xu, Y.; Guan, Z.; Cai, S. Perineural invasion in pT3N0 rectal cancer: The incidence and its prognostic effect. *Cancer* **2011**, *117*, 1415–1421. [\[CrossRef\]](#)
13. Bonert, M.; Zafar, U.; Maung, R.; El-Shinnawy, I.; Kak, I.; Cutz, J.C.; Naqvi, A.; Juergens, R.A.; Finley, C.; Salama, S.; et al. Evolution of anatomic pathology workload from 2011 to 2019 assessed in a regional hospital laboratory via 574,093 pathology reports. *PLoS ONE* **2021**, *16*, e0253876. [\[CrossRef\]](#)
14. Metter, D.M.; Colgan, T.J.; Leung, S.T.; Timmons, C.F.; Park, J.Y. Trends in the US and Canadian Pathologist Workforces from 2007 to 2017. *JAMA Netw. Open* **2019**, *2*, e194337. [\[CrossRef\]](#)
15. Bodalal, Z.; Trebeschi, S.; Beets-Tan, R. Radiomics: A critical step towards integrated healthcare. *Insights Imaging* **2018**, *9*, 911–914. [\[CrossRef\]](#)
16. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **2016**, *278*, 563–577. [\[CrossRef\]](#)
17. Kaiming, H.; Xiangyu, Z.; Shaoqing, R.; Jian, S. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
18. Echle, A.; Rindtorff, N.T.; Brinker, T.J.; Luedde, T.; Pearson, A.T.; Kather, J.N. Deep learning in cancer pathology: A new generation of clinical biomarkers. *Br. J. Cancer* **2020**, *124*, 686–696. [\[CrossRef\]](#)
19. Lu, M.Y.; Chen, T.Y.; Williamson, D.F., K.; Zhao, M.; Shady, M.; Lipkova, J.; Mahmood, F. AI-based pathology predicts origins for cancers of unknown primary. *Nature* **2021**, *594*, 106–110. [\[CrossRef\]](#)
20. Van der Laak, J.; Litjens, G.; Ciompi, F. Deep learning in histopathology: The path to the clinic. *Nat. Med.* **2021**, *27*, 775–784. [\[CrossRef\]](#)
21. Kartasalo, K.; Ström, P.; Ruusuvaari, P.; Samaratunga, H.; Delahunt, B.; Tsuzuki, T.; Eklund, M.; Egevad, L. Detection of perineural invasion in prostate needle biopsies with deep neural networks. *Virchows Arch.* **2022**, *481*, 73–82. [\[CrossRef\]](#)
22. Nateghi, R.; Pourakpour, F. Perineural invasion detection in multiple organ cancer based on deep convolutional neural network. *arXiv* **2021**, arXiv:2110.12283.
23. Ching, T.; Himmelstein, D.S.; Beaulieu-Jones, B.K.; Kalinin, A.A.; Do, B.T.; Way, G.P.; Ferrero, E.; Agapow, P.M.; Zietz, M.; Hoffman, M.M.; et al. Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* **2018**, *15*, 20170387. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Madabhushi, A.; Lee, G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med. Image Anal.* **2016**, *33*, 170–175. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albumentations: Fast and flexible image augmentations. *Information* **2020**, *11*, 125. [\[CrossRef\]](#)
26. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), Munich, Germany, 18–22 September 2015; Navab, N., Hornegger, J., Wells, W., Frangi, A., Eds.; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241.
27. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [\[CrossRef\]](#)
28. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12077–12090.
29. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Kai, L.; Li, F.-F. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
30. Egevad, L.; Delahunt, B.; Samaratunga, H.; Tsuzuki, T.; Olsson, H.; Ström, P.; Lindskog, C.; Häkkinen, T.; Kartasalo, K.; Eklund, M.; et al. Interobserver reproducibility of perineural invasion of prostatic adenocarcinoma in needle biopsies. *Virchows Arch.* **2021**, *478*, 1109–1116. [\[CrossRef\]](#)
31. Samek, W.; Wiegand, T.; Müller, K.R. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv* **2017**, arXiv:1708.08296.
32. Zhang, Y.; Weng, Y.; Lund, J. Applications of Explainable Artificial Intelligence in Diagnosis and Surgery. *Diagnostics* **2022**, *12*, 237. [\[CrossRef\]](#)

33. Ahmad, Z.; Rahim, S.; Zubair, M.; Abdul-Ghafar, J. Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: Present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review. *Diagn. Pathol.* **2021**, *16*, 1–16. [[CrossRef](#)]
34. Pantanowitz, L.; Sinard, J.H.; Henricks, W.H.; Fatheree, B.L.A.; Carter, A.B.; Contis, L.; Beckwith, B.A.; Evans, A.J.; Lal, A.; Parwani, A.V. Validating Whole Slide Imaging for Diagnostic Purposes in Pathology: Guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch. Pathol. Lab. Med.* **2013**, *137*, 1710–1722. [[CrossRef](#)]
35. Chong, Y.; Kim, D.C.; Jung, C.K.; Kim, D.C.; Song, S.Y.; Joo, H.J.; Yi, S.Y.; Medical Informatics Study Group of the Korean Society of Pathologists. Recommendations for pathologic practice using digital pathology: Consensus report of the Korean Society of Pathologists. *J. Pathol. Transl. Med.* **2020**, *54*, 437–452. [[CrossRef](#)]
36. Federal Association of German Pathologists Bundesverband Deutscher Pathologen (FAGP-BDP). *Guidelines Digital Pathology for Diagnosis on (And Reports of) Digital Images*; Federal Association of German Pathologists Bundesverband Deutscher Pathologen (FAGP-BDP): Berlin, Germany, 2018.
37. Digital Pathology Assessment Committee. *Technical Standards for Digital Pathology System for Pathologic Diagnosis*; Japanese Society of Pathology: Tokyo, Japan, 2015.
38. Schömig-Markiefka, B.; Pryalukhin, A.; Hulla, W.; Bychkov, A.; Fukuoka, J.; Madabhushi, A.; Achter, V.; Nieroda, L.; Büttner, R.; Quaas, A.; et al. Quality control stress test for deep learning-based diagnostic model in digital pathology. *Mod. Pathol.* **2021**, *34*, 2098–2108. [[CrossRef](#)]