Perineural Invasion Detection Using Deep Learning: A Survey

Hayagreev Jeyandran

Vidyashilp Academy, Bengaluru, India.

Contributing authors: hayagreevj@gmail.com;

Abstract

Perineural Invasion is the infiltration of tumor cells into the perineurium which surrounds peripheral nervous tissue. It is a finding commonly associated with pancreatic, rectal, head and neck, and prostate gland cancer. It is a sign of poor prognosis and has been related to recurrence and metastasis, increasing its prognostic value. However, identifying Perineural Invasion from Whole Slide Images is a laborious task for trained pathologists. Deep Learning has been appreciated for its contributions to medical imaging for many years now, and it can also assist in screening pathology samples for Perineural Invasion. This paper thoroughly surveys the use of Deep Learning for the assistive diagnosis of Perineural Invasion, covering various aspects of existing literature (such as dataset availability, pre-processing techniques, segmentation, detection, and classification models, evaluation metrics) and providing avenues for future work.

Keywords: Perineural Invasion, Deep Learning, Medical Imaging, Digital Pathology

1 Introduction

Perineural Invasion (PNI) was first identified as the invasion of neoplastic cancer cells into, around, and through nerves [1]. However, it has a more precise definition in relation to the peripheral nerve sheath.

1.1 Structure of the Peripheral Nerve Sheath

The layers are listed in order, from outer-most to inner-most [2].

- Epineurium: Dense elastin and collagen connective tissue. It surrounds the entire nerve and contains its vascular supply.

- Perineurium: Endothelial cells with basal laminae that tightly fit axons and form a selectively permeable membrane.
- Endoneurium: An uninterrupted sheath that contains single nerve fibres myelinated by Schawnn cells. It is made of delicate connective tissue running uninterrupted from the brain to the spinal cord. This layer also contains endoneurial fluid, a major component of the blood-nerve barrier.

PNI is the presence of tumor in any of these layers. Many pathologists classify the encircling of 33% of a nerve by tumor cells as PNI. Nevertheless, presence within the layers might be more clinically relevant [3]. There have been a multitude of earlier theories governing PNI, namely that it is an extension of lymphatic metastasis [4] and that the nerve sheath provides a conduit of low resistance for tumor spread [5]. However, both these theories were disproven subsequently. The current hypothesis is that the host and cancer share a mutual relationship in fuelling the extent of PNI [6]; the nerve micro-environment (macrophages, fibroblasts, Schwann cells), molecular signalling (such as neutrotrophic factors, extracellular matrices, chemokines, axonal guidance) might have a role to play. This is still under investigation even today [6], leaving us with a lack of targeted therapies for PNI.

1.2 Occurrence and Prognostic significance

Notwithstanding the molecular mechanisms of PNI still remaining largely unknown, there is plenty of evidence from various types of cancers collected over a long period of time to suggest that PNI is of prognostic significance, contributing to recurrence. For example, Zhang et al. evaluated the association between PNI and biochemical recurrence in cases of prostate cancer [7]. The prognostic role of PNI in surgically treated esophageal squamous cell carcinoma [8], gastric cancer [9] [10] and its impact on the 5-year survival of stage 1 - 3 rectal cancer patients [11] have been examined previously. Even for trained pathologists, detecting PNI from Whole-Slide Images (WSIs) is a laborious task because of the following reasons.

1. **Microscopic Structures** - Nerves on WSIs, especially small peripheral ones, are difficult to distinguish from benign glands, blood vessels, and muscle fibers.
2. **Morphology** - Tumor can invade nerves in various ways. The visual appearance of PNI varies depending on histology and organ type.
3. **Sparsity** - PNI can occur in isolated regions which are sparse and scattered throughout large WSIs, making them difficult to detect.

These challenges result in inter-observer and intra-observer variability while manually examining slides for PNI [12]. In a study on squamous cell carcinoma of the vulva, the authors stipulate that 27% of examined cases were reported as PNI absent while PNI was actually present [13]. Similarly, in a study involving colorectal cancer, the number of patients diagnosed as PNI positive increased from 7.5% to 24.3% following meticulous review by pathologists [14]. The intestinal wall is a dense structure whose complexity can easily overshadow subtle infiltration into the nerves. Interestingly, variability in diagnosis of oral squamous cell carcinoma was attributed to a lack of clarity on what constituted PNI in a study by Yan et al [15]. All of these lapses point us

toward the need for assistive technology to at least aid pathologists by automatically flagging a WSI in case of suspected PNI, so that the pathologist can undertake careful review further on.

1.3 Deep Learning in Medical Imaging

Deep Learning (DL), a subset of Artificial Intelligence, has revolutionized the field of medical imaging by enabling computers to automatically analyze complex patterns in medical data (CT, MRI, X-Rays, and Pathology Images) with remarkable accuracy. DL systems are capable of automatically extracting relevant features without the need for manual feature engineering [16]. Several studies over the years have tuned DL implementations to support diagnostic functionality. For example, a Convolutional Neural Network was used by Nirschl J.J et al. to detect signs of clinical heart failure from Hematoxylin and Eosin (H&E) stained WSIs [17]. Other advances include two-stage glomerular segmentation [18] and multi-instance gastric image segmentation [19]. DL has also been applied to PNI diagnosis before, although in a very limited number of studies. The following sections will thoroughly examine these studies. In section 2, the literature is introduced and grouped into various categories. Following this, the dataset and pre-processing methods are elucidated. Then, the neural network models, training strategies, post-processing techniques are explained. Following this, evaluation metrics for the models and error analysis of the algorithm's performance are examined. Lastly, future research scope is suggested before concluding.

2 Summary of Approaches

| Classification | Object Detection | Image Segmentation |
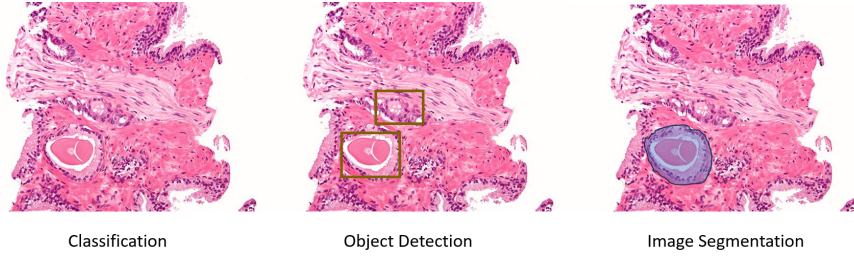
Fig. 1: DL Approaches: Courtesy of Wikimedia Foundation - CC-BY-NC2.

Most computer vision problems can be largely classified as either **classification**, **object detection**, or **image segmentation** [20]. In classification, the model only outputs probabilities for the entire image/input. In object detection, the model outputs precisely the bounding box co-ordinates of features within the image, such as tumor and nerve. Object detection can be combined with classification, where in the bounding-box contains one of multiple objects. Lastly, segmentation is when the model outputs pixel-wise probabilities for the entire image, producing a segmentation *mask*. This is a binary mask made from thresholding the pixel wise probabilities (eg:

regions containing PNI) and can be likened to shading relevant regions in the image. This system of classification is also maintained in the context of PNI detection. Most approaches directly utilize an end-to-end model and segment either the boundary between the tumor and nerve (known as the PNI Junction) or the PNI field (region) from the WSIs. The other approach is to use a sequential pipeline, where-in segmentation and object detection are used for tumor and nerve fields respectively. Then, a rule-based classifier identifies PNI based on the proximity or overlap of the outputs of the two stages. The first approach is less complex and requires less annotations, whereas the second approach is more sophisticated and allows for greater analysis due to the orthogonalization used.

Table 1: Reviewed Literature

Literature	Summary
Automated Hybrid Model for Detecting Perineural Invasion in the Histology of Colorectal Cancer [21]	Multi-Class Tumor and Nerve field Segmentation + rule based PNI classifier
Perineural Invasion Detection in Multiple Organ Cancer Based on Deep Convolutional Neural Network [22]	PNI Junction Segmentation
Efficient Perineural Invasion Detection of Histopathological Images Using U-Net (Multi-Organ Cancer) [23]	PNI Junction Segmentation
Perineural invasion detection in pancreatic ductal adenocarcinoma using artificial intelligence [24]	Hierarchical Contextual Analysis, Binary Tumor and Nerve field segmentation + rule based PNI classifier
Detection of perineural invasion in prostate needle biopsies with deep neural networks [25]	Binary PNI field segmentation
A multi-task deep learning framework for perineural invasion recognition in gastric cancer whole slide images [26]	Binary Tumor field segmentation and Nerve object detection + rule based PNI classifier

3 Dataset

In the modern DL era, visual training tasks such as traditional image classifiers and macroscopic object detectors use large corpora of 1 - 10 million training images to mitigate over fitting and learn nuanced features [27]. Working with a similarly large number of whole WSIs for PNI detection is infeasible due to the following reasons:

1. **Computational Resources** - WSIs are high-resolution images, often with gigapixel sizes. Processing them like traditional images would require extensive storage and powerful GPUs or TPUs, which are usually not accesible in a medical research environment.
2. **Annotation and Labelling** - Annotating WSIs requires the expertise of trained pathologists, who are limited in number. Precise annotations are time-consuming and expensive for a large dataset (¿ 1000 images) even with a large group of dedicated pathologists.
3. **Dimensionality** - WSIs show high morphological diversity. Each feature can be regarded as a separate pixel.

Thus, most works utilize a limited pool of WSIs and have a more unconventional train/validation/test split, with a larger proportion being used for the latter two (eg: 60/20/20) to achieve statistical significance and robust accuracy metrics. Moreover, these reasons also underscore the importance of pre-processing to make the data learnable [28].

3.1 PAIP 2021 - Grand challenge

The Pathology in AI Platform (PAIP) hosted a challenge in 2021: Perineural Invasion in Multiple Organ Cancer [29]. Authors of [22], [23] have taken part in this competition and utilized the dataset provided. The dataset consisted of **150 Training WSIs** (50 each of Colon, Prostate, and Pancreatic-Billiary duct cancer), **30 Validation WSIs**, **60 Testing WSIs** each scanned at **20X Magnification** by an **Aperio AT2 scanner** (Leica Biosystems, Nussloch, Germany). The raw resolution of each WSI was **60000 x 60000**. One-pixel thick binary masks annotating the PNI junction were also provided for all 240 WSIs.

3.2 STHLM3 - Prostate Biopsy cores

Between 2012 and 2014, a study was conducted in Stockholm county, Sweden, in which 59159 men were tested for high prostate-specific antigen (PSA) levels [30]. 7406 of these men were referred to a **10 - 12 core trans-rectal ultrasound guided biopsy**, resulting in **83,470 prostate cores**. The authors of [25] over sampled high grade cancers suitable for training a neural network to identify features of nervous and neoplastic tissue and randomized 80% of all resulting cores into a train set and the remaining 20% into a test set. In all, the **train set** included **1141 subjects and 7045 cores, including 379 PNI positive cores** and the **test set** included **286 subjects and 1758 cores, including 196 PNI positive cores**.

3.3 Hospital-specific datasets

- In [21], **77 WSIs** from **63 subjects** with **colorectal cancer** were collected from St. Mary's Hospital (Catholic Kwandong University, Incheon, Korea). Pathologists manually selected **530 regions of interest (ROI)** from the 77 WSIs. 4 classes were used - PNI, Non-PNI with nerve, Non-PNI with Tumor, and Normal, before slides were pre-processsed into patches. Each patch was assigned to one of four classes.
- In [26], 143 train, 12 validation, and 60 test WSIs from **gastric cancer** patients were gathered at **20x magnification** in Fujian Cancer Hospital, China.
- In [24], **260 training** (206 randomly selected, 54 manually assigned by pathologists due to abundant tumor and nerve features), **168 validation and testing** WSIs from patients with pancreatic ductal adenoma were gathered at **40x magnification** in Sourasky Medical Center, Tel-Aviv.

4 Pre-processing

4.1 Patching

To solve the problem of high dimensionality, most DL approaches include patching (tiling) as a mandatory pre-processing step [31]. Models are trained to make low-level predictions on each patch, and either max or average pooling is used over the entire slide to make higher level predictions.

4.1.1 Non-overlapping patches

In this approach, patches are given a fixed resolution of $p \times p$. A $p \times p$ window is slid over the entire $n_H \times n_W$ image with a stride of p in both the horizontal and vertical directions. If $n_H \nmid p$ or $n_W \nmid p$, there will be left-over regions that do not fully fit the window. This can be dealt with in two ways: either by applying a padding of $p - (n_H \mod p) \times p - (n_W \mod p)$ or by considering the smaller, remaining regions at the edge of the image as another patch, resulting in $\left\lceil \frac{n_H}{p} \right\rceil \times \left\lceil \frac{n_W}{p} \right\rceil$ patches. The first approach is preferred as it maintains consistency of input dimension into the model. Non-overlapping patches are used in [25] where p is specified as $0.25mm$ and [22] where $p = 512$px. This may result in loss of information from the boundary regions, which is dealt with using dilation techniques later on the pre-processing pipeline.

4.1.2 Overlapping Patches

In this second approach, the stride is set to $p(1 - \text{overlap proportion})$. Information is preserved in the boundary regions leading to more robust feature extraction at the cost of duplication. [26], [23], [21] all use this strategy with p ranging from 512 to 2048 and a overlap proportion of 0.50. In [26], the patches are cropped at equal intervals to minimize duplication.

4.2 Boundary Dilation

In approaches [22], [23] (which use PNI junction segmentation), morphological dilation is applied to enlarge the 1 pixel thick PNI-nerve boundary lines. By enhancing clarity in the region, the model can learn more complex transitional features between normal nerve tissue and PNI. Morphological dilation involves a structuring element of a particular radius (eg: [22] uses a disk of radius 3px). The operator is slid over the boundary, and the maximum value of the pixels within the operator is casted onto the remaining pixels within the operator.

4.3 Data Augmentation

Data augmentation is a commonly used pre-processing technique and decreases vulnerability to over fitting. It is important in medical imaging applications because of limited availability of training data and lesser margin for error during real-world implementation [33]. All works for PNI detection use some form of data augmentation, such as colour augmentation (HSV), brightness and contrast adjustment, geometric
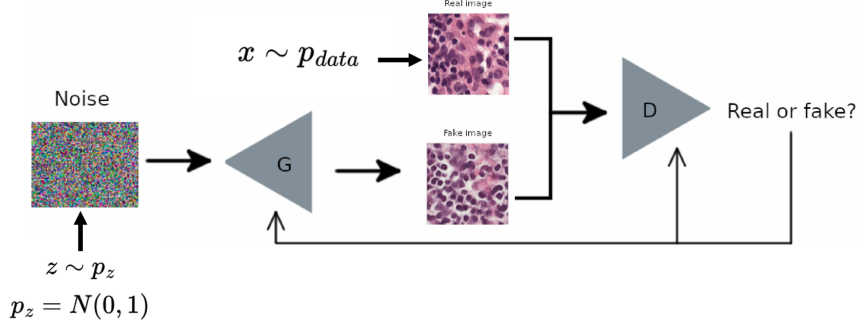
Fig. 2: GANs for WSIs [32]

augmentation (rotation, flip, shear) at the individual patch-level. In [24], synthetic data generation is utilized using **generative adversarial networks (GANs)** [34]. GANs consist of a generator G that produces synthetic images and a discriminator D) that distinguishes between real and synthetic images, optimizing through a minimax game to generate data that mimics the real distribution. These synthetic generation and augmentation methods can be seamlessly integrated into neural network pipelines using frameworks like TensorFlow [35] and PyTorch [36], enhancing model robustness against overfitting.

4.4 Background Image Elimination

Works [21] and [26] examine pixel intensities to infer the usability of patches. For example, [26] considers $\sigma_{pixelIntensity} < 0.6$ to be an indicator of a background image with no useful pathological features. [21] mandates that $39 < \mu_{pixelIntensity} < 235$ for all patches. Anything higher is considered background, while any patches below the range are regarded as low quality and discarded. σ and μ are the standard deviation and mean respectively.

5 Models and Training

5.1 Segmentation Models

5.1.1 U-Net

Proposed in [38] by Ronneberger et al., U-Net has been a seminal integration into the field of medical imaging. All works except [22] make use of the basic U-Net architecture, some of them modifying part of it. Its distinctive feature is the symmetric structure, consisting of an **encoder** and a **decoder** connected by skip connections. Here are the functions of each part of the architecture:

- **Encoder**: Takes as input a WSI patch of resolution n_H, n_W, n_C. The encoder follows a contracting path, wherein repeated 3×3 3D valid Convolutions are used to down sample n_H, n_W while increasing n_C. Max pooling operations with a window size of
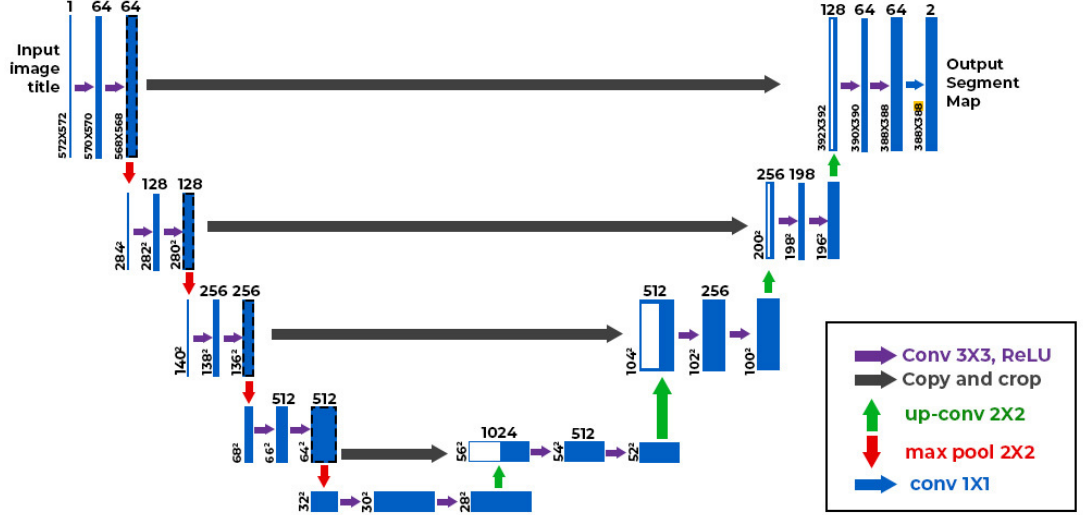
Fig. 3: UNet Architecture [37]

2×2 are also applied after every two convolution operations. The encoder allows the network to capture higher level features at multiple scales.

- **Bottleneck**: Deepest layer in the network. Images with the smallest spatial resolution and most abstract feature representations are processed using 3×3 convolutions.
- **Decoder**: Takes the higher-level features and restores their spatial resolution for precise segmentation of tumor/nerve regions. This is done through up sampling operations, such as the 2×2 transposed convolutions used in the original paper. Towards the end, a network in a network (1×1 convolution) [39] is applied to produce C channels, where C is the number of output classes. In the case of tumor/nerve binary segmentation, C = 2. If multiple segmentation is used, C = 3. The spatial dimensions are also resized to match the original patch dimensions.
- **Copy and Crop**: These are skip connections connecting each stage in the encoder to the corresponding stage in the decoder. The feature maps from the encoder are cropped to match that of the decoder and concatenated with it. Since the encoder has more information about the position of objects in the original image, whereas the decoder contains the abstraction of features, combining them produces a rich, accurate segmentation map.

[21], [23], [24], [25] use U-Net in some part of their architecture. U-Net is also very flexible and can facilitate learning paradigms like transfer learning [40], which is very useful given that most medical imaging tasks have a limited training data pool that is better suited to fine-tuning than training from scratch. [21] and [23] replace the U-Net encoder with different architectures, such as Inception-Resnet-v2 [41] and EfficientNetB0 through B4 [42]. There are three main reasons for this:

8

1. **Pre-training**: The original authors pre-train their networks on large, open-source databases, giving the network a head-start. This improves segmentation performance upon fine-tuning with ground truth segmentation maps from a limited training data corpus, and it is highly relevant for PNI detection as seen in 3.
2. **Complexity**: The UNet encoder is relatively shallow in comparison with Inception-Net and EfficientNet. This allows for capturing more intricate patterns and learning more complex hierarchies. EfficientNet is also more computationally efficient.
3. **Residual Learning**: Skip connections are key to network robustness and training given their depth, allowing the neural network to learn the identity function and preventing exploding or vanishing gradients.

One other strategy involves ensemble models, as seen in [25]. Pixel-wise probabilities are averaged across a set of U-Net models (soft-voting), improving accuracy and stability, and reducing sensitivities to anomalies. These increase the model's robustness against blurry sections of the WSI, for instance, while inferencing. [24] uses a fully unsupervised approach initially, implementing U-Net to approximate regions containing PNI in conjunctions with decision processes inspired by pathologist insight. This can be regarded as a pre-training step.

5.1.2 Other models

[22] uses Feature Pyramid Networks (FPNs) [43], which were originally designed for object detection. FPNs have a bottom-up pathway that produces feature maps at progressively lower resolutions (resembling a pyramid shape) using standard convolutions and pooling, followed by a top-down pathway where the feature maps are up-sampled and combined with the bottom-up pathway to learn features of the PNI junction at multiple scales. This allows the network to capture local information (precise junction boundary) while also considering broader context information. [26] uses U-Net++, which is a deeply supervised and more complex U-Net architecture [44] It introduces nested skip connections with dense-layers, allowing for features in skip pathways to be re-used as they are merged with the decoder's feature maps. The introduction of supervision in the hidden layers also improves gradient flow, stabilizing training and improving model convergence.

5.2 Object Detection Models

To detect nerve bounding boxes, [26] utilizes the Detection Transformer (DETR). It employs a Transformer architecture where object detection is framed as a sequence prediction problem. Self-attention is applied to compute the relationships between image features and object queries. An encoder takes in image features (usually an outcome of a backbone network, like ResNet101) and encodes them into keys $(\mathbf{K})$ and values $(\mathbf{V})$. A set number of queries (that can be likened to anchor boxes in traditional object detection models) are framed. The decoder inputs the queries, and uses $\mathbf{K}$ and $\mathbf{V}$ to produce pairs of bounding box co-ordinates and class labels. A special class label, ϕ, is used to denote a background bounding box. This is useful because the number of queries is much larger than the number of nerves in the patch, and excess queries are returned as bounding boxes.
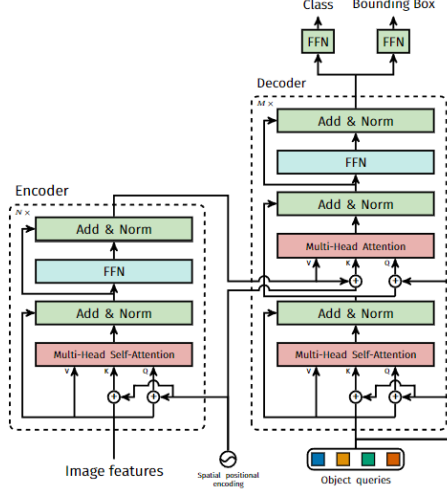
9

Fig. 4: Detection Transformer Model [45]

5.3 Classifiers

Upon obtaining segmented patches, they are post-processed (removing small, noisy predictions) and recombined into the WSI. After this, depending on the approach listed in 2, different methods are used to make either WSI-level or patient level predictions.

- **Junction segmentation**: [22] makes a simple prediction by taking the largest probability per patch and aggregating it across all WSIs to produce the image probability, which is then used for WSI-level classification. [23] is more sophisticated. It first finds matching pairs of predicted and ground truth lines by drawing bounding boxes at 150% scale around all instances within a WSI and looking for intersection between the enlarged predicted and ground truth boxes. Then, the average minimum distances between lines in matching pairs and the intersection over union (IoU) of their masks are used to calculate a single metric, $Dist\_Score$, whose value is compared with a fixed threshold to determine whether the prediction is a false positive, true positive, or false negative.
- **IoB and Area of Segmented Fields**: [21] counts $Area_{nervemap} \bigcap Area_{tumormap} \neq 0$ as a PNI. [26] is more sophisticated and uses the intersection over bounding box metric and predicts PNI positive only if $IoB > 0.12$ rather than flagging any intersection as positive. $IoB = \frac{A_{\text{nerveBox}} \bigcap A_{\text{tumorMap}}}{A_{\text{nerveBox}} \bigcup A_{\text{tumorMap}}}$
- **End-to-End Neural network classification**: [25] uses XCeption, a modified Inception Network with depth-wise separable convolutions within Inception blocks, simplifying model workflow [46]. Global max-pooling layers are added to the end of the standard XCeption network, followed by fully connected layers leading up to a sigmoid activation function. Max pooling is used over patches to assign WSI (core) and subject probabilities for PNI.

10

5.4 Loss Functions

Standard loss functions are used for all tasks:
Binary Cross Entropy $= -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right]$
Dice Loss $= 1 - \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}}$
Bipartite Loss $= \sum_{i=1}^{N} \max \left(0, \text{Margin} - \text{Sim}(\hat{y}_i, y_i) \right)$

Binary cross entropy measures the direct deviation between predicted probabilities and the ground truth PNI labels. Dice Loss penalizes overlap between the predicted and ground truth masks, with a lower Dice coefficient indicating more deviation. The version of the dice loss above is relevant only to binary segmentation problems. [26] uses Bipartite loss to tune DETR's bounding boxes. A similarity function is used to measure the deviation between predicted and actual nerve bounding boxes; A margin indicates the maximum similarity allowable on a single box to be able to contribute to the loss. All works except [25] and [26] utilize a loss combination technique to link the two different stages leading up to PNI classification together. [25] and [26] are orthogonalized and apply the loss functions independently.

6 Evaluation

6.1 Types of Evaluation Metrics

Prediction outcome (p)

		+	-	total
actual value (g)	**+**	True Positive (TP)	False Negative (FN)	P′
	-	False Positive (FP)	True Negative (TN)	N′
total		P	N	

6.1.1 Classification Metrics

Classification can be performed at different levels: patch-level, slide-level, or patient-level. Higher-level classification metrics, such as slide-level or patient-level, are more desirable because they better reflect the performance of the algorithm in a clinical context. This is because these metrics directly align with the inputs used by pathologists, who typically work with whole slides or regions of interest (ROIs) rather than individual patches. Here are the classification metrics used for PNI recognition:

1. Precision (PPV): $\frac{TP}{TP+FP}$. A low PPV implies that the algorithm is aggressively categorizing cases as PNI and raising false alarms, which may waste pathologists' time if used as a flag for further review.
2. Sensitivity: $\frac{TP}{TP+FN}$. A low sensitivity indicates that the algorithm is missing actual cases of PNI; it may benefit from decreasing regularization in the network or gathering more training patches.
3. Specificity: $\frac{TN}{TN+FP}$. Similar to PPV, but focuses on negative predictions. A low specificity indicates that the algorithm is classifying a large number of actual negatives as positives (unable to identify non-PNI regions).
4. Negative Predictive Value (NPV): $\frac{TN}{TN+FN}$. Akin to sensitivity, but focuses on negative predictions. A low NPV indicates that the model is classifying a lot of positives as negatives and missing cases of PNI.
5. AUC-ROC: The algorithm's confidence threshold is adjusted. This could involve adjusting the probability threshold across pixels for a positive prediction or, in the case of bespoke rule-based classifiers, adjusting the critical value of IoB/distance which classifies a sample of PNI. The precision and sensitivity (recall) are plotted with varying thresholds. AUC is the area under this curve. A low AUC shows poor class distinguishing abilities.
6. F1: $\frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$. Harmonic mean of Precision and Sensitivity, providing a single-valued evaluation metric.

From figure 5, we see that the highest F1 score is achieved by [26], which uses the rule-based IoB classifier at the end of a multi-stage segmentation and object detection approach. Closely following is [21], which categorizes any overlap in segmented nerve and tumor fields as PNI. [21] is more conservative, avoiding false positives and increasing precision in comparison to [26], even though it uses a less sophisticated classifier. This might be because segmentation masks are generally more precise [47] than bounding boxes, justifying the simpler approach to discriminating PNI from non-PNI that [21] uses in comparison to [26]. The XCeption ensemble used in [25] directly calculates patch-wise probabilities and extrapolates them to the patient and core level. This is an overtly orthogonalized approach which sacrifices spatial context derived from tumor and nerve segmentation, resulting in poorer performance. Interestingly, while moving to a higher context (from core to subject) might tend to introduce more random noise in the predictions, the F1 score for patient classification is actually higher by 0.04. Lastly, [22] and [23] achieve very low F1 scores likely due to the limited data pool and the inclusion of several types of cancer with different boundary apperances.
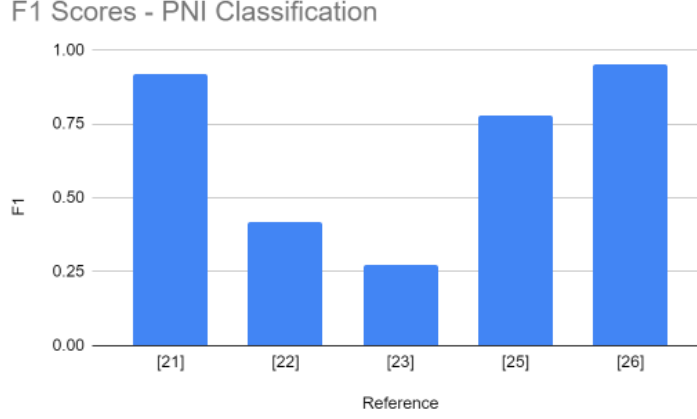
6.1.2 Segmentation Metrics

Segmentation metrics are measured at a pixel-level after thresholding prediction masks and recombining all patches into the WSI. Let P and G represent the prediction and ground truth WSI-level masks. IoU defines the common area between P and G as a ratio of their total areas.

$$\text{IoU} = \frac{P \bigcap G}{P \bigcup G}$$

Classification (PNI Recognition) Performance of the models					
Reference	[21]	[22]	[23]	[25]	[26]
Model	Rule-Based (Area of segmented fields)	FPN (segment), Rule (Classify) (Max-Pooling on segmented junctions)	U-net (segment), Rule (classify) (Max-Pooling on segmented junctions)	Xception Ensemble of 10 (PNI field)	Rule-Based (IoB)
Precision	0.95	-	-	Cores: 0.67 Subjects: 0.69	0.933
Sensitivity	0.9	-	-	Cores: 0.87 Subjects: 0.94	0.972
Specificity	0.95	-	-	Cores: 0.97 Subjects: 0.91	
NPV	0.9	-	-	Cores: 0.99 Patients: 0.99	-
AUC-ROC	0.92 ± 0.078	-	-	Cores: 0.98 ± 0.01 Subjects: 0.96 ± 0..03	-
F1	0.92	0.416	0.2747	Cores: 0.76 Subjects: 0.80	0.952

(a) Table of Classification metrics

(b) F1 Scores

Fig. 5: Classification metrics: Data

We define the pixel-wise segmentation error as follows:

$$
E(g, p, c) = \begin{cases} TP \text{ if } g = c, p = c \\ TN \text{ if } g \neq c, p \neq c \\ FP \text{ if } g \neq c, p = c \\ FN \text{ if } g = c, p \neq c \end{cases}
$$

For a particular class c, we can define the number of true positives as:

$$
TP_c = \sum_{g \in G, p \in P} |(g, p)| : E(g, p, c) = TP
$$

Thus, we can also define TN_c, FP_c, FN_c. Using these we can define Precision$_{\text{seg}}$, Sensitivity$_{\text{seg}}$ by summing over all classes C. For binary segmentation, C = 0

13

(Background) and 1 (Nerve/Tumor/Junction). For multi-class segmentation, C = 0 (Background) and 1 (Tumor) and 2 (Nerve). [‡]

$$\text{Precision}_{\text{seg}} = \sum_{c \in C} \frac{TP_c}{TP_c + FP_c} \qquad \text{Sensitivity}_{\text{seg}} = \sum_{c \in C} \frac{TP_c}{TP_c + FN_c}$$

From figure 6(a), it is generally observed that tumor segmentation is a harder task

Segmentation Performance of the models				
Reference	[21]	[24]	[25]	[26]
Model	Binary segmentation: UNet (Nerve) DeepLabv3+ (Tumor)	HCA + Binary Segmentation UNet (Nerve, Tumor)	Binary segmentation: UNet Ensemble of 10 (PNI field)	Binary segmentation: UNet++ (Tumor)
Sensitivity	Nerve: 0.943 Tumor: 0.903	Nerve: 0.85 Tumor: 0.57	-	0.81
Precision	Nerve: 0.937 Tumor: 0.839	-	-	0.87
F1	Nerve: 0.940 Tumor: 0.869	-	-	0.841
IoU	Nerve: 0.887 Tumor: 0.769	Nerve: 0.79 Tumor: 0.61	0.50 ± 0.045	0.815

(a) Table of Segmentation Metrics
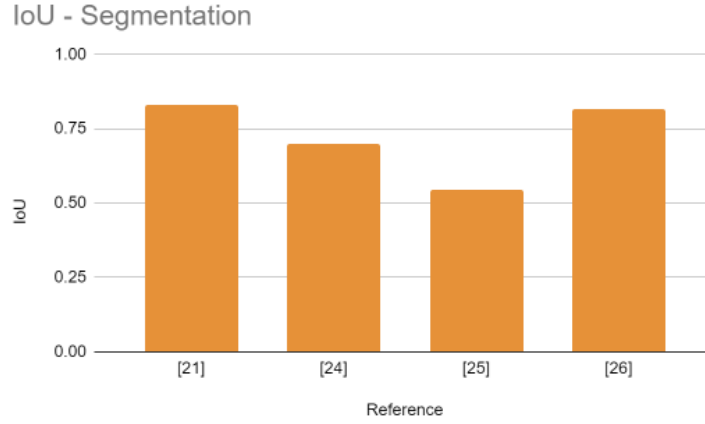
(b) IoU Scores

Fig. 6: Segmentation metrics: Data

than nerve segmentation, with lower IoU scores on the test set. This is because while nerves vary in shape and size across multiple organs, cancerous regions can be highly irregular and diffuse across multiple patches of the same type of cancer [48]. Complex pipelines with more novel segmentation models (eg: UNet++ and EfficientNet-B3 in [26] compared to the simpler UNet in [25]) demonstrate better performance. It would have been interesting to examine the segmentation performance on PNI junctions, observing the effect of their fine-grained nature compared to larger PNI, nerve, or

---

[‡]Please note that [24] also examines False alarm rate for segmentation. However, it has not been included in this paper as there is no comparison with other literature.

tumor fields. Unfortunately, [22] and [23] do not list it in as the rules of the PAIP 2021 challenge ranked teams based on the F1 score.

6.1.3 Bespoke Metrics - Bounding box detection

As [26] uses a DETR for predicting bounding boxes around nerves, it uses the **mean average precision (mAP)**[47] metric to evaluate the accuracy of the bounding boxes. Let κ represent the IoU threshold and C represent the class.

$$mAP_\kappa = \frac{1}{C} \sum_{c \in C} AP_{\kappa,c}$$

As discussed in 5, the decoder of the detection transformer outputs Q bounding boxes and softmax vectors containing the confidence scores (probabilities) for each class i.e. nerve and background. For each class, all bounding boxes at the slide-level are compared against their ground truth bounding boxes, and an IoU is calculated. If $IoU > \kappa$, then the predicted bounding box is counted as a true positive (TP). Otherwise, it is counted as a false positive (FP) because of insufficient overlap. A ground truth without any matching predictions will be counted as a false negative (FN). Based on TP, FP, FN, precision and recall scores can be calculated grouped by the confidence scores for that class. The minimum value for the confidence score is varied, leading to a receiver-operating curve with various values of precision and recall whose area will yield $AP_{\kappa,c}$. This process is repeated for both classes and their mean, mAP_κ is calculated. The authors in [26] use two values of κ, yielding $mAP_{0.75} = 0.587$ and $mAP_{0.50} = 0.775$ as the best performance. This metric cannot be compared with other works as it considers both the accuracy of bounding box classification (nerve, tumor) and localization, while all other metrics only denote performance in a single aspect.

6.2 Ablation Studies

[‡] An ablation study refers to further study undertaken by tweaking the components of the DL model and observing the effect it has on its performance. They are important in determining critical elements of the model, which must be sustained through any further iterations, and any points of improvement.

6.2.1 Effect of Transfer Learning

Transfer learning [40] is a learning paradigm wherein a neural network is pre-trained on a large dataset (for example, ImageNet [27]), where it learns to detect low-level features that can be useful in the segmentation of medical structures, including nerve and tumor. All authors have utilized pre-trained architectures and fine-tuned it using ground-truth annotations as seen in 5. [21] examined the effect of pre-training on the performance of Segformer's multi-class segmentation performance, showing a 29% and a 25% increase in IoU for Tumor and Nerve segmentation respectively; furthermore,

---

[‡]Please note that [22] also studies the effect of implementing TTA and extracting overlapping patches from test images. These are not discussed as they are used to improve rankings on the PAIP 2021 challenge and are computationally infeasible for integration into the inference pipeline.

the F1 scores also increased by 17.3% for tumor and 16.4% for nerve segmentation respectively. These metrics reinforce the importance of pre-training on a task with limited data availability.

6.2.2 Boundary Dilation and Loss combination

[23] examines the performance improvement brought about by changing the extent of boundary dilation (see section 4) and using both the standalone dice-loss and combining it with binary cross entropy. Changing the radius of the morphological dilation operator from $r = 1$ to $r = 2$ improves the detection F1-score by an average of only 2.6%, suggesting that the model captures most relevant junction information with the smaller radius. However, the inclusion of binary cross entropy improves the F1 score substantially, by an average of 41.5%. Integrating the end-goal (PNI detection) into the segmentation task's training allows the model to achieve balance between pixel-level accuracy and and region-level outcome.

6.2.3 Computational Complexity

[26] examines the performance of various segmentation and object detection models, documenting their FLOPs and number of parameters, opening up scope for implementing an accuracy-complexity trade-off scheme for integration into pathology labs where algorithms can flag images for review. For example, they show that using a UNet++ segmenter, a detection transformer model, and a shared EfficientNet backbone is 74% more efficient; the decrease in recognition F1 is only 15%. Similarly, traditional object detection algorithms, such as YOLO and Faster-RCNN detect nerves at $\sim$ 40 - 50% of the resource load as DETR with only a   5 - 15% reduction in mAP.

6.3 Algorithm Error Analysis

Due to the high stakes involved in medical imaging (and PNI detection especially so), carrying out error analysis is crucial [20]. The following systematic inaccuracies were observed across [21], [25], [26].

- **Misclassified Tumor**: Benign glands in the vicinity, Entrapped neural bundles resembling stroma, Inflammatory cells.
- **Misclassified Nerve**: Smooth muscle cells (resemble Schwann cells in their spindle shape), Mucinious Fibroplasia (in prostate adenocarcinoma only)

Histological tissue with similar structural features to nerve and tumor are bound to be misidentified when working with a small field like WSI patches. These can be dealt with using domain-specific filtering rules or ensemble models to refine the model's masks. However, consideration of increasing inference times is necessary here as well, keeping in mind the ease of distinguishing the above features from nerve/tumor manually if the algorithm is merely flagging images for further review.

7 Discussion

This section discusses the issues associated with current models and scope for future research with a focus on the ability to deploy these models to diagnose PNI from real-time patient data.

7.1 Combining Image analysis with blood markers

Inflammatory response in cancer is associated with several blood markers, for example, PSA has been shown to play a crucial role in the prognosis of prostate cancer and is associated with inflammation [49]. Likewise, it has been shown that Carcinoembryonic Antigens are a sign of inflammation, including PNI [50]. Using a multi-stage model to combine existing WSI analysis techniques with a separate machine learning algorithm could add value to PNI detection. For example, anomaly detection algorithms could flag patients with abnormal biomarker levels, refining predictions when combined with image-based features. Techniques like feature fusion could be applied to combine image features and biomarker data, leading to a more holistic prediction of PNI presence.

7.2 Integration into pathological workflows

In spite of utilizing state-of-the-art models and achieving high benchmarks, most of the models discussed will fail to operate when deployed in a pathology lab because of the large computational complexities involved in training and making predictions on them; this will be especially problematic when using real-time pathologist annotations on wrong predictions to improve the model. Based on the ablative experiments in 6.2.3, it might be more practical to implement a trade-off strategy between resource utilization and computational complexity. To do this, a systematic sensitivity analysis on a hyper parameter representing the objective complexity of the model will be required, with the level of trade-off stemming from external factors specific to the lab eg. the number of samples digitized into WSIs everyday, the number of reports submitted by pathologists on a daily basis. Moreover, strategies like model pruning [51] and quantization [52] could be considered to reduce the model's size and computational burden while still maintaining accuracy. Since most hospitals run AI algorithms on cloud GPUs, another avenue is examining how the models perform on cloud TPUs and GPUs to produce usage budget estimates for cost-benefit analysis by hospitals.

7.3 Prioritizing inter-observer agreement

A critical element of pathologists accepting deep learning algorithms to assist them is mutual agreement. Cohen's Kappa measures inter-observer agreement as follows [53]:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

p_o is a measure of the observed agreement between the two raters, while p_e is the expected agreement by chance, based on the marginal proportion of each of the rater's decisions. κ is a robust statistical measure that can be explicitly referenced in the loss

17

function, motivating the model to maximize it. The two raters would be the ground truth pathologist and the DL algorithm. The ground truth pathologist used in the calculation can also be representative of the average predictions of multiple expert pathologists, which would further work in favor of the optimizing objective. In fact, [25] performs a further experiment, having 4 pathologists (including the model) annotate the test set and documenting the mean pairwise κ value between each pathologist and the other three. For the human pathologists, this resulted in $0.684 \leq \kappa \leq 0.754$, and for the model, κ was 0.740, a value within the aforementioned range, demonstrating the calibre of these models and the extent of agreement that can be achieved when explicitly tuned to maximize κ.

7.4 Different staining techniques

While this section does not contain implementation avenues, it is still an interesting prospect to explore. The datasets used in most medical imaging papers, including this one, are stained with Hematoxylin and Eosin (H&E). It is commonplace and will distinguish between most structures clearly, but for PNI, it may not always provide the sufficient contrast to differentiate neural structures from adjacent tissues. Other stains, such as S100 (identification of Schwann cells), Neurofilament stain (general detection of nerve tissue), etc. can enhance the expression of PNI. This strategy is not practical to implement as the other stains are more expensive and not suited to general cancer recognition, but further study on this concept opens up avenues for exploring how specialized imaging techniques could complement AI-driven analysis, pushing the boundaries of precision.

8 Conclusion

This survey provides a comprehensive examination of PNI detection using DL techniques. It begins with a definition of PNI and an exploration of its molecular mechanisms. Approaches are categorized into three broad domains, with a focus on WSI datasets, pre-processing methods, neural network architectures, loss functions, and evaluation metrics, including detailed ablation studies and sources of algorithmic error. Finally, potential avenues for future research are outlined, emphasizing the integration of DL into digital pathology labs. The insights presented aim to guide ongoing and future advancements in PNI detection and its clinical applications. The integration of DL techniques into PNI detection represents a significant advancement in digital pathology. By leveraging sophisticated algorithms and vast datasets, these approaches have the potential to greatly enhance the accuracy and efficiency of PNI diagnosis, ultimately leading to more precise and timely patient care. This potential will be realized better when practical and implementation aspects are given more priority.

Acronyms

AI Artificial Intelligence
DL Deep Learning
FPN Feature Pyramid Networks
FLOPs Floating Point Operations per Second
GAN Generative Adversarial Networks
H and E Hematoxylin and Eosin Stain
IoU Intersection over Union
IoB Intersection over Bounding Box
mAP Mean Average Precision
PNI Perineural Invasion
PPV Positive Predictive Value
PSA Prostate-Specific Antigen
DETR Detection Transformer
WSI Whole Slide Image

References

[1] Batsakis, J.G.: Nerves and neurotropic carcinomas. Ann Otol Rhinol Laryngol **94**(4 Pt 1), 426–427 (1985)

[2] King, R.: 6. Anatomy of the peripheral nerve, pp. 32–37. John Wiley & Sons, Ltd, ??? (2014). https://doi.org/10.1002/9781118618424.ch6

[3] Amit, M., Binenbaum, Y., Trejo–Leider, L., Sharma, K., Ramer, N., Ramer, I., Agbetoba, A., Miles, B., Yang, X., Lei, D., Bjørndal, K., Godballe, C., Mücke, T., Wolff, K.-D., Eckardt, A.M., Copelli, C., Sesenna, E., Palmer, F., Ganly, I., Patel, S., Gil, Z.: International collaborative validation of intraneural invasion as a prognostic marker in adenoid cystic carcinoma of the head and neck. Head & Neck **37**(7), 1038–1045 (2015) https://doi.org/10.1002/hed.23710 https://onlinelibrary.wiley.com/doi/pdf/10.1002/hed.23710

[4] Larson, D.L., Rodin, A.E., Roberts, D.K., O'Steen, W.K., Rapperport, A.S., Lewis, S.R.: Perineural lymphatics: Myth or fact. The American Journal of Surgery **112**(4), 488–492 (1966) https://doi.org/10.1016/0002-9610(66)90309-6 . Papers of the Society of Head and Neck Surgeons

[5] Akert, K., Sandri, C., Weibel, E.R., Peper, K., Moor, H.: The fine structure of the perineural endothelium. Cell and Tissue Research **165**(3), 281–295 (1976) https://doi.org/10.1007/BF00222433

[6] Chen, S.-H., Zhang, B.-Y., Zhou, B., Zhu, C.-Z., Sun, L.-Q., Feng, Y.-J.: Perineural invasion of cancer: a complex crosstalk between cells and molecules in the perineural niche. Am. J. Cancer Res. **9**(1), 1–21 (2019)

19

[7] Zhang, L.J., Wu, B., Zha, Z.L., Qu, W., Zhao, H., Yuan, J., Feng, Y.J.: Perineural invasion as an independent predictor of biochemical recurrence in prostate cancer following radical prostatectomy or radiotherapy: a systematic review and meta-analysis. BMC Urol **18**(1), 5 (2018)

[8] Kim, H.E., Park, S.Y., Kim, H., Kim, D.J., Kim, S.I.: Prognostic effect of perineural invasion in surgically treated esophageal squamous cell carcinoma. Thorac Cancer **12**(10), 1605–1612 (2021)

[9] Duraker, N., man, S., Can, G.: The significance of perineural invasion as a prognostic factor in patients with gastric carcinoma. Surg Today **33**(2), 95–100 (2003)

[10] Tanaka, A., Watanabe, T., Okuno, K., Yasutomi, M.: Perineural invasion as a predictor of recurrence of gastric cancer. Cancer **73**(3), 550–555 (1994)

[11] Stojkovic Lalosevic, M., Milovanovic, T., Micev, M., Stojkovic, M., Dragasevic, S., Stulic, M., Rankovic, I., Dugalic, V., Krivokapic, Z., Pavlovic Markovic, A.: Perineural invasion as a prognostic factor in patients with stage I-III rectal cancer - 5-year follow up. World J Gastrointest Oncol **12**(5), 592–600 (2020)

[12] Chi, A.C., Katabi, N., Chen, H.S., Cheng, Y.L.: Interobserver Variation Among Pathologists in Evaluating Perineural Invasion for Oral Squamous Cell Carcinoma. Head Neck Pathol **10**(4), 451–464 (2016)

[13] Holthoff, E.R., Jeffus, S.K., Gehlot, A., Stone, R., Erickson, S.W., Kelly, T., Quick, C.M., Post, S.R.: Perineural Invasion Is an Independent Pathologic Indicator of Recurrence in Vulvar Squamous Cell Carcinoma. Am J Surg Pathol **39**(8), 1070–1074 (2015)

[14] 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 **117**(7), 1415–1421 (2011)

[15] Yan, F., Cheng, Y.L., Katabi, N., Nguyen, S.A., Chen, H.S., Morgan, P., Zhang, K., Chi, A.C.: Interobserver Variation in Evaluating Perineural Invasion for Oral Squamous Cell Carcinoma: Phase 2 Survey Study. Head Neck Pathol **15**(3), 935–944 (2021)

[16] Schneider, J., Vlachos, M.: A Survey of Deep Learning: From Activations to Transformers (2024). https://arxiv.org/abs/2302.00722

[17] Nirschl, J.J., Janowczyk, A., Peyster, E.G., Frank, R., Margulies, K.B., Feldman, M.D., Madabhushi, A.: A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of H&E tissue. PLoS One **13**(4), 0192726 (2018)

[18] Jha, A., Yang, H., Deng, R., Kapp, M.E., Fogo, A.B., Huo, Y.: Instance segmentation for whole slide imaging: end-to-end or detect-then-segment. Journal of Medical Imaging **8**(1), 014001 (2021) https://doi.org/10.1117/1.JMI.8.1.014001

[19] Feng, R., Liu, X., Chen, J., Chen, D.Z., Gao, H., Wu, J.: A deep learning approach for colonoscopy pathology wsi analysis: Accurate segmentation and classification. IEEE Journal of Biomedical and Health Informatics **25**(10), 3700–3708 (2021) https://doi.org/10.1109/JBHI.2020.3040269

[20] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Laak, J.A.W.M., Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical Image Analysis **42**, 60–88 (2017) https://doi.org/10.1016/j.media.2017.07.005

[21] Jung, J., Kim, E., Lee, H., Lee, S.H., Ahn, S.: Automated hybrid model for detecting perineural invasion in the histology of colorectal cancer. Applied Sciences **12**(18) (2022) https://doi.org/10.3390/app12189159

[22] Nateghi, R., Pourakpour, F.: Perineural Invasion Detection in Multiple Organ Cancer Based on Deep Convolutional Neural Network (2021). https://arxiv.org/abs/2110.12283

[23] Park, Y., Park, J., Jang, G.-J.: Efficient perineural invasion detection of histopathological images using u-net. Electronics **11**(10) (2022) https://doi.org/10.3390/electronics11101649

[24] Borsekofsky, S., Tsuriel, S., Hagege, R.R., Hershkovitz, D.: Perineural invasion detection in pancreatic ductal adenocarcinoma using artificial intelligence. Scientific Reports **13**(1), 13628 (2023) https://doi.org/10.1038/s41598-023-40833-y

[25] Kartasalo, K., m, P., Ruusuvuori, 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 **481**(1), 73–82 (2022)

[26] Hu, Z., Deng, Y., Lan, J., Wang, T., Han, Z., Huang, Y., Zhang, H., Wang, J., Cheng, M., Jiang, H., Lee, R.-G., Du, M., Tong, T., Gao, Q., Chen, G.: A multi-task deep learning framework for perineural invasion recognition in gastric cancer whole slide images. Biomedical Signal Processing and Control **79**, 104261 (2023) https://doi.org/10.1016/j.bspc.2022.104261

[27] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009). https://doi.org/10.1109/CVPR.2009.5206848

[28] Janowczyk, A., Madabhushi, A.: Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. J Pathol Inform **7**, 29

(2016)

[29] Challenge, G.: PAIP 2021 Challenge: Perineural Invasion in Multiple Organ Cancer (Colon, Prostate and Pancreatobiliary tract). Accessed: 2024-09-15 (2021). https://paip2021.grand-challenge.org/

[30] Eklund, M., m, T., Aly, M., Adolfsson, J., Wiklund, P., Brandberg, Y., Thompson, J., Wiklund, F., Lindberg, J., Presti, J.C., StLezin, M., Clements, M., Egevad, L., nberg, H.: The Stockholm-3 (STHLM3) Model can Improve Prostate Cancer Diagnostics in Men Aged 50-69 yr Compared with Current Prostate Cancer Testing. Eur Urol Focus **4**(5), 707–710 (2018)

[31] Janowczyk, A., Madabhushi, A.: Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. Journal of Pathology Informatics **7**(1), 29 (2016) https://doi.org/10.4103/2153-3539.186902

[32] Fernández Blanco, R., Rosado, P., Vegas, E., Reverter, F.: Medical image editing in the latent space of generative adversarial networks. Intelligence-Based Medicine **5**, 100040 (2021) https://doi.org/10.1016/j.ibmed.2021.100040

[33] Cossio, M.: Augmenting Medical Imaging: A Comprehensive Catalogue of 65 Techniques for Enhanced Data Analysis (2023). https://arxiv.org/abs/2303.01178

[34] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks (2014). https://arxiv.org/abs/1406.2661

[35] TensorFlow: Data Augmentation. Accessed: 2024-09-15 (2024). https://www.tensorflow.org/tutorials/images/data_augmentation

[36] PyTorch: Transforms. Accessed: 2024-09-15 (2024). https://pytorch.org/vision/0.9/transforms.html

[37] GeeksforGeeks: U-Net Architecture Explained. Accessed: 2024-09-15 (2024). https://www.geeksforgeeks.org/u-net-architecture-explained/

[38] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation (2015). https://arxiv.org/abs/1505.04597

[39] Lin, M., Chen, Q., Yan, S.: Network In Network (2014). https://arxiv.org/abs/1312.4400

[40] Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? (2014). https://arxiv.org/abs/1411.1792

[41] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going Deeper with Convolutions (2014). https:

//arxiv.org/abs/1409.4842

[42] Tan, M., Le, Q.V.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks (2020). https://arxiv.org/abs/1905.11946

[43] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature Pyramid Networks for Object Detection (2017). https://arxiv.org/abs/1612.03144

[44] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: UNet++: A Nested U-Net Architecture for Medical Image Segmentation (2018). https://arxiv.org/abs/1807.10165

[45] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-End Object Detection with Transformers (2020). https://arxiv.org/abs/2005.12872

[46] Chollet, F.: Xception: Deep Learning with Depthwise Separable Convolutions (2017). https://arxiv.org/abs/1610.02357

[47] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision – ECCV 2014, pp. 740–755. Springer, Cham (2014)

[48] Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., Li, S.: Breast cancer multi-classification from histopathological images with structured deep learning model. Scientific Reports **7**(1), 4172 (2017) https://doi.org/10.1038/s41598-017-04075-z

[49] Elgamal, A.A., Petrovich, Z., Van Poppel, H., Baert, L.: In: Petrovich, Z., Baert, L., Brady, L.W. (eds.) The Role of Prostate-Specific Antigen in the Management of Prostate Cancer, pp. 179–197. Springer, Berlin, Heidelberg (1996). https://doi.org/10.1007/978-3-642-60956-5_14

[50] Lee, T.-H., Kim, J.-S., Baek, S.-J., Kwak, J.-M., Kim, J.: Diagnostic accuracy of carcinoembryonic antigen (cea) in detecting colorectal cancer recurrence depending on its preoperative level. Journal of Gastrointestinal Surgery **27**(8), 1694–1701 (2023) https://doi.org/10.1007/s11605-023-05761-2

[51] Han, S., Pool, J., Tran, J., Dally, W.J.: Learning both Weights and Connections for Efficient Neural Networks (2015). https://arxiv.org/abs/1506.02626

[52] Krishnamoorthi, R.: Quantizing deep convolutional networks for efficient inference: A whitepaper (2018). https://arxiv.org/abs/1806.08342

[53] Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological Measurement **20**(1), 37–46 (1960) https://doi.org/10.1177/001316446002000104