Perineural Invasion Detection Using Deep Learning: A Systematic Review

Hayagreev Jeyandran [1*] and Dr. Veena R, MD [2]

[1]Undergraduate Researcher, Department of Electrical Engineering and Computer Science, College of Engineering, and Radiology Research Laboratory, Michigan Medicine, University of Michigan, Ann Arbor, USA.
[2]Senior Consultant Pathologist and Head of the Department of Histopathology and Digital Pathology, Triesta Sciences - A unit of HealthCare Global Enterprises Ltd., Bengaluru, India.

*Corresponding author(s). E-mail(s): hayagjey@umich.edu;
Contributing authors: drveena.r@hcgel.com;

Abstract

Perineural Invasion (PNI) is the infiltration of tumor cells into the perineurium, a protective sheath surrounding nerve tissue. It serves as a critical pathological marker associated with poor prognosis in various cancers, including pancreatic, rectal, head and neck, and prostate cancer. PNI is linked to cancer recurrence and metastasis, yet detecting PNI in whole-slide pathological images remains laborious and error-prone for pathologists, significantly impacting treatment planning. This paper surveys cutting-edge deep learning approaches for PNI diagnosis, offering a structured taxonomy for segmentation, object detection, and classification models. It begins with an overview of PNI and its diagnosis, the application of deep learning in medical imaging, and categorizes current literature approaches. The study then explores datasets, image pre-processing, synthetic generation, neural network architectures (e.g., U-Net, Detection Transformer), and analyzes segmentation and classification performance metrics reported in the literature. Gaps such as model robustness and clinical integration are identified, culminating in future research directions aimed at enhancing the accuracy and scalability of PNI diagnosis in clinical practice.

Keywords: Perineural Invasion, Histopathology, Whole Slide Images, Deep Neural Networks, Machine Learning

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 concerning the peripheral nerve sheath.

1.1 Structure of the Peripheral Nerve Sheath

The layers are listed 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 fibers myelinated by Schwann 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 a 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 fueling the extent of PNI [6]; the nerve micro-environment (macrophages, fibroblasts, Schwann cells), molecular signaling (such as neurotrophic 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 remaining largely unknown, plenty of evidence from various types of cancers collected over a long period suggests 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 laborious for 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 that 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 present [13]. Similarly, in a study involving colorectal cancer, the percentage of patients initially diagnosed as PNI positive increased from 7.5% to 24.3% following a meticulous second review by pathologists [14]. The intestinal wall is a dense structure whose complexity can easily overshadow subtle infiltration into the nerves. Interestingly, variability in the 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. A pathologist can review the case and the model's predictions more carefully subsequently.

1.3 AI for Healthcare and Deep Learning in Medical Imaging

In recent years, hospitals have seen increased digitization of electronic health record (EHR) data and consolidation of patient records, evidencing efforts to compile multi-modal, AI-ready datasets comprising medical images (CT, MRI, X-rays, and Pathology Images), ECG and EEG scans, genome sequences, lab test results, and doctors' notes [16].

Today, various categories of AI (artificial intelligence) are actively being investigated broadly in healthcare, ranging from NLP-powered expert systems to sophisticated patient positioning systems powered by reinforcement learning. For example, pipelines that integrate Bayesian optimization, expert systems, and neural networks have demonstrated accuracy levels of up to 98% in the screening of breast cancer [17]. Another study uses supervised learning methods (such as k-nearest neighbors and ensemble classifiers) along with signal processing techniques on ECGs to detect real-time and past myocardial infarction, achieving an accuracy of up to 92.5% in clinical trials [18]. Performance metrics in these and several other studies are comparable to those achieved by physicians, indicating the potential usefulness of AI as an assistive tool for physicians.

Many studies at the intersection of healthcare and AI involve deep learning (DL). Deep learning has revolutionized the field of medical imaging by enabling computers to automatically analyze complex patterns with remarkable accuracy. DL systems are capable of automatically extracting relevant features without the need for

manual feature engineering [19]. Several studies over the years have tuned DL implementations to support diagnostic functionality. For example, a Convolutional Neural Network was used by Nirschl J. et al. to detect signs of clinical heart failure from Hematoxylin and Eosin (H&E) stained WSIs [20]. Other advances include two-stage glomerular segmentation [21] and multi-instance gastric image segmentation [22]. 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 3, 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, and post-processing techniques are explained. Following this, evaluation metrics for the models and error analysis of the algorithm's performance are examined. Lastly, the future research scope is suggested before concluding.

2 Review Methodology

2.1 Search Strategy and Data Sources

To ensure a comprehensive survey of the current landscape, a multi-database search was conducted encompassing PubMed, IEEE Xplore, Google Scholar, and arXiv. The literature search prioritized studies published between January 2017 and **December 2024**, aiming to capture the transition from early convolutional neural networks (CNNs) to modern attention-based architectures. [‡]

The search strategy utilized a combination of Medical Subject Headings (MeSH) and Boolean keywords, including:

- *"Perineural Invasion"*,
- *"Deep Learning"*,
- *"Whole Slide Imaging"*,
- *"Computational Pathology"*,
- *"Semantic Segmentation"*,
- *"Perineural Spread.*

Given the rapid evolution of computer vision, specific emphasis was placed on retrieving recent preprints and conference papers (2022 and later), utilizing architectures such as the transformer to contrast emerging paradigms.

Inclusion and Exclusion Criteria

The selection process was informed by the PRISMA guidelines to ensure transparency. Articles were included if they:

1. Focused explicitly on PNI detection or nerve-tumor segmentation in histopathology images.
2. Reported quantitative performance metrics (e.g., F1-score, IoU, Accuracy) on defined datasets.
3. Detailed the deep learning architecture used (e.g., U-Net, Mask R-CNN, SegFormer).

Studies relying solely on handcrafted feature extraction (pre-deep learning) or lacking clear methodological descriptions were excluded. This filtering process resulted in the core selection of 6 studies analyzed in the subsequent sections, representing the state-of-the-art in automated PNI diagnosis as of late 2024.

3 Summary of Approaches

Most computer vision models can produce three types of output when performing inference on medical images [23]. Images can be **classified** to produce class-wise probabilities of various diseases occurring in the entire image. Secondly, a model might be trained to output bounding boxes for target features in an image, such as nerves and tumors, in an object detection approach. Lastly, models can perform classification at the pixel level. When these pixel-level probabilities are thresholded, they produce **segmentation** masks. Segmentation is akin to shading relevant regions in an image. Current PNI detection literature synthesizes various approaches from this taxonomy.

1. Most approaches directly utilize an end-to-end segmentation model to identify pixels containing the boundary between the tumor and nerve (the PNI junction) or the PNI field. This approach requires low-level, granular,

---

[‡]Though this manuscript was revised in Jan 2026, no new articles were included in the core literature review for reproducibility. The authors acknowledge that there have been notable advancements in PNI detection using deep learning, such as Vision Transformers (ViTs) and Graph Neural Networks (GNNs), since the end of the inclusion period, which have established even bolder performance metrics above CNN baselines.
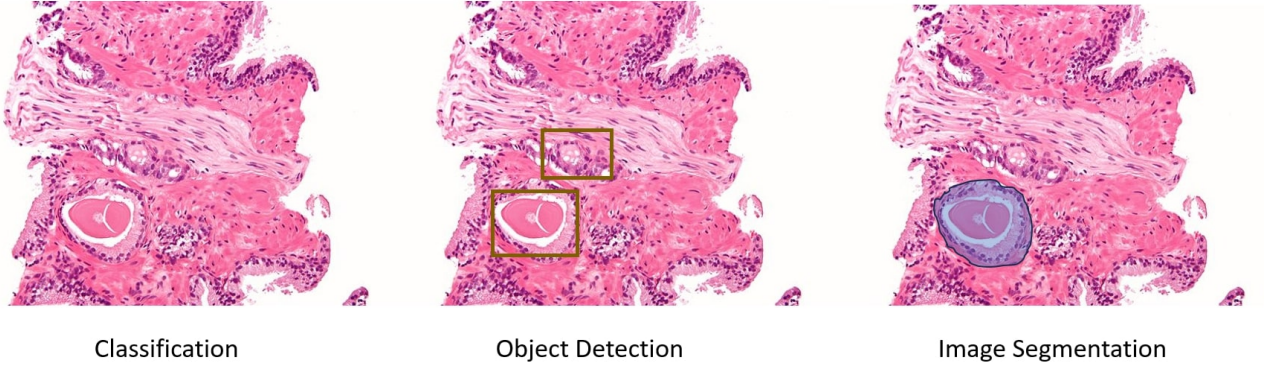
| Classification | Object Detection | Image Segmentation |

Fig. 1: Example of DL approaches applied to gland detection on a WSI. Image source: Wikimedia, licensed under CC BY-SA 3.0. Modifications made by the author.

and time-consuming annotation of WSI datasets but tends to perform better due to the microscopic scale of features being learned.

2. Part of the literature uses sequential pipelines and rule-based classification. Segmentation or object detection is used to identify tumors and nerve fields in the first two stages. Then, a rule-based classifier declares PNI based on the overlap of the outputs of the two stages. Though this approach is less complex and orthogonalized for better debugging and validation, its mesoscopic scale and rigidity make it less accurate.

Table 1: Reviewed Literature

Literature	Approach	Authors
Automated Hybrid Model for Detecting Perineural Invasion in the Histology of Colorectal Cancer [24]	Multi-Class Tumor and Nerve Field Segmentation + Rule-Based PNI Classifier	Jiyoon Jung et al.
Perineural Invasion Detection in Multiple Organ Cancer Based on Deep Convolutional Neural Network [25]	PNI Junction Segmentation	Ramin Nateghi et al.
Efficient Perineural Invasion Detection of Histopathological Images Using U-Net (Multi-Organ Cancer) [26]	PNI Junction Segmentation	Youngjae Park et al.
Perineural invasion detection in pancreatic ductal adenocarcinoma using artificial intelligence [27]	Hierarchical Contextual Analysis, Binary Tumor and Nerve Field Segmentation + Rule-Based PNI Classifier	Sarah Borsekofsky et al.
Detection of perineural invasion in prostate needle biopsies with deep neural networks [28]	Binary PNI Field Segmentation	Kimmo Kartasalo et al.
A multi-task deep learning framework for perineural invasion recognition in gastric cancer whole slide images [29]	Binary Tumor Field Segmentation and Nerve Object Detection + Rule-Based PNI Classifier	Ziwei Hu et al.

4 Dataset

In the modern DL era, visual training tasks such as traditional image classifiers and object detectors use large corpora of 1 - 10 million training images to mitigate over-fitting and learn nuanced features [30]. Working with a similarly large number of WSIs for PNI detection is infeasible due to the following reasons:

1. **Computational Resources** - WSIs are high-resolution, gigapixel images, often occupying over 5 GB per image. Processing a WSI dataset traditionally would require large amounts of GPU memory and compute, even with techniques like lazy loading, which is usually not accessible in a medical research environment.

2. **Annotation and Labeling** - 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.

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 [31]. Below, hospital-specific and public datasets are elucidated and analyzed, given these constraints.

4.1 PAIP 2021 - Grand challenge

The Pathology in Artificial Intelligence (AI) Platform (PAIP) hosted a challenge in 2021: Perineural Invasion in Multiple Organ Cancer [32]. Authors of [25], [26] 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.

4.2 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 [33]. 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 [28] oversampled 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**.

4.3 Hospital-specific datasets

- In [24], **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-processed into patches. Each patch was assigned to one of four classes.
- In [29], 143 train, 12 validation, and 60 test WSIs from **gastric cancer** patients were gathered at **20x magnification** in Fujian Cancer Hospital, China.
- In [27], **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.

These datasets are summarized in Table 2 on page 7.

Table 2: PNI Datasets: Summary

Source	Dataset Source	Organ and Cancer Type	Sample Size and Split (train/val/test)	Magnification	Annotation Type
[25], [26]	PAIP Grand Challenge 2021 [32]	Colon, Prostate, Pancreatic	240 WSIs (150/30/60)	20x	Pixel-level PNI Junction
[28]	STHLM3	Prostate (Biopsy Cores)	8,803 Cores (7045 train/1758 test)	(Not listed)	(Binary Core-Level Labels)
[24]	St. Mary's Hospital, Korea	Colorectal	77 WSIs (530 ROIs)	(Not listed)	4-class patch-level labels
[29]	Fujian Cancer Hospital, China	Gastric	215 WSIs (143/12/60)	20x	Pixel-level (Tumor) + Bounding Box (Nerve)
[27]	Sourasky Medical Center, Israel	Pancreatic	428 WSIs (260 train/168 val + test)	40x	Binary tumor and nerve segmentation

4.4 Dataset Limitations and Annotation Challenges

The descriptions of these datasets reveal important limitations in PNI datasets that impede research efforts in this space.

- **Large-scale datasets**: Historically, the field has lacked public datasets comparable to those in radiology and ophthalmology (eg, MIMIC-CXR Chest X-Rays [34] and EYEPACS variable resolution color fundus photographs [35]. WSI datasets have been traditionally fragmented due to annotation and computational constraints. However, recent initiatives are challenging this constraint. For instance, the **HISTAI** dataset [36] (released in May 2025) aggregates over 60,000 WSIs, signaling a shift towards large-scale, open-access repositories. Despite this breakthrough, the vast majority of reviewed literature still operates in the low data regime.

- **Domain shift**: Deep Learning relies on the critical assumption of independent, identical data (i.i.d). As aforementioned, WSI datasets are often single-center and proprietary (e.g., [24], [29], [27]), using different organ types with high morphological diversity, scanner magnifications, and true positive rates. This breaks the i.i.d assumption, hindering unification and the development of large-scale detection algorithms.
- **Annotation strategies**: These vary significantly, creating challenges for cross-study comparison. For example, the PAIP dataset [32] uses precise pixel-level masks for PNI junctions. Such precise annotations are laborious to generate. In contrast, studies like [28] rely on coarser core-level binary labels or bounding boxes [29]. These are easier to generate, but result in lower spatial supervision. The inconsistency in the scale of ground-truth generation remains a practical hurdle for developing unified PNI detection algorithms

To summarize, in the long run, the community should focus on integrating WSIs scanned using consistent parameters, with histological tissue resected from a wide variety of organs (outside just the digestive belt) into a large dataset with standardized PNI annotations [37]. [36] serves as a starting point for such an integrative effort.

5 Preprocessing

Given the issues with WSIs discussed in Section 4, such as dimensionality and class imbalance, both geometric and synthetic preprocessing are applied to ease training and inferential mechanics, making models more practical.

5.1 Patching

To solve the problem of high dimensionality, most DL approaches include patching (tiling) as a mandatory preprocessing step [38]. 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.

5.1.1 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 $p \nmid n_H$ or $p \nmid n_W$, there will be leftover regions that do not fully fit the window. This can be dealt with by applying a zero-padding of $p - (n_H \mod p) \times p - (n_W \mod p)$. Non-overlapping patches are used in [28] where p is specified as $0.25mm$ and [25] where $p = 512$px. This may lead to a loss of information from the boundary regions, which is dealt with using dilation techniques later on in the preprocessing pipeline.

5.1.2 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. [29], [26], [24] all use this strategy with p ranging from 512 to 2048 and a overlap proportion of 0.50. In [29], the patches are cropped at equal intervals to minimize duplication.

While addressing dimensionality, patching can also be tailored to address class imbalance. The overlapping technique generates more patches from the same region, resulting in a form of spatial oversampling for the PNI class. The probability that a sparse PNI junction or field is captured in the center of a patch is much higher, rather than being split at the edge (as in 4.1.1).

5.2 Boundary Dilation

In approaches [25], [26] (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, [25] uses a disk of radius 3px). The operator is slid over the boundary, and the maximum value of the pixels within the operator is cast onto the remaining pixels within the operator. Like in the case of overlapping patching, dilation improves feature learnability and makes the model more robust.

5.3 Data Augmentation

Data augmentation is a commonly used preprocessing technique to mitigate overfitting in machine learning applications, resulting from class imbalance [39]. Recent years have seen a proliferation in the use of synthetic data augmentation through Generative Adversarial Networks [40] across machine learning.GANs consist of a

generator G that produces synthetic images and a discriminator D) that distinguishes between real and synthetic images, optimizing through a min-max game so G generates data similar to the real distribution. This synthetic generation and augmentation method can be seamlessly integrated into neural network pipelines using frameworks like TensorFlow [41], and PyTorch [42]. The mechanism of GANs is illustrated in Figure 2. GANs
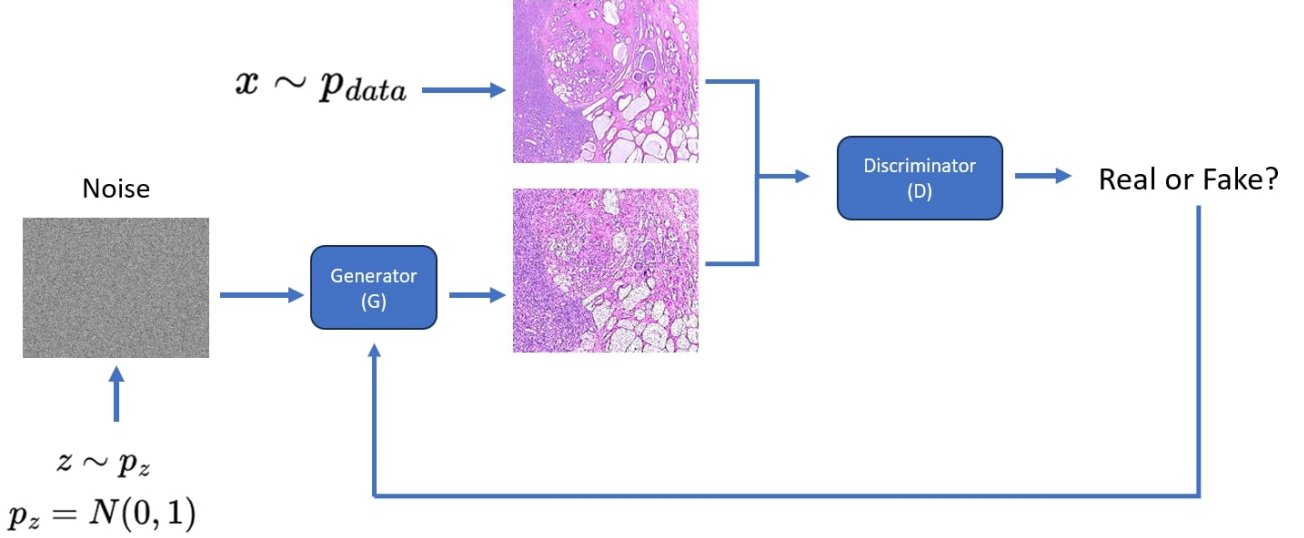

Fig. 2: GANs for WSIs: Framework. Source: Wikimedia Commons (licensed under CC-BY-SA 3.0). Modifications made by the author.

are widely used in detection tasks where the active class is sparse, for instance, in cybersecurity, where they are often used to generate data and threat examples to train anomaly detectors while preserving user privacy [43]. Another example is in wireless security, where a GAN was used to generate synthetic wireless signals to train authentication systems where real attack vectors were rare, achieving a 19% improvement in authentication accuracy [44]. In medical imaging and PNI detection, synthetic data augmentation is crucial. In the dataset used by [28], only 11% of the training set and 5% of the test set represent positive PNI cores, with further spatial sparsity observed within core-level pixels. While all examined literature use some form of data augmentation, such as color augmentation, brightness and contrast adjustment, flipping, shearing, etc., GANs, used in [27] can demonstrate superior performance as they extrapolate the data distribution rather than interpolating. GANs can mimic morphological diversity by producing new nerve shapes and interaction textures between nerves and tumors that do not exist in the dataset.

5.4 Background Image Elimination

Works [24] and [29] examine the standard deviation (σ) and mean (μ) of patch-wise pixel intensities to infer their usability. For example, [29] considers $\sigma_{pixelIntensity} < 0.6$ to be an indicator of a background image with no useful pathological features. [24] 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.

While primarily intended to reduce computational overhead by reducing slide surface area, this step also crudely functions as tissue segmentation. Eliminating empty glass using the mean filter increases the effective prevalence of PNI in the training set. However, rigid thresholding risks discarding faintly stained tissue (e.g., mucinous carcinoma), introducing a selection bias against subtle PNI features before the model sees them.

6 Models and Training

6.1 Segmentation Models

6.1.1 U-Net

Proposed in [45] by Ronneberger et al., the U-Net has been a seminal integration into the field of medical imaging. All works except [25] make use of the basic U-Net architecture, some of them modifying part of it.

The U-Net's distinctive features are its encoder and decoder networks. The encoder successively down-samples spatially at higher depths, capturing higher-level features, such as nerve and tumor fields. A decoder, combined with skip connections, restores spatial information through learned transpose convolutions, and the decoder's output is converted into a segmentation map using a 1x1 convolution [46]. Figure 3 shows the architecture of the U-Net. This symmetric structure allows the U-Net to produce detailed, feature-rich segmentation maps for PNI and nerve and tumor fields. [24], [26], [27], [28] use the U-Net in some part of their architecture. The U-Net is
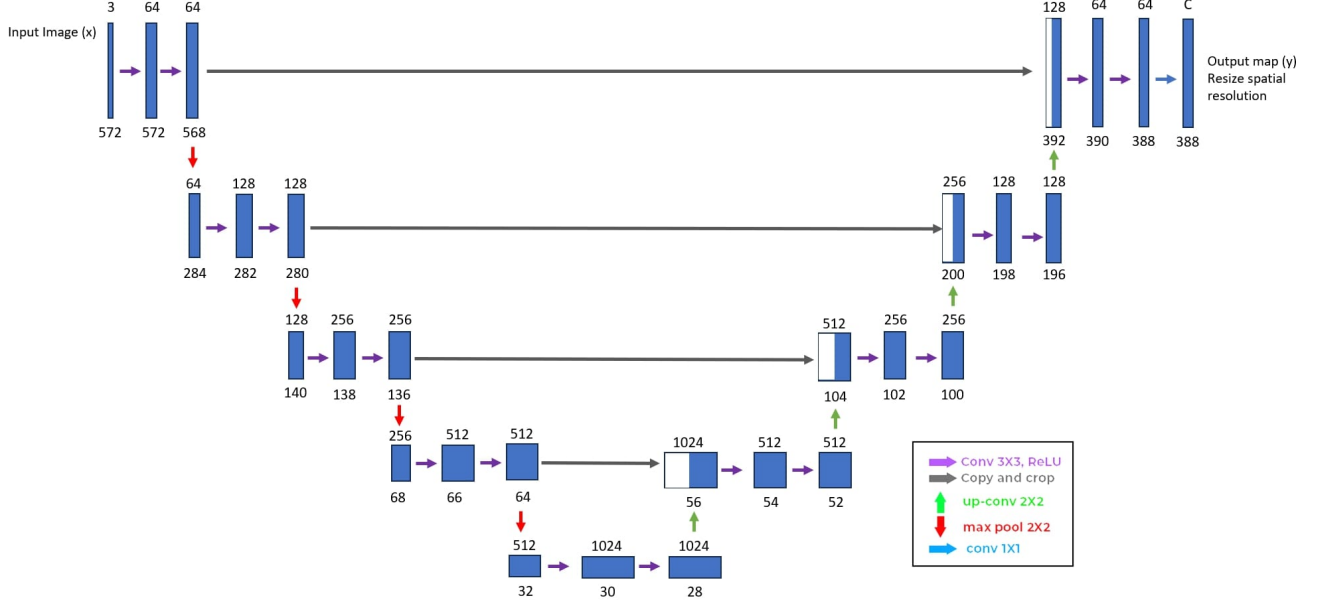

Fig. 3: U-Net Architecture

also very flexible and can facilitate learning paradigms like transfer learning [47], 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. [24] and [26] replace the U-Net encoder with different architectures, such as Inception-Resnet-v2 [48] and EfficientNetB0 through B4 [49]. There are three main reasons for this:

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 Section 4.
2. **Complexity**: The UNet encoder is relatively shallow in comparison with InceptionNet and EfficientNet. The substitution with deeper encoders allows for capturing more intricate patterns and learning more complex hierarchies. However, substituting deeper backbones like Inception-ResNet increases computational inference time, creating a trade-off between slight accuracy gains (e.g, +0.02 F1) and the feasibility of real-time deployment.
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 [28]. 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 or faint sections of the WSI, for instance, while inferencing. [27] uses a fully unsupervised approach initially, implementing U-Net to approximate regions containing PNI in conjunction with decision processes inspired by pathologist insight. This can be regarded as a pre-training step.

6.1.2 Other models

[25] uses Feature Pyramid Networks (FPNs) [50], 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. [29] uses U-Net++, which is a deeply supervised and more complex U-Net architecture [51]. It introduces nested skip connections with dense layers, allowing for features in skip pathways to be reused 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.

6.2 Object Detection Models

To detect nerve bounding boxes, [29] utilizes the Detection Transformer (DETR) [52]. Inspired by the transformer architecture, the DETR frames object detection as a sequence prediction problem. A backbone encoder network, like ResNet101, produces latent embeddings, which are encoded as values and keys of the transformer network. Bounding box queries are supplied in parallel, and the network is trained to output the image's class as well as bounding box vectors. Figure 4 shows the schematic of the DETR. Unlike convolutional networks restricted by
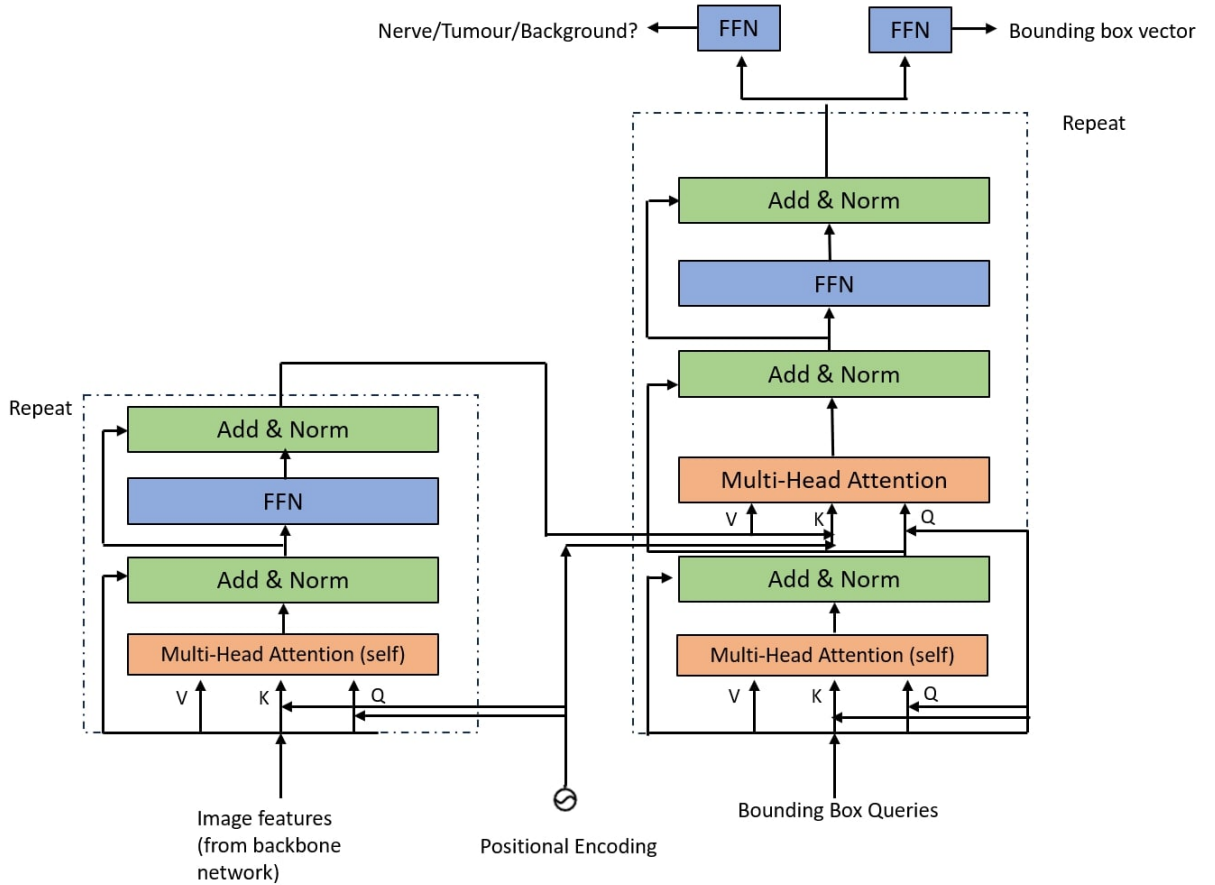

Fig. 4: Detection Transformer Model

local receptive fields, DETR's self-attention mechanisms can capture global WSI context. The DETR employs self-attention across the entire latent embedding, which is analogous to wide-angle vision, while a convolutional network has tunnel vision in its earlier layers. This helps analyze the surrounding tissue context to determine PNI early on. For instance, while a nerve, a blood vessel, and smooth muscle may look identical to each other in the early layers of a convolutional network.

But, for a DETR analyzing features across larger distances, factors like inflammation and gland alignment are factored into the classification of WSI artifacts, effectively filtering out nerve-mimics in the wrong biological neighborhood. However, the DETR is not widely used. Its large field of view may skip fine-grained features like thin nerve sheets that the FPN or U-Net are designed to handle, especially when the image is embedded by a ResNet and is already deprived of spatial information. The attention mechanism is also computationally expensive ($\in O(N^2)$) to run in hospital environments.

6.3 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 3, different methods are used to make either WSI-level (or patient-level predictions, in [28].

- **End-to-End Neural Network Classification**: [28] uses XCeption, a modified Inception Network with depth-wise separable convolutions within Inception blocks, simplifying model workflow [53]. The authors propagate patches directly through the XCeption model and extract sigmoid outputs interpreted as the probability that the patch contains PNI, given its segmentation map. Finally, max pooling is used over all patches to assign WSI probabilities, and in their case, patient-wise PNI probabilities.

$$p_{x,\max} = \text{MaxPool}(x) = \underset{i \in \text{patches}(x)}{\text{argmax}} \; p(y_i = 1)$$

  The end-to-end approach has the advantage of simplifying complex workflows and excluding the heavy segmentation step from a U-Net, at the cost of an explainability gap. The lack of an intermediate segmentation map hinderspathologists' trust, which will become an important factor as will be seen in Section 8. Though max pooling is biologically appropriate, a single false positive patch creates a false positive WSI and patient diagnosis, making it extremely brittle.

- **Junction Segmentation**: [25] makes a simple prediction by taking the largest probability per patch and applying average pooling across all patches to produce the WSI probability. This risks diluting the effect of a single patch strongly predicted as a positive, especially in non-overlapping patching, where sparse tumor-nerve interactions can straddle multiple patches, though it is not as volatile as max pooling. Global max-pooling can identify highly discriminating, dominant regions in WSIs. Although it ignores broader contextual information, it is more appropriate for PNI detection as nerve-tumor interactions leading to highly activated patch probabilities can be captured without diluting it with irrelevant background information.

$$p_{x,\text{avg}} = \text{AveragePool}(x) = \frac{1}{|\text{patches}(x)|} \sum_{i \in \text{patches}(x)} p(y_i = 1)$$

  [26] 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 an intersection between the enlarged predicted and ground truth boxes. Then, the average minimum distances between lines in matching pairs and the intersection over the 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**: [24] counts $Area_{nervemap} \bigcap Area_{tumormap} \neq 0$ as a PNI. [29] is more sophisticated and uses the intersection over bounding box (IoB) metric and predicts PNI positive only if $IoB > 0.12$ rather than flagging any intersection as positive. $IoB = \frac{A_{\text{nerveBox}} \bigcap A_{\text{tumorMap}}}{A_{\text{nerveBox}} \bigcup A_{\text{tumorMap}}}$

  A common theme among [24], [26], [29] is the problem of heuristic fragility. Though they do not seem as rudimentary as max- or average-pooling, they rely on specific, hard-coded values for the IoB or bounding box scale factors, or segmentation intersection to quantify PNI. These are likely overfit to the specific datasets used in the study; they do not adapt like learned parameters. The best classification strategy might be to find a middle-ground between a "black box" end-to-end classifier and the two-stage models, perhaps by preserving segmentation maps for pathologist review and using emerging architectures like Graph Neural Networks [54]. By constructing a tissue graph, where nodes represent segmented biological entities and edges represent spatial interactions, PNI can be classified while retaining segmentation maps and replacing hand-crafted logic.

6.4 Loss Functions

Given the heterogeneity of approaches used for PNI detection, a variety of loss functions from machine learning are used to train the U-Net, encoder backbones, XCeption network, and DETR. These are explained below.

$$\text{BCE}(y_i, p_i) = -\frac{1}{N} \sum_{i \in \text{patches}} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \tag{1}$$

$$\text{Dice Loss (Binary)} = 1 - \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \tag{2}$$

$$\text{Bipartite Loss} = \sum_{i=1}^{N} \max\left(0, \text{Margin} - \text{Sim}(\hat{y}_i, y_i)\right) \tag{3}$$

We can infer several important points from the choice of loss functions.

- **Task-specific objectives**: For end-to-end classification [28], (1) is the standard objective. However, given the core-level class imbalance seen in Section 4.2 ($\approx 5\%$ of PNI cores were positive), a successful implementation would require a weighted loss to penalize false negatives more heavily. For segmentation tasks, (2) optimizes the IoU of segmented masks with their ground truths directly, making it invariant to the overwhelming ratio of background pixels in a patch compared to PNI, ensuring the model focuses solely on the accurate delineation of the nerve-tumor boundary.
- **Orthogonalized vs Joint optimization**: All works except [28] and [29] utilize a loss combination technique to link the two different stages leading up to PNI classification together. [28] and [29] are orthogonalized and apply the loss functions independently. Decoupling allows tasks to converge independently without being influenced by each other's signals. This prevents gradient dominance [55], where the easier task (e.g., tumor segmentation) overwhelms the learning signal for the harder task (e.g., nerve segmentation or classification).
- **Bipartite Loss**: While Huber loss (smooth L1) is standard for bounding box prediction tasks, these cannot be utilized for DETRs. Standard object detection uses multiple anchor boxes for every target, and algorithms like non-max suppression are used to derive the final bounding box [56]. In a DETR, there is a one-to-one correspondence between queries and output bounding boxes. Using non-max suppression in a WSI can compress nerves, aggravating their existing sparsity and making them harder to predict.

7 Evaluation

In this section, an overview of the experimental data collected from the literature is presented. First, a comparative table with all metrics is presented, followed by individual meta-analyses of classification and segmentation metrics for each technique.

Table 3: Segmentation Performance of the models

Segmentation Performance of the models				
Reference (Identifier)	[24] (Seg1)	[27] (Seg2)	[28] (Seg3)	[29] (Seg4)
Dataset	• **77 WSIs** from 530 ROIs • Colorectal Cancer • St Mary's Hospital, Korea	• **428 WSIs** (61% train, 39% test) • Pancreatic Cancer • Sourasky Medical Center, Israel	• **8803 Cores** (80% train, 20% test) • Prostate Cancer • STHLM3 Study	• **215 WSIs** (67% train, 6% val, 27% test) • Gastric Cancer • Fujian Cancer Hospital, China
Preprocessing	• Overlapping patches (proportion: 50%) • Background patch elimination	• Synthetic data augmentation using GANs	• Non-overlapping patches (0.25mm)	• Overlapping patches (proportion: 50%) • Background patch elimination
Model and Method	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

Table 4: Classification (PNI Recognition) Performance of the models

Classification (PNI Recognition) Performance of the models					
Reference (Identifier)	[24] (C1)	[25] (C2)	[26] (C3)	[28] (C4)	[29] (C5)
Dataset	• **77 WSIs** from 530 ROIs • Colorectal Cancer • St Mary's Hospital, Korea	• **240 WSIs** (63% train, 12% val, 25% test) • Colon, Prostate, Pancreatic • PAIP Grand Challenge 021	• **240 WSIs** (63% train, 12% val, 25% test) • Colon, Prostate, Pancreatic • PAIP Grand Challenge 021	• **8803 Cores** (80% train, 20% test) • Prostate Cancer • STHLM3 Study	• **215 WSIs** (67% train, 6% val, 27% test) • Gastric Cancer • Fujian Cancer Hospital, China
Preprocessing	• Overlapping patches (proportion: 50%)	• Non overlapping patches (512px) • Boundary Dilation(3px)	• Overlapping patches (proportion: 50%) • Boundary Dilation	• Non-overlapping patches (0.25mm)	• Overlapping patches (proportion: 50%)
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** $\pm$ 0.01 Subjects: **0.96** $\pm$ 0.03	-
F1	0.92	0.416	0.2747	Cores: 0.76 Subjects: 0.80	**0.952**

7.1 Segmentation Metrics Analysis

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.

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

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 $\text{Precision}_{\text{seg}}$, $\text{Sensitivity}_{\text{seg}}$ by summing over all classes C. For binary segmentation, C = 0 (Background) and 1 (Nerve/Tumor/Junction). For multi-class

segmentation, C = 0 (Background), 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 5 below, it is generally observed that tumor segmentation is a harder task than nerve segmentation,
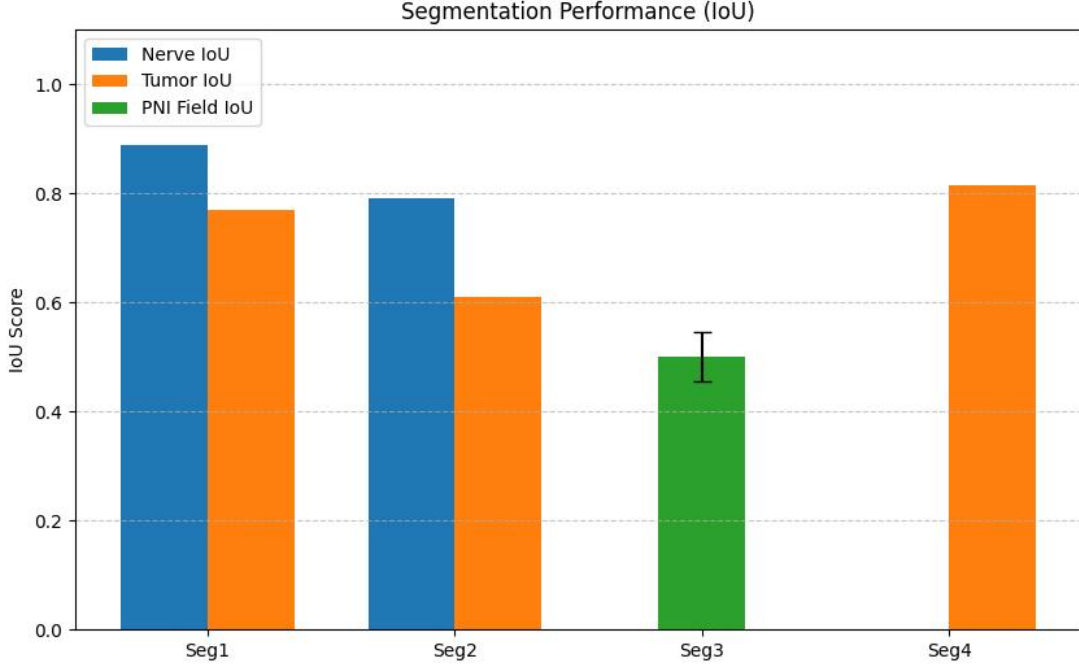

Fig. 5: Segmentation metrics: Please refer to the identifier in Table 3 for X-axis labels.

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 [57]. Complex pipelines with more novel segmentation models (eg, UNet++ and EfficientNet-B3 in [29] compared to the simpler UNet in [28]) 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 tumor fields. Unfortunately, [25] and [26] do not list it as the rules of the PAIP 2021 challenge ranked teams based on the F1 score. We cannot reliably compare the effect of changing the preprocessing methods or increasing the size of the training set on test set performance because these factors do not change in isolation between reviewed literature. Nevertheless, it seems like a drastic increase in training data in [28] does not significantly affect segmentation performance, suggesting that architecture and preprocessing outweigh the availability of raw training data at such scales of high variance and dimensionality. It is also worth noting that [28] was the only study to use non-overlapping patches, which might have contributed adversely to IoU due to the mentioned drawbacks of non-overlapping patching when compared to overlapping patching.

7.2 Classification Metrics Analysis

From Figure 6 below, we see that the highest F1 score is achieved by [29], which uses the rule-based IoB classifier at the end of a multi-stage segmentation and object detection approach. Closely following is [24], which categorizes any overlap in segmented nerve and tumor fields as PNI. [24] is more conservative, avoiding false positives and increasing precision in comparison to [29], even though it uses a less sophisticated classifier. This might be because segmentation masks are generally more precise [58] than bounding boxes, justifying the simpler

---

[‡]Please note that [27] also examines False alarm rate for segmentation. However, it has not been included in this paper as there is no comparison with other literature.
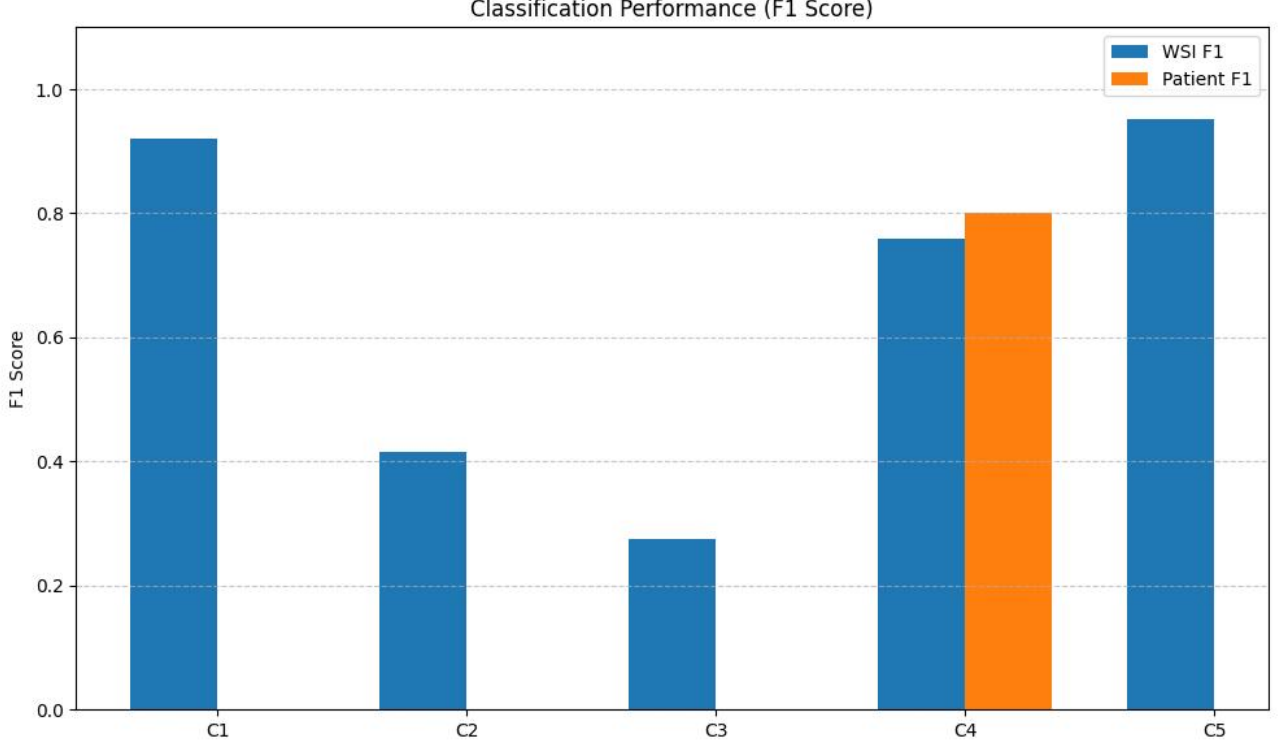
Fig. 6: Classification metrics: Please refer to the identifier in Table 4 for X-axis labels.

approach to discriminating PNI from non-PNI that [24] uses in comparison to [29]. The XCeption ensemble used in [28] directly calculates patch-wise probabilities and extrapolates them to the patient and core level. This is an overtly orthogonalized approach that 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 higher by 0.04. Lastly, [25] and [26] achieve very low F1 scores, likely due to the limited data pool and the inclusion of several types of cancer with different boundary appearances. There is a noticeable difference between the F1 scores of [59] and [25] of about 0.141 points, but a paired-sample t-test could not be performed due to the unavailability of slide-level variance data, the absolute difference (> 0.05) suggests a substantial architectural advantage for the FPN over the standard U-net, because of identical sample size ($n_{test} = 60$ and validation splits. This divergence likely stems from the FPN's superior multi-scale feature aggregation, which is critical for resolving PNI junctions of varying physical sizes in a diverse dataset with different organs.

7.3 Bespoke Metrics - Bounding box detection

As [29] uses a DETR for predicting bounding boxes around nerves, it uses the **mean average precision (mAP)**[58] 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 Section 6, 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 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 [29] 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

14

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.

7.4 Ablation Studies

[‡] An ablation study refers to a 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.

7.4.1 Effect of Transfer Learning

Transfer learning [47] is a learning paradigm wherein a neural network is pre-trained on a large dataset (for example, ImageNet [30]), 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 them using ground-truth annotations as seen in 6. [24] 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, 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.

7.4.2 Boundary Dilation and Loss combination

[26] examines the performance improvement brought about by changing the extent of boundary dilation (see section 5) 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 the 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 a balance between pixel-level accuracy and region-level outcome.

7.4.3 Computational Complexity

[29] examines the performance of various segmentation and object detection models, documenting their floating-point operations per second (FLOPs) and the number of parameters, opening up the 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++ segmentation algorithm, 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-R-CNN, detect nerves at $\approx$ 40 - 50% of the resource load as DETR with only a $\approx$ 5 - 15% reduction in mAP.

7.5 Algorithm Error Analysis

Due to the high stakes involved in medical imaging (and PNI detection, especially so), carrying out error analysis is crucial [23]. The following systematic inaccuracies were observed across [24], [28], [29].

- **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), Mucinous Fibroplasia (in prostate adenocarcinoma only)

Histological tissue with similar structural features to nerves and tumors is 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.

[‡]Please note that [25] 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.

8 Discussion

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

8.1 Combining Image analysis with blood markers

Inflammatory response in cancer is associated with several blood markers; for example, Prostate-specific antigen (PSA) has been shown to play a crucial role in the prognosis of prostate cancer and is associated with inflammation [60]. Likewise, it has been shown that Carcinoembryonic Antigens are a sign of inflammation, including PNI [61]. 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.

8.2 Integration into pathological workflows

Despite 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 Section 7.4.3, it might be more practical to implement a trade-off strategy between computational complexity and model accuracy. This tradeoff will depend on specific clinical scenarios:

- Frozen Sections: Speed is the most important factor here. Tissue is resected when the patient is still under anesthesia for intra-operative decision-making guidance, and a slightly lower F1 for meaningful improvements in inference time can be accepted [62].
- Routine Diagnosis: Diagnosis speed is not as critical here, since models can be run overnight in batches. Architecture that presents maximum sensitivity should be preferred here, at the cost of computational complexity.

Ultimately, there is no single optimal trade-off; the architecture must be tailored to the context in which assistive diagnosis is being conducted. In the second scenario, further refinements to the tradeoffs can be made depending on how many pathologists review the WSIs flagged by the deep learning model. The larger the scrutiny placed by physicians on the segmentation maps, the expected sensitivity can be lowered to headway to optimality, enabling the model to process more batches at a time.

To do this, a systematic sensitivity analysis on a metric representing the objective accuracy (accuracy bearing in mind complexity) of the model will be required, with the level of trade-off stemming from external factors specific to the lab, e.g., the number of samples digitized into WSIs every day, the number of reports submitted by pathologists daily, etc. For example:

$$C = f(L_{Conv}, U_{Conv}, L_{pool}, U_{pool}, ..., model\_size, memory\_usage)$$

$$M = \frac{A}{1 + \lambda C}$$

C, the complexity can be computed as a learned function $f(\cdot)$ of the number of layers of a specific type k, L_k, the number of units in the layer U_k, model storage and memory use, etc. The trade-off metric, M is designed so that $M \propto A$, the model's accuracy (eg: IoU/F1) and $M \propto \frac{1}{C}$. The hyperparameter λ indicates the relative importance of complexity with respect to accuracy, and can be tuned to the aforementioned external factors. $\frac{\delta M}{\delta A} = \frac{1}{1+\lambda C}$ and $\frac{\delta M}{\delta T} = -\frac{\lambda A}{(1+\lambda T)^2}$ can be calculated to evaluate the sensitivity of M when components of C and the accuracy A are varied, offering insight to converge at a workable operating point [63]. Moreover, established strategies like model pruning [64] and quantization [65] could be considered to traverse the tradeoff curve in the routine diagnosis scenario at the margin, trading slight decreases in accuracy for large speedups in continual training and inference. Another avenue would involve cost-benefit analyses of running PNI detection in GPU-accelerated environments. For instance, the authors of [66] perform cost-benefit analyses for brain MRI diagnosis using a variety of system-on-module chips and smartphones as edge inferencing devices.

8.3 Explainability and inter-observer agreement

A critical element of deployment in a pathology lab is pathologist-friendliness. The predictions of the model should be explainable through visual heat maps displaying areas that activated the model's output, in case of a PNI positive outcome. These insights are valuable even when pathologists can see the segmentation maps as specific edges, textures, and colours that the model diagnoses as nerve, tumour, or a PNI junction can be visualized. Saliency maps [67] are a good tool to use here.

$$S(x) = \left| \frac{\delta M(x)}{\delta x} \right| \qquad S_{norm}(x) = \frac{S(x) - min(S(x))}{max(S(x)) - min(S(x))}$$

$S(x)$ represents the saliency map of the patch x. This is the gradient of the output of the segmentation or classification model $M(x)$ with respect to the input image. The partial derivative helps us identify regions that are most sensitive to changes in input pixel values (and thus, have a high gradient). They are typically normalized into $S_{norm}(x)$. Grad-CAM [68] implements a similar technique, albeit by calculating gradients with respect to the last convolutional layer to produce intermediate weights, then multiplying them with the output of the last convolutional layer to produce the heat map. Grad-CAM provides better insight into how models use spatial arrangements from a higher level of abstraction (the last layer) and can offer highly relevant insight for PNI detection. Grad-CAM and explainable AI techniques can improve pathologists' trust in two key ways.

- Firstly, heatmaps allow pathologists to *sanity-check* the model's output, ensuring that it is activated by PNI junctions or nerve-tumor overlap, not artifacts from scanning the WSI.
- Secondly, the pathologist is immediately directed to the invasion site; the model acts as a digital pointer rather than a black box replacement for PNI detection.

Tuning inter-observer agreement is a strategy to improve concordance with pathologists and mimic their thought process. Cohen's Kappa is a measure of inter-observer agreement [69]:

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

p_o measures 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 raters' decisions. κ is a robust statistical measure that can be explicitly referenced in the loss 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 favour of the optimizing objective. In fact, [28] 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.

8.4 Noise in input data from preserved pathological samples

Most pathology labs use sequential sample processing. In high-volume labs, improper preservation and digitization after a non-trivial storage period may change the distribution of data, making it different from the training distribution. Sometimes, samples need to be re-digitized after a few months as well, which poses the same problem. The algorithm must be made noise-resistant. While data augmentation assists to an extent by introducing random noise through colour and contrast adjustment, more systematic solutions governing data collection and pre-processing must be considered. It would be beneficial to collect temporal data and curate a dataset containing PNI WSIs in different stages of degradation, which the model can familiarize itself with. Neural networks have also been applied to image restoration [70]. A small auxiliary model can be trained to restore images older than a pre-decided span. Denoising algorithms, such as Gaussian filters, are also feasible due to their efficient 1D convolution implementation [70].

8.5 Different staining and scanning techniques

While this section does not contain implementation avenues, it is still an interesting prospect. 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 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.

Moreover, WSIs stained with just (H&E) are also prone to domain shift, against which current methods are not robust. Switching scanner types can alter the histogram of a digital WSI, altering the appearance and distribution of data. Future studies should explore methods such as stain normalization [71] to improve protected models from shifting input distributions during training as well as inference.

8.6 Pathways to Clinical Translation

Translating high-performance PNI detection models from computational benchmarks to routine pathology workflows requires navigating distinct regulatory, economic, and validation hurdles.

8.6.1 Regulatory Frameworks and Life Cycle Management

Unlike static medical devices, AI algorithms drift as clinical practices evolve. Securing FDA clearance (e.g., via the 510(k) pathway) now increasingly relies on the *total product life cycle approach*. Recent FDA guidance [72] emphasizes the use of predetermined model updates, such as re-training on new scanner data, without requiring resubmission and approval for every iteration. This regulatory agility is critical for PNI models, which must adapt to the domain shift caused by varying staining protocols across institutions.

8.6.2 Economic Viability

For wide-scale adoption, PNI detection tools must demonstrate cost-effectiveness, not just accuracy. A key translational pathway is deploying AI as a "negative filter" rather than a primary diagnostician. Recent micro-simulation studies in prostate pathology suggest that AI-assisted workflows can reduce pathologist workload by up to 80% by automatically filtering benign cores [73], yielding significant cost savings without compromising patient safety. By positioning the PNI detector as a triage tool that flags only high-probability regions for manual review, the economic barrier to entry is substantially lowered.

8.6.3 Multi-Center Validation

A high F1-score for a single center is insufficient for clinical deployment due to the lack of scanner invariance. As noted in Section 8.5 and [71], algorithms often fail when transferred between scanner vendors (e.g., Aperio vs. Hamamatsu) due to subtle differences in color calibration and resolution. Future studies should prioritize multi-center prospective validation to ensure the model's decision boundary is robust to these hardware-induced variations, a prerequisite for regulatory approval.

9 Conclusion

This survey highlights the immense potential of DL in revolutionizing PNI detection, a critical component in cancer prognosis. By offering a structured analysis of existing models, datasets, and evaluation metrics, this paper lays the foundation for future research aimed at translating these models into clinical practice. Moving forward, efforts should focus on developing more comprehensive datasets encompassing a larger number of cancer types and sample preservation parameters, improving model interpretability and concordance with pathologists, ensuring that these AI tools can be integrated into real-world workflows with cost-benefit analyses for hospitals and scalable models, and seeking avenues for regulatory approval through multi-center studies conducted over large periods that validate the model invariance. The next step is for researchers and clinicians to collaborate closely, ensuring that AI-driven diagnostic tools are both effective and ethically implemented for the benefit of patient care.

Acronyms

AI Artificial Intelligence
AUC-ROC Area Under the Receiver Operating Characteristic Curve
BCE Binary Cross Entropy
CNN Convolutional Neural Network
CT Computed Tomography

CXR Chest X-Ray
DETR Detection Transformer
DL Deep Learning
ECG Electrocardiogram
EEG Electroencephalogram
EHR Electronic Health Records
FDA Food and Drug Administration
FLOPs Floating Point Operations per Second
FN False Negative
FP False Positive
FPN Feature Pyramid Networks
GAN Generative Adversarial Networks
GNN Graph Neural Network
GPU Graphics Processing Unit
Grad-CAM Gradient-weighted Class Activation Mapping
H and E Hematoxylin and Eosin Stain
HCA Hierarchical Contextual Analysis
IoB Intersection over Bounding Box
IoU Intersection over Union
mAP Mean Average Precision
MeSH Medical Subject Headings
MRI Magnetic Resonance Imaging
NLP Natural Language Processing
NPV Negative Predictive Value
PAIP Pathology in Artificial Intelligence Platform
PNI Perineural Invasion
PPV Positive Predictive Value
PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA Prostate-Specific Antigen
R-CNN Region-based Convolutional Neural Network
ROI Region of Interest
TN True Negative
TP True Positive
TTA Test Time Augmentation
ViT Vision Transformer
WSI Whole Slide Image
YOLO You Only Look Once

Declarations

Conflict of Interest statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical Approval and Consent to Participate

Not applicable

Consent for Publication

Not applicable

Authors' Contributions

- Author 1 (Hayagreev Jeyandran): Survey of literature and meta-analysis, content writing, image production, ideation of future work.
- Author 2 (Dr Veena R, MD): Topic proposal, overview, proofreading, ideation of future work, mentorship for survey and content writing.

Funding

Availability of Data and Materials

This article is a systematic review, and no primary datasets were generated or analyzed during this study. All data used for this review are publicly available through the articles cited in the references section. The studies referenced can be accessed via their respective publishers' platforms or repositories. Detailed information regarding each dataset, methodologies, and findings can be found in the cited literature.

References

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

[2] King, R.: 6. Peripheral Nerve Disorders, pp. 32–37. John Wiley & Sons, Ltd, Chichester, UK (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)

[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) https://doi.org/10.1111/1759-7714.13960

[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) https://doi.org/10.3390/jpm12060962

[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) https://doi.org/10.1002/1097-0142(19940201)73:3⟨550::aid-cncr2820730309⟩3.0.co;2-0

[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) https://doi.org/10.4251/wjgo.v12.i5.592

[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) https://doi.org/10.1007/s12105-016-0722-9

[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) https://doi.org/10.1097/PAS.0000000000000422

[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) https://doi.org/10.1002/cncr.25620

[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) https://doi.org/10.1007/s12105-021-01321-9

[16] Al-Waeli, S.H.A., Abdul Hameed, H.N.b.: AI in medical imaging & diagnostics: Review. Wasit Journal for Pure Science **4**(3), 35–45 (2025) https://doi.org/10.31185/wjps.752

[17] Pattanaik, S., Sen, R.K.: A review of multi-phase intelligent system (MPIS) for diagnosing and screening malignant breast cancer. In: 2023 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC) (2023). https://doi.org/10.1109/ICAIHC59020.2023.10431433 . https://ieeexplore.ieee.org/document/10431433

[18] Sondkar, S., Apte, S., Deshmukh, A., Chavan, P.: Machine learning based approach for myocardial infarction prediction using real time ecg signals. In: 2024 Intelligent Systems and Machine Learning Conference (ISML), pp. 383–388 (2024). https://doi.org/10.1109/ISML60050.2024.11007445

[19] Schneider, J., Vlachos, M.: A Survey of Deep Learning: From Activations to Transformers (2024). https://doi.org/10.48550/arXiv.2302.00722

[20] 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) https://doi.org/10.1371/journal.pone.0192726

[21] 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

[22] 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

[23] 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

[24] 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

[25] Nateghi, R., Pourakpour, F.: Perineural Invasion Detection in Multiple Organ Cancer Based on Deep Convolutional Neural Network (2021). https://doi.org/10.48550/arXiv.2110.12283

[26] 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

[27] 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

[28] 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) https://doi.org/10.1007/s00428-022-03326-3

[29] 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

[30] 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

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

[32] 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/

[33] 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) https://doi.org/10.1016/j.euf.2016.10.009

[34] Johnson, A.E.W., Pollard, T.J., Berkowitz, S.J., Greenbaum, N.R., Lungren, M.P., Deng, C.-y., Mark, R.G., Horng, S.: MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Scientific Data **6**(1), 317 (2019) https://doi.org/10.1038/s41597-019-0322-0

[35] Kaggle, EyePACS: Diabetic Retinopathy Detection. Kaggle Competition (dataset mirrored on Hugging Face by ctmedtech). Accessed 2026-02-01. (2015). https://huggingface.co/datasets/ctmedtech/EYEPACS

[36] Nechaev, D., Pchelnikov, A., Ivanova, E.: HISTAI: An Open-Source, Large-Scale Whole Slide Image Dataset for Computational Pathology (2025). https://arxiv.org/pdf/2505.12120

[37] Kondylakis, H., Kalokyri, V., Sfakianakis, S., Marias, K., Tsiknakis, M., Jimenez-Pastor, A., Camacho-Ramos, E., Blanquer, I., Segrelles, J.D., López-Huguet, S., Barelle, C., Kogut-Czarkowska, M., Tsakou, G., Siopis, N., Sakellariou, Z., Bizopoulos, P., Drossou, V., Lalas, A., Votis, K., Mallol, P., Marti-Bonmati, L., Alberich, L.C., Seymour, K., Boucher, S., Ciarrocchi, E., Fromont, L., Rambla, J., Harms, A., Gutierrez, A., Starmans, M.P.A., Prior, F., Gelpi, J.L., Lekadir, K.: Data infrastructures for ai in medical imaging: a report on the experiences of five eu projects. European Radiology Experimental **7**(1), 20 (2023) https://doi.org/10.1186/s41747-023-00336-x

[38] 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

[39] Cossio, M.: Augmenting Medical Imaging: A Comprehensive Catalogue of 65 Techniques for Enhanced Data Analysis (2023). https://doi.org/10.48550/arXiv.2303.01178

[40] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks (2014). https://doi.org/10.48550/arXiv.1406.2661

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

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

[43] Nadella, G.S., Addula, S.R., Yadulla, A.R., Sajja, G.S., Meesala, M., Maturi, M.H., Meduri, K., Gonaygunta, H.: Generative AI-enhanced cybersecurity framework for enterprise data privacy management. Computers **14**(2), 55 (2025) https://doi.org/10.3390/computers14020055

[44] Alhoraibi, L., Alghazzawi, D., Alhebshi, R.: Generative adversarial network-based data augmentation for enhancing wireless physical layer authentication. Sensors **24**(2), 641 (2024) https://doi.org/10.3390/s24020641

[45] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation (2015). https://doi.org/10.48550/arXiv.1505.04597

[46] Lin, M., Chen, Q., Yan, S.: Network In Network (2014). https://doi.org/10.48550/arXiv.1312.4400

[47] Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? (2014). https://doi.org/10.48550/arXiv.1411.1792

[48] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going Deeper with Convolutions (2014). https://doi.org/10.48550/arXiv.1409.4842

[49] Tan, M., Le, Q.V.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks (2020). https://doi.org/10.48550/arXiv.1905.11946

[50] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature Pyramid Networks for Object Detection (2017). https://doi.org/10.48550/arXiv.1612.03144

[51] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: UNet++: A Nested U-Net Architecture for Medical Image Segmentation (2018). https://doi.org/10.48550/arXiv.1807.10165

[52] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-End Object Detection with Transformers (2020). https://doi.org/10.48550/arXiv.2005.12872

[53] Chollet, F.: Xception: Deep Learning with Depthwise Separable Convolutions (2017). https://doi.org/10.48550/arXiv.1610.02357

[54] Tanis, J.H., Giannella, C., Mariano, A.V.: Introduction to Graph Neural Networks: A Starting Point for Machine Learning Engineers (2024). https://doi.org/10.48550/arXiv.2412.19419 . https://arxiv.org/abs/2412.19419

[55] Zhou, Y., Liang, Y.: Characterization of Gradient Dominance and Regularity Conditions for Neural Networks (2017). https://doi.org/10.48550/arXiv.1710.06910 . https://arxiv.org/abs/1710.06910

[56] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks (2016). https://doi.org/10.48550/arXiv.1506.01497

[57] 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

[58] 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). https://doi.org/10.48550/arXiv.1405.0312

[59] Bakst, R.L., Wong, R.J.: Mechanisms of perineural invasion. J. Neurol. Surg. B Skull Base **77**(2), 96–106 (2016) https://doi.org/10.1055/s-0036-1571835

[60] 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

[61] 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

[62] Jaafar, H.: Intra-operative frozen section consultation: Concepts, applications and limitations. The Malaysian Journal of Medical Sciences : MJMS **13**(1), 4–12 (2006)

[63] Fernández-Moreno, M., Lei, B., Holm, E.A., Mesejo, P., Moreno, R.: Exploring the trade-off between performance and annotation complexity in semantic segmentation. Engineering Applications of Artificial Intelligence **123**, 106299 (2023) https://doi.org/10.1016/j.engappai.2023.106299

[64] Han, S., Pool, J., Tran, J., Dally, W.J.: Learning both Weights and Connections for Efficient Neural Networks (2015). https://doi.org/10.48550/arXiv.1506.02626

[65] Krishnamoorthi, R.: Quantizing deep convolutional networks for efficient inference: A whitepaper (2018).

[10.48550/arXiv.1806.08342](https://doi.org/10.48550/arXiv.1806.08342)

[66] Memon, K., Yahya, N., Yusoff, M.Z., Remli, R., Mustapha, A.M.M., Hashim, H., Ali, S.S.A., Siddiqui, S.: Edge computing for AI-based brain MRI applications: A critical evaluation of real-time classification and segmentation. Sensors **24**(21), 7091 (2024) https://doi.org/10.3390/s24217091

[67] Zeiler, M.D., Fergus, R.: Visualizing and Understanding Convolutional Networks (2013). https://doi.org/10.48550/arXiv.1311.2901

[68] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision **128**(2), 336–359 (2019) https://doi.org/10.1007/s11263-019-01228-7

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

[70] Thakkar, J.D., Bhatt, J.S., Patra, S.K.: Self-supervised learning for medical image restoration: Investigation and finding. In: Singh, P., Singh, D., Tiwari, V., Misra, S. (eds.) Machine Learning and Computational Intelligence Techniques for Data Engineering, pp. 541–552. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-0047-3_46

[71] Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, X., Schmitt, C., Thomas, N.E.: A method for normalizing histology slides for quantitative analysis. In: Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI) (2009). https://www.cs.unc.edu/~mn/sites/default/files/macenko2009.pdf

[72] How the FDA Reviews AI and Machine Learning Medical Devices: Complete 2025 Guide. Complizen (online article). https://www.complizen.ai/post/fda-ai-machine-learning-medical-devices-review-2025

[73] Du, X., Hao, S., Olsson, H., Kartasalo, K., Mulliqi, N., Rai, B., Menges, D., Heintz, E., Egevad, L., Eklund, M., Clements, M.: Effectiveness and cost-effectiveness of AI-assisted pathology for prostate cancer diagnosis in sweden: A microsimulation study. European Urology Oncology **8**(1), 80–86 (2025) https://doi.org/10.1016/j.euo.2024.05.004