M2 IMA (Traitement avancé d'image et vision)

SORBONNE UNIVERSITÉ

Cerberus in Computational Pathology: Review, Optimization, and Transfer Learning Applications

Done by :
Oussama RCHAKI
Supervised by :
Pr. Vannary MEAS-YEDID HARDY

09/02/25

Contents

1 – BIBLIOGRAPHY

1.1 – Computational Pathology: Objectives and Challenges

Computational Pathology (CPath) is an emerging and transformative field that leverages machine learning and advanced computational techniques to analyze pathology data, aiming to enhance diagnostic accuracy, disease prognosis, and treatment planning. The primary objectives of computational pathology are to augment human expertise, improve the reproducibility of analyses, and reduce the variability inherent in manual interpretation of pathology images. However, despite its potential, CPath faces a range of significant challenges that must be addressed for its widespread and reliable adoption. Drawing on insights from recent studies, this detailed overview will explore the objectives and challenges of computational pathology, including its impact and potential in medical practice.

Objectives of Computational Pathology

1. Enhanced Diagnostic Accuracy and Reproducibility

One of the primary objectives of computational pathology is to increase the accuracy and consistency of diagnostic assessments. Traditional histopathological analysis relies on visual inspection by pathologists, which, while expert-driven, is subject to inter-observer variability. As described by *Gamper et al.* (2020) in their exploration of multi-task learning models in histopathology, the integration of computational approaches enables the development of models that can simultaneously perform multiple predictive tasks. These models contribute to widely generalizable frameworks capable of analyzing complex tissue structures and identifying pathologies that might be challenging for human observers.

2. Scalability and Efficiency in Pathology Workflows

Another key objective of CPath is the automation of labor-intensive tasks, enhancing efficiency and throughput in pathology workflows. *Graham et al.* (2023) highlight in their survey that computational tools can process vast amounts of histological data rapidly, facilitating high-throughput analysis and prioritizing cases that may require urgent attention. This capability is particularly valuable in large-scale screening programs where pathologist resources are stretched.

3. Personalized Medicine and Prognostic Insights

CPath has a significant role in advancing personalized medicine. According to the work by *Chen et al.* (2022), computational pathology not only aids in the accurate classification of tissue samples but also correlates image data with clinical outcomes. By leveraging tissue analysis algorithms, CPath can contribute to the development of tailored treatment plans based on patient-specific histological and molecular characteristics.

4. Integration with Multi-Modal Data for Comprehensive Analysis

A particularly compelling objective is the integration of pathology images with other data modalities, such as genetic and clinical data. *Zheng et al.* (2024) illustrate the importance of multi-modal fusion, using graph attention-based algorithms to merge pathology imaging data with gene expression profiles. This approach enhances the prediction of patient outcomes, demonstrating the potential of CPath to deliver more comprehensive and precise cancer prognoses.

5. Facilitating Multidisciplinary Collaboration

An essential objective of CPath is to foster collaboration among experts from various fields, such as pathologists, data scientists, and software engineers, to enhance the development and implementation of AI tools. Wahab et al. (2021) emphasize that leveraging the combined expertise of these disciplines can lead to the creation of robust and clinically relevant AI solutions. Such multidisciplinary approaches ensure that computational pathology tools are designed with practical insights, aligning with real-world medical needs and workflows, thus promoting more effective and impactful use in clinical practice.

Challenges in Computational Pathology

1. Data Heterogeneity and Generalizability

One of the most critical challenges in computational pathology is ensuring that models trained on specific datasets generalize well across diverse populations and institutions. *Gamper et al.* (2020) and *Graham et al.* (2023) emphasize the need for multi-task learning frameworks that enhance generalizability by sharing representations across related tasks. However, variability in staining techniques, scanning protocols, and patient demographics can hinder model performance when applied outside the training domain.

2. Labeling and Annotation Bottlenecks

The reliance on high-quality annotated data for supervised learning poses a significant challenge. *Bejnordi et al.* (2017) point out that obtaining expertly annotated data, such as those used in training deep learning algorithms for lymph node metastasis detection, is time-consuming and resource-intensive. This bottleneck can limit the scalability of training robust models capable of high diagnostic accuracy.

3. Complexity of Tissue Structures

The morphological complexity of tissues adds another layer of difficulty. *Veeling et al.* (2018) discuss how digital pathology tasks require models that account for variations in tissue orientation and structure. Their work on rotation-equivariant convolutional neural networks (CNNs) highlights a method to address this issue by enabling models to recognize tissue patterns regardless of orientation. However, this approach is still computationally intensive and requires further optimization for practical deployment.

4. Regulatory and Ethical Concerns

The deployment of CPath in clinical settings is also hindered by regulatory and ethical issues. *Graham et al.* (2023) underscore the necessity for comprehensive validation of these models in real-world clinical environments to ensure their safety and reliability. Moreover, the use of patient data, particularly when integrating multi-modal datasets, raises concerns about privacy and data governance.

5. Interpretability and Transparency: The Need for Explainable Models

The 'black box' nature of deep learning models remains a significant challenge, as noted by *Chen et al.* (2022). While these models achieve high accuracy, their decision-making processes are often opaque, making it difficult for pathologists and medical practitioners to trust their outputs. To enhance clinician acceptance, it is essential to develop interpretable models that provide insight into the features contributing to diagnostic decisions. Consequently, the need for explainable AI (XAI) within diagnostic applications has become another significant hurdle for the CPath domain.

Although deep learning models demonstrate high accuracy in identifying and classifying pathologies, their opacity can limit their acceptance in clinical settings. For diagnostic purposes, AI models must deliver clear, interpretable outputs that allow pathologists and medical practitioners to understand the rationale behind predictions. Such transparency not only builds trust but also helps in identifying potential errors or biases within the models. Therefore, developing XAI techniques that can highlight relevant tissue features or provide logical justifications for decisions is crucial for the successful integration of AI tools in real-world diagnostic workflows.

6. Application within Developing Countries

The application of CPath in developing countries presents unique challenges and opportunities. A major challenge facing the wider adoption of AI, particularly in resource-limited settings, is the robustness of algorithms due to the lack of well-curated, multi-centric, high-quality datasets. Studies have shown that even in developed regions, the clinical applicability of AI can be limited due to weak experimental design and the vulnerabilities of deep learning models (*Gamper et al.* , 2020b).

Importance and Promises of CPath

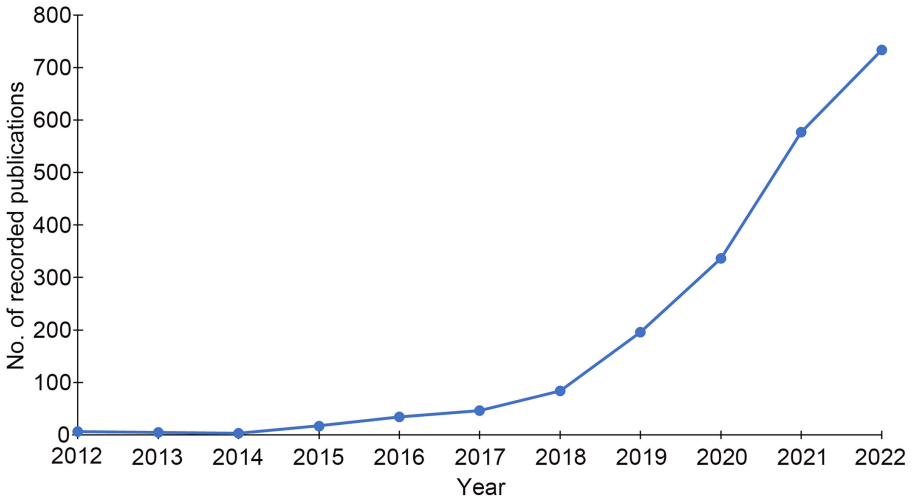


Figure 1 : Number of research publications in AI-based computational pathology recorded in PubMed from 2012 to 2022 (*Asif, A. (2023)*).

The interest in the field of CPath is increasing exponentially year after year (Figure 1). The future of computational pathology is promising, with potential benefits spanning both clinical and research domains. Research in this field is frequently combined with recent advancements and discoveries in medicine and biology (*Veeling et al.*, 2018), as well as in the deep learning field. For example, *Zheng et al.* (2024), as mentioned previously, underscore the transformative potential of CPath when combined with gene expression and other molecular data, leading to more nuanced insights into cancer survival predictions and treatment strategies. Additionally (in the same article), advances in graph-based algorithms and attention mechanisms illustrate how CPath can evolve to provide more personalized and detailed prognostic models.

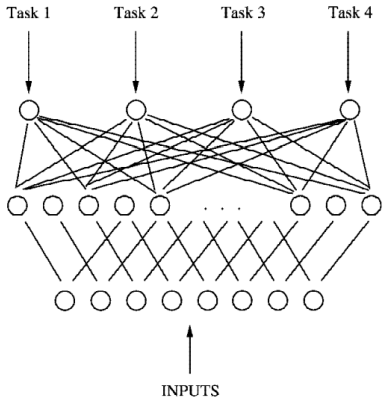
Despite its challenges, CPath is positioned to become a cornerstone of modern medical practice. Through continuous research, advancements in model architectures, data standardization, and regulatory frameworks will pave the way for robust, interpretable, and generalizable solutions that enhance diagnostic capabilities and patient outcomes. Addressing these challenges will be critical to ensuring that CPath fulfills its potential and becomes a seamless extension of clinical pathology.

1.2 – History of Multi-Task Learning (MTL)

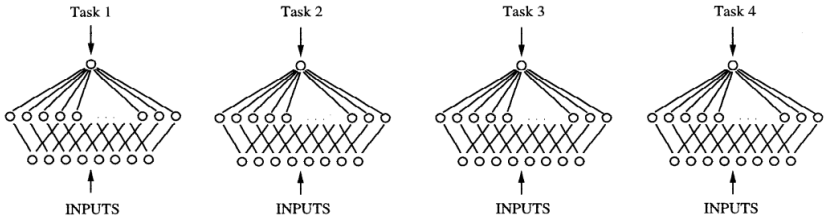
The model presented in this work (**Cerberus**) is a Multi-task Learning (MTL) model. Thus, it is important to define MTL, which has a rich history rooted in the evolution of machine learning methodologies aimed at enhancing model performance through simultaneous training on multiple related tasks. This approach takes advantage of shared information and similarities among tasks to improve generalization and performance across individual tasks. The history and development of MTL illustrate continuous refinement and optimization, leading to significant advancements in various domains, including computational pathology.

Early Foundations of MTL

The concept of MTL was formally introduced by Rich Caruana in 1997, who demonstrated that training a model to perform several tasks concurrently could lead to improved performance on each individual task due to the shared representations among them (*Caruana, 1997*). Caruana's pioneering work laid the groundwork for the application of MTL in machine learning by showing that the inductive bias inherent in multi-task learning helps models learn general features that are useful across tasks. This discovery



(A) Multitask Backpropagation (MTL) of four tasks with the same inputs.



(B) Single Task Backpropagation (STL) of four tasks with the same inputs.

FIGURE 2 : MTL and STL architecture (*Caruana, (1997)*).

highlighted the potential of MTL to reduce the risk of overfitting on smaller datasets, as knowledge transfer from related tasks could bolster the model's robustness.

In Figure 2, an example is provided that shows a single network with the same inputs as the four networks in Figure 2b. In Figure 2a, backpropagation is performed in parallel on the four outputs of the MTL network. Because the four outputs share a common hidden layer, the internal representations that arise in the hidden layer for one task can be utilized by other tasks. **Sharing what is learned across different tasks while training them in parallel is the central concept of multi-task learning**.

Evolution and Advancements in MTL

As machine learning evolved, MTL methodologies were refined and expanded. *Zhang and Yang (2017)* provided a comprehensive survey on the development of MTL, emphasizing that early implementations were primarily based on linear models or shallow neural networks. The advent of deep learning, however, introduced new dimensions to MTL, enabling the design of more complex and deeper shared architectures capable of learning from multi-modal and multi-scale data. *Ruder (2017)* further detailed the integration of MTL into deep neural networks, discussing how shared and task-specific layers could be structured to optimize performance for multiple outputs. The evolution of MTL has included improvements in network architectures, such as shared bottom layers with task-specific top layers, attention mechanisms for focusing on relevant task-specific information, and parameter-sharing techniques that adapt dynamically based on task relationships. In this sense *Tellez et al. (2020)* made significant contributions by exploring advanced MTL architectures, including the use of parallel pathways, cross-stitch networks, and configurations with shared and private layers. These sophisticated structures help mitigate task interference, where the learning of one task may negatively impact another. Their work demonstrated the potential of combining unsupervised learning with MTL for applications such as image compression and analysis, showcasing the flexibility and adaptability of MTL architectures.

In relation to computational pathology, *Chen et al. (2022)* discussed the application of MTL within this research domain, particularly in addressing the challenges associated with tissue analysis. The imple-

mentation of MTL in digital pathology has enabled models to handle multiple subtasks simultaneously (3), such as tissue segmentation and classification, fostering comprehensive models capable of delivering more holistic diagnostic processes.

Another noteworthy example of MTL in medical image analysis is presented by *Graham et al. (2019)*, who introduced Hover-Net, an MTL approach that simultaneously performs segmentation and classification of nuclei in histopathology images. This work exemplifies how MTL can be leveraged to enhance the granularity and comprehensiveness of diagnostic models, underscoring its importance in advancing medical imaging and computational pathology.

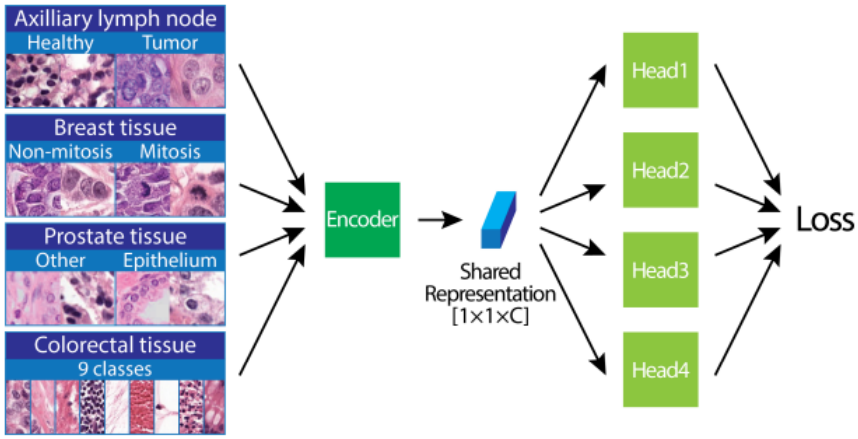


FIGURE 3 : Example of supervised multitask learning framework. Left: the full model is trained to solve four different tasks simultaneously. Center: the encoder provides a shared embedded representation for the images of all the tasks. Right: the head models perform each of the four classification tasks independently from each other (*Chen et al. (2022)*).

All these works show the importance of MTL techniques in CPath, underscoring their ability to handle complex, multi-dimensional tasks efficiently and effectively. The continuous evolution of MTL methodologies—from simple shared architectures to sophisticated networks capable of dynamic task balancing and parallel processing—has paved the way for breakthroughs in computational pathology and other fields. The integration of MTL in digital pathology has demonstrated how models can simultaneously perform tasks such as segmentation, classification, and anomaly detection, leading to more comprehensive and accurate diagnostic tools.

1.3 – MTL FOR TRANSFER LEARNING

In our case, the MTL model under study is evaluated on closely related tasks within the CPath framework, building on a foundation explored in prior work. Multi-Task Learning (MTL) and Transfer Learning are longstanding areas of interest in machine learning, with extensive research highlighting their complementary strengths and real-world applications.

Transfer Learning, in particular, empowers models to apply knowledge gained from one task—often trained on large, diverse datasets—to related tasks with limited data availability. For instance, in computer vision, models initially trained on comprehensive datasets like ImageNet can be adapted for more specialized applications, such as medical imaging, where labeled data is sparse. This technique is invaluable for its efficiency, reducing the need for extensive labeled data and the computational cost of training models from scratch. As a result, Transfer Learning accelerates model training and enhances performance, especially in contexts where data is scarce.

Transfer learning intersects with Multi-Task Learning (MTL) by enabling knowledge sharing across tasks to improve model generalization. MTL aims to train a single model on multiple tasks simultaneously, fostering shared representations that benefit each task. Through transfer learning, MTL can leverage knowledge from high-resource tasks to boost performance on related low-resource tasks. A notable example is *meta-transfer learning*, which combines transfer learning and meta-learning to enhance few-shot learning performance.

Several recent works highlight this synergy between transfer learning and MTL. For instance, *Sun et al. (2019)* proposed a framework for meta-transfer learning, allowing the transfer of meta-knowledge between tasks. Similarly, *Ruder (2021)* discussed how sharing information between tasks improves learning efficiency in MTL, with transfer learning techniques enhancing the transferability of learned representations across tasks. *Zhang and Yang (2022)* provided a comprehensive overview of MTL, emphasizing how transfer learning facilitates cross-task knowledge transfer, reducing training demands and optimizing model performance.

In medical imaging, *Graham et al. (2019)* demonstrated practical applications of MTL through simultaneous segmentation and classification tasks. Here, transfer learning helps handle diverse tissue types, improving the accuracy of both segmentation and classification tasks. Additionally, foundational research by *Glorot and Bengio (2010)* addressed challenges in training deep networks, noting that transfer learning overcomes data and training limitations by initializing models with pretrained weights.

1.4 – Multi-Label Learning (MLL) for CPath

Multi-Label Learning (MLL) is a machine learning approach where each instance in a dataset is associated with multiple labels simultaneously, meaning that each data point can belong to several categories at once. This approach is particularly effective in applications where multiple classes or conditions might co-occur within a single instance. For example, in image tagging, a single image might be tagged with several labels like "tree", "lion" and "elephant" (Figure 4), or in medical imaging, a scan might reveal multiple co-existing conditions.

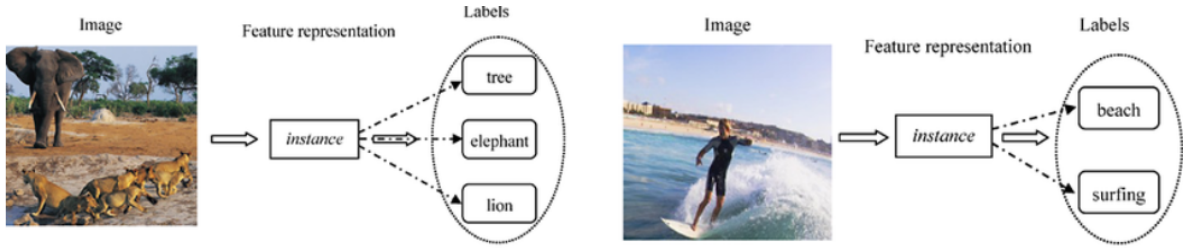


Figure 4 : Multi-Label Learning examples (*Hou, C. (2012)*).

In contrast to MTL, where tasks are typically distinct and may use separate datasets, MLL operates within a unified dataset where all labels apply to the same instance. Therefore, MLL is well-suited to tasks where label co-occurrence and interdependence significantly affect prediction outcomes. For instance, in complex image analysis scenarios, such as medical imaging, MLL can capture the nuanced relationships between structures within the same image. *Shephard et al. (2021)* applied MLL in the simultaneous nuclear instance and layer segmentation of histological images, where identifying each structure within the same data point was critical. Additionally, in the work by *Fraz et al. (2020)*, MLL proved advantageous in segmenting both microvessels and nerves in routine histology images of oral cancer. These studies show that MLL effectively leverages label co-existence to identify different tissue types simultaneously, underscoring its strength in handling multiple labels within a single dataset. This capacity to predict multiple labels per instance enables MLL to capture complex dependencies, which is especially valuable when these dependencies influence the accuracy of individual label predictions.

While MLL can be thought of as a special case of MTL when each label is treated as a task, it differs fundamentally in its data structure and approach. In MLL, all labels refer to the same instance, and the primary goal is to manage and predict the interrelated labels within that shared context. *Zhang and Zhou (2013)* note that this shared dataset structure is a defining feature of MLL, as all tasks (labels) relate to the same input data, unlike MTL, where tasks can span separate datasets.

1.5 – Datasets for CPath MLT: Nuclei, glands, lumina and tissue type classification dataset

Cerberus includes histological images, which are key resources for developing and training models for tasks such as segmentation and classification in medical imaging. Working with such datasets presents

distinct challenges, particularly in terms of finding suitable data and adapting it to meet the requirements of our model training. To aid understanding, definitions for each key biological structure targeted by our model are provided below, highlighting their biological roles and significance for readers who may be less familiar with this type of data.

Nuclei are the control centers of cells, containing genetic material (DNA) and coordinating cellular activities. In histopathology, nuclei segmentation is vital because variations in the shape, size, and density of nuclei can indicate disease progression, such as cancerous changes. Accurate identification and segmentation of nuclei are essential for understanding tissue organization and for diagnostic purposes in various cancers.

Glands are structures composed of a group of cells that produce and release substances, such as enzymes or hormones. In tissues like the colon (Figure 5), glands have a distinct tubular or circular organization. Gland segmentation is important in histology because abnormal glandular architecture, such as irregular shapes or disrupted arrangement, can signal the presence of diseases like colorectal cancer. Glands are often studied in conjunction with their surrounding structures to provide insights into tissue health and disease state.

Lumina are the hollow spaces within glands, formed by the arrangement of glandular cells around a central cavity. These spaces are crucial for the transport and secretion of glandular products. In histological images, the size and shape of lumina are often indicative of normal or abnormal gland function, and changes in luminal structure can be associated with various pathological conditions. Detecting lumina helps in identifying the orientation and integrity of gland structures, which is particularly useful in cancer diagnosis and tissue assessment.

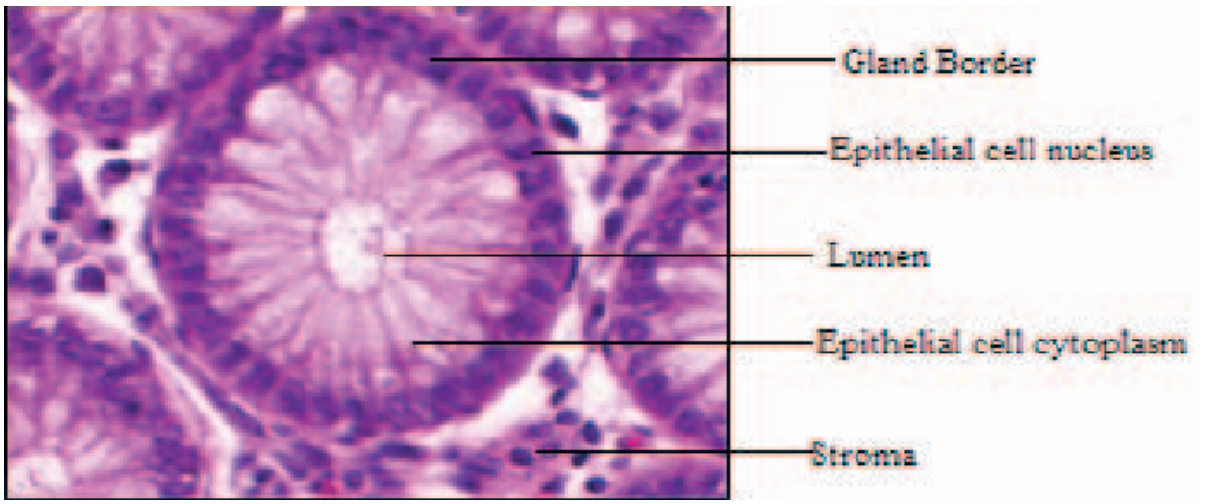


FIGURE 5 : Gland unit of colon tissue (*Banwari, M. (2016)*).

Datasets in histology, such as those used in gland segmentation and nuclei classification, are essential benchmarks in the field. For example, *Graham et al. (2019a)* proposed the MILD-net, a deep learning model designed to minimize information loss in gland segmentation tasks. Similarly, *Sirinukunwattana et al. (2017)* introduced the Gland Segmentation in Colon Histology (GlaS) dataset, which has become a significant resource in gland segmentation research and exemplifies the challenges in creating large, annotated datasets necessary for clinical implementation.

Adapting these datasets for model training is also challenging, as traditional segmentation methods may not be directly applicable to specific biological contexts in **Cerberus**. For instance, *Chen et al. (2017)* developed DCAN (Deep Contour-Aware Networks) to perform instance segmentation in histology images, a task that requires intricate contour information often not readily available in generic datasets. Also, the DigestPath dataset by *Da et al. (2022)* provides a benchmark for digestive-system pathology, illustrating the challenge in curating datasets with complex biological variation.

Moreover, *Graham et al. (2019c)* introduced Hover-Net, a convolutional neural network designed to simultaneously segment and classify nuclei across multi-tissue histology images. Hover-Net addresses the challenge of nuclear segmentation and classification by utilizing the instance-rich information within

the vertical and horizontal distances of nuclear pixels to their centers of mass, enabling it to distinguish between clustered nuclei and handle intra-class variability, such as that seen in tumor cells. This model exemplifies how specialized datasets and tailored architectures can overcome limitations in traditional data sources.

These studies emphasize the complexity and specific requirements for training on biological datasets, such as those present in **Cerberus**, highlighting the importance of specialized, high-quality, and well-annotated datasets for accurate model development and training.

1.6 – Cerberus

The **Cerberus** model is a fully convolutional neural network, structured to perform multi-task learning (MTL) by simultaneously handling multiple histological tasks in CPath. This model leverages Multi-Task Learning (MTL) to address several tasks within histological image analysis, such as segmentation, classification, and detection of biological structures, in a unified framework. By using a single model for multiple aligned tasks, Cerberus significantly improves efficiency and performance, eliminating the need for separate models for each task and thus optimizing computational resources.

As explained earlier, numerous MTL models have been developed over the years, and many have demonstrated strong performance, particularly in tasks involving transfer learning. So, **what makes Cerberus stand out?**

The uniqueness of **Cerberus** lies in its specific design to handle aligned tasks, ensuring that all tasks use compatible input data. This alignment minimizes issues of task entanglement and gradient conflicts, which are common challenges in MTL. Many MTL models in computational pathology (CPath) struggle with simultaneous predictions due to differing input assumptions across tasks. For instance, previous work by Mormont et al. (2020) highlighted the challenges of combining tasks with different staining methods—such as IHC for tumor and stroma identification and H&E for cell classification—within the same model. These differences in input assumptions can lead to conflicts during processing and ultimately produce inaccurate outputs.

In fact, perhaps surprisingly, none of the above-mentioned studies, in CPath, employ MTL with the primary goal of achieving simultaneous predictions across tasks. **Cerberus**, however, is specifically designed to ensure that all tasks share aligned inputs and objectives, enabling accurate simultaneous predictions. This alignment reduces gradient conflicts during network optimization, a problem that has often prevented other MTL models in CPath from achieving single-task performance levels. By focusing on tasks with compatible inputs, Cerberus overcomes the primary limitations seen in traditional MTL setups, delivering higher reliability and consistent performance across multiple tasks.

Model Architecture and Design

The **Cerberus** model incorporates a shared encoder-decoder architecture, where hard parameter sharing across tasks is applied in the encoder to learn a common representation of the input histology images. This shared representation is then passed through task-specific decoders, each tailored to a particular prediction task, such as nuclei segmentation, gland segmentation, lumina segmentation and tissue type classification.

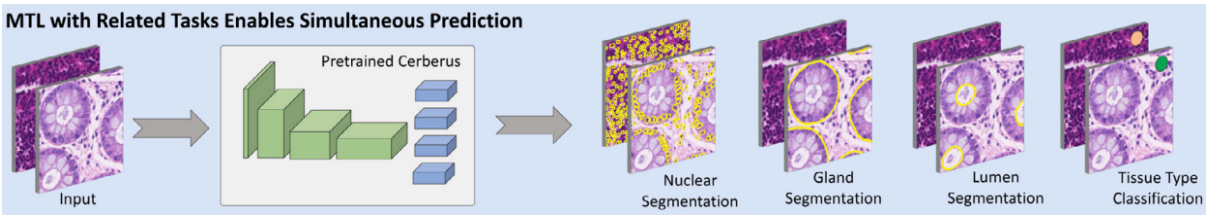


Figure 6 : Cerbus typical ouptuts (*Graham, S. (2023)*).

The architecture of **Cerberus** consists of:

- **Shared Encoder:** The encoder is based on a ResNet34 architecture, chosen for its robustness and **adaptability in transfer learning scenarios**. This shared encoder captures generalizable features from input histology images, benefiting each task by reinforcing shared representations. Using shared layers reduces the risk of overfitting and allows Cerberus to operate on large datasets from multiple sources, achieving consistent performance across tasks.
- **Task-Specific Decoders:** Each task handled by Cerberus has a dedicated decoder module that processes features from the shared encoder. For segmentation tasks, a U-Net style decoder with upsampling layers and skip connections (incorporating intermediate encoder features) is employed. The decoders incrementally upsample the features, and each upsampling operation is followed by two convolutional layers (3×3 kernels) with batch normalization, ensuring consistent feature scaling. For patch classification tasks, features are processed through a global average pooling layer followed by fully connected layers, outputting class probabilities. The setup allows flexibility across segmentation and classification tasks within the same network.

Training and Optimization Strategy

Another key element that sets **Cerberus** apart as a multi-task learning (MTL) model is its specialized training and optimization strategy. Designed to facilitate robust learning across diverse datasets while minimizing conflicts between tasks, the **Cerberus** model adopts several innovative approaches to achieve stable and high-performance MTL in computational pathology.

The training strategy of **Cerberus** employs a task sampler, which aggregates patches from each task-specific dataset into a batch during training. This task sampler uses two approaches:

- **Fixed Batch Sampling:** Each batch is sourced exclusively from a single task dataset, allowing focused learning on a specific task without interference from others.
- **Mixed Batch Sampling:** Batches contain patches from multiple tasks, fostering a richer feature representation that benefits all tasks through shared learning.

For optimization, **Cerberus** employs a technique called *dynamic weight freezing*, which adjusts the shared encoder weights with each batch but only updates task-specific decoder weights if data from that particular task is present in the batch. This approach effectively manages computational resources and mitigates potential conflicts in gradient updates across tasks, resulting in a balanced optimization process. To ensure consistent learning across both segmentation and classification tasks, a cross-entropy loss function is applied universally, providing flexibility across diverse objectives within the same network.

Additionally, **Cerberus** is trained on a large and diverse dataset, consolidated from independent sources and encompassing over 600,000 annotated objects for segmentation and 440,000 patches for classification. This extensive training set not only empowers high performance on primary tasks (such as nuclei, gland, and lumina segmentation) but also facilitates effective transfer learning, allowing the model to adapt its learned representations to new tasks with minimal retraining. The shared encoder, having absorbed features through comprehensive multi-task training, has proven particularly beneficial in downstream applications, showing notable improvement when transferred to related tasks, such as nuclear classification and object subtyping.

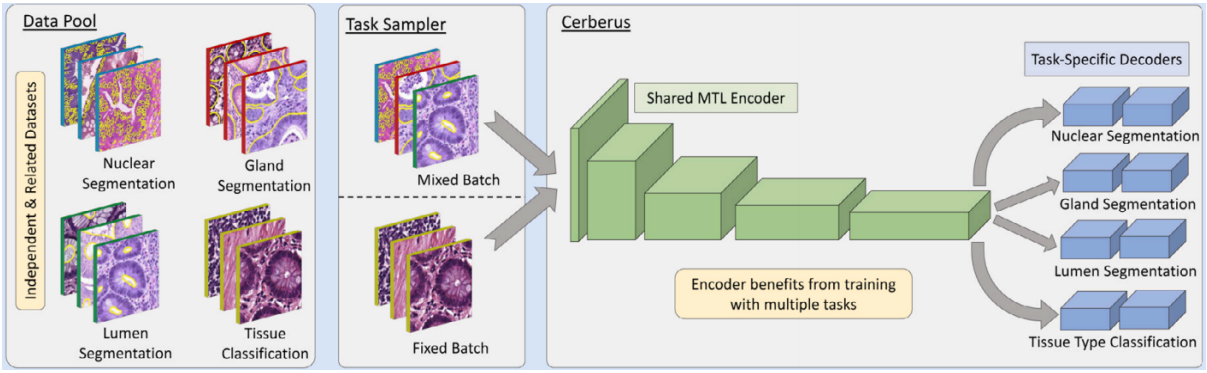


Figure 7 : MTL Training Schema (*Graham, S. (2023)*).

2 – Transfer Learning application: Results and discussion

2.1 – Dataset used for transfer learning

One of the main challenges of this project is to find well-annotated segmentation histological data that **Cerberus** was not trained on. The dataset presented in the article *EBHI-Seg: A Novel Enteroscope Biopsy Histopathological Haematoxylin and Eosin Image Dataset for Image Segmentation Tasks* (*Liyu Shi et al. (2022)*) was identified as an ideal dataset for our transfer learning objectives.

EBHI-Seg is a publicly available dataset designed for histopathological image segmentation tasks. It consists of 5,170 images of colorectal cancer tissue samples stained with Hematoxylin and Eosin (H&E), covering six different tumor differentiation stages. The dataset includes the corresponding ground truth images for segmentation, making it well-suited for training and evaluating deep learning models in computational pathology.

The dataset was constructed by a team of 12 biomedical researchers from the Microscopic Image and Medical Image Analysis Group at Northeastern University, in collaboration with the Department of Pathology at the Cancer Hospital of China Medical University and Shengjing Hospital.

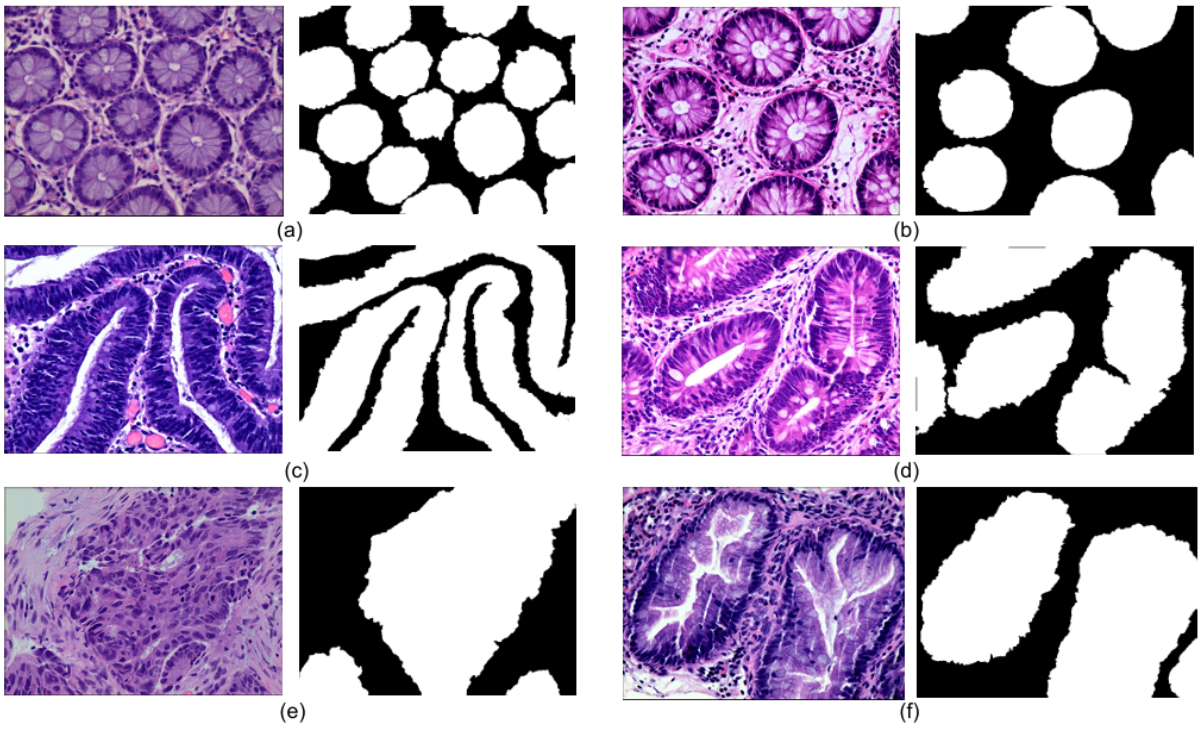


Figure 8 : (a) Normal (b) Polyp (c) High-grade Intraepithelial Neoplasia (d) Low-grade Intraepithelial Neoplasia (e) Adenocarcinoma (f) Serrated adenoma, *(Shi et al., (2023))*.

- **Normal Tissue**: Consistently ordered tubular structures without malignancy.
- **Polyp**: Benign growths on the mucosal surface with intact luminal structures.
- **Low-grade Intraepithelial Neoplasia (IN)**: Mildly altered glandular structure with nuclear enlargement.
- **High-grade Intraepithelial Neoplasia (IN)**: More pronounced nuclear enlargement and structural abnormalities.
- **Adenocarcinoma**: Malignant tumor with irregular luminal distribution.
- **Serrated Adenoma**: Uncommon lesion with characteristic crypt formations.

As seen in the class definitions above, certain biological structures, such as nuclei and lumina, play a key role in characterizing tissue types of the *EBHI-Seg* data set. These structures were also mentioned earlier in the description of the **Cerberus** training dataset (Section 1.5). Given that our transfer learning approach relies on a pre-trained **Cerberus** encoder, it is crucial to ensure that our data remains closely aligned with the features the model has previously encountered. Maintaining this similarity can significantly enhance the effectiveness of transfer learning and improve overall performance.

2.2 – Data augmentation methods

As mentioned before, the dataset used for our transfer learning application consists of only 5,170 images, including ground truth, meaning that we only have 2,855 images available for training. This number is considered relatively small for training deep learning models. However, the nature of histological datasets makes them highly suitable for augmentation using various data augmentation techniques. These methods not only artificially expand the dataset but also enhance model robustness by simulating variations that can occur in real-world histological samples:

- **Elastic Transformation**: Deforms the image using random elastic distortions (e.g., stretching, warping). This makes the model more robust to morphological variations in real histological structures.
- **Random Brightness/Contrast**: Adjusts brightness and contrast randomly to simulate variations in lighting conditions. This reflects differences in image acquisition settings in laboratories.
- **Rotation**: Rotates the image randomly within a range of $\pm 35°$. This technique improves the model's ability to recognize histological structures under different orientations.
- **Horizontal or Vertical Flip**: Flips the image horizontally or vertically. This simulates cases where tissue samples are sectioned and analyzed from different perspectives.
- **Gaussian Blur**: Applies a Gaussian blur with a kernel size between 3 and 7 to simulate blurring effects. This helps the model handle images acquired with imperfect focus.

During the training of any task (classification or segmentation), we apply data augmentation to the entire training set at each epoch, with each method applied at a certain probability. The probabilities are chosen based on the likelihood of each transformation occurring in real histological datasets. For instance, we assign a probability of 50% for elastic transformation and rotation since histological data does not have strict structural constraints. Meanwhile, Gaussian blur is assigned a probability of 15% as it is not heavily present in the test data, but remains a relevant transformation, especially for thick biological samples.

Needless to say that these data augmentation methods significantly improved the model's performance for both classification and segmentation tasks. For instance, in the classification task, applying data augmentation led to a performance improvement of 5%.

2.3 – Classification application

For our first application, classification, we use the encoder ResNet34 of **Cerberus** pretrained. We proceed by fine-tuning the model from each of its four layers, where fine-tuning from the 1st layer is equivalent to training the model from scratch. As for the SVM, we use the entire encoder of **Cerberus** and replace the ResNet classifier with an SVM.

With the same data augmentation methods mentioned earlier, a batch size of 32, and a scheduled learning rate starting from 0.1 and updated every 10 epochs, for a total of 30 epochs, we obtain the following results:

Fine-tuning from:	**Layer 1**	**Layer 2**	**Layer 3**	**Layer 4**	**SVM**
Accuracy	85.2%	82.5%	80.3%	79.4%	59.1%

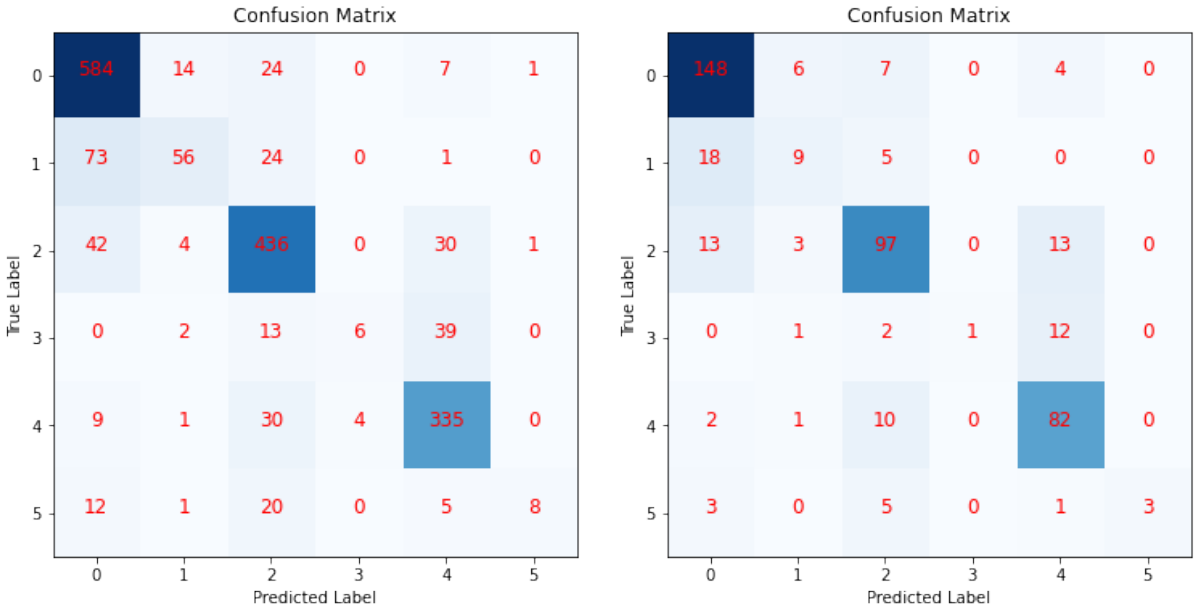
Table 1 : Classification results using transfer learning

The table shows that we obtain worse results as we go deeper in fine-tuning, meaning that the more the model learns by itself, the better the results. Relying solely on the feature representation given by the **Cerberus** encoder leads to suboptimal performance.

This can be explained by three main arguments. The first is that the encoder used for transfer learning was initially trained on four tasks, three of which were segmentation tasks and only one for classification. As a result, the encoder likely focused more on extracting features relevant to segmentation rather than classification, resulting in less effective classification performance when features are directly transferred.

The second argument is related to the class imbalance in the EBHI-Seg dataset. The normal and serrated adenoma classes contain significantly fewer images compared to other classes. This imbalance makes it more difficult for the model to generalize well to these underrepresented classes. Additionally, deep fine-tuning may lead to overfitting on the dominant classes, further worsening generalization to minority classes. Addressing this issue may require implementing strategies such as data balancing, weighted loss functions, or additional augmentation techniques to improve representation learning across all classes.

And the third is the significant similarity between certain classes such as the Normal and Polyp classes. If the entire encoder or a large portion of its weights is frozen, the model may struggle to learn sufficient distinguishing features between the similar classes.



(A) On the training data (B) On the test data

FIGURE 9 : Confusion matrix for the classification of the EBHI-Seg data set ["Adenocarcinoma": 0, "High-grade IN": 1, "Low-grade IN": 2, "Normal": 3, "Polyp": 4, "Serrated adenoma": 5]

This limitations highlight the importance of carefully selecting which encoder layers to fine-tune, ensuring that the model can effectively differentiate between similar categories while leveraging the benefits of transfer learning.

It is worth mentioning that the model was tested for classification using transfer learning on the NCT-CRC-HE-100K dataset, on which **Cerberus** was initially trained on only a portion. By applying the same transfer learning technique as with the EBHI-Seg data, we achieved an accuracy of 97% on 20,000 test images.

2.4 – Evaluation metrics for segmentation

Before presenting the segmentation results, we first define the evaluation metrics used in this work. These metrics assess the accuracy of the segmentation model by comparing predicted results with ground truth

annotations. Each metric provides different insights into the model's performance.

The Dice ratio measures the overlap between the predicted segmentation A and the ground truth B. It is defined as:

$$\text{Dice} = \frac{2 \cdot |A \cap B|}{|A| + |B|}$$

This metric is particularly useful in medical image segmentation because it emphasizes correctly identified regions and penalizes both false positives and false negatives equally.

The Jaccard Index, also known as the Intersection over Union (IoU), quantifies the similarity between the predicted and ground truth segmentations:

$$\text{Jaccard} = \frac{|A \cap B|}{|A \cup B|}$$

This metric is stricter than the Dice ratio as it penalizes incorrect predictions more heavily.

The Conformity Coefficient evaluates the consistency between predicted and actual segmentations:

$$\text{ConfmIndex} = \begin{cases} 1 - \frac{\theta_{AE}}{\theta_{TP}}, & \text{if } \theta_{TP} > 0, \\ \text{Failure}, & \text{if } \theta_{TP} = 0, \end{cases}$$

where $\theta_{AE} = \theta_{FP} + \theta_{FN}$ represents all errors and θ_{TP} is the number of correctly classified pixels. This index provides a direct and intuitive evaluation of the model's performance. A negative index indicates that there are more false predictions than true ones.

Precision measures how many of the predicted positive pixels are actually correct:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

High precision indicates fewer false positives, which is crucial when minimizing false alarms in medical applications.

Recall (or Sensitivity) assesses how many actual positive pixels were correctly identified:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

A high recall value ensures that most of the actual positive pixels are correctly segmented, as if you notice TP + FN = The All Positive Pixels, reducing missed detections in critical applications such as tumor detection.

2.5 – Segmentation application

For the segmentation task, we extend the previously constructed model, which consists of the **Cerberus** encoder and a classifier, by adding an additional head for segmentation. This results in a U-Net architecture with the **Cerberus** encoder. Consequently, the final model is capable of performing both segmentation and classification simultaneously, providing a complete analysis of the input image.

To evaluate the performance of our transfer learning for segmentation, we compare it with the performance of the three segmentation models used in the article by Liyu Shi et al. (2022) for the EBHI-Seg dataset. These models are: U-Net: A widely used convolutional neural network (CNN) for medical image segmentation. It follows an encoder-decoder architecture that captures both local and global contextual information, making it effective for delineating fine structures in histological images, Seg-Net: A deep learning-based segmentation network derived from the VGG16 architecture. It removes fully connected layers and introduces max-pooling indices for improved boundary delineation. Seg-Net performs well on large datasets and is particularly efficient in computational cost and memory usage and finally MedT (Medical Transformer): A transformer-based segmentation model designed for medical images. It incorporates an attention mechanism to capture long-range dependencies within an image, making it effective for segmenting complex histopathological structures.

With the training protocol used in the classification task, we obtain the following results:

Category	Model	Dice Ratio	Jaccard Index	CC	Precision	Recall
Normal	U-Net	0.411	0.263	-2.199	0.586	0.328
	Seg-Net	0.777	0.684	-0.607	0.895	0.758
	MedT	0.676	0.562	-0.615	0.874	0.610
	Cerberus (TL)	**0.967**	**0.937**	**0.933**	**0.972**	**0.963**
Polyp	U-Net	0.965	0.308	-1.514	0.496	0.470
	Seg-Net	0.937	0.886	0.858	0.916	**0.965**
	MedT	0.771	0.648	0.336	0.687	0.920
	Cerberus (TL)	**0.967**	**0.937**	**0.933**	**0.971**	0.963
High-grade IN	U-Net	0.895	0.816	0.747	0.847	**0.961**
	Seg-Net	0.894	0.812	0.757	0.881	0.913
	MedT	0.824	0.707	0.556	0.740	0.958
	Cerberus (TL)	**0.936**	**0.880**	**0.864**	**0.933**	0.940
Low-grade IN	U-Net	0.911	0.849	0.773	0.879	0.953
	Seg-Net	0.924	0.864	0.826	0.883	**0.977**
	MedT	0.889	0.808	0.718	0.876	0.916
	Cerberus (TL)	**0.964**	**0.932**	**0.927**	**0.970**	0.959
Adenocarcinoma	U-Net	0.887	0.808	0.718	0.850	0.950
	Seg-Net	0.865	0.775	0.646	0.792	**0.977**
	MedT	0.735	0.595	0.197	0.662	0.864
	Cerberus (TL)	**0.928**	**0.865**	**0.845**	**0.924**	0.931
Serrated adenoma	U-Net	**0.938**	**0.886**	0.865	0.899	**0.980**
	Seg-Net	0.907	0.832	0.794	0.859	0.963
	MedT	0.670	0.509	-0.043	0.896	0.544
	Cerberus (TL)	0.937	0.882	**0.867**	**0.965**	0.911

TABLE 2 : Segmentation results using transfer learning.

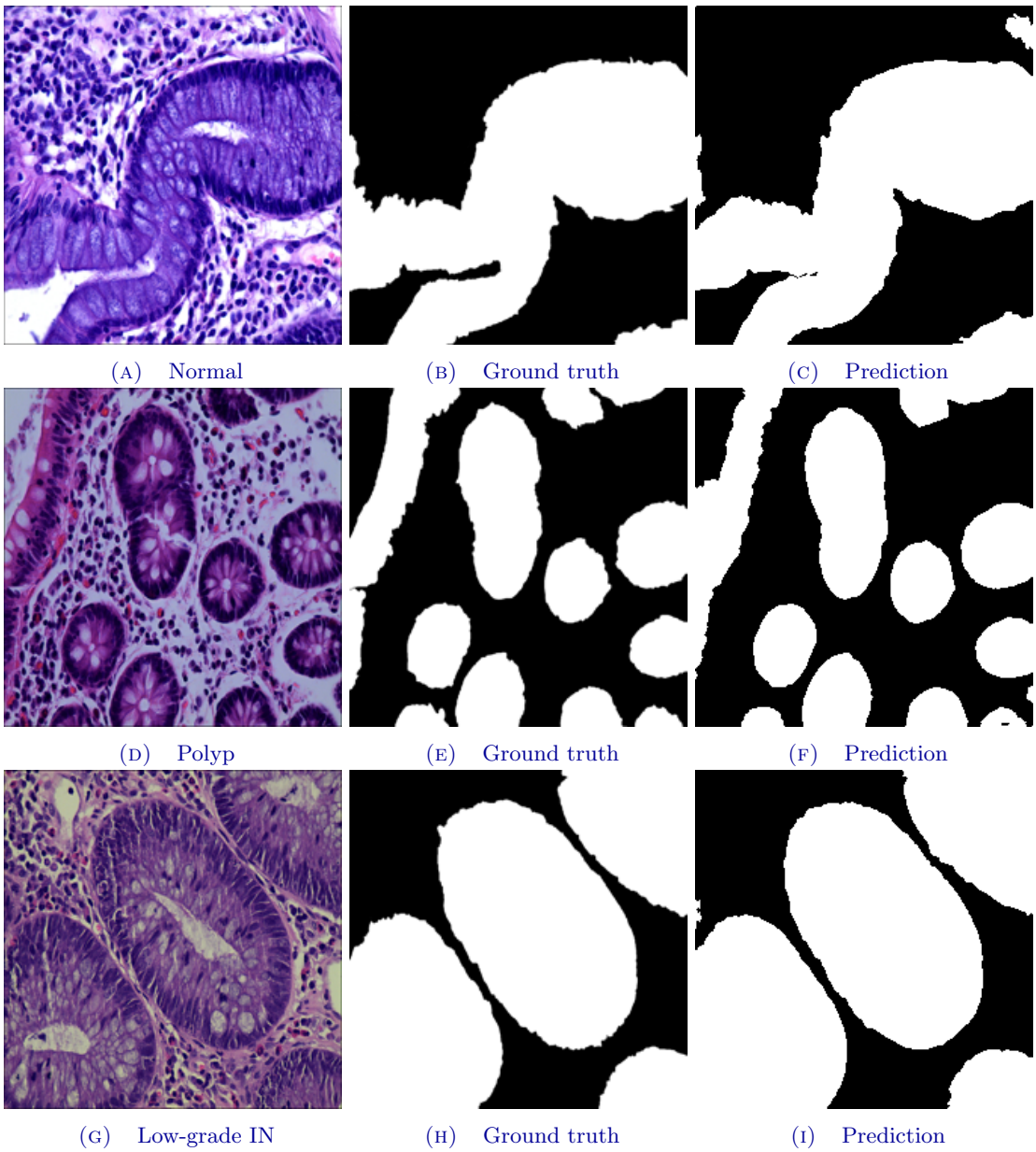
Across all six histopathological categories, **Cerberus** (TL) outperforms the other models in almost all metrics. However, there are still important points worth discussing based on this table.

One notable observation is that the 3 other models struggle with the Normal class, due to its smaller number of images. In contrast, the model using the **Cerberus** encoder does not face this issue at all. On the contrary, it achieves excellent performance across all metrics. This highlights the strength of transfer learning, demonstrating that the model does not require a large number of training images to achieve good or even outstanding results, as observed in our case.

Another remark regarding the performance of U-Net when segmenting the Polyp class is that while it achieves a high Dice ratio, all other metrics remain very low. This suggests that the model is generating segmented regions that overlap well with the ground truth but may suffer from imprecise boundaries, over-segmentation, or poor generalization across different samples. This serves as a strong example of why selecting the most appropriate evaluation metric is crucial, as relying solely on Dice could lead to misleading conclusions about the model's true performance in the context of the study.

Additionally, we can observe that MedT generally exhibits the worst performance among the models, despite being a transformer and theoretically expected to achieve superior results. This observation is likely due to the lack of sufficient training data relative to the complexity of MedT. As a transformer-based model, MedT requires a large amount of data to effectively learn spatial dependencies. However, with limited data, it may struggle to generalize, leading to overfitting, inefficient feature learning, and suboptimal segmentation performance compared to CNN-based architectures like U-Net and Seg-Net.

Finally, we observe that **Cerberus** fails to achieve the best performance in terms of the Recall index across most classes compared to the other models. This indicates that **Cerberus** tends to miss some positive pixels belonging to objects, whereas other models capture them more effectively. This limitation is particularly concerning in medical analysis, where under-segmentation can be problematic. In such applications, it is generally preferable to over-segment rather than to miss important structures, as overlooking critical regions could lead to potentially dangerous diagnostic errors.



(A) Normal (B) Ground truth (C) Prediction

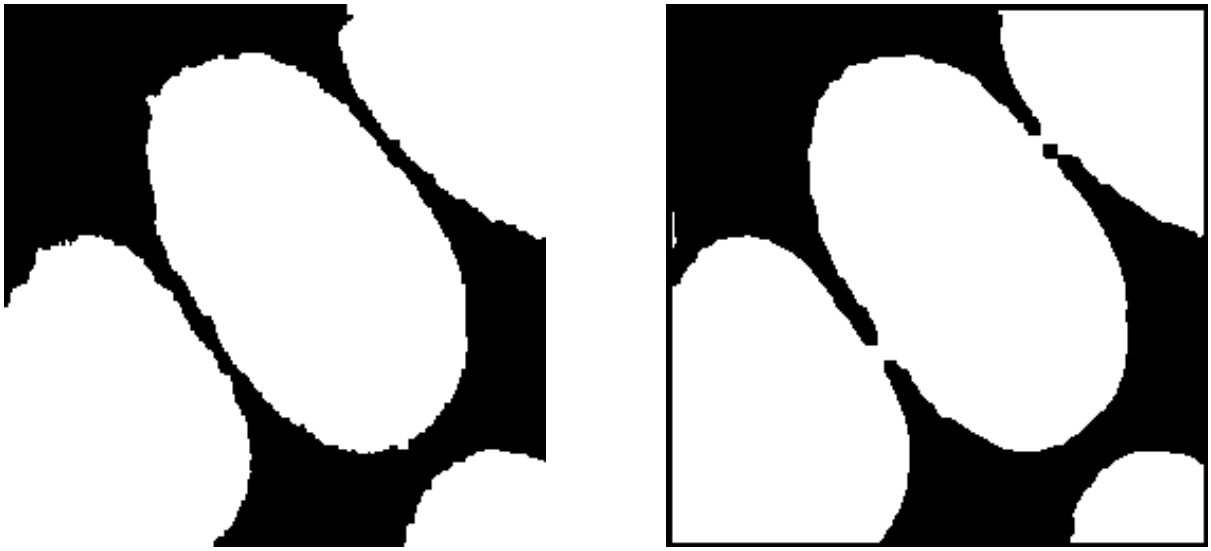
(D) Polyp (E) Ground truth (F) Prediction

(G) Low-grade IN (H) Ground truth (I) Prediction

FIGURE 10 : Examples of segmentation results with transfer learning by Cerberus

2.6 – POST-PROCESSING

In some prediction examples, we observed that **Cerberus** missed certain holes, and the edges were not smooth. This issue arose because the training images were compressed, causing the object boundaries to appear in grayscale due to compression artifacts. To address this, we applied a morphological closing operation. However, this led to an unintended consequence: some cells that should remain separate were erroneously connected.



(A) Before closing (B) After closing

FIGURE 11 : Impact of closing.

3 – CONCLUSION

In this work, we explored the domain of Computational Pathology (CPath), examining its objectives, challenges, and its growing role in medical diagnostics. We then investigated **Cerberus**, a Multi-Task Learning (MTL) model, and evaluated its performance in both classification and segmentation tasks through transfer learning. Our results highlight several critical aspects regarding the potential of MTL models, dataset requirements, and the future applications of **Cerberus** in medical imaging.

Cerberus as an MTL model demonstrated promising results, reinforcing the strength of multi-task learning in medical image analysis, by leveraging shared feature representations across tasks. The use of transfer learning further optimized its ability to generalize to new datasets, achieving superior results compared to traditional architectures.

Our findings underscore the critical need for large-scale histopathological datasets. Unlike general computer vision tasks, computational pathology lacks a standardized dataset equivalent to ImageNet. Developing an extensive, collaborative repository of annotated histological images would enable training of more powerful MTL models like **Cerberus**, further improving generalization, and would also facilitate innovation and collaboration within the research community, leading to more robust AI-driven pathology solutions.

Additionally, our study suggests that **Cerberus** could play a crucial role in automatic annotation of histological images. Given the high cost and time required for manual annotation, leveraging a well-trained MTL model to automate labeling could significantly accelerate dataset creation and improve AI-assisted diagnostic tools.

However, caution should be exercised when training or deploying MTL models to ensure optimal performance. The training tasks and datasets must be properly aligned, both within the multi-task learning framework and in the transfer learning phase. Failure to maintain this alignment can diminish the advantages of MTL.

References

[1] Gamper, J., Kooohbanani, N.A., Rajpoot, N., 2020b. Multi-task learning in histo-pathology for widely generalizable model. arXiv preprint arXiv:2005.08645.

[2] S. Graham et al., "Computational Pathology: A Survey Review and The Way Forward," *Medical Image Analysis*, vol. 83, pp. 102685, 2023.

[3] R. J. Chen et al., "Computational Pathology: Challenges and Promises for Tissue Analysis," *IEEE Transactions on Medical Imaging*, vol. 41, no. 4, pp. 757-770, 2022.

[4] Y. Zheng et al., "Graph attention-based fusion of pathology images and gene expression for prediction of cancer survival," *IEEE Transactions on Medical Imaging*, vol. 43, no. 9, pp. 3085-3097, 2024.

[5] Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J.A., Hermsen, M., Manson, Q.F., Balkenhol, M., et al., 2017. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 318 (22), 2199–2210.

[6] Veeling, B.S., Linmans, J., Winkens, J., Cohen, T., Welling, M., 2018. Rotation equivariant cnns for digital pathology. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 210–218.

[7] Caruana, R., 1997. Multitask learning. Mach. Learn. 28 (1), 41–75.

[8] Zhang, Y., Yang, Q., 2017. A survey on multi-task learning. arXiv preprint arXiv: 1707.08114.

[9] Mormont, R., Geurts, P., Marée, R., 2020. Multi-task pre-training of deep neural networks for digital pathology. IEEE J. Biomed. Health Inf. 25 (2), 412–421

[10] Wahab, N., Miligy, I.M., Dodd, K., Sahota, H., Toss, M., Lu, W., Jahanifar, M., Bilal, M., Graham, S., Park, Y., et al., 2021. Semantic annotation for computational pathology: Multidisciplinary experience and best practice recommendations. arXiv preprint arXiv:2106.13689.

[11] Tellez, D., Höppener, D., Verhoef, C., Grünhagen, D., Nierop, P., Drozdzal, M., Laak, J., Ciompi, F., 2020. Extending unsupervised neural image compression with supervised multitask learning. In: Medical Imaging with Deep Learning. PMLR, pp. 770–783.

[12] Ruder, S. "An Overview of Multi-Task Learning in Deep Neural Networks." arXiv preprint arXiv:1706.05098, 2017.

[13] Caruana, R. "Multitask Learning." Machine Learning, vol. 28, no. 1, pp. 41-75, 1997.

[14] Sun, Q., Liu, Y., Chua, T.S., and Schiele, B. "Meta-Transfer Learning for Few-Shot Learning." IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 403-412.

[15] Ruder, S. "Sharing-to-Learn and Learning-to-Share: Meta, Multi-Task, and Transfer Learning." arXiv preprint arXiv:2111.12146, 2021.

[16] Zhang, Y. and Yang, Q. "A Survey on Multi-Task Learning." IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 6, pp. 1-20, 2022.

[17] Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N., 2019c. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal. 58, 101563.

[18] Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings, pp. 249–256.

[19] Zhang, M.-L., Zhou, Z.-H., 2013. A review on multi-label learning algorithms. IEEE Trans. Knowl. Data Eng. 26 (8), 1819–1837.

[20] Zhang, Y., Yang, Q., 2017. A survey on multi-task learning. arXiv preprint arXiv: 1707.08114.

[21] Shephard, A.J., Graham, S., Bashir, S., Jahanifar, M., Mahmood, H., Khurram, A., Rajpoot, N.M., 2021. Simultaneous nuclear instance and layer segmentation in oral epithelial dysplasia. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 552–561.

[22] Fraz, M., Khurram, S.A., Graham, S., Shaban, M., Hassan, M., Loya, A., Rajpoot, N.M., 2020. Fabnet: Feature attention-based network for simultaneous segmentation of microvessels and nerves in routine histology images of oral cancer. Neural Comput. Appl. 32 (14), 9915–9928.

[23] Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.-A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., et al., 2017. Gland segmentation in colon histology images: The glas challenge contest. Med. Image Anal. 35, 489–502.

[24] Graham, S., Chen, H., Gamper, J., Dou, Q., Heng, P.-A., Snead, D., Tsang, Y.W., Rajpoot, N., 2019a. MILD-net: Minimal information loss dilated network for gland instance segmentation in colon histology images. Med. Image Anal. 52, 199–211.

[25] Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., Heng, P.-A., 2017. DCAN: Deep contour-aware networks for object instance segmentation from histology images. Med. Image Anal. 36, 135–146.

[26] Da, Q., Huang, X., Li, Z., Zuo, Y., Zhang, C., Liu, J., Chen, W., Li, J., Xu, D., Hu, Z., et al., 2022. DigestPath: a benchmark dataset with challenge review for the pathological detection and segmentation of digestive-system. Med. Image Anal. 102485.

[27] Banwari, M., & Sengar, S. S. (2014). Automated segmentation of colon gland using convolutional neural network-based deep learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (pp. 1-8).

[28] Hou, C., Zhang, Y., Zhou, Z.-H., & Wu, X. (2013). Bayesian multi-instance multi-label learning using Gaussian process prior. In Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI) (pp. 1888-1894).

[29] Shi, L., Li, X., Hu, W., Chen, H., Chen, J., Fan, Z., Gao, M., Jing, Y., Lu, G., Ma, D., Ma, Z., Meng, Q., Tang, D., Sun, H., Grzegorzek, M., Qi, S., Teng, Y., & Li, C. (2023). EBHI-Seg: A novel enteroscope biopsy histopathological hematoxylin and eosin image dataset for image segmentation tasks. Frontiers in Medicine, 10, 1114673.