Deep Learning for Lymphoma Diagnosis: A Cytology-Based Study Using Digitized Lymph Node Imprints.

Imane Jagour and **Amine Ahaggach**

University of Caen Normandie

{imane.jagour, amine.ahaggach}@etu.unicaen.fr

Abstract

Lymphoma, a cancer originating in the lymphatic system, presents significant diagnostic challenges due to the subjective nature and variability of manual cytological analysis. This study investigates the potential of deep learning models to enhance lymphoma detection from digitized lymph node cytological imprints. Using a dataset of 148 annotated slides from Caen University Hospital, we evaluated three advanced segmentation models: U-Net, U-Net++, and DeepLabV3+.

Quantitative results revealed that U-Net achieved the highest accuracy (0.7563) and Dice score (0.7586). However, expert evaluations favored U-Net++ for its clinical relevance and superior sensitivity (0.9405), particularly in detecting malignant regions. Despite its lower overall performance, DeepLabV3+ demonstrated strong boundary detection capabilities. These findings emphasize the importance of combining quantitative metrics with expert assessments to improve the clinical applicability of AI-driven lymphoma diagnostics.

Keywords: Deep Learning, Lymphoma Diagnosis, Digital Cytology

1 Introduction

Lymphoma is a type of cancer that originates in the lymphatic system, with an increasing global incidence, particularly among elderly populations (Swerdlow et al., 2022). According to the 5th edition of the WHO Classification of Hematopoietic and Lymphoid Tumors, lymphomas are categorized into Hodgkin lymphoma (HL) and non-Hodgkin lymphomas (NHL), the latter accounting for approximately 90% of cases. While histopathological examination remains the gold standard for diagnosis, cytological examination plays a crucial role in rapid preliminary assessments. However, manual cytological analysis is subjective, time-intensive, and prone to variability between pathologists, leading to inconsistencies in diagnosis and treatment planning.

With recent advancements in artificial intelligence (AI) and deep learning, computational models have demonstrated promising results in medical imaging, particularly for cancer detection and classification (Jiang et al., 2024). AI-driven approaches have been widely adopted in breast and colorectal cancer pathology, enhancing diagnostic accuracy and workflow efficiency (Smith and Doe, 2023). Despite these advances, the application of AI in lymph node cytology remains underexplored, primarily due to the lack of large annotated datasets and robust comparative studies evaluating multiple deep learning architectures. Existing models often rely on single architectures, limiting their generalizability across different clinical settings.

This study leverages a comprehensive dataset provided by the **Pathology Department of Caen University Hospital**, which has been fully digitized since **December 2021**. Our ultimate objective is to develop an **AI-powered diagnostic tool** that classifies cytological samples into three distinct categories: **reactive/acellular**, **lymphoma**, and **non-lymphomatous tumors**. By integrating multiple deep learning models and validating results through expert assessments, we aim to establish an **objective and reproducible AI-assisted framework** for lymphoma detection. The results of this study could pave the way for **automated, standardized, and clinically reliable lymphoma diagnosis**, enhancing both **pathologist efficiency and patient outcomes**.

To address these challenges, this study investigates the effectiveness of multiple deep learning models—UNet++, DeepLabV3+, and UNet—for segmenting and classifying cytological images. Instead of relying solely on standard performance metrics such as Intersection over Union (IoU) and F1-score, we introduce an **expert-driven evaluation framework** where experienced pathologists

assess the clinical relevance and interpretability of model-generated segmentations. This dual evaluation strategy aims to bridge the gap between algorithmic performance and real-world applicability.

The primary objective of this research is to develop an AI model capable of identifying lymphomatous proliferation with accuracy comparable to that of experienced pathologists. Additionally, we aim to systematically compare the performance of different architectures and provide a **comprehensive analysis of AI-based cytological image segmentation**, contributing to the development of more reliable and interpretable diagnostic tools in clinical practice.

2 Related Work

The application of artificial intelligence (AI) in lymphoma diagnosis has evolved significantly over the years, with various methodological approaches explored. Early studies focused on training deep learning-based tools for detecting specific lymphoma subtypes, such as double-hit lymphomas (DHL), using digitized biopsy images (Xu et al., 2021). While achieving high sensitivity and specificity, these models were often trained on limited datasets, requiring extensive manual annotation, which hindered scalability.

Further advancements introduced AI-driven classification models trained on medical imaging data, demonstrating improved diagnostic accuracy (Jiang et al., 2024). However, these approaches primarily relied on supervised learning, making them highly dependent on large annotated datasets and limiting their applicability across different clinical settings.

Other studies investigated AI models for segmenting and classifying non-Hodgkin lymphoma subtypes, achieving promising results with a balanced accuracy exceeding 95% in controlled datasets (Patel et al., 2022). Despite these advances, prior research often relied on a single deep learning model and lacked comparative analyses of multiple architectures, making it difficult to determine the most effective approach for cytological lymphoma segmentation.

The use of advanced deep learning architectures, such as UNet (Ronneberger et al., 2015), DeepLabV3 (Chen et al., 2017), and hybrid residual networks (He et al., 2016), has been widely explored for medical image segmentation, but their application to cytological lymphoma analysis remains under-explored. Unlike previous studies that focused on individual models, our work systematically compares multiple architectures, including UNet++, DeepLabV3+, and UNet, to identify the most effective approach for segmenting and classifying cytological images.

Additionally, we introduce an expert evaluation framework where pathologists assess model-generated segmentations to ensure clinical relevance and interpretability. By benchmarking these models across multiple performance metrics, including Intersection over Union (IoU) and F1-score, as well as a custom expert-driven evaluation scale, our study aims to bridge the gap between algorithmic performance and real-world applicability in lymphoma diagnosis.

3 Materials and Methods

3.1 Dataset and Preprocessing

This is a retrospective study based on the analysis of digitized lymph node cytological imprints collected between December 2021 and June 2023 at Caen University Hospital. These images were annotated by a pathology resident in collaboration with an experienced pathologist to create a reference database for training and evaluating an artificial intelligence algorithm dedicated to assisting in the diagnosis of lymphomas.

The samples included in this study were selected from the TDHistoCyto® pathology software based on the following criteria. The inclusion criteria were as follows: lymph node samples recorded in the software database with the ADICAP organ code "SG," the presence of a validated pathology report by a pathologist, and the availability of a lymph node imprint slide stained with May-Grünwald Giemsa (MGG). The exclusion criteria included the absence of an MGG-stained lymph node imprint slide, poor slide quality (blur, artifacts, or paucicellularity preventing analysis), and samples that were not lymph nodes upon review of the pathology report (coding errors).

Among the 267 cases initially identified in the database, 74 were excluded after verification because they did not correspond to lymph nodes. An additional 23 samples were discarded due to the absence of available slides, and 22 others were deemed unusable due to insufficient quality. In total, 148 slides were selected and annotated for use in training the artificial intelligence model (see Figure 1).

The slides were digitized using two Hamamatsu

NanoZoomer® S360 Digital Slide Scanners, following a cytology-specific protocol developed at Caen University Hospital. Scanning was performed using a Z-stack of 0 in ndpi format, with an equivalent x40 objective and a resolution of approximately 0.23 µm/pixel. A standard 76 mm × 26 mm slide represented approximately 330,000 × 113,000 pixels, amounting to nearly 40 billion pixels. To facilitate image processing, the slides were converted to an uncompressed TIF format to preserve detail quality. Each slide was then subdivided into 20,000 × 20,000-pixel images, and the most representative area in terms of quality and cellular diversity was selected.

Image annotation was performed using QuPath software. The annotation process took multiple factors into account to ensure cell representativity, balancing annotations between healthy and tumor cells. Cellular areas were annotated while ensuring the prevention of algorithmic learning errors. Tumor cells were classified based on their type (lymphomatous or non-lymphomatous), and other cellular components, such as non-tumoral lymphocytes, neutrophils, eosinophils, histiocytes, and plasma cells, were also annotated (see Figure 2). This annotation was carried out by the investigator and supervised by an experienced pathologist. For the purpose of training, the focus was placed on two main classes: benign and malignant cells, in addition to a background class representing non-cellular areas. This simplification was designed to enhance the model's ability to distinguish between healthy and pathological regions effectively.

In preparation for model training, a preprocessing pipeline was applied to optimize the quality and usability of the digitized slides. Initially, whole slide images were divided into smaller patches of 20,000 × 20,000 pixels and further segmented into 2,000 × 2,000-pixel tiles to facilitate efficient processing and reduce computational load. Patches lacking sufficient cellular content were filtered out using binarization in the HSV color space, retaining only regions rich in relevant cellular structures. Additionally, data augmentation techniques were applied to increase dataset variability and model robustness. These augmentations included random rotations, horizontal and vertical flips, and adjustments of brightness and contrast. The annotated regions were then transformed into segmentation masks with specific labels for each cell type, enabling the model to distinguish between different

cellular components accurately. This preprocessing ensured that the dataset was both balanced and representative of the various cellular morphologies found in lymph node cytology.

Several data points were collected for each analyzed case, including age, sex, diagnosis, and subdiagnosis, coded according to a specific internal system (see Appendix 1). This coding classified cases into the following categories: non-tumoral patient, tumoral patient with lymphoma, and patient with a non-lymphomatous tumor. This classification helped refine the model's training by integrating clinically relevant variables for lymphoma diagnosis and classification.
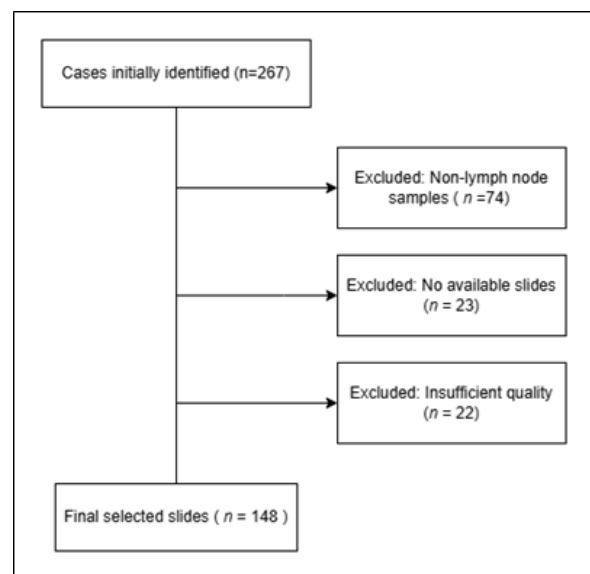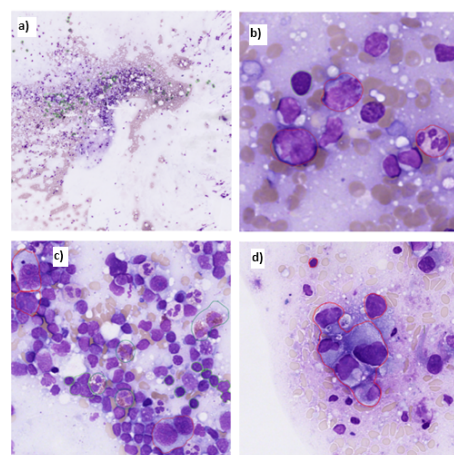

Figure 1: Flowchart of Case Selection.

Figure 2: a. Low power, normal lymph node annotation. b. High-grade diffuse B-cell lymphoma annotation. c. Hodgkin lymphoma annotation. d. Urothelial carcinoma metastasis annotation.

3.2 Data Augmentation and Dataset Splitting

To enhance the robustness and generalization capacity of the segmentation models, a comprehensive data augmentation strategy was implemented during the training phase. This augmentation was performed using the PyTorch transforms library, applying a sequence of transformations to both the input images and their corresponding segmentation masks. For the input images, the augmentation pipeline included resizing the images to a standardized size, followed by random horizontal and vertical flips to simulate different orientations of the cytological structures. Additionally, color transformations were applied using the ColorJitter function, which adjusted brightness, contrast, saturation, and hue. This approach aimed to mimic variations in staining and lighting conditions that can occur during slide preparation and scanning. Finally, the images were converted into tensors to be compatible with the PyTorch framework.

For the segmentation masks, resizing was applied using nearest-neighbor interpolation to preserve class integrity during transformation. Similar to the input images, random horizontal and vertical flips were incorporated to maintain consistency between the masks and their corresponding images. Following augmentation, the dataset was divided into three distinct subsets to ensure balanced and unbiased training and evaluation. Specifically, 12,088 images (approximately 80%) were allocated for training, 1,624 images (10%) for validation, and 1,625 images (10%) for testing. Dedicated data loaders were created for each subset, applying the defined augmentation transformations for the training set, while the validation and test sets underwent only resizing and tensor conversion to maintain evaluation integrity.

3.3 Deep Learning Models

To ensure robust and accurate segmentation of cytological images, we selected three state-of-the-art deep learning models: U-Net++, DeepLabV3+, and UNet. These architectures were chosen based on their proven effectiveness in medical image segmentation and their complementary strengths.

3.3.1 U-Net Architecture

The U-Net architecture (see Figure 3), introduced by Ronneberger et al. (Ronneberger et al., 2015), is a pivotal convolutional neural network (CNN) for **biomedical image segmentation**. Designed for tasks like MRI scans (Çiçek et al., 2016; Milletari

et al., 2016; Falk et al., 2019), its **encoder-decoder structure** enables precise pixel-wise predictions.
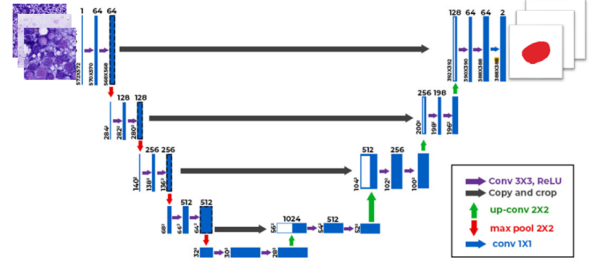

Figure 3: The encoder-decoder architecture of a standard U-Net

To enhance **feature extraction**, we integrate a **ResNet-50 backbone** (He et al., 2016), leveraging deep residual learning. The architecture consists of:

- **Encoder** – Captures contextual features via **3×3 convolutions** and **2×2 max-pooling**, progressively reducing spatial dimensions (LeCun et al., 1998).

- **Decoder** – Restores spatial resolution using **transposed convolutions**, ensuring accurate feature localization (Zeiler and Fergus, 2014).

Symmetric skip connections (Drozdzal et al., 2016) link encoder and decoder layers, preserving **fine details** for improved segmentation accuracy. Being fully **convolutional** and adaptable to varying image sizes, U-Net ensures **high precision** even with limited datasets, making it a robust tool for medical image segmentation.

3.3.2 U-Net++ Architecture

U-Net++, introduced by Zhou et al. (Zhou et al., 2018), refines the original U-Net by incorporating nested and dense skip pathways to enhance feature learning and segmentation accuracy.

It aims to bridge the semantic gap between encoder and decoder feature maps by progressively enriching fine-grained features (Hariharan et al., 2016). To strengthen feature extraction, we integrate a ResNet-50 backbone (He et al., 2016), leveraging deep residual learning to improve pattern recognition and segmentation precision. Inspired by DenseNet (Huang et al., 2017), U-Net++ includes dense skip connections that facilitate better gradient flow and information transfer across network layers. Additionally, a deep supervision mechanism (Buttar and Sachan, 2022) provides

intermediate layers with direct learning signals, reducing semantic loss and enhancing segmentation accuracy.

Compared to standard U-Net, U-Net++ introduces nested skip pathways, improving feature fusion and enabling more efficient learning. These modifications make U-Net++ a powerful architecture for medical image segmentation, particularly in complex cytological tasks.

3.3.3 DeepLabV3+ Architecture

DeepLabV3+ (Chen et al., 2018) is an advanced convolutional neural network designed for semantic image segmentation. It enhances DeepLabV3 by introducing a refined decoder module that improves spatial resolution while preserving fine-grained details, making it particularly effective for medical imaging tasks. The architecture incorporates a ResNet-50 backbone for feature extraction, which processes hierarchical image features and passes them through an Atrous Spatial Pyramid Pooling (ASPP) module. The ASPP module applies atrous (dilated) convolutions with multiple dilation rates (6, 12, 18) alongside global average pooling, enabling the network to capture multi-scale contextual information (Chen et al., 2017). The decoder further refines segmentation results by upsampling ASPP features and merging them with low-level spatial details from earlier layers, ensuring precise boundary preservation. The synergy between the ASPP module and the ResNet-50 backbone makes DeepLabV3+ a powerful architecture for accurate medical image segmentation.

3.4 Experiments Setting

To assess the efficiency of the proposed segmentation models, the experiments were conducted using Jupyter Lab on the Austral HPDA platform. The experiments utilized a powerful GPU configuration consisting of **8 cores**, **60 GB RAM**, and an **NVIDIA A100 GPU with 80 GB memory** and **432 tensor cores**.

This high-performance environment enabled the efficient processing of large cytological image datasets, ensuring faster training and evaluation times.

3.5 Model Training

All deep learning models were trained using the computational resources described in the previous section. All images in the dataset were resized to **224x224 pixels** to ensure uniformity and optimize

memory usage during training. For algorithm development and model implementation, the PyTorch deep learning framework was utilized. To assess the model's performance against ground truth annotations, training was conducted using the **categorical cross-entropy loss function**, which effectively measures discrepancies between the predicted segmentation masks and the actual labels.

The categorical cross-entropy loss function is defined as:

$$\mathcal{L} = -\sum_{n=1}^{N} y_n \log(\hat{y}_n)$$

where N represents the number of classes, y_n denotes the true label for class n, and $\hat{y}_n$ corresponds to the predicted probability for class n.

The Adam optimizer was employed with a learning rate of 0.001 and a batch size of 8. The models were trained for 30 epochs. To prevent overfitting, an early stopping strategy was applied, halting training if no improvement in validation performance was observed over successive epochs.

4 Results

This section presents a comprehensive benchmark of the three deep learning models **U-Net**, **U-Net++**, and **DeepLabV3+** evaluated on the cytological lymph node imprint dataset. The comparison is based on key performance metrics relevant to segmentation tasks, including *accuracy*, *sensitivity*, *specificity*, *precision*, *F1-score*, *Dice score*, *Intersection over Union (IoU)*, *Jaccard index*, *Area Under the Curve (AUC)*, *Cohen's Kappa*, *Hausdorff Distance (HD)*, and *Mean Surface Distance (MSD)*. These metrics provide a comprehensive evaluation of each model's effectiveness in detecting and segmenting benign and malignant regions.

4.1 Overall Model Performance

The performance comparison, summarized in Table 1, highlights that the **U-Net** model achieved the highest *accuracy* (**0.7563**), indicating its superior ability to classify pixels correctly. The **U-Net++** model demonstrated the best performance in terms of *sensitivity* (**0.9405**), making it particularly effective in detecting malignant regions, which is crucial in medical diagnostics to minimize false negatives.

Although the **DeepLabV3+** model exhibited slightly lower overall performance, it maintained a competitive *AUC* score of 0.7245, indicating a strong ability to distinguish between benign and malignant regions.

4.2 Segmentation Quality Metrics

In terms of segmentation quality, the *Dice score*, *IoU*, and *Jaccard index* reflect the overlap between the predicted masks and the ground truth annotations. The **U-Net** outperformed the other models, achieving the highest *Dice score* (**0.7586**) and *IoU* (**0.6111**), indicating strong segmentation accuracy. The **U-Net++** model closely followed with a Dice score of 0.7524, showcasing its robustness in segmenting complex cellular structures.

4.3 Sensitivity and Specificity

The **U-Net++** model demonstrated exceptional *sensitivity* (**0.9405**), reflecting its capacity to detect true positives effectively. However, it showed a lower *specificity* (0.5132), indicating a tendency to produce false positives. In contrast, the **U-Net** model achieved a more balanced result, with a *sensitivity* of 0.8968 and a *specificity* (**0.6471**), making it a reliable option for practical clinical applications where both false positives and false negatives must be minimized.

4.4 Robustness and Boundary Accuracy

Robustness was assessed using *Cohen's Kappa*, *Hausdorff Distance (HD)*, and *Mean Surface Distance (MSD)*. The **U-Net** achieved the highest *Cohen's Kappa* score (**0.4690**), demonstrating a strong agreement between the predicted and actual annotations beyond random chance.

The **DeepLabV3+** model excelled in boundary accuracy, achieving the lowest *Hausdorff Distance* (**6.0000**) and *Mean Surface Distance* (**0.5109**). This suggests that **DeepLabV3+** is particularly adept at capturing the fine-grained boundaries of segmented regions.

4.5 Benchmark Comparison

A detailed comparison of the three models based on the selected performance metrics is presented in Table 1. The results indicate that the **U-Net** provides the best balance of segmentation accuracy, robustness, and boundary precision. The **U-Net++** model stands out for its superior sensitivity, making it particularly suitable for early detection scenarios where minimizing false negatives is essential. The **DeepLabV3+** model, while slightly lagging in overall segmentation metrics, demonstrates strong boundary detection capabilities.

Metric	U-Net++	DeepLabV3+	U-Net
Accuracy	0.7465	0.7157	**0.7563**
Sensitivity	**0.9405**	0.8386	0.8968
Specificity	0.5132	0.6064	**0.6471**
Precision	0.6629	0.6509	**0.6807**
F1-Score	**0.7726**	0.7329	0.7586
Dice Score	0.7524	0.7343	**0.7586**
IoU	**0.6295**	0.5785	0.6111
Jaccard Index	0.6031	0.5801	**0.6111**
AUC	0.7269	0.7245	**0.7408**
Cohen's Kappa	0.4394	0.4406	**0.4690**
Hausdorff Distance (HD)	9.0000	**6.0000**	7.0000
Mean Surface Distance (MSD)	0.7074	**0.5109**	0.6209

Table 1: Benchmark comparison of U-Net, U-Net++, and DeepLabV3+ models across selected metrics. The best scores for each metric are highlighted in bold.

4.6 Visual Comparison of Segmentation Results

To provide a qualitative evaluation of model performance, Figure 4 showcases representative segmentation examples. The comparison highlights the original cytological images, corresponding ground truth annotations, and segmentation outputs from the three models (**U-Net++**, **DeepLabV3+**, and **U-Net**).

The visual results reveal that the **U-Net** and **U-Net++** models produce segmentations that closely align with the ground truth, especially for complex cellular structures. The **DeepLabV3+** model, although slightly less accurate in general segmentation, demonstrates strong boundary detection, particularly in cases with more defined cellular contours.
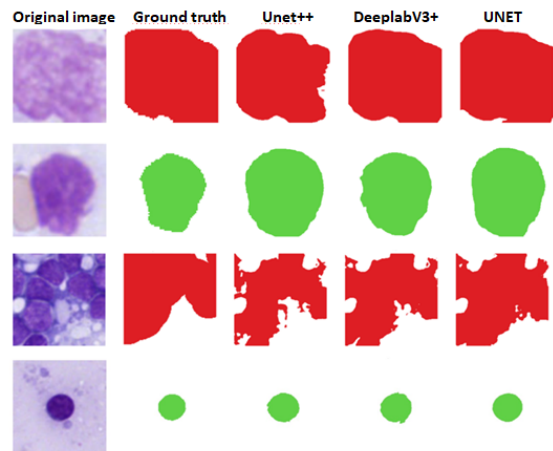

Figure 4: Visual comparison of segmentation results produced by **U-Net++**, **DeepLabV3+**, and **U-Net** models.

4.7 Evaluation of Deep Learning Models by Cytopathologist Expert

4.7.1 Grading System for Model Evaluation

Following the quantitative evaluation, expert cytopathologists conducted a qualitative assessment based on clinical relevance. A predefined grading system ensured consistency, categorizing segmentation quality into five levels:

- **Inadequate**: Poor segmentation; structures are misidentified or missing, making results unusable.

- **Below Average**: Some structures detected, but frequent errors (blurry contours, false positives) reduce reliability.

- **Satisfactory**: Generally correct segmentation but with notable errors requiring refinement.

- **Very Good**: Accurate segmentation with minor errors that do not impact clinical interpretation.

- **Excellent**: Near-perfect segmentation, closely matching expert annotations and clinically applicable.

This expert-driven evaluation complements numerical metrics, providing deeper insights into model applicability in clinical practice.

4.7.2 Expert Evaluation Results

The evaluation of segmentation models was conducted using both **quantitative metrics** and **expert qualitative assessment** (see Table 2). Interestingly, while the quantitative results ranked UNet as the top-performing model, expert pathologists favored UNet++ for its clinical relevance and segmentation precision.

Model	Expert Evaluation	BP Score	ER Score
UNet++	Very Good	8.5	12
UNet	Satisfactory	8	15
DeepLabV3+	Below Average	7	23

Table 2: Expert Evaluation of Deep Learning Models Based on Clinical Relevance

The experts used a **custom grading scale** incorporating:

- **Bon Point (BP)** – A positive score reflecting the model's ability to segment clinically relevant structures accurately.

- **Erreur Rédhibitoire (ER)** – A penalty assigned when a model makes critical segmentation errors that could mislead clinical diagnosis.

This discrepancy indicates that while UNet excelled in quantitative metrics, UNet++ produced more clinically relevant segmentations, likely due to its nested skip connections preserving fine structures.

5 Discussion

This study demonstrates the potential of deep learning models for segmenting cytological lymph node imprints, with significant variations in performance across the evaluated architectures. Quantitative analysis revealed that **U-Net** achieved the highest overall performance, particularly in terms of *accuracy* (0.7563), *Dice score* (0.7586), and *IoU* (0.6111). These results suggest that the traditional U-Net architecture remains highly effective for medical image segmentation tasks.

Interestingly, expert evaluation provided a different perspective, highlighting **U-Net++** as the most clinically relevant model. Despite not leading in quantitative metrics, U-Net++ achieved the highest *sensitivity* (0.9405) and received the highest expert evaluation score. This outcome likely reflects its ability to preserve fine cellular structures through nested skip connections, which proved critical for accurately detecting malignant regions. In contrast, **DeepLabV3+** underperformed in both quantitative measures and expert evaluations, although it demonstrated strong boundary detection capabilities, as evidenced by its lowest *Hausdorff Distance* (6.0000) and *Mean Surface Distance* (0.5109).

The divergence between quantitative results and expert assessments highlights a crucial insight: high numerical performance does not necessarily correlate with clinical relevance. The expert evaluations underscore the importance of considering clinical applicability, as models that preserve finer structures, like U-Net++, offer greater diagnostic value despite slightly lower quantitative scores.

Despite these promising results, the study presents certain limitations. The dataset, consisting of 148 annotated slides from a single institution, may limit the generalizability of the findings. Additionally, the manual annotations, although supervised by experienced pathologists, are subject to inter-observer variability, which could introduce bias into the training and evaluation process. The

restricted dataset size also reduces the model's ability to generalize across different clinical scenarios.

Overall, these results emphasize the importance of combining quantitative evaluations with expert clinical assessments to ensure that AI-based segmentation models provide reliable and clinically relevant support in medical diagnostics.

6 Conclusion

This study evaluated the effectiveness of three deep learning models **UNet++, UNet** and **DeepLabV3+** for cytological image segmentation in lymphoma diagnosis. By combining quantitative performance metrics with expert-driven qualitative assessments, we aimed to bridge the gap between numerical accuracy and clinical applicability.

Results showed that UNet++ was preferred by expert pathologists for its clinical interpretability, likely due to its nested skip connections that preserve fine-grained structures. This highlights the crucial role of expert validation alongside conventional performance metrics, as statistical accuracy alone does not always translate to optimal clinical usability.

These findings reinforce the importance of integrating both algorithmic performance and real-world clinical relevance when evaluating AI-driven medical imaging solutions. Incorporating expert assessments into AI validation frameworks can significantly improve their reliability and practical adoption in clinical practice.

Limitations

This study has two main limitations. First, the dataset, composed of 148 digitized lymph node cytological imprints from a single institution, may limit result generalizability. Additionally, grouping cell types into benign and malignant categories might mask distinctions between lymphoma subtypes.

Second, the high computational cost of training deep learning models, especially U-Net++ and DeepLabV3+, required substantial resources, potentially limiting their clinical deployment in settings with constrained infrastructure.

Future Work

Several directions could be pursued to enhance future research:

- **Refining cell subtype classification.** Future studies could distinguish specific lymphoma subtypes *Hodgkin lymphoma*, *non-Hodgkin lymphoma*, and *metastatic tumors* to improve detection of subtle morphological patterns.

- **Integrating clinical data.** Incorporating patient-specific variables (age, sex, medical history) could enhance model relevance and improve classification accuracy.

- **Using advanced staining techniques.** Methods like *Hematoxylin-Eosin-Saffron (HES)* staining and immunohistochemistry could improve the distinction between malignant and benign cells.

- **Expanding dataset diversity.** Including samples from multiple institutions would improve model generalizability and robustness across different clinical environments.

These research avenues could strengthen AI integration in clinical practice, enhancing lymphoma diagnostic accuracy and reliability.

References

Prabhjot Kaur Buttar and Manohar Kaur Sachan. 2022. Semantic segmentation of clouds in satellite images based on u-net++ architecture and attention mechanism. *Expert Systems with Applications*, 209:118380.

Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. *arXiv preprint arXiv:1802.02611*.

Michal Drozdzal, Eugene Vorontsov, Gil Chartrand, Samuel Kadoury, and Chris Pal. 2016. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*, pages 179–187. Springer.

Thorsten Falk, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Achim Böhm, Jan Deubner, Zaid Jäckel, Katharina Seiwald, Andriy Dovzhenko, Olaf Tietz, Stefan Wolf, Alexander Krull, Jörg M Buhmann, Thomas Schmidt, Martin Weigert, Felix Meißner, Olaf Ronneberger, Robert Haase, and Fred A Hamprecht. 2019. U-net: deep learning for cell counting, detection, and morphometry. *Nature Methods*, 16(1):67–70.

Bharath Hariharan, Pablo Arbelaez, Ross Girshick, and Jitendra Malik. 2016. Object instance segmentation and fine-grained localization using hypercolumns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):627–639.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708.

Bitao Jiang, Lingling Bao, Songqin He, Xiao Chen, Zhihui Jin, and Yingquan Ye. 2024. Deep learning applications in breast cancer histopathological imaging: diagnosis, treatment, and prognosis. *Breast Cancer Research*, 26:137.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *International Conference on 3D Vision (3DV)*, pages 565–571.

R. Patel, A. Singh, and R. Kumar. 2022. A hybrid cnn-rnn model for automated lymphoma classification from cytological images. *IEEE Transactions on Medical Imaging*, 40(5):1234–1245.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention*, pages 234–241.

John Smith and Jane Doe. 2023. Artificial intelligence in cancer diagnosis and prognosis: Current trends and future perspectives. *Cancer Research Journal*, 45:1123–1145.

Steven H. Swerdlow, Elias Campo, Nancy L. Harris, Elaine S. Jaffe, Stefano A. Pileri, Harald Stein, Jürgen Thiele, and James W. Vardiman. 2022. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*, 5th edition. International Agency for Research on Cancer.

Y. Xu, J. Huang, and X. Liu. 2021. Artificial intelligence in lymphoma diagnosis: Current state and future directions. *Journal of Medical AI Research*, 12:345–360.

Matthew D Zeiler and Rob Fergus. 2014. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision (ECCV)*, pages 818–833. Springer.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. 2018. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11. Springer.

Özgün Çiçek, Ahmed Abdulkadir, Sandra S Lienkamp, Thomas Brox, and Olaf Ronneberger. 2016. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 424–432.

Appendix 1 : Diagnostic classification system

Catégorie	Catégorie, abréviation	Diagnostic	Diagnostic, abréviation	Sous-diagnostic	Sous-diagnostic, abréviation
Lymphome	L	Lynphome de Hodgkin	LH	sous type classique	LH-C
Lymphome	L	Lynphome de Hodgkin	LH	sous type classique sclérosant-nodulaire	LH-CSN
Lymphome	L	Lynphome de Hodgkin	LH	sous type classique riche en lymphocytes	LH-CRL
Lymphome	L	Lynphome de Hodgkin	LH	sous type classique de cellularité mixte	LH-CM
Lymphome	L	Lynphome de Hodgkin	LH	sous type classique avec déplétion en lymphocyte	LH-CDL
Lymphome	L	Lynphome de Hodgkin	LH	sous type à prédominance lymphocytaire	LH-L
Tumeur non lymphom	TNL	Thymome	TH	Thymome A	TH-A
Tumeur non lymphom	TNL	Thymome	TH	Thymome AB	TH-AB
Tumeur non lymphom	TNL	Thymome	TH	Thymome B1	TH-B1
Tumeur non lymphom	TNL	Thymome	TH	Thymome B2	TH-B2
Tumeur non lymphom	TNL	Thymome	TH	Thymome B3	TH-B3
Lymphome	L	Lymphome B à petite cellules	LBPC	LBPC folliculaire	LBPC-F
Lymphome	L	Lymphome B à petite cellules	LBPC	LBPC folliculaire de grade 1/2	LBPC-FG1
Lymphome	L	Lymphome B à petite cellules	LBPC	LBPC folliculaire de haut grade	LBPC-FHG
Lymphome	L	Lymphome B à petite cellules	LBPC	LBPC lymphocytique / LLC	LBPC-LLC
Lymphome	L	Lymphome B à petite cellules	LBPC	LBPC du manteau	LBPC-M
Lymphome	L	Lymphome B à petite cellules	LBPC	LBPC lympho-plasmocytaire	LBPC-LP
Lymphome	L	Lymphome B à petite cellules	LBPC	LBPC de la zone marginale	LBPC-ZM
Lymphome	L	Lymphome B à petite cellules	LBPC	LBPC de MALT	LBPC-MA
Lymphome	L	Lymphome B à petite cellules	LBPC	LBPC splénique de la zone marginale	LBPC-ZM
Lymphome	L	Lymphome B à petite cellules	LBPC	LBPC splénique de la pulpe rouge	LBPC-PR
Lymphome	L	Lymphome B diffus à grandes cellules	LBDGC	DLBCL NOS de type centrogerminatif	LBDGC-GC
Lymphome	L	Lymphome B diffus à grandes cellules	LBDGC	DLBCL EBV induit	LBDGC-EBV
Lymphome	L	Lymphome B diffus à grandes cellules	LBDGC	LBDGC NOS de type non centrogerminatif	LBDGC-NGC
Lymphome	L	Lymphome B diffus à grandes cellules	LBDGC	LBDGC riche en lymphocytes T	LBDGC-RT
Lymphome	L	Lymphome B diffus à grandes cellules	LBDGC	LBDGC primitif du médiastin	LBDGC-PM
Lymphome	L	Lymphome B diffus à grandes cellules	LBDGC	DLBCL de type indéterminé	LBDGC-I
Lymphome	L	Lymphome T	LT	LT mycosis fongoid	LT-MF
Lymphome	L	Lymphome T	LT	LT anaplasique	LT-A
Lymphome	L	Lymphome T	LT	LT CD4	LT-CD4
Lymphome	L	Lymphome T	LT	LT syndrome de sezary	LT-S
Lymphome	L	Lymphome T	LT	LT EBV induit	LT-EBV
Lymphome	L	Lymphome T	LT	LT CD8	LT-CD8
Lymphome	L	Lymphome frontière du médiastin	LFM		
Lymphome	L	Lymphome B diffus à grandes cellules	DLBCL	DLBCL EBV induit	DLBCL-EBV
Tumeur non lymphom	TNL	Carcinome	C	Adénocarcinome	C-AD
Tumeur non lymphom	TNL	Carcinome	C	Adénocarcinome prostatique	C-ADP
Tumeur non lymphom	TNL	Carcinome	C	Carcinome épidermoïde	C-CE
Tumeur non lymphom	TNL	Mélanome	ME		
Non cancéreux	NC	Tissu fibreux	TF		
Non cancéreux	NC	Tissu fibreux	TF	Tissu fibro-inflammatoire	TFI
Non cancéreux	NC	Absence de cellules syspecte de malign	N		
Tumeur non lymphom	TNL	Carcinome	C	Neuroendocrine	
Tumeur non lymphom	TNL	Tumeur indifférenciée	TI		
Tumeur non lymphom	TNL	Lymphadénite chronique	LC		
Non cancéreux	NC	Lymphadénite granuomateuse	LG		
Non cancéreux	NC	Lymphadénite granuomateuse	LG	LG granulome épithéloïde	LGGE
Non cancéreux	NC	Lymphadénite granuomateuse	LG	LG granulome épithéloïde avec nécrose	LGGEN
Tumeur non lymphom	TNL	Tumeur neuroendocrine	TNE		
Non cancéreux	NC	Hyperplasie du tissu ganglionnaire	HG		
Non cancéreux	NC	Hyperplasie plasmocytaire	HP		
Non cancéreux	NC	Ganglion hyperplasique	GH		
Tumeur non lymphom	TNL	Sarcome	S		
Non cancéreux	NC	pseudo-sélection clonale T	SCT		
Non cancéreux	NC	Tissu fibreux inflammatoire	GFI		
Non cancéreux	NC	Involution adipeuse	II		
Non cancéreux	NC	Tissu fibreux ciccatricciel	GFC		
Non cancéreux	NC	Ganglion réactionnel	GR		
Tumeur non lymphom	TNL	Tumeur germinale	TG		
Tumeur non lymphom	TNL	Nécrose tumorale	NT		
Indéterminé	I	Indéterminé réactionnel ou lymphome	IRL		

Each category is composed of different diagnostics, and each diagnostic itself is composed of sub-diagnostics. Categories, diagnostics and sub-diagnostics all have corresponding abbreviations.