Report on the DigiLut Data Challenge Detection of Graft Rejection Following Lung Transplantation

Arijana Bohr and Emmanuelle Salin
Department Artificial Intelligence in Biomedical Engineering
Machine Learning and Data Analytics Lab
Erlangen, Germany
arijana.bohr@fau.de, emmanuelle.salin@fau.de

***Abstract*—Lung transplantation, the sole treatment for certain respiratory failures, requires early detection of acute rejection (type A rejection) to prevent graft failure and patient mortality. This study presents a methodology for developing a diagnostic algorithm to identify pathological regions in transbronchial biopsy images. Key steps include data preprocessing, artifact detection, and model training using a light-weight DINO vision transformer with optimized focal loss and soft labels as well as our proposed algorithm to efficiently detect lesions. The proposed approach shows promise in detecting type A rejection, potentially improving patient outcomes.**

***Index Terms*—Digital Pathology, Lung Transplants, Data Challenge**

I. Introduction

Lung transplantation is the only treatment available for certain types of respiratory failure. The success of transplantation depends on the occurrence of acute rejection episodes (type A rejection), which can lead to irreversible graft rejection and, ultimately, the patient's death.
The DigiLut Challenge, a collaborative project between Trustii.io and Foch Hospital and supported by the Health Data Hub and Bpifrance, aims to develop a medical decision support algorithm to diagnose graft rejection episodes in lung transplant patients by detecting pathological regions (A lesions) on transbronchial biopsies.
This paper presents an approach for using vision-transformer based models for diagnosing graft rejection in lung transplant patients by detecting pathological regions on transbronchial biopsies.

II. Methodology

In this section, we outline the methodology employed to develop a medical decision support algorithm for detecting graft rejection episodes in lung transplant patients. The process begins with an overview of the dataset, followed by the preprocessing of the data. We then detail the model training and optimization steps to accurately identify pathological regions on transbronchial biopsies. Finally, we discuss the inference process and the evaluation metrics used to assess the model's performance.

A. Data Overview

The dataset includes Whole slide images (WSIs), also known as virtual slides, which are high-resolution images used in digital pathology [1]. The anonymized database constructed from digitized biopsy slides includes annotations of the zones of interest, created by an international panel of expert pathologists. Only 25% of images have been annotated. The images that contain at least one bounding box contain at least one lesion (= graft rejection). The images that do not present bounding box can contain rejection zones (lesion = 1) or no rejection zones at all (no lesion = 0).
Due to the WSIs being quite large, we use the open-source Python library OpenSlide [1].
The available pathology images are provided at different magnification levels, ranging from zero to five. Level five is the most zoomed-out, while level zero offers the highest detail. Based on our expertise in mathematics and computer science, rather than digital pathology, our opinion is that magnification levels two and three appear to be the most suitable. These levels provide sufficient zoom to observe cells clearly without losing essential context, while also being more time-efficient as they require fewer patches for processing. On the other hand, levels four and five seem too zoomed-out, making it challenging to discern individual cells, whereas levels zero and one are overly detailed, lacking adequate context.
Because it is not with our resources computationally feasible, and probably not helpful in general to have the whole image as input to the ML model, we generate patches from the image with a patch size of 224×224 pixels, which aligns with the typical requirements of most machine learning models, minimizing the need for additional adaptation.

B. Data Preprocessing

In this section, we describe the preprocessing techniques applied to the dataset, which include removing irrelevant background regions, detecting and eliminating artifacts, and augmenting the data to standardize and enhance the visual features. These steps are essential for accurately improving

the model's ability to identify pathological regions in transbronchial biopsy images.

a) Removing background: Regions that are predominantely white are removed from the image.

b) Artifact detection: We encountered some artifacts in the pathology images, including a grey line near the top of many images. Since this artifact appears in varying positions, cropping the top of the image indiscriminately wouldn't be practical, as some regions of interest (ROIs) are located near or above the artifact.

Upon closer inspection, we observed that the ROIs generally display shades of pink, purple, or blue. To address the artifacts, we applied an algorithm to detect and highlight these colors. By inverting the resulting mask, we isolated and removed the artifacts, rendering those areas white while preserving the true ROIs.

To efficiently manage artifact detection, we implemented a multi-step approach. Initially, we perform artifact detection at a higher magnification, allowing us to quickly eliminate many unwanted regions. After narrowing down the areas of interest, we conduct another round of artifact detection at a lower magnification until we reach the desired magnification.

c) Data Normalization and Augmentation: We standardize the appearance of image patches before input into the model. The process begins by ensuring the image patch is in the RGB color space. Following this, the patch's contrast is enhanced, amplifying the visual features and potentially improving the model's ability to detect lesions. The patch undergoes luminosity standardization to achieve consistent lighting conditions across all patches. The patch is subsequently normalized using a pre-fitted normalizer. For this, we use the GitHub library 'stainlib' [2] and the Reinhard stain normalization technique [3], aligning the color distribution of the patches with that of a reference target image. Finally, an image processor converts the patch into a tensor.

C. Model Training and Optimization

The objective of the training process is to enable the model to accurately classify whether a given patch is associated with a lesion.

a) Selecting suitable patches for the training process: Each annotated image contains one or more bounding boxes that indicate the presence of type A lesions. As mentioned, we divided all images into patches and applied preprocessing steps, such as artifact removal and patch augmentation. To maintain a balanced dataset, we randomly selected a number of patches for the training process such that there is a 2:1 ratio of patches outside and within of bounding boxes.

b) Machine Learning Models: We employed a computer vision-transformer-based model (ViTs) to classify patches. Initially, we experimented with two models: a Swin Transformer (tiny) [5] and a DINO model (small) [4]. Both models demonstrated comparable performance; however, we ultimately decided to proceed with the DINO model.

The DINO model is a self-supervised learning method which is a form of self-distillation with no labels [4]. The 'small' variant of the DINO model is a more compact model with fewer parameters. We trained from a pre-trained checkpoint from hugging face: "facebook/dinov2-small-imagenet1k-1-layer" [8]. Initially, a binary classification approach was employed, but the introduction of soft labels later on proved to be a key strategy, significantly improving the model's performance of the validations metrics (precision and recall).

Given the unbalanced nature of the data, we trained the model using focal loss [6]. We conducted a limited parameter search to optimize the focal loss settings, settling on alpha = 0.05 and gamma = 2.0. Additionally, we adapted the loss function to accommodate the soft labels by using KL divergence in place of the traditional binary cross-entropy.

D. Inference and Evaluation

We propose an algorithm designed to efficiently detect lesions by processing the WSIs through leveraging the amount of detail in the different magnification levels of the images. The algorithm utilizes a combination of preprocessing, probability predictions, and clustering techniques to accurately identify potential lesions. The pseudocode for our proposed method is outlined below. The steps will be explained in detail in the following.

Algorithm 1 Lesion Detection Algorithm

Step 1: Create CSV with preprocessed level 3 Patches
Step 2: Predict level 3 lesion probability using the CSV
Step 3: Create level 3 bounding boxes:
    a. Select bounding boxes where prediction > 0.6
    b. Select bounding boxes with n non-overlapping regions (n is the number of bounding boxes)
Step 4: Create level 2 patches from the selected Level 3 Bounding Boxes
Step 5: Predict lesion probability from level 2 patches
Step 6: Create bounding boxes:
    a. Select the top $20 \times n$ patches
    b. Remove outliers (those with fewer than two neighbors)
    c. Perform clustering on the remaining patches

a) Step 1-3: We used a sliding window approach to go over the whole slide image at magnification level three to detect whether acute rejection episodes occur in an image. The patch size is 224×224 pixels and the step size is 180 pixels. Each patch (unless it is background/artifact) is preprocessed and then classified as whether it is part of a lesion.

After applying the sliding window approach to classify relevant patches from level three, we selected those with predictions exceeding 0.6. Additionally, in cases where the number of created bounding boxes at level three was insufficient, we included more patches to meet the required number.

b) Step 4-6: The next steps involve closely examining the predicted bounding boxes at a higher magnification level (level two) to refine the classification and achieve more precise delineation. After generating and classifying the

level two patches from the predicted level three bounding boxes, we concentrated on those with the highest prediction scores. Specifically, we select patches that rank within the top $20 \times n$ predictions, where n represents the desired number of bounding boxes. To reduce the risk of false positives, we exclude outliers—specifically, patches that do not have at least two neighboring patches.
Following this, we employed clustering.
We observed that the bounding box generation process is critical, especially given our per-patch classification approach. Enhancing the bounding box generation technique significantly improved our overall results.

To evaluate the accuracy of the final predicted bounding boxes on the validation set, we utilize the 'Generalized Intersection over Union' (GIoU) metric [7], aiming for a GIoU score greater than 0.5. The GIoU is defined as:

$$\text{GIoU} = \text{IoU} - \frac{\text{AC} - \text{AU}}{\text{AC}} \tag{1}$$

where IoU is the Intersection over Union, AC is the area of the smallest enclosing box that contains both the predicted and ground truth bounding boxes, and AU is the area of their union.

III. Discussion

This study demonstrates the potential of applying lightweight computer vision-transformer models to digital pathology data for detecting graft rejection following lung transplantation. Our results provide a foundation for further exploration and refinement.
Given our background in mathematics and computer science rather than digital pathology, we encountered several important questions. For instance, how do we best determine which model performs optimally? Are the observed improvements genuine, or might bias influence them? A key example is the GIoU threshold, which was set during training to seperate positive from negative patches at 0.01. This decision was made to account for contextual factors and the limited number of positive patches available. However, we recognize that this threshold may not fully align with medical significance, potentially influencing the accuracy of model validation.
Moreover, our per-patch classification approach posed difficulties in correctly aggregating patches into bounding boxes due to the inherent noisiness of the boxes.
Additionally, the presence of noisy labels can complicate the training and evaluation of models. Indeed, as the Trusti Team (Yaniss Hamiche) highlighted in the forum, the intra-observer reproducibility is around 0.4.
The size of our model was constrained by the available computing power (A100 GPU). With access to greater computational resources, it would be valuable to explore several areas further:

- **Pre-training methods:** Investigating different pre-training strategies, such as reconstruction versus image classification, could yield improvements. The methods we tested so far did not show a significant enhancement in performance.
- **Larger model sizes:** The small and tiny versions of the models we used may not perform as well as the base or large-sized transformer models, which could offer better results with increased capacity.
- **More fine-grained parameter search:** A more extensive search for optimal parameters could potentially lead to better model performance, which was limited by the current computational constraints.

We discovered that optimizing the bounding box generation greatly enhanced our results. However, due to time constraints, we were unable to further refine our technique.

IV. Acknowledgments

We would like to thank the Trustii.io and Foch Hospital with the support of the Health Data Hub and Bpifrance for this interesting dataset and research question.
We would love to hear back from you regarding our approach and to receive feedback on our model/preprocessing choices from a medical point of view.

References

[1] Gilbert, B., Kamentsky, L., and Bourdev, L., 2024. OpenSlide: A C library for reading whole slide image files. Version 3.4.1. Available at: http://openslide.org/.
[2] Cazares, S., 2023. Stainlib: Stain Normalization Library for Histopathology. Available at: https://github.com/sebastianffx/stainlib.
[3] Reinhard, E., Adhikhmin, M., Gooch, B., and Shirley, P., 2001. Color transfer between images. IEEE Computer Graphics and Applications, 21(5), pp.34–41. IEEE.
[4] Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., and Joulin, A., 2021. Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.9650–9660.
[5] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B., 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp.10012–10022.
[6] Lin, T.Y., Goyal, P., Girshick, R., He, K., and Dollár, P., 2017. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp.2980–2988.
[7] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S., 2019. Generalized intersection over union: A metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.658–666.
[8] Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al., 2023. Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193.