Dealing With Strong Imbalance to Classifying Glomerular Lesions via Self-Supervised Learning

Gabriel de Azevedo[1], Alexsandro Silva Santos[1],
Luciano Rebouças de Oliveira[2], Washington Luís Conrado dos Santos[3], Angelo Amâncio Duarte[1]
[1]Universidade Estadual de Feira de Santana (UEFS), Feira de Santana, Brazil
[2]Universidade Federal da Bahia (UFBA), Salvador, Brazil
[3]Fundação Oswaldo Cruz (FIOCRUZ), Salvador, Brazil
Emails: {gabrielsilvadeazevedo, alexsandrossantos75}@gmail.com,
Lrebouca@ufba.br, washington.santos@fiocruz.br, angeloduarte@uefs.br

***Abstract*—The histopathological analysis of renal glomeruli is fundamental for diagnosing kidney diseases, but it is hampered by challenges such as inter-observer variability, the high cost of expert annotation, and the inherent complexity of renal lesions, which often coexist. Self-Supervised Learning (SSL) offers a promising strategy to mitigate the reliance on large-scale annotated datasets. This work-in-progress paper presents a systematic comparison of SSL methods for classifying glomerular lesions on a challenging real-world dataset of 11,928 histopathology images. Our preliminary results demonstrate that the DINO method, significantly outperforms supervised baselines and other SSL techniques. An interesting finding from our data analysis is that several images depict glomeruli with multiple, co-occurring lesions, meaning that only normal glomeruli consistently represent a single, isolated class. This suggests that a standard multi-class classification approach may be suboptimal. Based on this finding, we propose the development of a novel two-stage, binary-multilabel pipeline as the main direction for our future work. This pipeline will first distinguish between normal and Lesion images and then apply a multi-label model to identify all specific lesions present, better reflecting clinical reality.**

I. Introduction

Chronic kidney diseases affect approximately 10% of the global population [1]. A key step in diagnosis is the histopathological analysis of renal glomeruli, central to identifying nephropathies and guiding treatment. However, this process faces challenges such as inter-observer variability, lengthy analysis, and difficulty in identifying rare lesions.

Computational Pathology (CPath) develops computational systems to support pathologists in routine tasks. In nephropathology, a major challenge is handling the large variability in biopsy staining, which alters image appearance and hinders the creation of standardized, generalizable models [2]–[4]. These models, based on Machine Learning (ML), require large, well-annotated datasets—costly to produce due to the need for expert annotation—and are particularly difficult to obtain for rare diseases [5], [6].

Self-Supervised Learning (SSL) can help address these issues by learning robust feature representations from unlabeled data, reducing dependence on manual annotation and enabling use of large medical imaging archives [7]–[10]. While SSL has been applied in medical imaging, its use in digital nephropathology, especially for stain-robustness and rare lesion detection, remains limited [11], [12].

Few-shot learning (FSL) tackles extreme label scarcity by enabling models to generalize to new classes with few labeled examples, often through meta- or metric-learning [13], [14]. A review of studies from 2018–2023 [15] shows FSL improves performance in data-scarce medical tasks, with meta-learning being prominent in areas like tumor segmentation and disease classification [16].

SSL and FSL offer complementary benefits: SSL extracts features from large unlabeled datasets, while FSL adapts models to underrepresented classes with minimal labeled data. Although explored separately in medical imaging, their combined use in nephropathology is unexplored.

We investigate whether integrating SSL and FSL can improve classification of glomerular lesions characterized by stain variability and severe class imbalance. The dataset is multi-label—glomeruli may show multiple lesions—and contains a high proportion of normal samples. If trained uniformly, the imbalance can hinder detection of rare lesion classes.

We propose a two-stage classification pipeline. First, a binary SSL-based model distinguishes normal from lesioned samples, reducing the impact of the majority class. Second, a multi-label classifier is trained only on lesioned glomeruli, with FSL mitigating the scarcity of annotated examples for certain lesions.

This work evaluates whether this SSL–FSL binary–multilabel pipeline can address class imbalance, lesion co-occurrence, and stain variability in renal glomerular histopathology.

II. Methodology

This study employs an experimental pipeline to investigate how different self-supervised learning (SSL) methods and backbone architectures affect the classification performance of models trained on an imbalanced dataset of renal glomerular lesions. While this study does not introduce a novel core methodology, its primary objective is to systematically analyze how architectural choices and SSL strategies interact with the specific challenges of renal histopathology, particularly severe class imbalance and multi-lesion co-occurrence. This

analysis then informs the development of a novel, two-stage classification pipeline designed to better reflect clinical reality [4], [17], [18].

The pipeline's design was centered on evaluating SSL methods. It consists of a pretraining phase in which models learn visual representations from unlabeled histopathological images, followed by a downstream classification task using a standard linear probe. This structure allows for isolated evaluation of SSL pretraining strategies, using consistent datasets and evaluation protocols.

To extend this pipeline, an additional fine-tuning stage was incorporated to explore the use of few-shot learning (FSL) techniques. This stage is decoupled from the SSL pipeline and can be independently applied after SSL pretraining. Instead of relying on full supervision for fine-tuning, this variant assumes a limited number of labeled examples per class, mimicking realistic annotation constraints. The FSL fine-tuning is applied only to the subset of Lesion samples, where label scarcity and class imbalance are most severe.

This modular design enables comparative analysis between conventional SSL pipelines and SSL combined with FSL strategies, allowing us to evaluate the potential benefits of integrating few-shot adaptation into self-supervised models in the context of multi-label classification and stain heterogeneity.

A. Dataset

The study utilizes a dataset of glomeruli images curated in collaboration with the Oswaldo Cruz Foundation (FIOCRUZ), comprising 11,928 JPEG images of renal glomeruli stained with five histological techniques: Hematoxylin-Eosin (HE), Periodic Acid-Schiff (PAS), Azan trichrome (AZAN), Periodic Acid Methenamine Silver (PAMS), and Preferential Solvation Index (PSI). These images span twelve diagnostic categories—eleven types of glomerular lesions and normal tissue—as detailed in Table I. The dataset is characterized by significant class imbalance, with sample counts ranging from 2,695 for normal glomeruli to only 34 for fibrinoid necrosis. In terms of resolution, the images exhibit substantial variability, with widths ranging from 219 to 4,128 pixels and heights from 205 to 3,286 pixels.

A critical characteristic revealed during analysis is the presence of co-occurring pathologies: 471 images (3.95%) are associated with two diagnostic classes and 111 images (0.93%) with three classes. This reflects the clinical reality where multiple Lesions can manifest simultaneously in glomeruli.

B. SSL Frameworks and Architectures

Three SSL paradigms were evaluated for representation learning from unlabeled histopathology images, selected based on their demonstrated efficacy in medical imaging domains [11], [19]:

- **SimCLR (Simple Framework for Contrastive Learning of Visual Representations)**: A contrastive framework maximizing agreement between differently augmented views of the same image while repelling dissimilar pairs [7]. Particularly effective for learning discriminative texture features.
- **BYOL (Bootstrap Your Own Latent)**: A non-contrastive approach using online and target networks where the online network predicts target network representations without negative samples [8]. Reduces computational complexity while maintaining performance.
- **DINO (Self-Distillation with no Labels)**: A self-distillation method training a student network to match output distributions of a momentum teacher across different image crops [9]. Excels at capturing global semantic patterns.

TABLE I: Ditribuition of glomeruli images in the dataset

Lesion Class	Total	HE	PAS	AZAN	Other
Total Samples	11928	5959	3537	1171	1254
Normal	2695	1585	542	223	345
Amyloidosis	374	145	102	31	96
Hypercellularity	3134	1890	987	257	0
Pure Hypercellularity w/o Cres.	224	0	164	60	0
Crescent	1104	467	359	121	157
Membranous	1539	712	367	136	324
Sclerosis	617	276	219	0	122
Podocytopathy	505	65	244	90	106
Neutrophil Accumulation	127	76	44	0	7
Hyaline Deposits	94	51	29	14	0
Fibrinoid Necrosis	34	20	8	6	0

C. Experimental Design

The framework implements a four-stage pipeline (Fig. 1). To ensure statistical reliability given the dataset size (11,928 images), we employed stratified 5-fold cross-validation. This approach maximizes data utilization while providing stable performance estimates, particularly important given the class imbalance where some categories contain fewer than 50 samples.

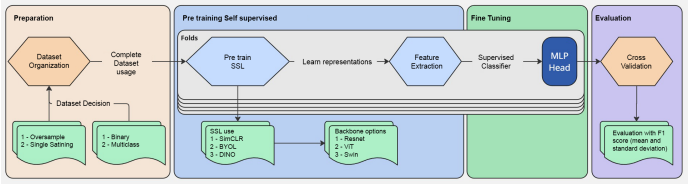

Fig. 1: Pipeline used for the experiments. The process is divided into four main stages: Preparation, Self-Supervised Pretraining, Fine-Tuning, and Evaluation. In the Preparation stage, the dataset is selected and organized, including options for oversampling and choosing between binary or multiclass classification. The entire dataset is then used in the Self-Supervised Pre-training stage, where one of the SSL methods (SimCLR, BYOL, or DINO) is applied across multiple folds to learn feature representations using a selected backbone architecture (ResNet, ViT, or Swin). In the Fine-Tuning stage, a supervised classifier (MLP Head) is trained on the extracted features. Finally, in the Evaluation stage, cross-validation is performed and the model's performance is measured using the F1-micro, reporting both the mean and standard deviation.

Data preparation involved organizing the images by histological stain (HE, PAS, AZAN, PAMS, PSI) and diagnostic class. The dataset was structured to support two classification schemes from the beginning: a 12-class multiclass task and multiple binary classification tasks, each focused on identifying specific lesion types. Additionally, configurations with and without oversampling of minority classes—performed through duplication—were defined during the preparation stage, prior to pretraining. These variations were intended to assess how different model configurations would generalize under varying class distributions.

Following the dataset preparation, which defined the classification schemes and oversampling strategies, the next step involved self-supervised pretraining. This phase was carried out using the unlabeled dataset, allowing the model to learn generalizable feature representations without relying on manual annotations. Different SSL methods were applied, each with specific data augmentation strategies: SimCLR and BYOL used random cropping, color jittering, and horizontal flipping, while DINO employed a multi-crop strategy that combined global views (covering more than 50% of the image) with smaller local patches (96×96 pixels).

After self-supervised pretraining, the learned representations were transferred to a supervised fine-tuning stage. In this phase, the feature extractor—pretrained and frozen—was used to generate embeddings, which were then passed to trainable classification modules composed of multilayer perceptrons (MLPs). Distinct MLP classifiers were employed for each task: one tailored for the 12-class multiclass setting and separate ones for each binary lesion classification. Fine-tuning was conducted using the Adam optimizer with an initial learning rate of 0.0005, a batch size of 32, and cross-entropy loss. Training was monitored using early stopping, with a patience of 5 epochs and a minimum improvement threshold of $\sigma = 0.001$ to prevent overfitting.

Given significant class imbalance (Normal: 2,695 samples vs. Fibrinoid Necrosis: 34 samples), the primary performance measure was the F1-micro, which accounts for overall correctness without class frequency bias. Secondary analyses included per-class F1-micro scores and t-SNE visualizations.

The experiments were structured to evaluate the impact of different self-supervised learning strategies and classification setups. Comparisons between SSL methods—SimCLR, BYOL, and DINO, all using a ResNet-18 backbone—were conducted within the 12-class multiclass classification framework to assess their ability to capture comprehensive diagnostic representations. In parallel, binary one-vs-all experiments were performed exclusively with the DINO method, also using a ResNet-18 backbone, to focus on the detection of specific lesion types. This experimental design enabled the analysis of both general classification performance across multiple classes.

The second line of investigation focused on assessing the impact of different neural network architectures, using the best-performing self-supervised learning method identified in the initial comparison (DINO). This evaluation included ResNet variants with 18, 50, and 101 layers, as well as transformer-based models such as ViT-B/16 and Swin Transformer in Tiny, Small, and Base configurations. All architectures were evaluated exclusively within the 12-class multiclass classification task, in order to determine whether performance gains from specific backbones contribute to improved diagnostic representation learning.

The third line of investigation aimed to evaluate the robustness of the models to variations in dataset configuration, using the optimal combination of SSL method and backbone architecture (DINO with ResNet-18). This set of experiments tested how different training conditions—specifically, using either multi-stain data or only PAS-stained images, and applying either the original class distribution or minority-class oversampling—affect model performance. Both multiclass and binary classification tasks were used to assess whether these variations influenced general diagnostic classification or lesion-specific detection. In the binary setup, a strict one-vs-all scheme was adopted, where each lesion type was treated as a separate classification problem against all other samples.

III. Preliminary Results and Discussion

Our initial experiments evaluated the effectiveness of self-supervised learning (SSL) techniques on the glomeruli image dataset through a classification task. These analyses were structured to investigate key aspects of the learning pipeline, including the quality of learned representations, the impact of data imbalance, the influence of different neural network architectures, and the relevance of the models to clinically meaningful scenarios.

A. Visualization of Learned Representations with t-SNE

We performed t-SNE visualization to qualitatively assess the structure of learned representations from SSL models, particularly after fine-tuning. This step was crucial to understand how the model internally organizes histological patterns and to identify potential areas of inter-class confusion or overlap.

The visualization indicates patterns in the learned feature space that relate to class representation and separability. The Normal class, shown in green on the right side of the plot, forms a more concentrated region, possibly associated with its higher prevalence in the dataset and lower morphological variability. Minority classes, including Fibrinoid Necrosis (red), Hyaline Deposits (orange), and Hypercellularity (pink), appear scattered throughout the space, without clearly defined groupings. This dispersion reflects the difficulty in learning consistent representations for less frequent lesions. Overlapping areas between certain classes, particularly those with similar morphological traits, can be observed near the center of the plot. These overlaps suggest challenges in distinguishing between classes with subtle visual differences. Overall, the visualization served to clarify how difficult the separation between classes is in this setting.

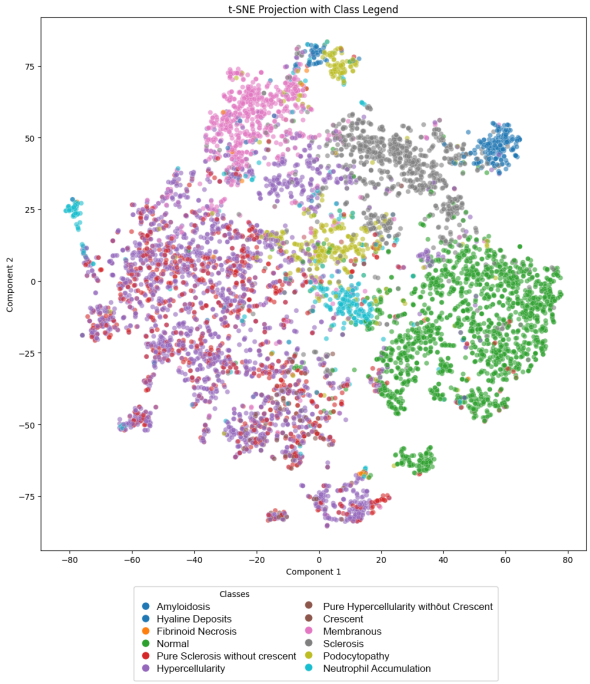

Fig. 2: t-SNE visualization of feature representations obtained from the SimCLR model after supervised fine-tuning on the multiclass classification task. Each point represents a glomerulus image, and colors indicate its corresponding histopathological class. Distinct clusters are observable for some classes, such as 'Normal' (green, right side), 'Amyloidosis' (dark blue, top right), and 'Podocytopathy' (cyan, top center), suggesting high intra-class similarity and separability. In contrast, classes such as 'Hypercellularity' (pink), 'Hyaline Deposits' (orange), and 'Fibroid Necrosis' (red) exhibit significant spatial overlap in the central region of the plot, indicating semantic or visual similarity that may contribute to classification confusion.

B. Comparison: Multiclass vs. One-vs-All Without Rebalancing

Two different classification paradigms were compared: a standard multiclass classification approach and a one-vs-all reformulation, both without any data balancing techniques. The multiclass setup follows a 5-fold cross-validation strategy to mitigate data sparsity issues in certain classes. The complete results for multiclass classification are shown in Table II. Notably, the DINO model consistently outperforms both the supervised baseline and SimCLR across most classes, including rare ones such as Amyloidosis.

To investigate whether rare-class detection could be improved, the same task was reformulated into independent binary problems (one-vs-all), allowing a focused evaluation of each lesion's separability. The binary classification results for DINO and SimCLR, also without balancing, are shown in Table III.

Overall, DINO achieves the best results in the multiclass setting, while SimCLR surpasses DINO in 6 out of 9 binary cases. This behavior is consistent with SimCLR's contrastive nature, which inherently favors sharper binary decision boundaries, whereas DINO maintains more balanced performance across the full label set.

TABLE II: F1-micro scores (mean ± std) per class across 5-fold cross-validation for multiclass classification task. Comparison between supervised baseline, SimCLR, and DINO methods.

Class	Baseline	SimCLR	DINO
Amyloidosis	46.80 ± 17.62	47.44 ± 5.00	78.25 ± 3.25
Normal	86.22 ± 1.11	80.70 ± 1.28	91.50 ± 0.97
Hypercellularity	55.20 ± 4.09	64.58 ± 1.23	69.91 ± 1.00
Hip. P.S.C.	0.51 ± 1.01	0.00 ± 0.00	2.99 ± 2.75
Crescent	65.30 ± 3.96	60.11 ± 1.32	78.81 ± 0.97
Membranous	64.55 ± 2.49	55.19 ± 1.73	77.28 ± 1.62
Sclerosis	40.60 ± 3.07	26.50 ± 2.14	62.01 ± 3.60
Podocytopathy	52.55 ± 5.05	52.75 ± 2.50	70.75 ± 3.40
Other Lesions	37.73 ± 10.62	9.37 ± 7.84	56.42 ± 7.69

TABLE III: Binary classification performance per lesion class using DINO and SimCLR models without balancing. Each row shows the F1-micro scores (mean ± standard deviation) across 5-fold cross-validation.

Class	DINO	SimCLR
Amyloidosis	20.86 ± 4.28%	32.77 ± 5.74%
Normal	63.26 ± 2.78%	81.37 ± 1.82%
Hypercellularity	54.55 ± 2.44%	54.13 ± 2.75%
Hypercellularity P. S. C.	12.56 ± 0.80%	6.76 ± 3.32%
Crescent	37.93 ± 2.92%	55.09 ± 4.98%
Membranous	37.86 ± 2.64%	44.96 ± 4.14%
Sclerosis	19.90 ± 2.52%	29.00 ± 6.41%
Podocytopathy	26.49 ± 3.43%	42.11 ± 2.59%

C. Impact of Data Balancing Techniques

In this stage, we fixed the classification setup to the **binary one-vs-all formulation** using the DINO model, and evaluated multiple data balancing strategies to address the class distribution disparity present in the dataset. The tested strategies included random oversampling and stain-specific training, such as using only PAS-stained images. The latter was motivated by the fact that PAS is the only stain represented across all lesion classes, providing a more consistent visual basis for learning. Each strategy was applied independently to assess its effect on classification performance. Table IV presents the comparative results, enabling an analysis of how each method influenced the detection of underrepresented classes.

TABLE IV: Impact of data balancing techniques on DINO (F1-micro for the positive class) in the one-vs-all setting.

Class	All Stains	PAS Only	All Stains + OS	PAS + OS
Normal	50.68 ± 11.20%	41.98 ± 7.95%	63.26 ± 2.78%	56.25 ± 3.93%
Amyloidosis	0.00 ± 7.57%	0.00 ± 8.85%	20.86 ± 4.28%	38.71 ± 13.34%
Hypercellularity	39.17 ± 20.33%	51.40 ± 4.08%	54.55 ± 2.44%	57.14 ± 3.28%
Hypercellularity P. S. C.	0.00 ± 0.00%	0.00 ± 0.00%	12.56 ± 0.80%	18.48 ± 2.38%
Crescent	12.14 ± 11.75%	13.64 ± 10.55%	37.93 ± 2.92%	43.90 ± 3.96%
Membranous	4.88 ± 11.11%	7.79 ± 4.62%	37.86 ± 2.64%	33.49 ± 4.36%
Sclerosis	0.00 ± 2.51%	4.35 ± 9.67%	19.90 ± 2.52%	28.79 ± 5.25%
Podocytopathy	8.40 ± 9.60%	19.35 ± 5.40%	26.49 ± 3.43%	29.60 ± 4.39%

The results demonstrate that oversampling (OS) enhances F1-micro by 15-40% across classes, while PAS-only train-

ing specifically benefits certain classes like Hypercellularity; combining PAS with oversampling yields the best results for minority classes, and variance reduces significantly with these balancing techniques. This analysis was essential to quantify the improvement in rare-class recall from balancing, identify the role of staining standardization in learning discriminative features, and demonstrate the variance reduction achieved through oversampling, highlighting its regularization effect. The results confirm that preprocessing choices and sampling strategies critically affect model fairness and reliability, especially for underdiagnosed conditions.

D. Insights from Prior Few-Shot Learning Experiments

To contextualize our current investigation into self-supervised learning (SSL), compare the results of SSL against Few-Shot Learning FSL obtained on the same dataset [20].

Table V summarizes the F1-scores obtained across multiple lesion types and datasets. Notably, FSL achieved the highest performance in the classification of *HE-amyloidosis* ($92.60\% \pm 3.11\%$) and *Curated-HE-membranous nephropathy* ($87.73\% \pm 3.16\%$). In other lesion types, such as *PAS-sclerosis* and *PAS-podocytopathy*, FSL also achieved competitive results compared to convolutional neural network (CNN) baselines. However, for classes like *ROS-PAS-pure hypercellularity* ($27.81\% \pm 1.39\%$), CNN approaches performed better, highlighting that FSL is not uniformly superior across all lesion types.

These results underscore three main limitations: (i) reduced generalization to classes with higher intra-class variability, (ii) sensitivity to the staining protocol used in training, and (iii) reliance on limited support samples, which can introduce prototype bias. Such constraints motivate the exploration of SSL methods that can learn more robust, stain-invariant representations from large and diverse datasets. Integrating SSL embeddings into FSL pipelines may help reduce variance sensitivity and improve generalization, as suggested in recent work [21], [22].

TABLE V: F1-scores by class and dataset from prior experiments.

Class	Dataset	F1-Score
Amyloidosis	HE-amyloidosis	92.60% (±3.11%)
Normal	ROS-HE-normal	96.41% (±0.62%)
Pure Sclerosis	Baseline-pure-sclerosis	50.42% (±1.31%)
Hypercellularity	ROS-Curated-HE-hypercellularity	78.79% (±1.51%)
Pure Hypercellularity	ROS-PAS-pure-hypercellularity	27.81% (±1.39%)
Crescentic Glomerulonephritis	Curated-glomerulonephritis	84.02% (±0.73%)
Membranous Nephropathy	Curated-HE-membranous	87.73% (±3.16%)
Sclerosis	PAS-sclerosis	73.91% (±7.93%)
Podocytopathy	PAS-podocytopathy	75.76% (±5.76%)

E. Discussion: Reframing the Problem

DINO achieved the highest average F_1-micro score (58.27%), outperforming the supervised baseline (43.47%) and other SSL methods such as SimCLR (36.33%) and BYOL (8.62%). SimCLR performed relatively better in binary tasks, e.g., *Hypercellularity* detection (64.58% vs. DINO's 69.91%), reflecting its advantage in scenarios with clearer decision boundaries. SSL improved results by learning from unlabeled data, mitigating the scarcity of annotations and enabling better generalization to underrepresented classes, unlike supervised learning, which tends to overfit dominant categories in imbalanced datasets.

Task structure also impacted performance. Multiclass classification struggled with extreme imbalance, as seen in *Necrosis* ($F_1 < 9.37\%$ across methods). Reformulating the task as binary one-vs-all boosted rare-class detection, with DINO's *Neutrophil Accumulation* score rising from 0% to 56.42%.

Dataset characteristics further shaped outcomes. About 5% of samples had co-occurring lesions, challenging the assumption of mutually exclusive classes and supporting multilabel strategies. Training only on PAS-stained images improved some lesions ($+12.23\%$ for *Hypercellularity*) but harmed others, indicating a need for stain-robust methods. Oversampling also helped, e.g., *Fibrinoid Necrosis* improved from 0% to 36.55% when combined with PAS-specific training.

Overall, a pipeline with an initial binary screening to filter normal cases, followed by multilabel classification for overlapping lesions, may better reflect dataset heterogeneity. SSL's strong performance in low-data conditions also makes it a promising foundation for enhancing few-shot learning.

IV. Proposed Future Work and Schedule

Building on the preliminary results, future work will refine and validate a two-stage classification pipeline for renal pathology, combining self-supervised pretraining with few-shot learning to address limited annotations and lesion diversity.

A. Binary–Multilabel Pipeline with Few-Shot Adaptation

The pipeline comprises:

1) **Stage 1: Binary Screening (Normal vs. Lesioned):** A DINO-pretrained Swin Transformer will distinguish "Normal" from "Lesioned" tissue, leveraging its strong performance on the Normal class (F1-micro: 91.5%) and ability to generalize visual patterns [11], [12]. This step prioritizes computational resources for diagnostically relevant cases.
2) **Stage 2: Multi-Label Lesion Classification with Few-Shot Learning:** Lesioned samples will be processed by a multilabel classifier detecting 11 renal lesions [19]. Few-shot techniques will mitigate class imbalance and improve detection of rare or co-occurring pathologies [23].

Planned experiments will evaluate the pipeline's overall performance, robustness in detecting infrequent lesions, and the added value of few-shot learning for underrepresented classes, assessing its potential for real-world diagnostic support.

V. Conclusion

This work-in-progress establishes the significant potential of Self-Supervised Learning for the automated classification of renal histopathology images. Our preliminary results not only demonstrate that SSL methods like DINO can substantially

outperform traditional supervised approaches but also reveal crucial insights about the nature of the data itself.

The key contribution of our ongoing research is the data-driven reframing of the problem—from a simple multi-class task to a more clinically relevant binary-multilabel framework. The proposed two-stage pipeline is designed to accommodate the biological reality of co-existing lesions and is expected to yield a more robust and accurate diagnostic support tool. Our next steps are clearly defined, focusing on the implementation of this pipeline, its rigorous statistical and clinical validation, and the dissemination of our findings.

ACKNOWLEDGMENT

The authors would like to thank Dr. Washington Luis Conrado dos Santos, a pathologist at FIOCRUZ, for providing the invaluable dataset of histopathological images that made this study possible. Angelo Duarte is sponsored by FAPESB and UEFS, under the grants PET 0017/2024 and FINAPESQ 115/2024. Luciano Oliveira is sponsored by CNPq under the grant 301789/2025-8. Washington LC dos-Santos is supported by CNPq, grant No. 406141/2023.

REFERENCES

[1] M. Haas *et al.*, "Consensus definitions for glomerular lesions by light and electron microscopy: recommendations from a working group of the renal pathology society," *Kidney international*, vol. 98, no. 5, pp. 1120–1134, 2020.

[2] A. Tiard *et al.*, "Stain-invariant self supervised learning for histopathology image analysis," *arXiv preprint arXiv:2211.07590*, 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2211.07590

[3] D. Tellez *et al.*, "Whole-slide mitosis detection in h&e breast histology using phh3 as a reference to train distilled stain-invariant convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 636–647, 2019.

[4] M. Gadermayr *et al.*, "Stain-adaptive self-supervised learning for histopathology image analysis," *Medical Image Analysis*, vol. 77, p. 102385, 2022.

[5] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," *Medical image analysis*, vol. 63, p. 101693, 2020.

[6] W. He, T. Liu, Y. Han, W. Ming, J. Du, Y. Liu, Y. Yang, L. Wang, Z. Jiang, Y. Wang *et al.*, "A review: The detection of cancer cells in histopathology based on machine vision," *Computers in Biology and Medicine*, vol. 146, p. 105636, 2022.

[7] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[8] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar *et al.*, "Bootstrap your own latent-a new approach to self-supervised learning," *Advances in neural information processing systems*, vol. 33, pp. 21 271–21 284, 2020.

[9] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby *et al.*, "Dinov2: Learning robust visual features without supervision," *arXiv preprint arXiv:2304.07193*, 2023.

[10] A. Taleb, W. Loetzsch, N. Danz, J. Severin, T. Gaertner, B. Bergner, and C. Lippert, "3d self-supervised methods for medical imaging," *Advances in neural information processing systems*, vol. 33, pp. 18 158–18 172, 2020.

[11] S.-C. Huang, A. Pareek, M. Jensen, M. P. Lungren, S. Yeung, and A. S. Chaudhari, "Self-supervised learning for medical image classification: a systematic review and implementation guidelines," *NPJ Digital Medicine*, vol. 6, no. 1, p. 74, 2023.

[12] J. Wang, H. Quan, C. Wang, and G. Yang, "Pyramid-based self-supervised learning for histopathological image classification," *Computers in Biology and Medicine*, vol. 165, p. 107336, 2023.

[13] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/hash/cb8da6767461a14cfe3c4e4aa4f2ec9f-Abstract.html

[14] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proceedings of the 34th International Conference on Machine Learning (ICML)*, vol. 70. PMLR, 2017, pp. 1126–1135. [Online]. Available: https://proceedings.mlr.press/v70/finn17a.html

[15] E. Pachetti and S. Colantonio, "A systematic review of few-shot learning in medical imaging," *Artificial Intelligence in Medicine (preprint / arXiv)*, 2023.

[16] N. Schiavone, J. Wang, S. Li, R. Zemp, and X. Li, "Myriadal: Active few shot learning for histopathology," in *Proceedings of IEEE Conference on Artificial Intelligence (or relevant workshop)*, 2024, combines contrastive encoder and pseudo-label active few-shot learning in histology.

[17] K. Stacke *et al.*, "A closer look at domain shift for deep learning in histopathology," *Medical Image Analysis*, vol. 73, p. 102152, 2021.

[18] Y. Jin *et al.*, "Histossl: Unleashing the power of self-supervised learning for histopathology," *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3652–3663, 2022.

[19] O. Ciga, T. Xu, and A. L. Martel, "Self supervised contrastive learning for digital histopathology," *Machine Learning with Applications*, vol. 7, p. 100198, 2022.

[20] —, "Addressing class imbalance in renal amyloidosis classification: A comparative study of few-shot learning and conventional machine learning," in *IMPROVE 2025 – Proceedings of the 5th International Conference on Image Processing and Vision Engineering*.

[21] Z. Chen, J. Ge, H. Zhan, S. Huang, and D. Wang, "Pareto self-supervised training for few-shot learning," *arXiv preprint*, 2021, arXiv:2104.07841.

[22] Y. Lu, L. Wen, J. Liu, Y. Liu, and X. Tian, "Self-supervision can be a good few-shot learner," *arXiv preprint*, 2022, arXiv:2207.09176.

[23] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging properties in self-supervised vision transformers," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 9650–9660. [Online]. Available: https://arxiv.org/abs/2104.14294