POLITECNICO
MILANO 1863

AN2DL - Second Challenge Report
NoNameTeam

Pietro Ghersetich, Alessandro Mattiazzi, Matteo Mugnai, Tommaso Tron

pietroghersetich, alemattiazzi, mugna0990, thomastronz

303190, 302419, 306252, 299596

December 16, 2025

1 Introduction

This project addresses the domain of *image classification* for predicting the molecular subtype of diseased human tissue samples. The dataset pairs each image with a binary mask that helps to identify the areas that are likely to contain the diseased tissue.

The primary objective is to classify the actual molecular subtype of the tissue into four distinct classes: [Luminal A, Luminal B, HER2(+), Triple negative]. Our goal is to develop a deep learning model that **accurately predicts these subtypes**, assessed by the weighted **F1-score**. Our approach focuses on using Convolutional Neural Networks (CNNs), specifically utilizing a **Transfer Learning** strategy with a pretrained architecture to effectively capture the visual characteristics and irregularities within the tissue samples.

2 Problem Analysis

Our models are evaluated according to the **weighted F1-score**, which is the primary metric for this challenge. It is defined as:

$$F1_{weighted} = \sum_{c \in C} w_c \cdot 2 \cdot \frac{\text{precision}_c \cdot \text{recall}_c}{\text{precision}_c + \text{recall}_c} \quad (1)$$

This metric's sensitivity to class distribution makes handling data imbalance a critical aspect of the project, especially in a medical context where certain subtypes may be rare.

The given dataset contains 1,272 high-resolution images of diseased human tissue, split into training and test sets. A unique feature of this challenge is the inclusion of **binary masks** for each image, identifying the relevant regions for the analysis. This adds a layer of complexity since the model must learn to focus on the biological morphology within these regions of interest while ignoring the sterile background.

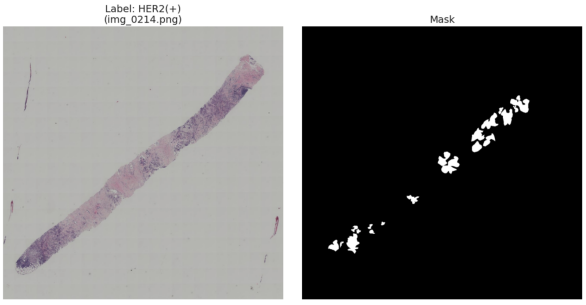

Figure 1: Tissue sample with its paired binary masks.

We began with a data inspection phase to identify potential outliers and duplicate images. Through a simple script we found and removed 224 images, reducing the training set down to 464 images.

During the analysis of input images we also noted that images are of different sizes, we then required a robust preprocessing pipeline to standard-

ize inputs avoiding a loss of morphological details.

Another significant characteristic identified was, as is common in medical datasets, a **significant class imbalance** in the training labels. As shown in Table 1, the 'Luminal A' subtype is the most prevalent (35.13%), whereas the aggressive 'Triple Negative' type is significantly underrepresented, accounting for only 11.85% of the data. This imbalance directly guided our methodological choices to ensure the model does not become biased toward the majority classes.

Table 1: Cleaned training_set label distribution.

Class	Occurrences	Percentage
0: Luminal A	163	35.13%
1: Luminal B	126	27.16%
2: HER2(+)	120	25.86%
3: Triple Negative	55	11.85%
Total	464	100.00%

3 Method

Our methodology was structured to first address the dataset's specific characteristics through preprocessing, then systematically experiment with different models and techniques to overcome the problem's main challenges.

Data Preprocessing and Augmentation

Our first task was transforming the raw high-resolution tissue images into a format suitable for our CNNs. This involved:

1. **Mask Usage**: Since the original images are large and contain significant blank background areas, we utilized the masks to reduce redundant information. This strategy allowed us to furnish the model with informative, high-resolution tissue patches while discarding the redundant information.

2. **Augmentation**: To improve generalization and simulate staining variability that is characteristic of histological images, we implemented multiple geometric and color augmentations during the training phase.

3. **Normalization**: All input patches were normalized using standard ImageNet mean and standard deviation values to ensure compatibility with transfer learning architectures.

Class Imbalance

As noted, the dataset suffers from a significant class imbalance. To address this, we first tried making use of a **Weighted Cross-Entropy Loss** (equation 3), where class weights were inversely proportional to their frequency in the training set.

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i) \quad (2) \qquad \mathcal{L}_{WCE} = -\frac{1}{N}\sum_{i=1}^{N} w_{c_i} \cdot \log(\hat{y}_i) \quad (3)$$

Later we also tried balancing the dataset by using a **Weighed Image Augmentation** to bring the number of occurrences per class to the level of the maximum represented one.
This approach proved essential for preventing the model from ignoring the minority classes, directly optimizing for the weighted F1-score metric.

Model Choices

We initially started developing custom CNNs to establish a baseline. However, after early experiments confirmed the superior potential of pretrained weights, we quickly switched to **Transfer Learning**. This strategy allowed us to use representations learned from large-scale datasets, which substantially improved performance compared to training from scratch.

4 Experiments

Preprocessing and Augmentation Search

We initially attempted to apply binary masks to the full images to black out the background. However, shrinking these masked full images to fit standard input sizes resulted in a loss of resolution critical for identifying cellular subtypes. We therefore shifted to a **tiling strategy**, generating N random high-resolution crops per image focused on the masked regions. This preserved morphological details and significantly boosted performance. Regarding data augmentation, we initially implemented standard geometric transformations but achieved superior generalization by adopting **RandAugment** [2].

Class Imbalance

To handle class imbalance, we compared two approaches: **Weighted Cross-Entropy Loss** and a weighted sampling strategy (generating more tiles for minority classes). The sampling approach showed no improvement, whereas the weighted loss proved effective and was retained.

Table 2: Performance comparison of key models on our validation set.

Model	Accuracy	Precision	F1 (weighted)
EfficientNet [7]	34.98%	37.53%	0.4548
ResNet50 [4]	35.90%	37.94%	0.5052
Final Model (Phikon) [3]	35.04%	37.69%	0.4613

Optimizer Choices

Regarding optimization, we experimented with **Adam**, **AdamW**, and the **Lion** optimizer [1]. AdamW consistently provided the most stable convergence and best final results, leading us to discard Lion, which necessitated lower learning rates that negatively impacted convergence speed without offering significant performance improvements.

Architecture Evolution and Final Model

Our architectural search began with lightweight models like **MobileNetV3** [5] to establish a baseline. We then scaled up to larger general-purpose architectures, including **ResNet50** [4], **EfficientNet** [7], and **ConvNext** [6]. Among these standard CNNs, ResNet50 achieved the highest performance but still struggled to differentiate between the subtler molecular subtypes.

Consequently, we moved to domain specific foundation models available via HuggingFace and selected **Phikon** [3]. This model, pre-trained on histology data, demonstrated a superior ability to extract relevant features from our tissue tiles.

Our final successful configuration combined the **Phikon** backbone with our **10-view tiling strategy**, the **AdamW** optimizer and **RandAugment**. For inference, we employed a **multi-view voting strategy**, feeding multiple tiles of a test image to the model and averaging the output probabilities to determine the final class.

In the last few days we also attempted an **ensemble** of Phikon and EfficientNet models, but this did not show the expected results.

5 Results

As detailed in Table 2, our final architecture, built upon the **Phikon ViT** backbone [3], achieved a **weighted F1-score** of 0.4613 on the validation set. This model achieved our most successful configuration on the test set, although the ResNet50 model achieved the strongest overall performance on the validation set.

Despite these improvements, the performance plateau highlights the inherent difficulty of the challenge. While the **Weighted Cross-Entropy Loss** successfully forced the model to attend to minority classes, distinguishing the subtler molecular subtypes remains difficult. This indicates that while our model captures general visual patterns well, it struggles to detect the subtle differences needed to distinguish rare classes, such as *Triple Negative*.

6 Discussion

Our **tiling strategy** was a key factor in performance, as it successfully directed the model's focus toward significant tissue details rather than the uninformative background (Figure 2), while still keeping a high resolution image.

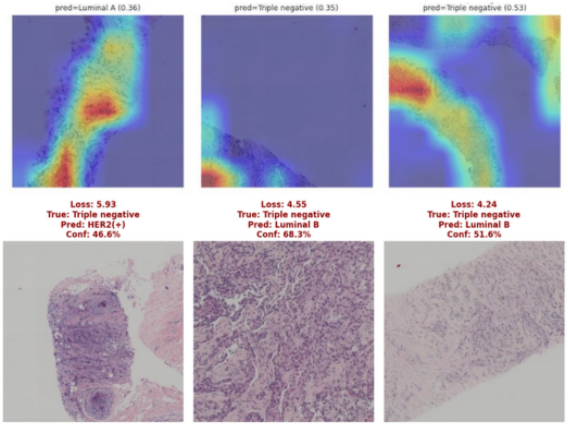

Figure 2: Tiling focus CAM (top) and highest loss samples (bottom).

However, the Triple Negative class proved difficult to learn. The model exhibited the highest losses on this subtype (Figure 2), indicating that its specific morphological features were significantly harder to distinguish compared to other classes.

7 Conclusions

We successfully tackled the classification of molecular subtypes using a **Transfer Learning** approach

based on the Phikon backbone and a specific **tiling strategy**. However, performance could still be boosted significantly, by testing more advanced and optimized **Data Augmentation** techniques.

References

[1] X. Chen, C. Liang, D. Huang, E. Real, K. Wang, H. Liu, H. Pham, X. Dong, T. Luong, C.-J. Hsieh, et al. **Symbolic discovery of optimization** algorithms. *Advances in Neural Information Processing Systems*, 36, 2023. Link.

[2] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le. **Randaugment**: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020. Link.

[3] A. Filiot, R. Giret, A. Ben Cheikh, and et al. Scaling **Self-Supervised Learning for Histopathology** with masked image modeling. *medRxiv*, 2023. Link.

[4] K. He, X. Zhang, S. Ren, and J. Sun. **Deep residual learning** for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. Link.

[5] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, et al. Searching for **MobileNetV3**. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1314–1324, 2019. Link.

[6] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie. A **ConvNet** for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11976–11986, 2022. Link.

[7] M. Tan and Q. Le. **EfficientNet**: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019. Link.