Proceedings of Machine Learning Research – Under Review:1–6, 2024

Multi Instance Learning for Lymphocytosis classification

David Iagaru[*1,2] DAVID.IAGARU@STUDENT-CS.FR

Louise Durand–Janin[†2] LOUISE.DURAND–JANIN@ENS-PARIS-SACLAY.FR
[1] *CentraleSupélec*
[2] *MVA M2*

Editors: Under Review for the Deep Learning for Medical Imaging course

Abstract

Lymphocytosis manifests as an elevation in lymphocyte levels within the bloodstream. Detecting the underlying cause of this condition involves conducting a comprehensive analysis, which includes examining blood smear images and accompanying clinical data like age and lymphocyte concentration. This approach aids in pinpointing the specific disease responsible for the observed lymphocytosis. We propose a method that melds insights from two different data sources - images and clinical attributes - using minimal computational resources for diagnosing lymphocytosis within a multi-instance learning framework. Initially, we focus on extracting image features through self-supervised learning techniques. This enables us to condense each image into a compact, low-dimensional representation.In the next phase, we integrate information derived from these image embeddings alongside clinical attributes to diagnose lymphocytosis effectively.

Keywords: Multi Instance Learning Lymphocytosis, Self-Supervised, Deep Learning

1. Introduction

Lymphocytosis, a common occurrence, can arise either as a reactive response to factors like infection or acute stress. Currently, diagnosing whether it is reactive or tumoral relies on visually inspecting blood cells under a microscope and considering clinical factors such as age and lymphocyte count. This assessment also involves examining the texture and size of lymphocytes in the blood smear to determine the subtype of lymphoid malignancy. While this approach is quick and cost-effective, it lacks consistency. Further clinical tests, particularly flow cytometry, are necessary to definitively confirm lymphocyte malignancy, but these tests are expensive and time-consuming, making them impractical for every patient. Thus, the development of automated and precise methods could assist clinicians in identifying which patients require further analysis with flow cytometry, enhancing diagnostic accuracy and streamlining patient care.

To construct a dataset for this study, blood smear samples and patient characteristics were gathered from 204 individuals at the Lyon Sud University Hospital's routine hematology laboratory. Blood smears were generated automatically using a Sysmex automat tool, while nucleated cells were photographed using a DM-96 device.

[*] Contributed equally
[†] Contributed equally

2. Data analysis

The dataset comprises 163 subjects, including 50 cases of reactive lymphocytosis and 113 cases of malignant lymphocytosis for training, with an additional 42 subjects reserved for testing purposes. The positive samples represent 69% of the patients in the training set which is hence imbalanced. For the model training, we split the initial training set into training and validation subsets while the test set remains unchanged. The validation subset represents 20% of the initial training. To handle class imbalanced, we make sure to have the same proportion of both classes in the sets using stratify argument in the splitting. The clinical annotations have low dimensionality in comparison to the images that are of the size 224×224. They may however contain useful information.

To verify it, we trained an AdaBoost Classifier on the clinical annotations exclusively. It managed to achieve 0.81 ± 0.08 of balanced accuracy on the validation set.

This result shows that the clinical annotations contain rich information, even though it is not sufficient for our objectives. So we better take them in account in the proposed methods.

Each patient indexed by i is associated with $c_i, a_i \in \mathbb{R}, y_i \in \{0, 1\}$ (count of lymphocyte, age, label) and images $(X_i^j)_{1 \leq j \leq N_i} \in (\mathbb{R}^{224 \times 224})^{N_i}$. Note that the number of of images available varies for each patient.

First, since the data is limited in quantity, we augment it performing rotations of $90°, 180°$ and $270°$. It allows to improve the robustness of the models as well. Our resources being limited, the training procedure took too long when we took the raw images. So we decided to downsample them to the dimension of $d_{im}^2 = 112 \times 112$.

3. Methodology

We are facing a problem falling under the category of Multiple Instance Learning (MIL). Specifically, the labels we aim to predict pertain to the entire group or "bag" rather than individual instances. This implies that while we lack knowledge of which specific lymphocytes are malignant, we understand that within a given set of lymphocytes, there exists some malignant ones. Thus, a patient containing at least one malignant lymphocyte is classified as a positive example, while a patient devoid of malignant lymphocytes is classified as a negative example.

One first approach would then consist in training a classifier taking as input an individual image, and deciding a patient is positive if one prediction is positive. The main drawback of this approach is that the error rate of this classifier would be amplified exponentially with respect to the number of images in the patient's bag. So we decided to discard such a simplistic approach.

The method we chose used decomposes in 3 parts : image encoder, encoding aggregation at bag level and a classifier on top of that.

3.1. Encoder

The image encoder is meant to extract the relevant features of an image and map them into a compact representation. For that task we found CNN architecture to be the most adapted. The one we used is represented in 1. The associated decoder has a symmetrical architecture.

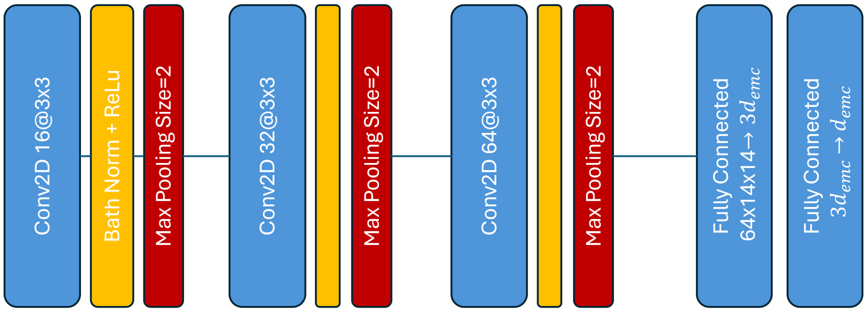

Figure 1: Architecture of the encoder

We denote the feature extraction and the compact representation of an image as $h_i^j := \phi_{CNN}(X_i^j) \in \mathbb{R}$. And the set of patient i's encoding as $E_i := (h_i^j)_{1 \leq j \leq N_i}$

We chose to set the encoding dimension to $d_{enc} = 60$ After having tested values in $\{45, 60, 75, 100, 125\}$ as reported in 1. The performance seems to be the best on the validation set for $d_{enc} \in \{60, 100, 125\}$. Since they are however close to each other, we chose $d_{enc} = 60$ providing hence a lighter model which is still performant. It also gave the best results on the test set. The model has been trained with the Mean squared error loss, using the Adam optimizer with a batch size of 5, a learning rate of 10^{-4} on T4 GPUs provided by Kaggle, taking 1 hour.

3.2. Bag-level attention-based aggregation

After having computed the embedding, of each image of one bag, the next target is to aggregate them into an unique representation which represents the associated patient. For this purpose, we seek symmetry invariant pooling functions. Commonly used examples of such a function are the mean and the max

$$f_{mean}(E_i) = \frac{1}{N_i} \sum_{j=1}^{N_i} \phi_{CNN}(X_i^j)$$

$$f_{max}(E_i) = (max_{1 \leq j \leq N_i}(\phi_{CNN}(X_i^j)_k))_{1 \leq k \leq d_{enc}}$$

However, we wanted to capture more precisely the relative importance of the encodings. We therefore chose the attention mechanism to aggregate the representations of a patient. By capturing fine-grained relationships within the input sequence, attention-based pooling can lead to improved performance in tasks requiring understanding of long-range dependencies or subtle nuances in the data (Er et al., 2016; Bhattacharjee et al., 2021) . Patient

i's image information is then summarized in D_i as proposed by (Ilse et al., 2018)

$$D_i = \sum_{j=1}^{N_i} \frac{exp(w^T \tanh(V(h_i^j)^T))}{\sum_{l=1}^{N_i} exp(w^T \tanh(V(h_i^l)^T))} h_i^j$$

We tried both the attention-based and basic mean pooling function in the architecture. The performances were higher using the first possibility, so we chose to keep the attention based pooling function.

3.3. Final Classifier

Now that the each patient has a compact representation for its associated images, it remains to associate them with the clinical data as an input of a classifier that would output a number in $[0, 1]$ as a prediction. For that, we used two fully connected layers (of dimensions 256 and 128). Formally,

$$y_{pred} = \sigma(\psi_{fc}(D_i, a_i, c_i))$$

Where $\psi_{fc} : \mathbb{R}^{d_{enc}+2} \rightarrow \mathbb{R}$ is the fully connected network and σ is the sigmoid function.

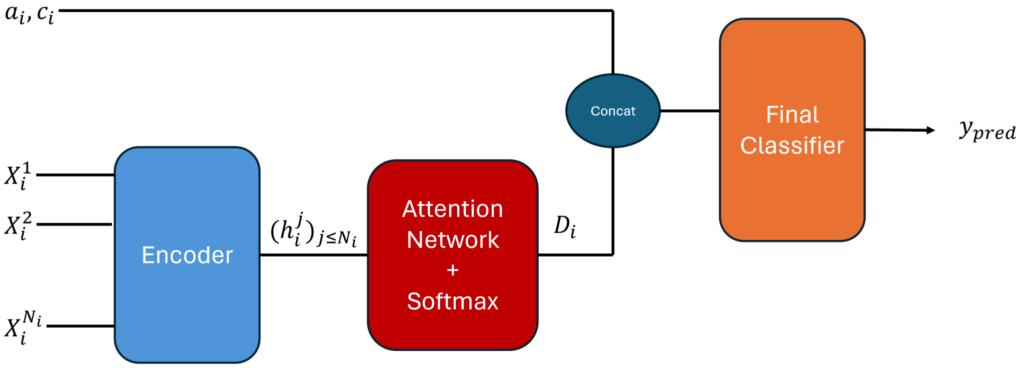

Figure 2: Global structure of our method

3.4. Other experienced methods

Another option that we tried was motivated by the fact that either images or attributes can have a different contribution to the prediction among the patient. The model we tried was inspired by the Mixture of Experts Model from (Sahasrabudhe, 2021). It uses a linear layer to calculate the contribution π_{CNN} of the images and then defined the contribution of the attributes as $\pi_{MLP} = 1 - \pi_{CNN}$. The final prediction is then the weighted sum of both outputs. It performed a balanced accuracy of 0.61 on the test which was not really satisfying.

4. Results and validation

We evaluate our methods on the validation set. The model's performances on the Validation set depends of the encoding dimension as reported in 1. The performance on the testset was the best setting $d_{enc} = 60$ and it was reported to be 0.82, which means we slightly over-fit on the train and validation set.

We visualize the Precision-Recall curve and compute the associated AUPRC a well

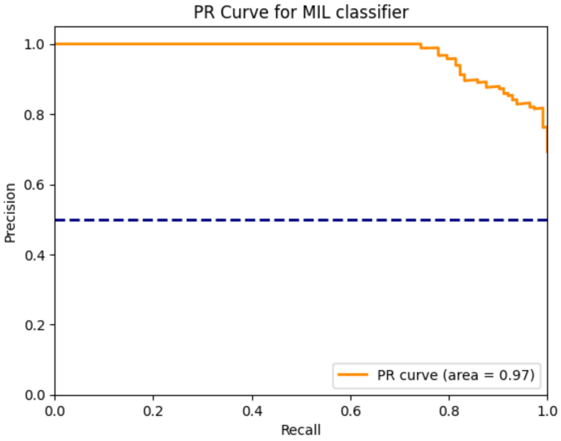

Figure 3: Precision-Recall Curve, AURPC = 0.93, $d_{enc} = 60$

Table 1: Results on the validation set for different values of d_{enc}

d_{enc}	**Balanced Accuracy**	**F1 Score**
45	0.86	0.87
60	0.90	0.889
75	0.84	0.81
100	0.90	0.889
125	0.92	0.91

The precision-recall curve is close to ideal with an area of 0.93. A resulting threshold for deciding whether a patient is considered as positive would be corresponding to a recall of 0.7, right before the precision starts decaying.

References

Kamanasish Bhattacharjee, Arti Tiwari, Millie Pant, Chang Wook Ahn, and Sanghoun Oh. Multiple instance learning with differential evolutionary pooling. *Electronics*, 10:1403, 06 2021. doi: 10.3390/electronics10121403.

Meng Joo Er, Yong Zhang, Ning Wang, and Mahardhika Pratama. Attention pooling-based convolutional neural network for sentence modelling. *Information Sciences*, 373:

388–403, 2016. ISSN 0020-0255. doi: https://doi.org/10.1016/j.ins.2016.08.084. URL https://www.sciencedirect.com/science/article/pii/S0020025516306673.

Maximilian Ilse, Jakub M. Tomczak, and Max Welling. Attention-based deep multiple instance learning. 2018.

Sujobert P. Zacharaki E. I. Maurin E. Grange B. Jallades L. Paragios N. Vakalopoulou M. Sahasrabudhe, M. Deep multi-instance learning using multi-modal data for diagnosis of lymphocytosis. *IEEE journal of biomedical and health informatics*, 25:2125–2136, 2021. doi: 10.1109/JBHI.2020.3038889. URL https://doi.org/10.1109/JBHI.2020.3038889.