# Expectation-Maximization vs Deep Sets for Histopathology Data

By

Anastasiia Livochka, Maryam Yalsavar, Cassandra Wong, and David Evans

STAT 841 Final Project
University of Waterloo
Fall 2021

**Abstract**

Histopathological classification of lung cancer is one of the routine pathological diagnosis tasks for pathologists. Using image classification techniques would be beneficial in easing the workloads on pathologists, especially in regions that have shortages in access to pathological diagnosis services. In this project we investigate and compare the performance of two previously proposed approaches named Expectation-Maximization inspired algorithm [10] and Deep Sets [22] in recognizing three types of lung cancer on small and heavily unbalanced histopathology dataset. While training a neural network on this data is a very challenging task we show that it is possible to obtain a reasonable accuracy with both methods: 60% with EM-based approach and $70\%$ with Deep Sets one.

# 1 Introduction

Despite the fact that the convolutional neural network (CNN) technique is well-known for its success in image classification, it is computationally impossible to use them for cancer classification due to the sheer size of a high-resolution whole-slide image (WSI) [11]. One of the main challenges in computational pathology is the aggregation of patch-level classification results. When applying a deep learning classifier, the WSI is divided into several thousand patches, with the classifier then applied independently on each patch. The result of each individual patch is to be aggregated to obtain the final classification output. Another main challenge is the existence of non-discriminative patches, referring to the patches in WSI's that do not contain the biomarker identified in the classification process [11]. In this project, we are given biopsy histopathology whole-slide images (WSIs) of various bronchus and lung samples. We compared the performance of two set classification models, Expectation Maximization, and Deep Sets. The training datasets in this project was obtained from Kaggle, which consists of 72 lung adenocarcinoma (LUAD) WSI's, 156 lung squamous cell carcinoma (LUSC) WSI's, and 9 mesothelioma (MESO) WSI's. Each WSI contains a folder with its respective patched images. Note that the dataset is strongly unbalanced.

# 2 Related Works

The formulation of the Expectation Maximization (EM) Algorithm using a patch-level CNN and training a decision fusion model as a two-level model is detailed by the following paper [10]. Below are notable results/works and breakthroughs that made this design apparent: 1) Majority of Whole Slide Tissue Images classification methods fixate on classifying or obtaining features on patches [1, 9, 20, 21, 8, 3, 16, 4]. 2) Multiple Instance Learning (MIL) based classification utilizes unlabeled patches to predict labels [6, 19, 18]. 3) MIL-based CNNs have been applied to object recognition [13] and semantic segmentation [5]. 4) Finally, max-pooling and voting (average pooling) were applied [12, 17, 1]. The Deep Sets architecture was implemented to maintain the permutation invariance property when performing set classification. Several recent works that contributed to the Deep Sets architecture had studied the equivariance and invariance in deep networks [7, 2, 15].

Reviewing related works such as KimiaNet expanded the knowledge on patch-level CNNs built on top of DenseNet. KimiaNet "employs the topology of DenseNet with four dense blocks, fine-tuned and trained with histopathology images in different configurations. KimiaNet boasts superior results compared to the original DenseNet and smaller CBR networks using a feature extractor to represent histopathology images" [14].

# 3 Model Architecture

## 3.1 Set Classification using Expectation Maximization (EM) Algorithm

We denote $X = \{X_1, \ldots, X_N\}$ to be a dataset that consists of $N$ images with corresponding labels $Y = \{y_1, \ldots, y_N\}$. Subsequently, every i-th image is a set of corresponding $N^i$ patches $X_i = \{X_{i,1}, \ldots X_{i,N^i}\}$. We denote $H_{i,j}$ to be a binary label of patch $X_{i,j}$ with $H_{i,j} = 1 \iff X_{i,j}$ has the same label as $X_i$, meaning patch $X_{i,j}$ is discriminative for the image $X_i$, $X_{i,j}$ contains a cancerous area. By the nature of the challenge patch binary labels $H_{i,j}$ are unknown.

Assuming that sets of patches (images) are independent and identically distributed we have:

$$P(X,H) = \prod_{i=1}^{N} \prod_{j=1}^{N^i} P(X_{i,j}|H_{i,j})P(H_{i,j})$$

We want to maximize the data likelihood $P(X)$ and labels $H$ are hidden, suggesting the EM algorithm's use. We use M- and E- steps as proposed in [10]. We assume a statistical model that generates images and patches $X$ and hidden labels $H$. The model is parameterized by $\theta$. The maximum likelihood estimate (MLE) of the unknown parameters is determined by maximizing the marginal likelihood of the observed data. $L(\theta;X) = P(X,H|\theta)$. The EM algorithm seeks to find the MLE by iteratively applying the following. **M-step:** we update the parameters $\theta$ to maximize the data likelihood. $D$ denotes the set of discriminative patches. Assuming a uniform generative model for all non-discriminative instances we allow the simplification:

$$\theta \leftarrow argmax_\theta \prod_{x_{i,j} \in D} P(x_i, y_i|\theta) \tag{1}$$

Please note that all the discriminative patches have the same label as their image, which allowed for simplification in eq. (1). **E-step:** re-estimate all hidden labels $H$ by model with updated parameters.
Fig.1. depicts the general steps of employing the EM algorithm for set classification of histopathology data. In the beginning, the Resnet18 model pre-trained on ImageNet fine-tunes on the whole train set with weighted cross-entropy loss and Adam optimizer. Then we have a prediction on the whole train set that classifies patches into two groups named discriminative and non-discriminative. The thresholds for classifying patches into two groups is generated in each epoch by using the procedure that is mentioned in [10]. After this classification, just the discriminative train data will be used in fine-tuning the model, and this procedure continues until an acceptable train, and validation accuracy is observed.

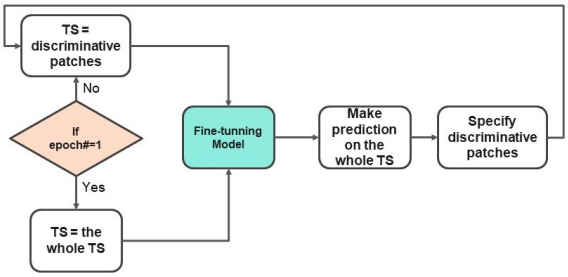



Figure 1: Expectation Maximization (EM) Steps.

After fine-tuning the Resnet18 model, second-level models, logistic regression (LR), or Support Vector Machines (SVM) has been used for classifying images into three classes. For training a second-level model, we create a train set in two ways: 1) after passing each patch through the Resnet18 model and assigning a label to each patch, we count the number of patches that correspond to each class of cancer in each image, so for each image, we will have a 1*3 vector that shows [#patches in class1,#patches in class2,#patches in class3]. 2) after passing the patches through the model, we take the average over the probabilities of all patches in each dimension for each image. As the output size of Resnet18 is 3, for each image, there will be a 1*3 vector that includes the average of probabilities over the whole image's patches in each dimension. We also used max-pooling on the output of the Resnet18 model after fine-tuning for assigning a label to an image. For the histopathology data, both the second-level model and max-pooling show the same performance, while the superiority of second-level models is mentioned in the literature.

### 3.2 Set Classification using Deep Sets

Because patches themselves do not have a known label, it is natural to approach the challenge from the set classification perspective. It is essential for the models that perform set the classification to have the permutation invariance property: the target value for a given set should be the same regardless of the order of objects in the

set. The main inspiration for this project came from [22] later referred to as Deep Sets. A simple example of a permutation invariant model is:

$$model(\{X_{i,1}, \ldots, X_{i,N^i}\}) = \rho(pool(\{\phi(X_{i,1}), \ldots, \phi(X_{i,N^i})\}))$$

So in Fig. 2: (1) each patch $X_{i,j}$ is transformed (encoded) into some representation $\phi(X_{i,j})$; (2) the representations in the set are aggregated by sum/max/avg operation; (3) this aggregated representation of the set $X_i$ is used by another model $\rho$ to produce image-level prediction $\hat{y}_i$.

For this project, we used Resnet-34 without the classifier layer pre-trained on ImageNet as the encoder network

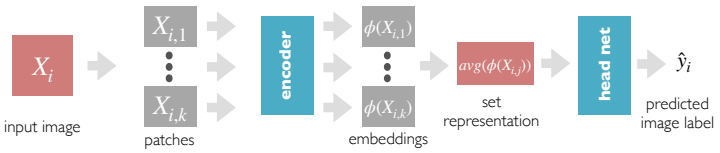



Figure 2: The approach inspired by Deep Sets: in each iteration, we sample random $k$ patches from randomly selected images, feed the patches to encoder model to obtain patches representations $\phi(X_{i,j})$ which are later aggregated (e.g. averaged over dimensions) to form a set representation on which the main set classifier is trained.

and a neural network with four fully-connected layers (each is followed by BatchNorm and ReLU non-linearities) as set classifier - head model. Both networks were trained simultaneously with weighted cross-entropy loss and Adam optimizer. In each epoch, we randomly select 7 patches for every image, obtain patches embeddings from the encoder, and take maximum over each embedding dimension to obtain image representation. Image representation is used as an input to the head model, which was trained to predict image class. Experiments showed that color and affine image augmentation combined with weighted loss make training very unstable and terminate with lower accuracy, so image augmentation was not used for training.

## 4 Experiments

We performed an additional random split of provided train data into the train ($80\%$) and test($20\%$). In both approaches, models were trained only on train parts, and we performed the hyper-parameters selection (e.g. learning rate, number of epochs) based on test performance. No models were ever trained, and no parameters were selected based on dev performance to make it as close to the real-world scenario as possible. For the EM algorithm, we applied random color and affine augmentation Fig. 3 for every batch of patches in training mode, which helped to prevent overfitting. The best classification accuracy we were able to reach with EM-based approach on dev set is

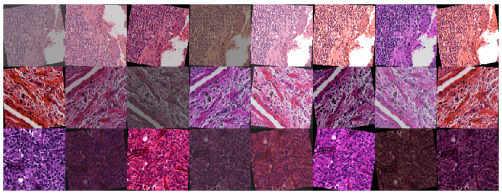
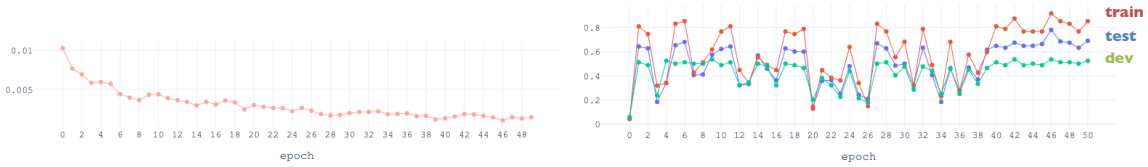


Figure 3: Example of random augmentation applied to three patches. The augmentation includes random vertical flip, random rotation in the range $(-10°, 10°)$, color jitters (affecting brightness, contract and hue)

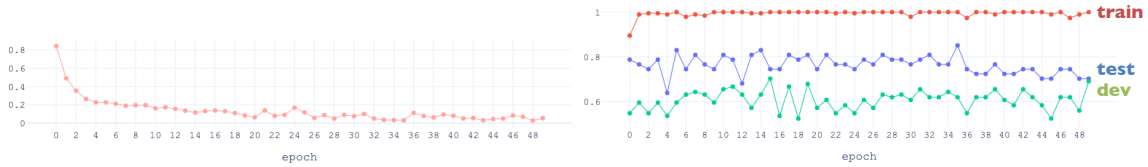
$60\%$ during experiment depicted at Fig. 4. Please note that both graphs visualize training of the first-level model, which output on test set was later used as input to train a second-level model which performs decision fusion. Fig. 6 (a) shows the confusion matrix for predictions of second-level on dev set.

In the deep sets approach we trained both encoder and head model simultaneously. No augmentation was applied to patches in the final version of the algorithm because it allowed us to arrive at greater accuracy on both train and test. As encoder was chosen to be a pre-trained model we needed to apply data normalization. Because of no augmentation we see visual signs of overfitting in Fig. 5. Nevertheless, at the end of this experiment we arrived at $70\%$ classification accuracy on dev set.




(a) (b)

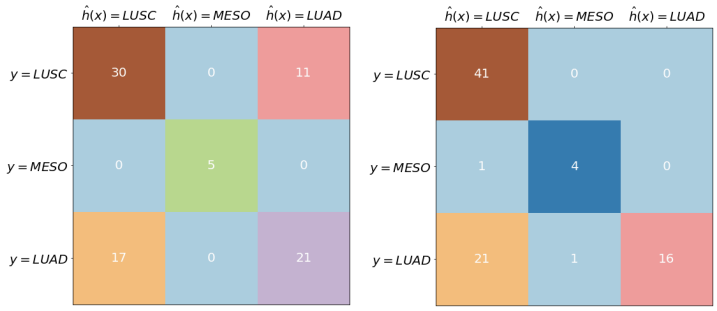
Figure 4: The EM experiment that achieved 60% accuracy (a) training loss, weighted cross-entropy; (b) train/test/dev image classification accuracy calculated without the use of ML second-level model, the final class for image is decided as argmax of predictions for image patches




(a) (b)

Figure 5: The Deep Sets experiment that achieved 70% accuracy (a) training loss, weighted cross-entropy; (b) train/test/dev image classification accuracy calculated on the set representation from all the patches

## 5 Conclusion




(a) (b)

Figure 6: Confusion matrices from experiments that resulted in best-attained accuracy after 50 epochs (a) training with EM approach that resulted in 60% accuracy; (b) training with Deep Sets approach that resulted in 70 % accuracy.

The small number of data samples and the differences between the distribution of dev set and train set (data has been sampled from different patients by different devices) makes applying set classification on histopathology data challenging. As the experiments show, both EM algorithm and deep sets were unsuccessful at arriving at high accuracy, although we performed many experiments to arrive at reasonable results. Both approaches turned to be very unstable during training. Gradient optimization was frequently stacked in local minima (giving the same labels to all instances) or jumping from one local minima to another. Although on average, it was easier to get better accuracy with Deep Sets. We think that using models that are pre-trained on similar data sets to histopathology data and employing some methods that can alleviate the distribution shifts due to the change of device or patient might help improve the accuracy and tackle the problem in this case. Due to the lack of time, we have many ideas that are left for future work. 1. EM algorithm: experiments with percentiles for image- and class-level thresholds and with network architecture and image augmentations. 2. Deep Sets: experiments with both encoder and head model architecture, training regime (learning rate, freezing layers), experiments with loss function and class weights.

## References


[1] A. Cruz-Roa, A. Basavanhally, F. Gonzalez H. Gilmore M. Feldman S. Ganesan N. Shih J. Tomaszewski and A. Madabhushi. 2014. "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks." *Medical Imaging* .

[2] Cohen, Taco S and Max Welling. 2016. "Group equivariant convolutional networks." *arXiv preprint arXiv:1602.07576* .

[3] D. Altunbay, C. Cigir, C. Sokmensuer and C. Gunduz-Demir. 2010. "Color graphs for automated cancer diagnosis and grading." *J Biomed Eng.* .

[4] D. C. Ciresan, A. Giusti, L. M. Gambardella and J. Schmidhuber. 2013. "Mitosis detection in breast cancer histology images with deep neural networks." *MICCAI* .

[5] D. Pathak, E. Shelhamer, J. Long and T. Darrell. 2014. "Fully convolutional multi-class multiple instance learning." *arXiv* .

[6] E. Cosatto, P.-F. Laquerre, C. Malon H.-P. Graf A. Saito T. Kiyuna A. Marugame and K. Kamijo. 2013. "Automated gastric cancer diagnosis on he-stained sections; Training a classifier on a large scale with multiple instance machine learning." *Medical Imaging* .

[7] Gens, Robert and Pedro M Domingos. 2014. "Advances in neural information processing systems." pp. 2537–2545.

[8] H. Chang, Y. Zhou, A. Borrowsky K. Barner-P. Spellman and B. Parvin. 2014. "Stacked predictive sparse decomposition for classification of histology sections." *IJCV* .

[9] H. S. Mousavi, V. Monga, G. Rao and A. U. Rao. 2015. "Automated discrimination of lower and higher grade gliomas based on histopathology image analysis." *JPI* .

[10] Hou L, Samaras D, Kurc TM Gao Y-Davis JE and Saltz JH. 2016. "Patch-based Convolutional Neural Network for Whole Slide Tissue Image Classification." *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit* .

[11] Litjens, Geert, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken and Clara I Sánchez. 2017. "A survey on deep learning in medical image analysis." *Medical image analysis* 42:60–88.

[12] M. H. Nguyen, L. Torresani, F. De La Torre and C. Rother. 2009. "Weakly supervised discriminative localization and classification: a joint learning process." *ICCV* .

[13] M. Oquab, L. Bottou, I. Laptev and J. Sivic. N.d. "Weakly supervised object recognition with convolutional neural networks." *NIPS*. Forthcoming.

[14] Riasatian, Abtin, Morteza Babaie, Danial Maleki, Shivam Kalra, Mojtaba Valipour, Sobhan Hemati, Manit Zaveri, Amir Safarpoor, Sobhan Shafiei, Mehdi Afshari, Maral Rasoolijaberi, Milad Sikaroudi, Mohd Adnan, Sultaan Shah, Charles Choi, Savvas Damaskinos, Clinton JV Campbell, Phedias Diamandis, Liron Pantanowitz, Hany Kashani, Ali Ghodsi and H. R. Tizhoosh. 2021. "Fine-Tuning and Training of DenseNet for Histopathology Image Representation Using TCGA Diagnostic Slides.".

[15] Siamak Ravanbakhsh, Jeff Schneider and Barnabas Poczos. 2017. "Equivariance through parameter-sharing." *arXiv preprint arXiv:1702:08389* .

[16] T. H. Vu, H. S. Mousavi, V. Monga U. Rao and G. Rao. 2015. "Discriminative feature-oriented dictionary learning for histopathological image classification." *arXiv* .

[17] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner. 1998. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* .

[18] Y. Xu, J.-Y. Zhu, I. Eric C. Chang M. Lai and Z. Tu. 2014. "Weakly supervised histopathology cancer image segmentation and classification." *Medical Image Analysis* .

[19] Y. Xu, T. Mo, Q. Feng P. Zhong M. Lai E. I. Chang et al. 2014. "Deep learning of feature representation with multiple instance learning for medical image analysis." *ICASSP* .

[20] Y. Xu, Z. Jia, Y. Ai F. Zhang M. Lai E. I. Chang et al. 2015. "Deep convolutional activation features for large scale brain tumor histopathological image classification and segmentation." *ICASSP* .

[21] Y. Zhou, H. Chang, K. Barner P. Spellman and B. Parvin. 2014. "Classification of histology sections via multispectral convolutional sparse coding." *CVPR* .

[22] Zaheer, Manzil, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov and Alexander J Smola. 2017. Deep Sets. In *Advances in Neural Information Processing Systems*, ed. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett. Vol. 30 Curran Associates, Inc.
**URL:** *https://proceedings.neurips.cc/paper/2017/file/f22e4747da1aa27e363d86d40ff442fe-Paper.pdf*