MyriadAL: Active Few Shot Learning for Histopathology

Nico Schiavone*†   Jingyi Wang*   Shuangzhi Li   Roger Zemp   Xingyu Li

Electrical and Computer Engineering, University of Alberta, Edmonton AB, Canada

{nschiavo, jingyi28, shuangzh, rzemp, xingyu}@ualberta.ca

Abstract—**Active Learning (AL) and Few Shot Learning (FSL) are two label-efficient methods which have achieved excellent results recently. However, most prior arts in both learning paradigms fail to explore the wealth of the vast unlabelled data. In this study, we address this issue in the scenario where the annotation budget is very limited, yet a large amount of unlabelled data for the target task is available. We frame this work in the context of histopathology image classification, where labelling is prohibitively expensive. To this end, we introduce an active few shot learning framework, Myriad Active Learning (MAL), including a contrastive-learning encoder, pseudo-label generation, and novel query sample selection in the loop. Specifically, we propose to massage unlabelled data in a self-supervised manner, where the obtained data representations and clustering knowledge form the basis to activate the AL loop. With feedback from the oracle in each AL cycle, the pseudo-labels of the unlabelled data are refined by optimizing a shallow task-specific net on top of the encoder. These updated pseudo-labels serve to inform and improve the active learning query selection process. Furthermore, we introduce a novel recipe to combine existing uncertainty measures and utilize the entire uncertainty list to reduce sample redundancy in AL. Extensive experiments on two public histopathology datasets show that MAL has superior test accuracy, macro F1-score, and label efficiency compared to prior works, and can achieve a comparable test accuracy to a fully supervised algorithm while labelling only 5% of the dataset. Code is available at https://github.com/mesophil/MyriadAL**

Index Terms—**artificial intelligence, deep learning, computer vision, data efficient learning, active learning, histopathology**

I. INTRODUCTION

Deep learning [20] has achieved numerous successes in supervised settings, producing state of the art accuracy and generalization [22], [26], [34]. However, for tasks with a scarcity of labelled data, deep learning is not nearly as effective [14], [37]. With the ever increasing amount of unlabelled data, and the growing annotation cost, innovation has shifted towards more label efficient strategies [19], [35]. In recent years, two label efficient learning paradigms have emerged: Active Learning (AL) and Few Shot Learning (FSL). AL tackles the data scarcity problem by selecting only the most informative data for labelling [27]. However, conventional AL models require moderate annotation budgets and often underperform otherwise [23]. On the other hand, FSL takes

a different approach, utilizing transfer learning or model adaptation/generalization techniques and only a handful of labelled samples from the target dataset [28], [31].

Despite the promising performance, it should be noted that both AL and FSL utilize only a few pieces of annotated data in model training, leaving the unlabelled samples as untapped potential. We argue that the effective use of the unlabelled data would further improve the performance, especially under the scenario where only a small annotation budget is available. To tackle this challenge, we propose a novel framework for Active Few Shot Learning: Myriad Active Learning (MAL). Our first contribution stemming from MAL is a framework that incorporates self-supervised learning, pseudo-labels, and active learning in a positive feedback loop. The pseudo-labels of the unlabelled data are updated every AL cycle and supplement the uncertainty measurement for a more precise and diverse query selection. Our second contribution is designing an algorithm to make the most efficient use of the pseudo-labels in the query selection. The new recipe combines classic uncertainty measures to precisely define sample types based on their comparative uncertainty. We then sample evenly from these types using a self-regulating algorithm, facilitating pseudo-label updates in the next cycle.

We frame the target problem in the context of digital histopathology, where expert pathologists are required to annotate samples in a prohibitively expensive and time-consuming process [2]. In contrast, the number of unlabelled histopathology images is extremely high, as a single scan can produce hundreds of unique images due to the underlying tissue structures. Under the setting of a very limited annotation budget, we evaluate MAL on two histopathology image sets, and show its superiority to classical active learning techniques via comparison and ablation. Notably, MAL can achieve a comparable test accuracy to a fully supervised learning algorithm while labelling only 5% of a target dataset, demonstrating its potential for effective label efficient learning.

To the best of our knowledge, no benchmark currently exists for histopathology in the limited budget active learning setting; therefore, MAL also functions as a new benchmark for future works to be compared against. For the purpose of comparing our method against the best available alternatives, we have reworked several well known active learning and few shot learning methods to fit the limited budget setting.

Our contributions are summarized as follows:

1) We formulate a new problem: active few-shot learning to

address high annotate cost in digital histopathology. The proposed framework utilizes the abundant unlabelled data for more label-efficient model learning.

2) We develop a novel uncertainty-based active learning algorithm, Myriad Active Learning (MAL). MAL defines a new type of sample diversity, effectively supplementing the existing annotated data for rapid classification accuracy increases.

3) We show that the proposed method achieves state-of-the-art performance on histopathology datasets in the target setting. As well, with a higher annotation budget, MAL can quickly obtain test accuracies comparable to that of a fully-supervised model.

II. RELATED WORK

Self-Supervised Learning (SSL) aims to learn generic representations from unlabelled data as a preliminary step to train encoders to efficiently solve subsequent traditional supervised learning tasks [3]. SSL explores data relations by constructing self-supervised pretext tasks based on the unlabeled inputs to produce these representations [1]. A prominent type of SSL is contrastive learning [13], including the recent variant Momentum Contrastive Learning (MoCo) By He et al. [12], which has produced state of the art results on many image-based datasets. SSL algorithms have also produced results competitive with supervised learning on histopathology datasets [5], [16] when the datasets are large. However, most recent efforts to incorporate a limited annotation budget in computational histopathology have been unsuccessful in terms of label efficiency [5], [21].

Deep Active Learning (DAL) is another option to tackle the data hungry nature of Deep Learning (DL). The key proposition of DAL is that the majority of the benefits can be obtained from a minority of the samples. In DAL, the training set is unlabelled, optionally with a small number of labelled samples, called the seed. A portion of the unlabelled samples is selected for labelling every cycle, and then included in the training set.

Sample selection strategies [25], [37] are the subject of much innovation, and can be broadly classified into two categories: uncertainty based, and diversity based. Uncertainty based algorithms evaluate samples based on predefined criteria, such as Entropy [29]. CEAL [35], a recent sampling method, labels the least uncertain samples with their predicted labels, possibly greatly increasing the labelled set size for no additional labelling cost. However, all of these classical tactics underperform expectations on histopathology datasets [11], [37]. Diversity sampling methods, such as k-means clustering, or VAAL [30], posit that the model cannot predict uncertainties accurately enough to be useful. Instead, these methods gather as many types of samples as possible, in order to label, and expose the model to, as many classes as possible. These methods are broadly more effective on histopathology datasets, but are more inconsistent due to their random selection nature within clusters or buckets.

Few Shot Learning (FSL) attempts to tackle the data scarcity problem by learning through only a handful of labelled samples (often <1% of the dataset) in the task domain. This is often accompanied by a well pre-trained model on a large source dataset. FSL has seen great success in histopathology [28], but is still in a primitive stage with regards to label efficiency research. Active Few Shot Learning provides an potential avenue to solve this problem [23], [36], but the results in many applications, especially histopathology, have not met expectations. This may be a result of the underperformance of active learning on low annotation budgets in general.

Semi-Supervised Learning (SemiSL) is another method dedicated to maximizing performance on a limited number of labels. Halfway between supervised learning and unsupervised learning, the label information given is often in the form of pre-labelled data, or a randomly selected labelled subset. Active SemiSL has seen success in natural image classification, such as the work by Gao et al. [10], and in label effective methods utilizing GANs, such as the VAAL method in active learning [30].

III. METHODS

A. Motivation

In this work, we aim to investigate active few-shot learning solutions under very low annotation budgets. An example of such a scenario is computational histopathology, where annotation budgets are small, and labelling is prohibitively expensive, but unlabelled data is relatively abundant.

When constructing an effective and efficient solution for this setting, we considered many paradigms and components. We could not utilize many conventional methods, such as knowledge distillation [8], due to the lack of a well-labelled source dataset. Although it is still technically possible to use these methods, they would have to be trained on a more general image dataset, such as ImageNet [6]. Such a dataset would contain primarily natural images, and this would introduce a large domain gap between the source and target datasets, biasing the model for the downstream tasks.

Therefore, we decided to pursue active-few shot learning, a paradigm which has not yet been explored in the context of histopathology, but has the potential to operate effectively without nearly as many resources.

B. Problem Formulation

Mathematically, the problem can be formulated as follows: we are given a large data set $\mathcal{D} = \{(\mathbf{x}_j)\}_{j=1}^{N_{ul}}$ consisting of N_{ul} unlabelled samples $\mathbf{x}$ from K categories, and a very small annotation budget. In this case, data sample $\mathbf{x} \in R^{H \times W \times C}$ is an image, where W, H are the image width and height, and $C = 3$ represents the colour channels. Under the annotation budget, up to K samples can be selected from D for annotation in each active learning cycle. The target is to learn a model $\mathcal{M}$ to predict the categories of queries, i.e., image classification.

Fig. 1 provides a visualization of MAL. Initially, two sets are constructed: one unlabelled set $U_0 = D$, and one empty set $\mathcal{T}_0 = \{\emptyset\}$ to record the labelled samples from the downstream
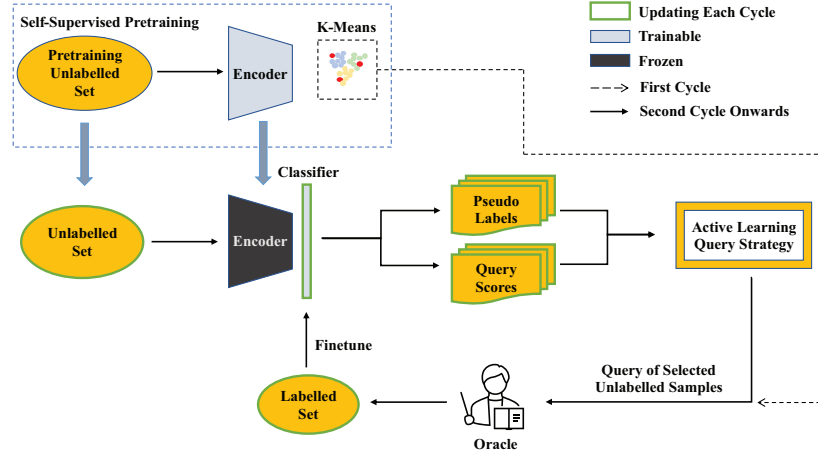
Fig. 1. Diagram of the proposed framework Myriad Active Learning. Pseudo-labels of the unlabelled set are updated and explored for query sample selection.

AL cycles. To explore the rich information in the unlabelled set D, we utilize a constrastive learning algorithm to train an encoder, mapping raw histopathology images into numerical features, $f_x = E(x)$. Then, the encoder is frozen and a one-layer classifier is added on top. In the tth active learning cycle, a batch of unlabelled samples $\mathcal{Q}$ are removed from $\mathcal{U}_{t-1}$ for annotation, and the newly labelled samples, $\{(\mathbf{x}_i, y_i)\}_{i=1}^K$, are added into the labelled set, i.e. $\mathcal{U}_t = \mathcal{U}_{t-1} - \{(\mathbf{x}_i)\}_{i=1}^K$ and $\mathcal{T}_t = \mathcal{T}_{t-1} \cup \{(\mathbf{x}_i, y_i)\}_{i=1}^K$, where $y_i \in (0,1)^K$ is the one-hot label vector for $\mathbf{x}_i$ over the K classes. The classifier is then optimised on $\mathcal{T}_t$ and generates pseudo-labels P_t for all samples in $\mathcal{U}_t$, preparing information for the $t+1^{th}$ cycle. The active learning cycles continue until the budget is exhausted.

C. Self Supervised Learning

In the Myriad Active learning framework, we first train an encoder with self-supervised learning on unlabelled samples in $\mathcal{U}_0$. The encoder is then frozen and the extracted numerical features from the unlabelled samples form the basis of the down-stream active learning classification problem. Particularly, the initial clustering effect of SSL encoder helps generate the initial pseudo-labels, which remedies the cold start problem in subsequent active learning. This self-supervision stage is visualized in the top portion of Fig. 1.

For our framework, we utilize the SSL algorithm known as Momentum Contrastive Learning (MoCoV2), by Chen et al. [4], which has shown state-of-the-art results on many image-based datasets, including in histopathology [5]. Mo-CoV2 learns positive/negative (similar/dissimilar) representations from the data, which becomes a list of positive/negative pairs. MoCoV2 formulates this as a dictionary lookup problem, with keys for the representations. Given an unlabelled sample $x \in \mathcal{U}_0$, we perform two different data augmentations on x and denote the data augmented versions as x' and x'' respectively. Then a query representation $q = f_{x'} = E(x')$ and corresponding key representation $k^+ = f_{x''} = E(x'')$

form the positive pair in MoCoV training. Representations from other images constitute a set of negative samples k^-. The loss function for encoder optimization is formulated in Eq. (1).

$$\mathcal{L}_{q,k^+,\{k^-\}} = \\ -\log \frac{\exp(q \cdot k^+/\tau)}{\exp(q \cdot k^+/\tau) + \Sigma_{k^-} \exp(q \cdot k^-/\tau)} \quad (1)$$

where τ is the temperature hyper-parameter. The large, dynamic dictionary utilized by MoCoV2 is also more efficient than many SSL algorithms [4], increasing its usability in real-world settings.

Note that in this study, we chose not to use more specialized SSL algorithms, such as HistoSSL by Jin et al. [16], to show that it is not necessary to use a histopathology specific pretraining algorithm for our solution to be effective.

D. Pseudo-Label Generation

Once the SSL model is pretrained on the unlabelled dataset, the learned features are used as an input to a shallow network - a one layer classifier. It is important to use a shallow network in this case to reduce the chance of overfitting, as we will only have a few labelled samples per class in the low budget setting.

Initially, there is no training data for the target task, as we do not use an initial seed; therefore, the shallow network cannot provide any meaningful information to the active learning algorithm. Instead, we use K-Means clustering on the numerical features from the frozen encoder to form K clusters in the first cycle. The first query is composed of one sample selected from each of the clusters. From the second cycle onwards, the framework proceeds in a closed-loop fashion, as depicted in Fig. 1. The classifier is updated using the labelled samples in $\mathcal{T}_t$, and generates the set of pseudo-labels P_t to use in the next cycle. The pseudo-labels $\hat{y}$ are generated based on the

predicted class probabilities for each sample $\mathbf{x}_j \in \mathcal{U}_t$; defining $\hat{y}_j$ as the most likely class label of $\mathbf{x}_j$, the set of pseudo-labels $P_t = \{\hat{y}_j\}_{j=1}^{N_{ul}}$ is formed. Excellent pseudo-labels are crucial to the function of our framework, as the pseudo-labels will inform the diversity of the active learning query and be the best mode of prevention against redundant samples.

E. Myriad Active Learning

Myriad Active Learning (MAL) provides a solution to the goal of collecting a diverse query, minimizing redundancy, and maximizing the accuracy in each cycle. To accomplish this, we utilize conventional active learning techniques in a novel way, allowing for more precise and deliberate sample selection. MAL uses a modified version of conventional uncertainty-based strategies, the pseudo-labels obtained using the methods in Sec. III-D, and a novel sample selection strategy. First, we combine margin sampling and entropy sampling into Margin-Entropy (M-E) Sampling to allow for a more precise mapping of the uncertainties (using $\mathcal{M}$ to denote the model):

$$\sigma_{\mathbf{m\text{-}e}}(\mathbf{x}, \mathcal{M}) = \frac{1}{\sigma_{\mathbf{margin}}(\mathbf{x}, \mathcal{M})} + \sigma_{\mathbf{entropy}}(\mathbf{x}, \mathcal{M}), \quad (2)$$

$$\sigma_{\mathbf{margin}}(\mathbf{x}, \mathcal{M}) = p_{\mathcal{M}}(\hat{y}_1|\mathbf{x}) - p_{\mathcal{M}}(\hat{y}_2|\mathbf{x}), \quad (3)$$

$$\sigma_{\mathbf{entropy}}(\mathbf{x}, \mathcal{M}) = -\Sigma_{i=1}^{k} p_{\mathcal{M}}(y_i|\mathbf{x}) \log p_{\mathcal{M}}(y_i|\mathbf{x}), \quad (4)$$

where $p_{\mathcal{M}}(\hat{y}|\mathbf{x})$ represents the class probability of sample $\mathbf{x}$ under model $\mathcal{M}$. The aforementioned classical uncertainty sampling strategies are defined as follows: **Margin Sampling** [24] gives the uncertainty based on the difference between the probabilities of a sample's most and second most likely labels ($\hat{y}_1$ and $\hat{y}_2$, respectively). In this case, a lower value means a larger uncertainty. **Entropy Sampling** [29] selects data based on the maximal entropy. When combined, the $1/\sigma_{\mathbf{margin}}$ is used rather than $\sigma_{\mathbf{margin}}$ to ensure larger values map to larger uncertainties.

Margin sampling favours samples closer to the decision boundaries [24], regardless of how many classes intersect at those boundaries. Entropy sampling favours the most uncertain samples close to the decision boundary of *as many* classes as possible, and therefore often picks out noisy or difficult samples. Classically, the samples will be selected from the most uncertain to the last uncertain until the per cycle quota K is filled.

The reason for this is that these sampling methods only provide concrete information on where the *highest* uncertainty samples will be found. As a result, similar samples are often labelled, potentially wasting label information. A visualization of these conventional methods can be found on the left side of Fig. 2.

Comparatively, M-E has a much more predictable structure: a low M-E uncertainty sample is likely to be found near the centre of a cluster, due to a low margin score *and* a low entropy. Conversely, a high M-E uncertainty sample will

---

Algorithm 1 Myriad Active Learning

Require: $N_{ul} \geq K$, $K \geq 1$, $t > 1$
  $\alpha \leftarrow [\ \ ], \beta \leftarrow [\ \ ], \mathcal{Q} \leftarrow [\ \ ], S \leftarrow [\ \ ], n \leftarrow 0$
  for all $\mathbf{x}_i \in \mathcal{U}_t$ **do:** **append:** $\sigma_{\mathbf{m\text{-}e}}(\mathbf{x}_i, \mathcal{M})$ **onto** α
  $\beta \leftarrow$ **ArgSort:** α in **descending** order
  Split: β into K approximately equally sized **subarrays** $\beta_0, ..., \beta_{K-1}$
  while len($\mathcal{Q}$) $< K$ **do**
    for i **in** β_n **do**
      if $P_t(i)$ **not in** S **then**
        append: $\mathcal{U}_t(i)$ **onto** $\mathcal{Q}$
        append: $P_t(i)$ **onto** S
        break
      end if
    end for
    if $n \geq K - 1$ **then** $n \leftarrow 0$ **else** $n \leftarrow n + 1$
  end while

---

almost certainly be found near a decision boundary. In this way, the entire list of uncertainties can be deliberately utilized. The overconfident predictions at the low end of the list serve to establish anchors, while the top end of the list samples the usual suspects from conventional methods. We utilize these characteristics by splitting up the sorted uncertainty list into K sub-arrays, so that no more than one sample will be selected from the same area. A visual representation of this is given in Fig. 2, intuitively showing the difference between MAL and the current techniques.

Combining M-E sampling with the diversity presented by the pseudo-labels, we create the novel selection algorithm Myriad Active Learning (MAL). The pseudocode for MAL can be found in Alg. 1. A summary of the algorithm follows: MAL takes in $\mathcal{U}_t$, P_t, and K as inputs, calculates the uncertainty σ for every sample in $\mathcal{U}_t$, and stores them in order in an array α. This array is argument sorted into β, such that $\beta[i]$ is the index of the ith most uncertain sample in $\mathcal{U}_t$. β is then split into K sub-arrays $\beta_0, ..., \beta_{K-1}$, and one-by-one the most uncertain sample from each sub-array is added to $\mathcal{Q}$. After each sample is selected, its corresponding pseudo-label from P_t is appended to S, and subsequent samples added to $\mathcal{Q}$ must have a pseudo-label which is not in S. This is to ensure that a sample with a different pseudo-label is selected from each sub-array. This portion loops until $\mathcal{Q}$ contains K samples, at which point they are sent to the oracle for labelling.

IV. EXPERIMENTS

A. Datasets

The proposed framework, MAL, is evaluated on two public histopathology datasets.

NCT-CRC-HE-100K (NCT) [17] provides 100,000 non-overlapping histopathology image patches cropped from 86 H&E stained tissue slides by the National Center for Tumor Diseases and the University Medical Center Mannheim pathology archive. All images in NCT are color-normalized with 224x224 pixels at 0.5 microns per pixel. The 9 tissue
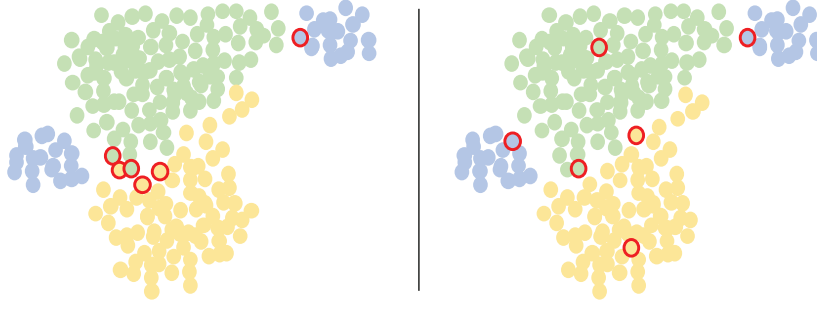
Fig. 2. Abstracted t-SNE plot of example data (3-class). Left: 6 samples selected by classical active learning methods with entropy sampling, notably selecting many samples from the same area which are highly likely to be redundant. Right: 6 samples selected using MAL, finetuning several of the borders simultaneously, while providing anchor samples for two of the classes.

Dataset	PrT LR	PrT Batch Size	PrT Epochs	LR	Batch Size	Epochs
NCT	0.015	128	200	0.0004	128	200
BreaKHis	0.0005	32	20	0.0012	128	100

TABLE I

HYPER-PARAMETERS (I.E. LEARNING RATE, BATCH SIZE, AND TRAINING EPOCH NUMBER) FOR SSL-PRETRAINING (DENOTED PRT) ENCODER AND THE ACTIVE LEARNING LOOP. A WEIGHT DECAY OF 0.0005 IS USED IN ALL CASES.

categories included in this dataset are adipose, background, debris, lymphocytes, mucus, smooth muscle, normal colon mucosa, cancer-associated stroma, and colorectal adenocarcinoma epithelium.

Breast Cancer Histopathological Database (BreaKHis) [32] is composed of 9,109 H&E stained breast tumor images from 82 patients, with 2,480 benign and 5,429 malignant samples. Aside from the binary labels of benign and malignant, images in BreakHis are further categorized into 8 categories, which are adenosis, fibroadenoma, phyllodes tumor, and tubular adenoma, carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma. All images in BreaKHis is with 700X460 pixels in the PNG format of 3-Channel RGB, 8-bit depth in each channel. Different from the NCT dataset, BreakHis is a very difficult dataset since it includes images with different magnifying factors (40X, 100X, 200X, and 400X) which makes computational image diagnosis more challenging.

B. Experimental Protocol

In our study, NCT and BreaKHis constitute 9-category classification and 8-class diagnosis tasks, respectively. For each dataset, images are divided into unlabelled training set and test set with a ratio of 80:20. The samples were divided randomly, as we operate under the assumption that we do not know the underlying distribution of the datasets (i.e. we have no label information at this stage).

This study focuses on the scenario of active few-shot learning where a very low annotation budget is available.

That is, we allow only n-shot samples for data annotation in experimentation, where $n = 1, 5, 10$.

The self-supervised learning model is composed of a ResNet-50 backbone and a 2-layer MLP head (2048-dimensional hidden layer with a ReLU activation). We follow the study in Chen et al. [4] for data augmentation. One exception is that for the BreaKHis dataset, a 460x460-pixel crop was used for the randomly resized images. The one-layer classification network is initialized using Xavier uniform distribution and is optimized with ADAM [18]. The hyper-parameters including learning rate, batch size, and training epoch number, are specified in Table IV-A.

C. Comparison Baseline

To the best of our knowledge, no prior works address the active few-shot learning scenario in histopathology, so we will make our comparisons over several different paradigms. Specifically, we divide our comparison experiments into two main parts. First, we compare MAL to a recent few-shot learning benchmark, FHIST [28]. Here, we use classification accuracy and macro F1 scores as the performance metrics. Then, we compare our methods to popular active learning methods, summarized by Zhan et al. [37].

Note that both the few-shot learning results and active learning results reported in this paper were reproduced using the official code published by their respective authors. The active learning methods required an initial seed, as they do not utilize pretraining, so they are given K randomly chosen labelled samples initially. All results are reported with the mean and sample standard deviation from 3 seeds.

Datasets	Method ↓	Accuracy			F1 Score		
		1-shot	5-shot	10-shot	1-shot	5-shot	10-shot
NCT	MAL	48.7±6.7	**77.9**±2.0	**87.1**±0.5	**51.2**±9.0	**79.5**±2.3	**88.0**±0.4
	FHIST [28]	**56.2**±10.8	75.4±8.1	80.9±7.2	30.3±6.2	41.8±4.5	44.9±4.0
BreaKHis	MAL	**33.9**±6.6	51.6±4.0	**65.1**±1.7	16.2±5.5	**29.7**±3.1	**37.2**±0.9
	FHIST [28]	33.8±7.36	**53.1**±8.54	62.0±7.44	**18.0**±4.1	29.1±4.9	34.2±4.2

TABLE II

TEST ACCURACIES (%) AND MACRO F1 SCORES (%) ON THE NCT AND BREAKHIS DATASETS IN THE FEW SHOT LEARNING SETTING. FHIST BY SHAKERI ET AL. [28] IS A RECENT STUDY PRESENTING A FEW-SHOT LEARNING BENCHMARK ON HISTOPATHOLOGY IMAGES.

Dataset	Method [37] →	MAL	Rand.	Margin	Entropy	VarRatio	CEAL	KMeans
NCT	1-shot	**48.7**±6.7	29.0±3.3	32.5±1.5	24.5±0.9	27.0±5.0	30.2±2.2	28.4±7.7
	5-shot	**77.9**±2.0	35.7±4.1	39.9±4.5	33.7±2.4	32.0±7.7	31.7±10.5	33.7±3.9
	10-shot	**87.1**±0.5	47.7±3.8	42.5±2.3	38.0±10.0	42.7±10.0	37.9±5.0	33.7±3.0
BreaKHis	1-shot	33.9±6.6	25.6±13.4	38.3±2.8	31.9±7.3	44.3±0.5	39.5±2.1	**43.7**±0.6
	5-shot	**51.6**±4.0	44.5±0.9	44.6±0.8	44.6±6.3	45.7±2.2	43.1±2.9	42.6±1.4
	10-shot	**65.1**±1.7	43.4±0.9	49.1±0.3	46.1±2.2	48.9±0.8	44.2±2.1	46.2±2.1

TABLE III

TEST ACCURACIES (%) FOR THE NCT AND 8-CLASS BREAKHIS DATASETS IN THE ACTIVE LEARNING SETTING. MAL IS COMPARED TO CONVENTIONAL DEEP LEARNING METHODS EVALUATED IN ZHAN ET AL. [37].

Dataset	Method [37] →	MAL	Rand.	Margin	Entropy	VarRatio	CEAL	KMeans
NCT	1-shot	**51.2**±9.0	22.0±7.1	24.1±2.5	17.4±2.4	21.3±7.1	23.0±5.0	20.5±8.9
	5-shot	**79.5**±2.3	31.3±2.1	36.2±4.7	27.5±3.2	24.6±6.9	25.8±9.8	27.1±3.3
	10-shot	**88.0**±0.4	44.5±4.8	40.6±1.8	34.6±12.0	34.6±12.0	34.8±6.1	27.0±3.6
BreaKHis	1-shot	**16.2**±5.5	6.0±1.9	11.0±1.2	10.1±2.3	9.3±1.2	9.5±1.0	11.9±1.6
	5-shot	**29.7**±3.1	9.1±0.9	10.4±1.3	15.6±1.4	13.0±1.8	13.9±0.6	9.9±1.3
	10-shot	**37.2**±0.9	8.9±0.9	15.7±0.7	15.3±0.7	14.3±0.9	14.9±2.2	12.5±1.4

TABLE IV

MACRO F1 SCORES (%) FOR THE NCT AND 8-CLASS BREAKHIS DATASETS IN THE ACTIVE LEARNING SETTING. MAL IS COMPARED TO CONVENTIONAL DEEP LEARNING METHODS EVALUATED IN ZHAN ET AL. [37].

D. Main Results and Discussion

Few-Shot Learning Comparisons: FHIST [28] is a recent few shot learning benchmark particularly designed for histopathology images. FHIST uses a large neural network pretrained on a well-annotated pathology image set, which is then transferred to and finetuned with the few-shot samples from the target dataset (NCT and BreaKHis in this case). Table II shows the n-shot learning results on NCT/BreaKHis with FHIST and MAL.

It is reasonable that the 1-shot accuracies are lower, as the classifier in MAL trains only on K samples, or a possible one sample per class from K-means, while FHIST takes advantage of the knowledge transferred from the extra annotated data. However, when more data is selected and annotated, for example, in 5-shots and 10-shots, MAL substantially improves the test accuracy and macro F1 score of the model, and significantly outperforms FHIST. Notably, in all 10-shot cases, MAL outperforms FHIST.

Higher budget settings are also explored in Table V. A CNN

trained from scratch achieves 96.16% accuracy on NCT [9], which can be boosted to 99.76% using transfer learning [7]. MAL achieves comparable results at only 5% labels, with a test accuracy of 95.9%.

Active Learning Comparisons: Next, we compare MAL against other classical deep active learning methods, using the methodology described by Zhan et al. [37] on the 8-class BreaKHis dataset and the NCT dataset. As shown in Tables III and IV, MAL improves upon popular deep active learning methods in the few shot setting. Specifically, macro F1 score is improved by 4.3%, 14.1%, and 21.5% at 1, 5, and 10-shots, respectively for the BreaKHis dataset. For NCT, a similar trend is observed with a 27.1%, 43.3%, and 43.5% increase in macro F1 score at 1, 5, and 10-shots, respectively. Similar trends are observed for the test accuracies.

Higher Budget Settings: In this setting, we relax the few-shot condition and investigate the performance of MAL with a higher annotation budget and report its performance in Table V. For a reference, a fully supervised CNN-based

	10-shot (0.11% Lbl)	0.5% Lbl	5% Lbl	20% Lbl	50% Lbl	100% Lbl
Acc	87.1±0.5	91.1±0.5	95.9±0.3	96.8±0.1	97.2±0.1	97.3±0.0
F1	87.3±0.4	91.2±0.4	95.4±0.2	96.9±0.1	97.3±0.1	97.4±0.0

TABLE V

MAL IN HIGHER ANNOTATION BUDGET SETTINGS ON THE NCT DATASET. TEST ACCURACIES (%) AND MACRO F1 SCORES (%) ARE SHOWN FOR ALL LABEL AMOUNTS. FOR A REFERENCE, A FULLY SUPERVISED CNN-BASED CLASSIFIER TRAINED ON THE ENTIRE NCT DATASET ACHIEVES 96.16% ACCURACY [9], [33].

Pseudo-Labels	SSL Pretraining	Sub-arrays	M-E	1-shot	5-shot	10-shot
×	×	×	×	21.2±5.1	49.5±9.2	59.4±3.5
✓	×	×	×	35.3±3.6	55.3±3.5	68.4±2.7
✓	✓	×	×	42.0±7.5	64.7±2.8	75.6±1.4
✓	✓	✓	×	45.4±6.4	71.2±3.2	78.5±0.2
✓	✓	✓	✓	48.7±6.7	77.9±2.0	87.1±0.5

TABLE VI

ABLATION ON THE COMPONENTS OF MAL. THE NUMERICAL NUMBERS ARE TEST ACCURACY (%) ON THE NCT DATASET.

classifier trained on the entire NCT dataset achieves 96.16% accuracy [9], [33]. That is, MAL is able to achieve similar performance with only 5% annotation by always selecting the most informative samples to supplement model learning.

Discussion and limitations: Our experiments show that MAL outperforms prior FSL and AL methods on histopathology images [28], [37] in terms of accuracy, macro F1 score, and label efficiency.

There is a notable large gap in macro F1 score between MAL and the other FSL and AL methods on the NCT dataset. This is due to the nature of MAL to select a more balanced query, which results in more even performance increases across the classes, and thus a higher macro F1 score for a comparable test accuracy.

One **limitation** of MAL is relatively low performance in the 1-shot setting, which is due to the use of K-Means clustering to circumvent the lack of information in the first cycle. This could be remedied, at the cost of budget, by using an initial seed. In addition, the use of pseudo-labels in MAL is simple and straightforward. A more sophisticated design would potentially improve the performance. For example, one can assign multiple pseudo-labels to each sample, and reduce overlap on the whole set of pseudo-labels, rather than the most likely one. We leave the pursuit of these directions for future work.

E. Ablation Studies

MAL Components: In this ablation study, we gradually remove the four essential components of MAL to measure their impact on the overall performance on the NCT dataset. The four components are pseudo-label generation, the SSL pretrained encoder, the segmentation of the uncertainty list into sub-arrays, and margin-entropy sampling. These are denoted as pseudo-labels, SSL pretraining, Sub-array, and M-E in Table VI, respectively. In each case, the relevant feature is either omitted, or replaced by a conventional version. Specifically,

where the SSL encoder is not used, a few-shot learning model pretrained on the CRC-TP dataset (280 000 images, 7-classes of colorectal cancer) [15] is used; and when M-E sampling is not used, the conventional entropy sampling is used instead. As seen in Table VI, when each piece of MAL is removed, the test accuracies drop a significant amount at 5 and 10 shots. The 1-shot performance is once again hampered by the lack of target dataset information, but generally decreases as parts are removed. This high variance can be attributed to the lack of knowledge in the early cycles - in the first cycle, the algorithm has no information on the target dataset, so the samples selected are generally low quality, and do not accurately represent the underlying distribution of data.

V. CONCLUSIONS AND FUTURE WORK

In this work, we proposed Myriad Active Learning (MAL), a framework to efficiently and effectively increase the classification accuracy of an active few shot learning model by utilizing unlabelled data. MAL exploits the nature of uncertainty-based active learning sample selection by combining classical uncertainty estimation techniques for a more precise and deliberate query selection strategy. Pseudo-labels generated by an SSL encoder and classifier informed the active learning queries, allowing each sample to be "aware" of the others for a consistently more diverse and less redundant query. MAL produced excellent results, achieving comparable classification accuracy of a fully supervised model on the NCT dataset with only 5% annotation. MAL also outperforms current few-shot learning methods at 5 and 10 shots, and outperforms common active learning methods in the limited budget setting.

One may notice that in our few-shot active learning paradigm, a SSL encoder is trained on the unlabelled data for the initial pseudo-label generation, avoiding cold start in active learning. Recently, we witness the surge of foundation models. We hypothesize the knowledge in these foundation models would be another good source for pseudo-labels. We will validate this hypothesis in future studies.

REFERENCES

[1] Randall Balestriero, Mark Ibrahim, Vlad Sobal, Ari Morcos, Shashank Shekhar, Tom Goldstein, Florian Bordes, Adrien Bardes, Gregoire Mialon, Yuandong Tian, Avi Schwarzschild, Andrew Gordon Wilson, Jonas Geiping, Quentin Garrido, Pierre Fernandez, Amir Bar, Hamed Pirsiavash, Yann LeCun, and Micah Goldblum. A cookbook of self-supervised learning, 2023.

[2] Samuel Budd, Emma C Robinson, and Bernhard Kainz. A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Medical Image Analysis*, 71(102062):102062, July 2021.

[3] Vivien Cabannes, Leon Bottou, Yann Lecun, and Randall Balestriero. Active self-supervised learning: A few low-cost relationships are all you need, 2023.

[4] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning, 2020.

[5] Ozan Ciga, Tony Xu, and Anne Louise Martel. Self supervised contrastive learning for digital histopathology. *Machine Learning with Applications*, 7:100198, 2022.

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[7] Hamida et al. Deep learning for colon cancer histopathological images analysis. *Computers in Biology and Medicine*, 136:104730, 2021.

[8] Iliopoulos et al. Weighted distillation with unlabeled examples. In *NeurIPS*, 2022.

[9] Sourodip et al. Colorectal histology tumor detection using ensemble deep neural network. *Engineering Applications of Artificial Intelligence*, 100:104202, 2021.

[10] Mingfei Gao, Zizhao Zhang, Guo Yu, Sercan O. Arik, Larry S. Davis, and Tomas Pfister. Consistency-based semi-supervised active learning: Towards minimizing labeling cost, 2020.

[11] Marc Gorriz, Axel Carlier, Emmanuel Faure, and Xavier Giro i Nieto. Cost-effective active learning for melanoma segmentation, 2017.

[12] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning, 2020.

[13] Ashish Jaiswal, Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Debapriya Banerjee, and Fillia Makedon. A survey on contrastive self-supervised learning, 2021.

[14] Dietmar Jannach, Gabriel de Souza P. Moreira, and Even Oldridge. Why are deep learning models not consistently winning recommender systems competitions yet? a position paper. In *Proceedings of the Recommender Systems Challenge 2020*, RecSysChallenge '20, page 44–49, New York, NY, USA, 2020. Association for Computing Machinery.

[15] Sajid Javed, Arif Mahmood, Naoufel Werghi, Ksenija Benes, and Nasir Rajpoot. Multiplex cellular communities in multi-gigapixel colorectal cancer histology images for tissue phenotyping. *IEEE Transactions on Image Processing*, 29:9204–9219, 2020.

[16] Xu Jin, Teng Huang, Ke Wen, Mengxian Chi, and Hong An. Histossl: Self-supervised representation learning for classifying histopathology images. *Mathematics*, 11(1), 2023.

[17] Jakob Nikolas Kather, Niels Halama, and Alexander Marx. 100,000 histological images of human colorectal cancer and healthy tissue, Apr. 2018.

[18] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[19] Bogdan Kwolek, Michal Koziarski, Andrzej Bukala, Zbigniew Antosz, Boguslaw Olborski, Pawel Wkasowicz, Jakub Swadzba, and Boguslaw Cyganek. Breast cancer classification on histopathological images affected by data imbalance using active learning and deep convolutional neural network. In Igor V. Tetko, Vvera Krurkova, Pavel Karpov, and Fabian Theis, editors, *Artificial Neural Networks and Machine Learning – ICANN 2019: Workshop and Special Sessions*, pages 299–312, Cham, 2019. Springer International Publishing.

[20] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015.

[21] Ammara Masood, Adel Al-Jumaily, and Khairul Anam. Self-supervised learning model for skin cancer diagnosis. In *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 1012–1015, 2015.

[22] Nimmagadda Satyanarayana Murthy and Chaitanya Bethala. Review paper on research direction towards cancer prediction and prognosis using machine learning and deep learning models. *J. Ambient Intell. Humaniz. Comput.*, Mar. 2021.

[23] Thomas Müller, Guillermo Pérez-Torró, Angelo Basile, and Marc Franco-Salvador. Active few-shot learning with fasl, 2022.

[24] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y. Ng. Reading digits in natural images with unsupervised feature learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011*, 2011.

[25] Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Brij B. Gupta, Xiaojiang Chen, and Xin Wang. A survey of deep active learning. *ACM Comput. Surv.*, 54(9), oct 2021.

[26] Iqbal H Sarker. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput. Sci.*, 2(6):420, Aug. 2021.

[27] Burr Settles. Active learning literature survey. 2009.

[28] Fereshteh Shakeri, Malik Boudiaf, Sina Mohammadi, Ivaxi Sheth, Mohammad Havaei, Ismail Ben Ayed, and Samira Ebrahimi Kahou. Fhist: A benchmark for few-shot classification of histological images, 2022.

[29] C. E. Shannon. A mathematical theory of communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, 5(1):3–55, jan 2001.

[30] Samarth Sinha, Sayna Ebrahimi, and Trevor Darrell. Variational adversarial active learning. *arXiv preprint arXiv:1904.00370*, 2019.

[31] Yisheng Song, Ting Wang, Subrota K Mondal, and Jyoti Prakash Sahoo. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities, 2022.

[32] Fabio A. Spanhol, Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462, 2016.

[33] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, 2019.

[34] Jeroen van der Laak, Geert Litjens, and Francesco Ciompi. Deep learning in histopathology: the path to the clinic. *Nat. Med.*, 27(5):775–784, May 2021.

[35] Keze Wang, Dongyu Zhang, Ya Li, Ruimao Zhang, and Liang Lin. Cost-effective active learning for deep image classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(12):2591–2600, dec 2017.

[36] Mark Woodward and Chelsea Finn. Active one-shot learning, 2017.

[37] Xueying Zhan, Qingzhong Wang, Kuan hao Huang, Haoyi Xiong, Dejing Dou, and Antoni B. Chan. A comparative survey of deep active learning, 2022.