

Few-Shot Histopathology Image Classification: Evaluating State-of-the-Art Methods and Unveiling Performance Insights

Ardhendu Sekhar, Ravi Kant Gupta and Amit Sethi

Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India

Keywords: Deep Learning, Few-shot Classification, Medical Image.

Abstract: This paper presents a study on few-shot classification in the context of histopathology images. While few-shot learning has been studied for natural image classification, its application to histopathology is relatively unexplored. Given the scarcity of labeled data in medical imaging and the inherent challenges posed by diverse tissue types and data preparation techniques, this research evaluates the performance of state-of-the-art few-shot learning methods for various scenarios on histology data. We have considered four histopathology datasets for few-shot histopathology image classification and have evaluated 5-way 1-shot, 5-way 5-shot and 5-way 10-shot scenarios with a set of state-of-the-art classification techniques. The best methods have surpassed an accuracy of 70%, 80% and 85% in the cases of 5-way 1-shot, 5-way 5-shot and 5-way 10-shot cases, respectively. We found that for histology images popular meta-learning approaches is at par with standard fine-tuning and regularization methods. Our experiments underscore the challenges of working with images from different domains and underscore the significance of unbiased and focused evaluations in advancing computer vision techniques for specialized domains, such as histology images.

1 INTRODUCTION

Traditional deep learning models often require large amounts of labeled data for training. These models learn representations and patterns from a substantial dataset to generalize well to unseen examples. The learning process involves adjusting numerous parameters through backpropagation to minimize the difference between predicted and actual outputs. Transfer learning is commonly used in traditional deep learning, where models pre-trained are fine-tuned for specific tasks. The knowledge gained from the pre-training on large dataset helps in solving related problems with smaller dataset. In medical imaging, traditional deep learning methods have been successful for tasks such as image classification and segmentation, but they often require extensive labeled dataset. But traditional models might struggle when faced with new tasks or limited data, requiring substantial retraining or fine-tuning. This is where few-shot learning can be useful. Few-shot learning, as the name implies, is designed to make accurate predictions when only a small number of examples per class are available for training. This can be achieved through various techniques such as meta-learning, where the model is trained on a va-

riety of tasks and adapts quickly to new tasks with minimal examples by leveraging knowledge gained from previous tasks. One would think that meta learning would be particularly valuable in medical imaging due to the scarcity and cost of labeling medical data. However, such techniques have only been evaluated on natural images. If successful on medical images, this would enable demonstrate learning even when only a handful of annotated medical images are available.

In this work, we have evaluated certain state-of-the-art few-shot classification techniques on histopathology medical datasets. A dataset prepared by (Komura and Ishikawa, 2021) and FHIST (Shakeri et al., 2022) were considered for the experiments. FHIST dataset comprises of many histology datasets. These are : CRC-TP (Javed et al., 2020b), NCT-CRC-HE-100K (Kather et al., 2019), LC25000 (Borkowski et al., 2019) and BreakHis (Spanhol et al., 2016). For our experiments, we have considered CRC-TP, NCT and LC25000 datasets. CRC-TP is a colon cancer dataset with six classes. NCT is also a colon cancer dataset with nine classes. LC25000 consists of both lung and colon cancer images with five classes. We have also used a histology dataset proposed by (Komura and Ishikawa, 2021). It has around 1.6 million

cancerous image patches of 32 different organs in the body. The classes in this dataset are defined according to the different organ sites. Dataset proposed by (Komura and Ishikawa, 2021) is used to train the few-shot classification models. The trained model is then evaluated on various FHSIT dataset.

In our comprehensive exploration of few-shot classification techniques, we have meticulously incorporated various state-of-the-art methodologies to ensure a thorough evaluation. The methods employed in our experiments includes Prototypical Networks (Snell et al., 2017), Model-Agnostic Meta-Learning (MAML) (Finn et al., 2017), SimpleShot (Wang et al., 2019), LaplacianShot (Ziko et al., 2020), and DeepEMD (Zhang et al., 2020). Prototypical Networks, a benchmark technique for few-shot learning, leverage prototypes as representative embeddings for each class. By minimizing the distance between query examples and class prototypes, this model excels in adapting quickly to new classes with limited labeled data. MAML adopts a meta-learning paradigm, training a model to quickly adapt to new tasks with minimal examples. This approach proves invaluable in scenarios where prompt adaptation to novel classes is paramount. As its name implies, SimpleShot emphasizes simplicity in few-shot learning. This approach often relies on straightforward yet effective techniques, showcasing the power of simplicity in addressing complex classification tasks with limited data. LaplacianShot introduces Laplacian regularization to enhance few-shot learning performance. By incorporating this regularization technique, the model aims to improve generalization and robustness across diverse classes. DeepEMD learns the image representations by calculating the discrepancy between the joint characteristics of embedded features and product of the marginals.

2 RELATED WORK

In recent years, there has been a notable surge in research endeavors for addressing the challenges of few-shot learning within specific domains of medical imaging. This section outlines some contributions in the field, shedding light on how researchers have applied various few-shot learning techniques to tackle different problems in medical diagnostics. One such study (Mahajan et al., 2020), focuses on skin-disease identification. Here, the researchers applied two prominent few-shot learning methods, Reptile (Nichol and Schulman, 2018) and Prototypical Networks (Snell et al., 2017), showcasing their efficacy in the task of distinguishing various skin diseases. The

task of transferring knowledge across different tissue types was addressed in (Medela et al., 2019), where a deep Siamese neural network was trained to transfer knowledge from a dataset containing colon tissue to another encompassing colon, lung, and breast tissue. Meanwhile, (Teng et al., 2021) proposed a few-shot learning algorithm based on Prototypical Networks, specifically tailored for classifying lymphatic metastasis of lung carcinoma from Whole Slide Images (WSIs). In another work in (Chen et al., 2021), a two-stage framework was adopted for the crucial task of COVID-19 diagnosis from chest CT images. The initial stage involved capturing expressive feature representations using an encoder trained on publicly available lung datasets through contrastive learning. Subsequently, in the second stage, the pre-trained encoder was employed within a few-shot learning paradigm, leveraging the Prototypical Networks method. In (Yang et al., 2022), the authors have incorporated contrastive learning (CL) with latent augmentation (LA) to build a few-shot classification model. Here, contrastive learning learns important features without needing labels, while Latent Augmentation moves semantic variations from one dataset to another without supervision. The model was trained on publicly available colon cancer datasets and evaluated on PAIP (Kim et al., 2021) liver cancer Whole Slide Images. Shakeri et al. (Shakeri et al., 2022) introduced the FHIST data coupled with assessments of certain few-shot classification methodologies serves as a catalyst for additional exploration in this domain.

The exploration of few-shot learning is not confined to image classification alone. In (Ouyang et al., 2020), a few-shot semantic segmentation framework was introduced, aiming to alleviate the dependency on labeled data during the training phase. Similarly, (Yu et al., 2021) ventured into the domain of medical image segmentation, employing a prototype-based method known as the location-sensitive local prototype network, which strategically incorporates spatial priors to enhance segmentation accuracy.

3 METHODOLOGY

Few shot classification deals with the cases of limited data. For example, a medical dataset has five rare classes and in each class there are five images. It is very difficult to train and test DL models with such small number of images. Training the model with such small dataset my lead to over-fitting. Since the classes are rare, there is a high probability that the pre-trained models might not have seen those images while being trained. Instead of tradi-

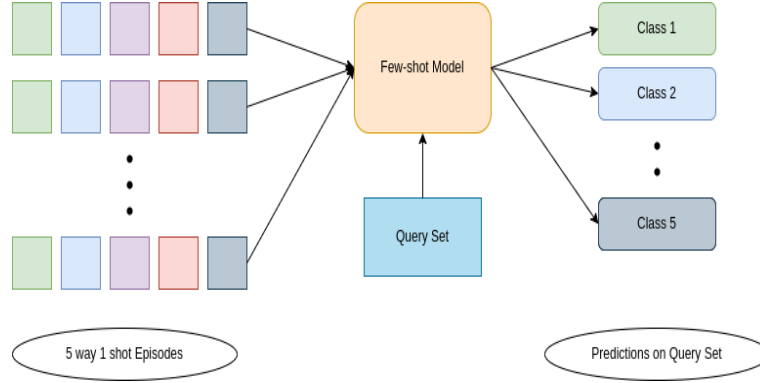


Figure 1: The diagram illustrates a few-shot learning model, showcasing its ability to effectively generalize and recognize classes for unlabeled query set with only a limited number of support examples. The 5 different colors in support set represent 5 different classes (ways) having 1 sample (shot) each.

tional transfer-learning, the few-shot models follow episodic-training paradigm.

In few-shot training, the train and test set are disjoint. The training set is denoted as D_{Train} . It consists of $(X_{Train}, Y_{Train})_{i=1}^M$. X_{Train} are the images. Y_{Train} are the labels. M is the number of classes. In episodic training, the huge labelled train dataset is broken into many episodes. In each episode, there is a support set (S) and a query set (Q). Each episode is defined by K -way N -shot Q -query. It means, for each episode, K classes are randomly selected from M classes. From each K class, N images and Q images are selected. $N \times K$ images form the support set and $Q \times K$ images form the query set. The few shot model is trained in such a way that it must learn from the support set to predict the labels of the query set. Once the model is trained, it is tested on the test set.

$$D_{Train} = (X_{Train}, Y_{Train}) \quad (1)$$

$$S := (X_i, Y_i)_{i=1}^{K \times N} \quad (2)$$

$$Q := (X_i, Y_i)_{i=1}^{K \times Q} \quad (3)$$

$$D_{Test} = (X_{Test}, Y_{Test}) \quad (4)$$

$$S := (X_i, Y_i)_{i=1}^{K \times N} \quad (5)$$

$$Q := (X_i) \quad (6)$$

$$Y_{Train} \cap Y_{Test} = \Phi \quad (7)$$

3.1 Prototypical Networks

The central idea of Prototypical Networks (Snell et al., 2017) is rooted in the observation that data points exhibiting proximity to a singular prototype representation for each class contribute to a meaningful embedding. To operationalize this concept, a non-linear mapping transforms the input data into a specialized embedded space. Within this embedded space, the prototype representation for each class was

derived by calculating the mean of its respective support set. The process is initiated by the formation of a non-linear mapping, effectively transforming the original input data into a specialized embedded space. This mapping is crucial for capturing intricate relationships and patterns within the data. Within the embedded space, a class prototype was defined for each category. The prototype representation for a specific class was computed by determining the mean of its corresponding support set. This means that the prototype becomes a central point representing the class in the embedded space. To perform classification on a given query, the approach relies on identifying the nearest distance between the embedded query and the prototype representation of each class. The class associated with the closest prototype is then assigned to the query, effectively determining its classification.

The small support set consists of K labeled examples which is represented by $S = [(X_1, Y_1), \dots, (X_K, Y_K)]$. X_i is a D -dimensional feature vector for each image and Y_i represents the corresponding label of X_i . There are N classes within a support set. The number of examples within a class N is represented by S_N .

With the help of an embedding function, i.e. a CNN, f_Φ , prototypical networks estimate an M dimensional representation C_N of each class. C_N is the mean vector of the support points of each class. Φ are the learnable parameters.

$$f_\Phi : R^D \rightarrow R^M \quad (8)$$

$$C_N \in R^M \quad (9)$$

$$C_N = \frac{1}{|S_N|} \sum_{(X_i, Y_i) \in S} f_\Phi(X_i) \quad (10)$$

Prototypical networks generate a probability distribution across classes for a given query example X . This distribution is computed using a softmax func-

tion applied to the distances between the query example and other prototypes in the embedding space.

$$P_{\Phi}(Y = N|X) = \frac{\exp(-d(f_{\Phi}(X), C_N))}{\sum_{N'} \exp(-d(f_{\Phi}(X), C_{N'}))} \quad (11)$$

The model learns by minimizing the negative of the log probability of $J(\Phi)$ of the true class N through Adam solver.

$$J(\Phi) = -\log(P_{\Phi}(Y = N|X)) \quad (12)$$

3.2 Model Agnostic Meta Learning(MAML)

MAML (Finn et al., 2017) introduces a meta-learning objective that involves training a model on a distribution of tasks. The goal is to learn a good initialization of the model's parameters that facilitates quick adaptation to new, unseen tasks. The focus of MAML is on few-shot learning scenarios, where the model is required to generalize well from a small number of examples per class. This is particularly important in real-world applications where collecting extensive labeled dataset can be impractical. During the meta-training phase, the model is exposed to a variety of tasks. Each task consists of a support set (small labeled dataset) and a query set (unlabeled examples). The model is trained to adapt quickly to new tasks by updating its parameters through gradient descent. MAML introduces a two-step gradient descent update during meta-training. The first update involves computing the gradients on the support set and using them to update the model's parameters. The second update fine-tunes the model on the query set to improve its performance on the specific task. The model's initialization is learned in a way that allows it to adapt quickly to new tasks. The meta-training process encourages the model to learn parameter initialization that generalize well across different tasks. During meta-testing, the model is fine-tuned quickly on new tasks with limited examples. This is achieved by applying a small number of gradient descent steps using the updated initialization. The goal is to improve the model's performance on the specific task.

The neural network is parameterized by a set of weights θ . The weights are to be updated in such a way that they can be rapidly adapted towards different solutions. For instance, if there are 3 episodes E_1, E_2, E_3 then through a few gradient steps the weights should be able to move from a specific point θ to another configuration of weights θ_1^*, θ_2^* and θ_3^* . It should be done in such a way that the configuration of weights are now well adapted to the task to be solved. This can be achieved by optimizing three different

losses on three different support sets at the same time. In this case, 2 losses will be optimized. The first loss is across all the support to find θ . Another query set specific loss will be optimized in θ_1^*, θ_2^* and θ_3^* directions.

$$S = [(X_1, Y_1), (X_2, Y_2) \dots (X_5, Y_5)] \quad (13)$$

$$Q = [(X_1, Y_1)] \quad (14)$$

The two above equations represent a particular episode. Suppose there are 3 different episode. In each of the support sets of these episodes, there are five different images and five labels. And in each query set, there is one image and one label. The objective is finding out the one particular class to which the query image belongs. It is a standard few shot setting. The below equations will depict how gradient descent happens in one episodic training.

$$E_1 : f_{\theta}(S_1) = V_1 \quad (15)$$

$$\theta_1^* = \theta - \alpha \Delta_{\theta} \mathcal{L}(S_1, V_1) \quad (16)$$

$$f_{\theta_1^*}(Q_1) = Z_1 \quad (17)$$

$$\theta = \theta - \beta \Delta_{\theta} \mathcal{L}(Q_1, Z_1) \quad (18)$$

Considering episode 1, the model is parameterized by f_{θ} . The S_1 is passed through f_{θ} , V_1 is obtained as output. The θ is fine tuned towards a specific set of parameters θ_1 . This θ_1 is used to do another forward pass on the query set. This is the prediction on the query image. The loss is calculated from the query set which is used to do a backward pass on the θ . The task is repeated for the rest other episodes E_2 and E_3 . Then θ is updated by taking the gradient of accumulated losses of all the query sets in all the tasks.

3.3 SimpleShot

SimpleShot (Wang et al., 2019) is a simple non-episodic few-shot learning method. It uses the concept of transfer-learning and nearest-neighbour rule. A large scale deep neural network is trained on training classes. Then nearest-neighbour rule is carried out on the images of the test episodes by using the trained deep neural network as a feature encoder.

Training set is defined as D_{Train} .

$$D_{Train} = [(X_1, Y_1), (X_2, Y_2), \dots, (X_N, Y_N)] \quad (19)$$

X are the images. Y are the corresponding labels. The train set is trained on a CNN with cross entropy loss.

$$\underset{\theta}{\operatorname{argmin}} \sum_{(X,Y) \in D_{Train}} l(W^T f_{\theta}(X), Y) \quad (20)$$

f_θ represents the CNN. W represents the weights of the classification layer. l represents the cross entropy loss function.

Classification on the test set is done using the nearest neighbour rule. The features of an image are obtained by passing it through the CNN. The features are denoted by Z . The nearest neighbour classification is carried out using the Euclidean distance.

$$Z = f_\theta(X) \quad (21)$$

In one shot setting, support set S of test set has 1 labelled example from each of the N classes.

$$S = ((\hat{X}_1, 1), \dots, (\hat{X}_N, N)) \quad (22)$$

Using the Euclidean distance measure, the nearest neighbour rule classifies the query image \hat{X} to the most similar support image.

$$Y(\hat{X}) = \underset{N \in \{1, 2, \dots, N\}}{\operatorname{argmin}} d(\hat{Z}, \hat{Z}_N) \quad (23)$$

\hat{Z} and \hat{Z}_N are the CNN features of the query and support images respectively. In multi shot setting, \hat{Z}_N is the average of feature vector of each class in the support set.

3.4 LalaplacianShot

LaplacianShot (Ziko et al., 2020) introduces a new approach for few-shot tasks - a transductive Laplacian regularized interference. It minimises a quadratic binary assignment function comprising of two essential terms. The first one is a unary term that allocates query samples to their closest class prototype. The second term is a pairwise-Laplacian term that promotes consistent label assignments among nearby query samples.

Similar to SimpleShot, train set is defined as D_{Train} . The train set is trained on a CNN with a simple cross-entropy loss. It does not involve any episodic or meta learning strategy.

$$\underset{(X, Y) \in D_{Train}}{\operatorname{argmin}}_\theta \sum l(W^T f_\theta(X), Y) \quad (24)$$

The regularization equations involved at the time of few-shot test inference are described below.

$$E(Y) = N(Y) + \lambda \frac{1}{2} L(Y) \quad (25)$$

$$N(Y) = \sum_{q=1}^N \sum_{c=1}^C y_{q,c} d(z_q - m_c) \quad (26)$$

$$L(Y) = \frac{1}{2} \sum_{q,p} w(z_q, z_p) \|y_q - y_p\|_2^2 \quad (27)$$

In this objective, the first term $N(Y)$ is minimized by assigning each query point to the class of the nearest prototype m_c , from the support set, using a distance metric such as Euclidean distance. The second term, $L(Y)$, represents the Laplacian regularizer and is expressed as $\operatorname{tr}(Y^T L Y)$. L is a laplacian matrix that corresponds to the affinity matrix $W = w(z_q, z_p)$. It measures the similarity between the feature vectors z_q and z_p with the help of a kernel function. z_q and z_p are feature vectors of the query images x_p and x_q .

3.5 DeepEMD

In DeepEMD (Zhang et al., 2020), Earth Mover's Distance is used as a metric to calculate the structural distance dense image representations, determining image similarity. The Earth Mover's Distance produces optimal matching flows between structural elements, minimizing the matching cost. The minimized cost is then utilized to signify the image distance for classification. To derive crucial weights for elements in the EMD, a cross-reference mechanism is introduced which mitigates the impact of clustered backgrounds and intra-class variations. For K-shot classification, a structured fully connected layer capable of directly classifying dense image representations using EMD is used.

The Earth Mover's Distance serves as a distance metric between the two sets of weighted objects or distributions, leveraging the fundamental concept of distance between individual objects. It adopts the structure of the Transportation Problem in Linear Programming. A set of supplier $S = (S_i | i=1, 2, \dots, m)$ must transport goods to a set of demanders $D = (d_j | j=1, 2, \dots, k)$. S_i represents the supply unit of supplier and d_j indicates the demand of demander j . The expense for transporting one unit from supplier i to demander j is represented by c_{ij} . And the quantity of units transported is represented by x_{ij} . The objective of the transportation problem is to find the least-expensive flow of goods $\tilde{x} = (\tilde{x}_{ij} | i=1, 2, \dots, m, j=1, 2, \dots, k)$ from supplier i to demander j :

$$\underset{x_{ij}}{\operatorname{minimize}} \sum_{i=1}^m \sum_{j=1}^k c_{ij} x_{ij} \quad (28)$$

In few shot classification, to calculate the similarity between the support and query images, the images are passed through a fully convolutional network to generate image features S and Q .

$$S \in \mathbb{R}^{H \times W \times C} \quad (29)$$

$$Q \in \mathbb{R}^{H \times W \times C} \quad (30)$$

Each image feature is a collection of vectors i.e. $S = [s_1, s_2, \dots, s_{HW}]$ and $Q = [q_1, q_2, \dots, q_{HW}]$. The matching cost between the 2 set of vectors is denoted as the similarity of images. The cost between the 2 embeddings s_i and q_j is given by:

$$c_{ij} = 1 - \frac{s_i^T q_j}{\|s_i\| \|q_j\|}, \quad (31)$$

The optimal matching flow is represented by $\tilde{X} \in R^{HW \times HW}$. The distance between the support and the query image is represented by:

$$d(S, Q) = \sum_{i=1}^{HW} \sum_{j=1}^{HW} (1 - c_{ij}) \tilde{x}_{ij} \quad (32)$$

In a support set, when the shot is more than 1 then the learnable embedding becomes a group of image features for a class rather than one vector. Then the mean of it is taken to give a single image feature. It is similar to calculating the prototype of a class in a support set. The fully connected network is a feature extractor and SGD optimizer is used to update the weights by sampling few-shot episodes from the dataset.

4 EXPERIMENTS AND RESULTS

4.1 Dataset

(Komura and Ishikawa, 2021) have created a dataset of histology images from uniform tumor regions in The Cancer Genome Atlas Program whole slide images (tcga, 2005). TCGA consist of tissue slides from 32 cancer types at different sites of a human body. From this dataset, (Komura and Ishikawa, 2021) created a set of 1,608,060 image patches with six different magnification levels i.e. $0.5\mu/\text{pixel}$, $0.6\mu/\text{pixel}$, $0.7\mu/\text{pixel}$, $0.8\mu/\text{pixel}$, $0.9\mu/\text{pixel}$, $1.0\mu/\text{pixel}$. As these images are obtained from 32 different types of cancer, they are categorized into 32 classes. NCT (Kather et al., 2019) dataset comprises of 100,000 image patches of human colorectal cancer extracted from Hematoxylin and Eosin stained histological images and normal tissue. The resolution of images are 224×224 . These images are categorized into seven classes which are Adipose (ADI), background (BACK), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR) and colorectal adenocarcinoma epithelium (TUM). LC25000 (Borkowski et al., 2019) or Lung and Colon Histopathological dataset contains 25000 image patches, As the name suggests, it comprises of

images from lung and colon cancer. The resolution of the images is 768×768 . they are categorized into five classes. Three classes belong to lung cancer and the remaining two classes belong to colon cancer. The classes are benign colon tissues, colon adenocarcinoma, lung squamous cell carcinoma and benign lung tissues. CRC-TP (Javed et al., 2020b) is also a colon cancer dataset. It consists of 280,000 image patches categorized into six classes. They are tumor, stroma, complex stroma muscle, debris, inflammatory and benign. The images dimensions are 150×150 . The dataset prepared by Komura et al. is used as training dataset. As the dataset is huge, a formidable few-shot backbone network can be expected after training on it. Rest other dataset are used as testing set. In few-shot learning, the test and training set are disjoint.

4.2 Results and Discussion

All the experiments are conducted on an NVIDIA A100 in pytorch. Two types of training regimes are followed: Episodic training and standard training. Standard training trains across an entire dataset with abundant examples per class. In contrast, episodic training, designed for few-shot learning, adapts a model to quickly generalize from small episodes containing very few examples of new classes. Models such as Protonet (Snell et al., 2017), MAML (Finn et al., 2017), DeepEMD (Zhang et al., 2020) follow episodic training. SimpleShot (Wang et al., 2019) and LaplacianShot (Ziko et al., 2020) follow the standard training procedure.

In episodic training, the methods are trained for 120 epochs. In each epoch, 600 episodes are randomly selected from the training set. The models are trained on 5-way 1-shot, 5-way 5-shot and 5-way 10-shot scenarios separately. In all these cases, the number of query images are set to 15. The initial learning rate is set to $1e-3$. γ is set to 0.1. The batch size is set to 1 episode. During 5-way 1-shot, 5-way 5-shot and 5-way 10-shot training, the number of images in a batch are $(5 \times 1 + 5 \times 15)$ 80, $(5 \times 5 + 5 \times 15)$ 100 and $(5 \times 10 + 5 \times 15)$ 125 respectively. Few-shot methods that follow standard CNN training have been trained for 150 epochs. The batch size is set to 512. The initial learning rate, weight decay are initialized as 0.05 and $5e-4$. Resnet18 is used as a backbone network in both episodic and training procedure.

All the trained models follow meta-testing. In this procedure, 5000 episodes are randomly sampled from CRC-TP, NCT and LC25000 dataset. In all K-shot testing scenarios, the number of query images are fixed as 15 in each episode. The results on 1-shot, 5-shot, 10-shot on different datasets are re-

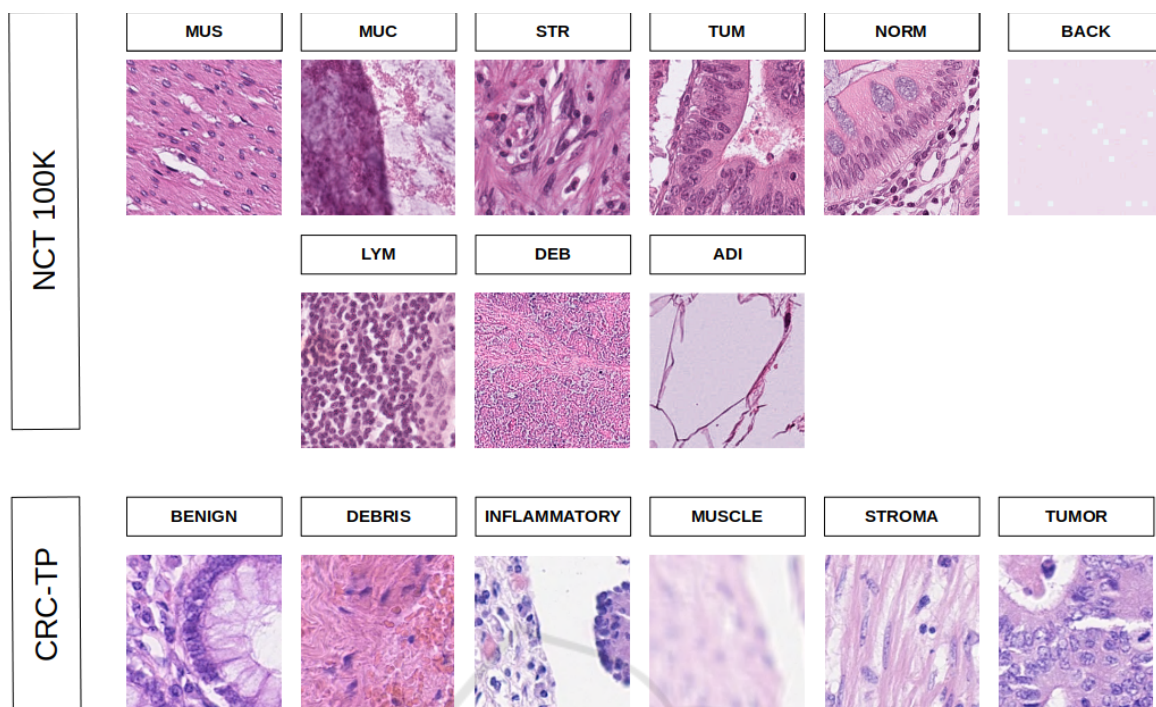


Figure 2: Snapshot of sample images of each class from NCT (top 2 rows) and CRC-TP (bottom row) of FHIST dataset.

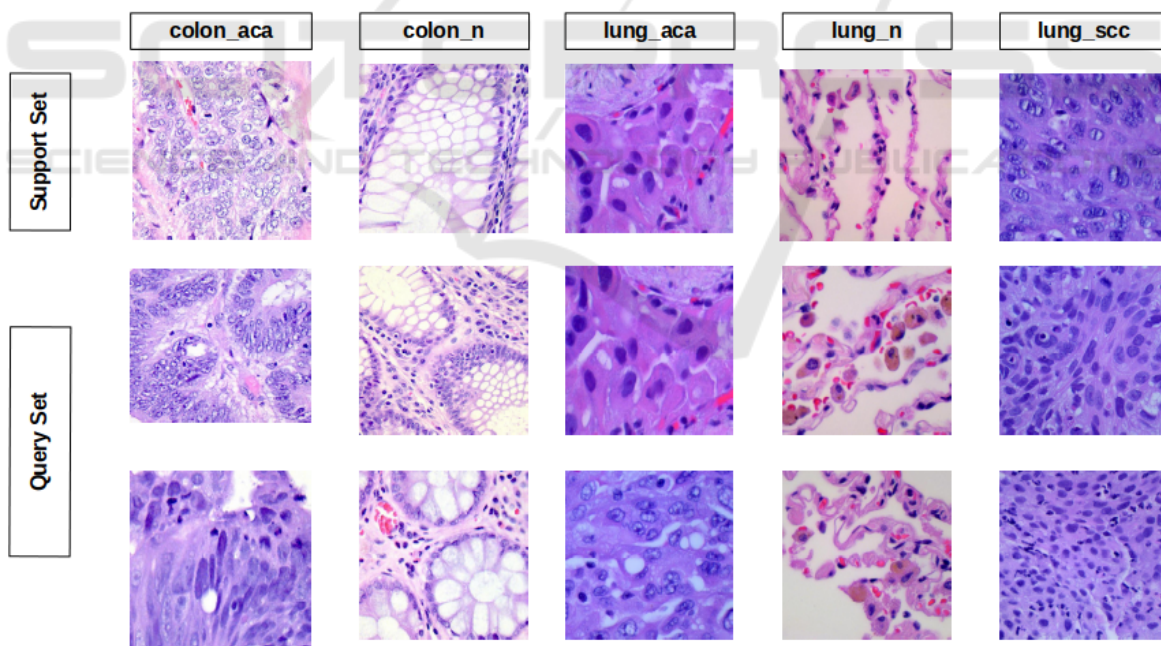


Figure 3: An example of 5-way 1-shot 2-query episode. The first row represents the support set from 5 different classes of LC25000 dataset. The last two rows represent the query set.

ported in the Table 1. The results we achieved by employing few-shot methods on the histology dataset align to a great extent with the outcomes obtained from the datasets of natural images, but there were

a few surprises as well. As anticipated, few-shot techniques trained using standard fine-tuning such as SimpleShot and LaplacianShot yielded better results compared to approaches that adhered to episodic

Table 1: Accuracy(%) on three different datasets; CRC-TP (Javed et al., 2020b), NCT (Kather et al., 2019) and LC25000 (Borkowski et al., 2019). The few-shot models are trained on the dataset proposed by (Komura and Ishikawa, 2021).

		CRC-TP		
Method	Training Method	5-way 1-shot	5-way 5-shot	5-way 10-shot
MAML (Finn et al., 2017)	Episodic	38.5	58.7	63.0
ProtoNet (Snell et al., 2017)	Episodic	43.8	63.6	68.3
DeepEMD (Zhang et al., 2020)	Episodic	47.3	64.6	68.6
SimpleShot (Wang et al., 2019)	Standard	47.9	66.9	71.4
LaplacianShot (Ziko et al., 2020)	Standard	48.5	68.0	72.8
		NCT		
Method	Training Method	5-way 1-shot	5-way 5-shot	5-way 10-shot
MAML (Finn et al., 2017)	Episodic	57.2	65.5	69.2
ProtoNet (Snell et al., 2017)	Episodic	62.6	80.9	84.9
DeepEMD (Zhang et al., 2020)	Episodic	68.5	84.0	86.0
SimpleShot (Wang et al., 2019)	Standard	71.2	85.5	88.2
LaplacianShot (Ziko et al., 2020)	Standard	71.8	86.9	89.5
		LC25000		
Method	Training Method	5-way 1-shot	5way 5-shot	5-way 10-shot
MAML (Finn et al., 2017)	Episodic	58.5	69.2	74.8
ProtoNet (Snell et al., 2017)	Episodic	67.2	84.8	86.2
DeepEMD (Zhang et al., 2020)	Episodic	73.8	85.3	86.4
SimpleShot (Wang et al., 2019)	Standard	66.4	83.6	87.2
LaplacianShot (Ziko et al., 2020)	Standard	67.5	84.2	87.9

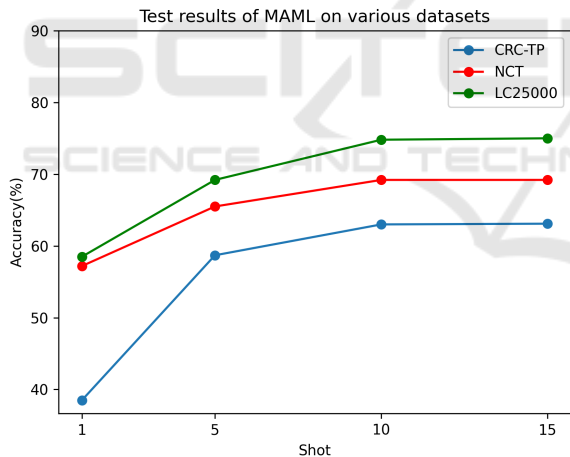


Figure 4: Accuracy plot of MAML on different datasets.

training such as MAML, ProtoNet and DeepEMD. The possible explanation for standard training procedures outperforming episodic training procedures is due to training of these methods on the substantial size of the dataset proposed by Komura et al., which includes cancer image patches from 32 diverse classes or organs. For histopathology images, the results obtained from ProtoNet are comparable to those achieved with DeepEMD. Considering that for natural images, DeepEMD exhibits significantly superior performance compared to ProtoNet. MAML, a few-shot learning-to-learn paradigm designed to learn and

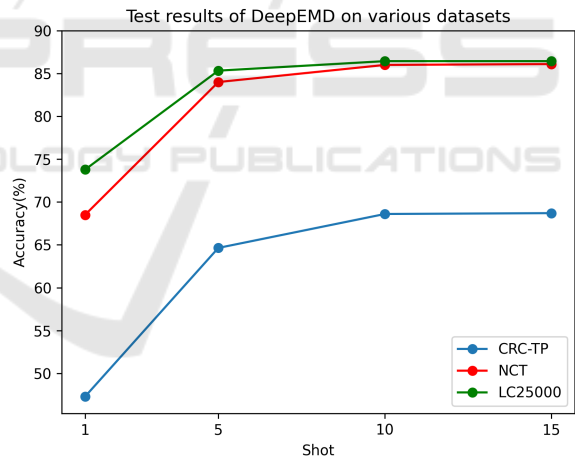


Figure 5: Accuracy plot of DeepEMD on different datasets.

adapt quickly, does not appear to be well-suited for histopathology images. As the number of shots are increased in all few-shot methods, there is an observed tendency for the results to reach a saturation point after 10 shots, as observed from the above plots. In 10-shot test scenarios, the accuracies of all methods, except MAML, across all datasets, excluding CRC-TP, fell within the range of 85% to 90%. In the context of complete training, (Javed et al., 2020a) reported an accuracy of 84.1% on the CRC-TP dataset. Likewise, (Kather et al., 2019) achieved an accuracy of 94.3% on the NCT dataset, and (Sarwinda et al.,

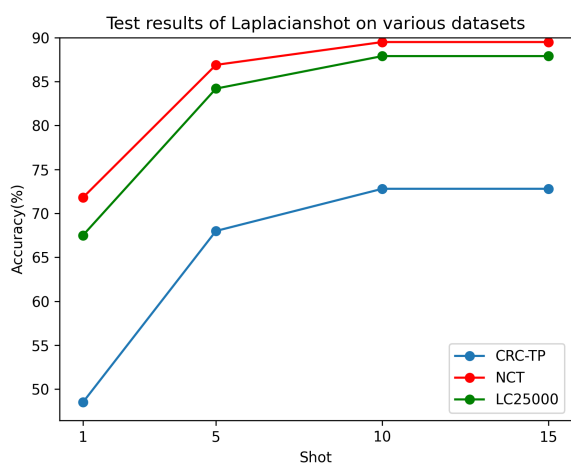


Figure 6: Accuracy plot of Laplacianshot on different datasets.

2020) reported an accuracy of 98.5% on the LC25000 dataset. The highest accuracy attained through the 10-shot method is comparable to that achieved through complete training.

5 CONCLUSION

In conclusion, this research has explored the application of few-shot classification in the domain of histopathology images. Despite the growing prominence of few-shot learning in image classification, its application in histopathology has remained relatively uncharted. This study has specifically addressed the challenges arising from the scarcity of labeled data in medical imaging. Through the utilization of certain state-of-the-art few-shot learning methods, we evaluated their performance across various scenarios within the histology data domain. The consideration of four histopathology datasets has contributed significantly to our broader understanding of few-shot histopathology image classification. The evaluation process, scrutinizing 5-way 1-shot, 5-way 5-shot, and 5-way 10-shot scenarios, employing state-of-the-art classification techniques, has yielded remarkable results. The selected methods demonstrated exceptional capabilities, achieving accuracies close to 90% in the 5-way 10-shot scenarios. The top accuracy obtained through the 10-shot method closely aligns with the accuracy achieved through full training, emphasising the effectiveness of Few-Shot Learning particularly when ample pre-training data is not accessible. These insights significantly contribute to the ongoing discourse surrounding the optimization of classification techniques for medical imaging, particularly in scenarios where labeled data is limited.

Future work in the application of few-shot classification in histopathology images could focus on multi-modal image classification, transfer learning, and pre-training, as well as the development of explainability and interpretability methods. Few-shot learning in multi-modal medical image classification involves training models to accurately classify medical images with limited examples per pathology or disease, leveraging information from various modalities like visual (X-rays or MRIs), textual (medical reports or descriptions), and potentially sensor data. In medical scenarios where obtaining a large labeled dataset for every pathology is challenging, this approach becomes crucial. By combining information from different modalities and addressing the challenges posed by limited examples, the model aims to make accurate predictions in medical image classification tasks, particularly where traditional deep learning models may struggle due to the scarcity of labeled medical data. Additionally, researchers can explore innovative data augmentation techniques and active learning strategies. Improving robustness to image variability, fostering collaboration with domain experts, establishing benchmark datasets, and addressing real-time implementation challenges are also crucial for advancing the field. Overall, these directions aim to enhance diagnostic accuracy and efficiency in histopathology through the effective use of limited labeled data and the integration of machine learning into clinical workflows.

REFERENCES

- Borkowski, A. A., Bui, M. M., Thomas, L. B., Wilson, C. P., DeLand, L. A., and Mastorides, S. M. (2019). Lung and colon cancer histopathological image dataset (lc25000).
- Chen, X., Yao, L., Zhou, T., Dong, J., and Zhang, Y. (2021). Momentum contrastive learning for few-shot covid-19 diagnosis from chest ct images. *Pattern Recognition*, 113:107826.
- Finn, C., Abbeel, P., and Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning (ICML)*.
- Javed, S., Mahmood, A., Fraz, M. M., Koohbanani, N. A., Benes, K., Tsang, Y.-W., Hewitt, K., Epstein, D., Snead, D., and Rajpoot, N. (2020a). Cellular community detection for tissue phenotyping in colorectal cancer histology images. *Medical Image Analysis*, 63:101696.
- Javed, S., Mahmood, A., Werghi, N., Benes, K., and Rajpoot, N. (2020b). Multiplex cellular communities in multi-gigapixel colorectal cancer histology images for tissue phenotyping. *IEEE Transactions on Image Processing*, 29:9204–9219.

- Kather, J. N., Krisam, J., Charoentong, P., Luedde, T., Herpel, E., Weis, C.-A., Gaiser, T., Marx, A., Valous, N. A., Ferber, D., Jansen, L., Reyes-Aldasoro, C. C., Zörnig, I., Jäger, D., Brenner, H., Chang-Claude, J., Hoffmeister, M., and Halama, N. (2019). Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS Med*, 16(1):e1002730.
- Kim, Y. J., Jang, H., Lee, K., Park, S., Min, S.-G., Hong, C., Park, J. H., Lee, K., Kim, J., Hong, W., et al. (2021). Paip 2019: Liver cancer segmentation challenge. *Medical Image Analysis*, 67:101854.
- Komura, D. and Ishikawa, S. (2021). Histology images from uniform tumor regions in tcga whole slide images. *Cell Reports*, 38(9):110424.
- Mahajan, K., Sharma, M., and Vig, L. (2020). Meta-dermdiagnosis: Few-shot skin disease identification using meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 730–731.
- Medela, A., Picon, A., Saratxaga, C. L., Belar, O., Cabezón, V., Cicchi, R., Bilbao, R., and Glover, B. (2019). Few-shot learning in histopathological images: reducing the need of labeled data on biological datasets. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1860–1864. IEEE.
- Nichol, A. and Schulman, J. (2018). Reptile: A scalable metalearning algorithm. *arXiv preprint arXiv:1803.02999*.
- Ouyang, C., Biffi, C., Chen, C., Kart, T., Qiu, H., and Rueckert, D. (2020). Self-supervision with superpixels: Training few-shot medical image segmentation without annotation. In *European Conference on Computer Vision*, pages 762–780. Springer.
- Sarwinda, D., Bustamam, A., Paradisa, R. H., Argyadiva, T., and Mangunwardoyo, W. (2020). Analysis of deep feature extraction for colorectal cancer detection. In *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, pages 1–5. IEEE.
- Shakeri, F., Boudiaf, M., Mohammadi, S., Sheth, I., Havaei, M., Ben Ayed, I., and Kahou, S. E. (2022). Fhist: A benchmark for few-shot classification of histological images. *arXiv*.
- Snell, J., Swersky, K., and Zemel, R. (2017). Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- Spanhol, F. A., Oliveira, L. S., Petitjean, C., and Heutte, L. (2016). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7):1455–1462.
- tcga (2005). The cancer genome atlas. <https://www.cancer.gov/ccg/research/genome-sequencing/tcga>. Accessed on: 2023-11-30.
- Teng, H., Zhang, W., Wei, J., Lv, L., Tang, L., Fu, C.-C., Cai, Y., Qin, G., Ye, M., and Fang, Q., e. a. (2021). Few-shot learning on the diagnosis of lymphatic metastasis of lung carcinoma. *Research Square*.
- Wang, Y., Chao, W.-L., Weinberger, K. Q., and van der Maaten, L. (2019). Simpleshot: Revisiting nearest-neighbor classification for few-shot learning. *arXiv preprint arXiv:1911.04623*.
- Yang, J., Chen, H., Yan, J., Chen, X., and Yao, J. (2022). Towards better understanding and better generalization of few-shot classification in histology images with contrastive learning. *arXiv preprint arXiv:2202.09059*.
- Yu, Q., Dang, K., Tajbakhsh, N., Terzopoulos, D., and Ding, X. (2021). A location-sensitive local prototype network for few-shot medical image segmentation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 262–266. IEEE.
- Zhang, C., Cai, Y., Lin, G., and Shen, C. (2020). Deepemd: Few-shot image classification with differentiable earth mover’s distance and structured classifiers. In *Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Ziko, I. M., Dolz, J., Granger, E., and Ben Ayed, I. (2020). Laplacian regularized few-shot learning. In *International Conference on Machine Learning (ICML)*.