

Magnification Prior: A Self-Supervised Method for Learning Representations on Breast Cancer Histopathological Images

Prakash Chandra Chhipa^{1,*}, Richa Upadhyay¹, Gustav Grund Pihlgren¹, Rajkumar Saini¹, Seiichi Uchida² and Marcus Liwicki¹

¹*Machine Learning Group, EISLAB, Luleå Tekniska Universitet, Luleå, Sweden*
²*Human Interface Laboratory, Kyushu University, Fukuoka, Japan*

*Corresponding author - prakash.chandra.chhipa@ltu.se

Abstract—

This work presents a novel self-supervised pre-training method to learn efficient representations without labels on histopathology medical images utilizing magnification factors. Other state-of-the-art works mainly focus on fully supervised learning approaches that rely heavily on human annotations. However, the scarcity of labeled and unlabeled data is a long-standing challenge in histopathology. Currently, representation learning without labels remains unexplored in the histopathology domain. The proposed method, Magnification Prior Contrastive Similarity (MPCS), enables self-supervised learning of representations without labels on small-scale breast cancer dataset BreakHis by exploiting magnification factor, inductive transfer, and reducing human prior. The proposed method matches fully supervised learning state-of-the-art performance in malignancy classification when only 20% of labels are used in fine-tuning and outperform previous works in fully supervised learning settings for three public breast cancer datasets, including BreakHis. Further, It provides initial support for a hypothesis that reducing human prior leads to efficient representation learning in self-supervision, which will need further investigation. The implementation of this work is made available online on GitHub¹.

Keywords: *self-supervised learning, contrastive learning, representation learning, breast cancer, histopathological images, transfer learning, medical images*

I. INTRODUCTION

Cancer diagnosis by analyzing histopathological whole-slide images (WSI) is an active research field in machine learning [33].

A challenge for the supervised learning approaches applied to histopathological WSI is the scarcity of labeled data. Furthermore, label information for digital WSI is also limited and does not provide details of the affected region at different magnifications, as in dataset BreakHis [52]. Representations learned through supervised learning might suffer as such methods typically require a large amount of labeled data. This can lead to sub-optimal performance on downstream tasks.

Exploring efficient representation learning on small-scale histopathological WSI data using objectives that do not require

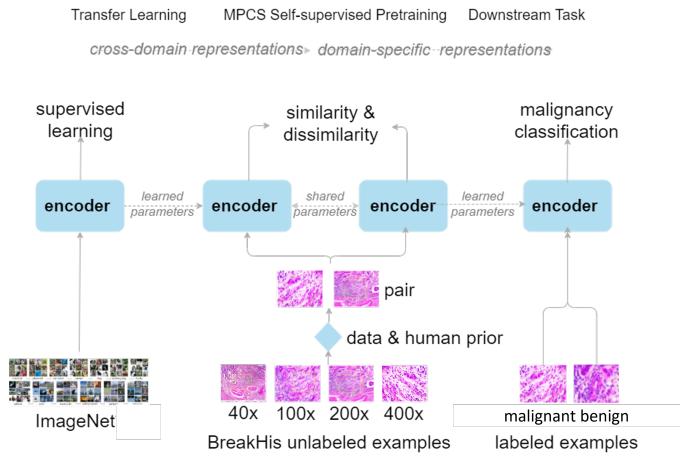


Fig. 1: The proposed approach comprises three steps: (1) Parameters initialization with supervised ImageNet weights. (2) Self-supervised pre-training on unlabeled BreakHis histopathology images using proposed method **Magnification Prior Contrastive Similarity** to provide positive pairs by exploiting supervision signal, e.g., magnification from data and reducing human prior. (3) Fine-tuning on histopathology images for the downstream task.

labels is a promising approach as it requires a lower amount of labeled data needed to learn successful downstream models.

This work proposes a novel self-supervised learning (SSL) method based on contrastive joint embedding called Magnification Prior Contrastive Similarity (MPCS), to learn efficient representations without labels. The proposed method uses a magnification factor (a signal from data) to construct positive pairs for contrastive similarity. MPCS uses magnification factors to enable SSL on the small-scale dataset. This work also hypothesizes that reducing human inducted prior in SSL methods enhances representation learning.

The proposed method MPCS conducts self-supervised pre-training on histopathological WSI on the BreakHis [52] dataset with two different backbone encoders Efficient-net [55] and dilated ResNet-50 [61], pre-trained on ImageNet [15]. The

¹<https://github.com/prakashchhipa/Magnification-Prior-Self-Supervised-Method>

TABLE I: Methodologies Used on the BreakHis Dataset.

Work	Model	Augmented	Extra Training	Ensemble	Evaluation
Deep [51]	AlexNet [35] variant	No	No	No	5 trials
MI [6]	Custom CNN	Simple	Custom	No	5-fold
GLPB [3]	TCNN [1]	Custom	No	No	5-fold
MIL [53]	Various	MIL [31]	No	No	5 trials
A-MIL [42]	Custom CNN	Simple	No	No	Unclear
MRN [24]	ResNet [27] variant	Simple	No	No	Unclear
MV [21]	Various	No	Voting	No	5 trials
SM [22]	DenseNet [29]	Simple	Pretrained	XGBoost [10]	3 trials
PI [23]	ResNet [27]	Simple	Pretrained	Custom	3 trials
TL [16]	AlexNet [35] & VGG16 [48]	No	Pretrained	Two networks	5-fold
RPDB [39]	DenseNet [29]	Simple	AnoGAN [45]	No	5-fold CV
MIM [7]	Various	Simple	Pretrained	No	Unclear
MPCS	Efficient-net b2 [55]	Simple	Custom	No	5-fold stratified CV

effectiveness of the learned representations is assessed by fine-tuning for the downstream task of malignancy classification on multiple public datasets e.g. BACH [2], Breast Cancer Cell Dataset [18] including BreakHis dataset. The complete approach is depicted in Figure 1. Following are the main contributions:

- 1) Enables self-supervised learning on small-scale histopathology dataset by exploiting a data prior
- 2) Hypothesizes a relation between human induced prior and data prior for self-supervised representation learning

With contributions mentioned above, this work demonstrates significantly improved performance on downstream tasks on three public datasets. It provides preliminary empirical support that 1) reducing human-prior leads to efficient representation learning and 2) learning being magnification invariant by cross-magnification evaluation.

II. RELATED WORK

Most efforts in machine learning for histopathological analysis have been using supervised learning. However, current supervised learning methods struggle when labeled data is scarce [33]. Other methods, e.g. pseudo-label and transfer learning, are often used with supervised learning to make up for the lack of labeled data. Examples of such methods are augmentation [37] for the first category and feature-extraction and selection [50], [58] for the second.

One public histopathological image dataset that poses such a challenge of data- and label scarcity is BreakHis [52]. Many different methods have been applied to the BreakHis [52] dataset, most of which utilize a Convolutional Neural Network (CNN) in conjunction with one or both of the methods for handling data scarcity above.

Table I contains a summary of a few such approaches along with some of the strategies used on BreakHis [52]. In the table, custom means a specialized or novel method introduced. Simple augmentations refer to one or more operations among rotation, flipping, cropping, shifting, and zooming. Another breast cancer dataset is BACH [2], on which previous works in transfer learning TL [57], patch-based PT [44], and hybrid networks HN [60] shows performance improvements. Further, the small-scale dataset Breast Cancer Cell Dataset [18] is also being evaluated for binary tasks in various previous works based on shearlet transform ST [43] and attention methods are ATN [30] and MATN [34].

Another learning paradigm to effectively counter earlier stated data scarcity challenges is self-supervised learning. Representation learning from self-supervised learning paradigms for computer vision can be categorized majorly as (i) Joint

Embedding Architecture & Method (JEAM) ([11], [20], [9], [62]), (ii) Prediction Methods ([56], [41], [17]), and loosely (iii) Reconstruction Methods ([32], [19]). Specifically, JEAM can be divided further with each subdivision providing many interesting works; (i) Contrastive Methods (PIRL [40], SimCLR [11], SimCLRV2 [12], MoCo [26]), (ii) Distillation (BYOL [20], SimSiam [13]), (iii) Quantization (SwAV [9], DeepCluster [8]), and (iv) Information Maximization (Barlow Twins [62], VICReg [5]). Of these divisions, this work focuses on contrastive methods.

Recently, contrastive JEAM has been tailored for medical images. Contrastive learning on digital pathology DPCL [14] applies contrastive learning and shows improvement in the breast cancer dataset [2]. In MICLE [4], which is based on SimCLR [11], multiple instance contrastive learning is applied by enabling input views from several image instances of the same patient. Another work from contrastive methods on histopathology is DRL [59]. Other applications making use of contrastive JEAM are chest X-rays [49], [38], CT scans for COVID-19 [28], 3d-Radiomic [36], and Radiograph [63]. A work using contrastive JEAM on the BreakHis dataset is SMSE [54], which trains the network using pair and triplet losses. SSL methods including JEAM require large-size data. Thus applying the contrastive JEAM paradigm on small datasets with the reduced human dependency of prior is an open challenge and interest of this work.

,

III. METHODOLOGY

The primary focus of this work is to introduce a novel self-supervised pre-training method with the aim is to learn representations from data without labels, while using supervision signals from data, e.g., magnification factor and using inductive transfer from ImageNet [15] pre-trained weights.

A. Inductive Transfer Learning

Given the fact that BreakHis [52] is a small-scale and class-imbalanced dataset, this work hypothesizes a constraint case of inductive transfer for representation learning by initializing encoder ImageNet [15] pre-trained weights. In this work, the inductive transfer (i) helps to obtain improved performance on the downstream task of malignancy classification and (ii) enables self-supervised pre-training using the proposed method on the small-scale dataset.

B. Self-supervised Method - Magnification Prior Contrastive Similarity

Magnification Prior Contrastive Similarity (MPCS) method formulates self-supervised pre-training to learn representations on microscopic Histopathology WSI without labels on small-scale data. The main objective of MPSC is to lower the amount of labeled data needed for the downstream task to address challenges in supervised learning.

MPSC construct pairs of distinct views considering characteristics of microscopic histopathology WSI (H-WSI) for contrastive similarity-based pre-training. Microscopic H-WSI

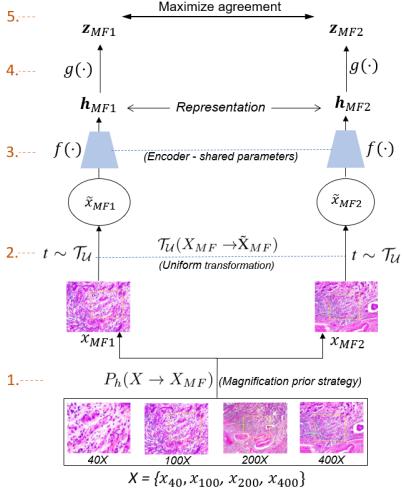


Fig. 2: Magnification Prior Contrastive Similarity method explained

structural properties are different from natural visual macroscopic images [15] (vehicles, cats, or dogs) in terms of location, size, shape, background-foreground, and concrete definition of objects. Unlike SimCLR [11] where pairs of distinct views from the input image is constructed by human-centered augmentations, MPCS constructs pair of distinct views using pair sampling methods based on the signal from data itself i.e. magnification factor in BreakHis [52]. Two H-WSI from different magnification factors of the same sample makes a pair. Utilizing prior from data (magnification factor) enables meaningful contrastive learning on histopathology H-WSI and reduces dependency over human induced prior. Further, tumor-affected regions in H-WSI are characterized by format and highly abnormal amounts of nuclei. Such affected regions are promising in all the H-WSIs of different magnification for the same sample. Thus, affected regions being common and size invariant in positive pair of a sample allow learning contrastive similarity by region attentions.

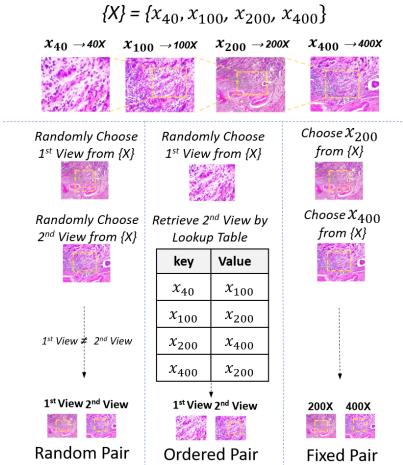


Fig. 3: Strategies for pair sampling based on inducted Human Prior(HP). Added measure to prevent mode collapse in pre-training by $1^{st} view \neq 2^{nd} view$ in all strategies.

Current work also hypothesizes that reduced human-prior

in the pre-training method provides an enhanced degree of freedom to the method that can increase the potential of the network to learn efficient representations in a self-supervised approach. To investigate, three strategies for pair sampling are formulated based on inducted human prior. The number of human decisions defines the level of inducted human-prior (HP) during pair sampling. As explained in Figure 3, in Fixed Pair, decisions to choose magnification factor for both views are by human, thus making strong human-prior. In Ordered Pair, only the second view of the pair is chosen by a human using a look-up table, making weaker human prior. In Random Pair, no human prior inducted and magnification factor for both views are sampled randomly. Further, Figure 4 demonstrates the Degree of Freedom (DoF) for the method where the Fixed Pair strategy provides no DoF, Order Pair provides one DoF, and Random Pair provides 2 DoF to the method.

In MPCS, to formulate a batch of $2N$ views, randomly sampled batch of N sets of input $X = \{X^{(1)}, X^{(2)}, \dots, X^{(N)}\}$ are considered where each set of input $X^{(i)} = \{x_{40}^{(i)}, x_{100}^{(i)}, x_{200}^{(i)}, x_{400}^{(i)}\}$ contains the images corresponding to four magnification factors. Positive pair of views is constructed based on a selection strategy of pair sampling which contains two views from the same example of different magnification. Further, similarity maximizes (loss minimizes) by an objective defined by the contrastive loss in Eq. 1). MPCS is demonstrated in Figure 2 and components are explained below.

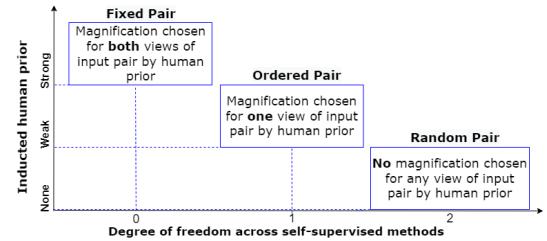


Fig. 4: Relation between inducted Human Prior (HP) of magnification and Degree of Freedom (DoF) for method

- A *domain specific human prior* module $P_h(X \rightarrow X_{MF})$: $X = \{x_{40}, x_{100}, x_{200}, x_{400}\}$ that exploits supervision signal from data i.e. magnification and samples the two views $X_{MF} = (x_{MF1}, x_{MF2})$ of different magnifications to construct pair based on employed strategy of pair sampling shown in step 1 of Figure 2.
- A *uniform stochastic transformation* based module $\mathcal{T}_U(X_{MF} \rightarrow \tilde{X}_{MF})$ that uniformly transforms both views from $X_{MF} = (x_{MF1}, x_{MF2})$ to $\tilde{X}_{MF} = (\tilde{x}_{MF1}, \tilde{x}_{MF2})$ of positive pair by sampled augmentation transformations scheme shown in step 2 of Figure 2.
- A neural network *base encoder* $f(\cdot)$ which yields representations from transformed views of pair. It obtains $h_{MF1} = f(\tilde{x}_{MF1}) = \text{encoder-network}(\tilde{x}_{MF1})$ and $h_{MF2} = f(\tilde{x}_{MF2}) = \text{encoder-network}(\tilde{x}_{MF2})$ where $h_{MF1}, h_{MF2} \in \mathbb{R}^d$ are the output after the respective average pooling layers, shown in step 3 of Figure 2.
- A small-scale *MLP projection head* $g(\cdot)$ that maps representations to the latent space where contrastive loss is applied, shown in step 4 of Figure 2. A multi-layer

- perceptron with single hidden layer to obtain $z_{MF1} = g(h_{MF1}) = W^{(2)}\sigma(W^{(1)}h_{MF1})$ and $z_{MF2} = g(h_{MF2}) = W^{(2)}\sigma(W^{(1)}h_{MF2})$ where σ is ReLU.
- A *contrastive loss function*, normalized temperature-scaled cross entropy loss (NT-Xent) from SimCLR is defined for a contrastive prediction, shown in step 5 of Figure 2. For given a set \tilde{x}_k including a positive pair examples \tilde{x}_{MF1} and \tilde{x}_{MF2} , the contrastive prediction task tends to find \tilde{x}_{MF2} in $\{\tilde{x}_k\}_{k \neq MF1}$ for the given \tilde{x}_{MF1} .

The loss function for a positive pair of examples (MF1, MF2) is defined as

$$L_{MF1,MF2} = -\log \frac{\exp(sim(\mathbf{z}_{MF1}, \mathbf{z}_{MF2})/\tau)}{\sum_{k=1}^{2N} 1_{[k \neq MF1]} \exp(sim(\mathbf{z}_{MF1}, \mathbf{z}_k)/\tau)} \quad (1)$$

In Eq. (1), where $1_{[k \neq MF1]} \in 0, 1$ is an indicator evaluating to 1 if $k \neq i$.

IV. EXPERIMENTAL EVALUATIONS

This section investigates the representation learning capabilities of the proposed method MPCS on two encoder networks with detailed experimentation on three public datasets.

A. Datasets

1) *BreakHis*: BreakHis [52] dataset consists of 2,480 benign and 5,429 malignant histopathological microscopic images from 82 patients at four magnification levels (40x, 100x, 200x, 400x). Each image in the BreakHis dataset is of size 700x460, stained with hematoxylin and eosin (HE). Following the previous works, two evaluation metrics are used, image-level accuracy(ILA) and patient-level accuracy (PLA). PLA shows patient-wise classification performance, calculated as the mean over total no. of patients using patient-score. Patient-score is correctly classified images of the patient over a total number of images of that patient. ILA disregards patient-level details and thus serves as standard image classification accuracy.

2) *BACH*: The second dataset, Breast Cancer Histology Images (BACH) [2] is from the ICIAR2018 Grand Challenge and contains 400 histopathology slides. The BACH dataset has four classes, normal, benign, in-situ, and invasive. The slide size is relatively large, 2048 × 1536 pixels; thus, patches of size 512x512. Two evaluation metrics, patch-wise accuracy and image-wise accuracy, are used, whereas image-wise accuracy is calculated based on majority voting over the patches of the respective image.

3) *Breast Cancer Cell Dataset*: The third dataset, Breast Cancer Cell Dataset [18] is from the University of California, Santa Barbara Biosegmentation Benchmark. This dataset contains 58 HE-stained histopathology 896x768 size images of breast tissue, of which 26 are malignant, and 32 are benign. Patches of size 224x224 were created, and image-wise accuracy was calculated using majority voting over patches of the respective image.

B. Encoder Architectures

In this current work, the proposed method MPCS investigated two different CNN encoder architectures. ResNet-50 [61] and Efficient-net b2 [55] are used for pre-training and fine-tuning. SSL-specific MLP projection head used for Efficient-net b2 is three layers network of 2048-1204-128 units, whereas ResNet-50 is the most common backbone encoder, used projection head adapted from SimCLR, having 1024-128 units. encoder and projection heads are demonstrated in Fig. 2.

C. Training Protocol

This section shares parameter configurations used in pre-training and fine-tuning.

1) *SSL pre-training*: Self-supervised pre-trainings of both encoders take place on the BreakHis dataset for 1000 epochs with temperature parameter 0.01, learning rate 1e-05, and a set of augmentation methods such as color-jitter, flip, and rotation. Efficient-net b2 encoder pre-trained using Adam optimizer with a batch size of 128 and image input of (341, 341). However, ResNet-50 adapted standard configurations of self-supervised practices and pre-trained using the LARS optimizer with a batch size of 1024 and input image size of 224x224.

2) *Fine-tuning*: The common training configurations across datasets for both encoders are learning rate of 2e-05, batch size of 32, image input of 224x224, augmentations methods such as random crop, flip, affine, rotation, and color-jitter, and using adam as optimizer. A dropout of 0.3 is used in the fully connected layer.

D. Experimentation Details

To ensure the reliability and consistency of the models, this work follows 5-cross validation data split scheme. This is applied to all three datasets in which each fold contained 20% data, following class distribution from whole data. Four out of five folds are used for training & validation, and the remaining one for testing. Thus all the results reported are in terms of mean value with standard deviation. In the above-stated 5-cross validation settings, both backbone encoders, ResNet-50 and Efficient-net b2, are being pre-trained on the first dataset BreakHis with all three variants (ordered pair, random pair, and fixed pair) of the proposed SSL method MPCS for learning domain-specific representations. Further, downstream-task-specific fine-tuning experiments are carried out to investigate the impact of learned representations for all three datasets e.g., BreakHis, BACH, and Breast Cancer Cell. Following are the details of experimentation for each dataset. Table II describes fine-tuning experiments for the first dataset BreakHis. All the mentioned experiments for malignancy classification are conducted for both encoders, Efficient-net b2 and ResNet-50 for all four magnifications (40X, 100X, 200X, and 400X). Experiments (Exp-1 to Exp-4) in *Limited Labeled Data Setting* evaluate model performance while using only 20% labels, whereas experiments (Exp-5 to Exp-8) in *Fully Supervised Setting* use all labels. The sole objective is to compare the performance of models pre-trained

TABLE II: Experiment details for BreakHis dataset

No.	Pre-training Method	BreakHis data (%) for SSL	Finetuning on BreakHis Dataset	
			Train (%)	Test (%)
Limited Labeled Data Setting				
Exp-1	ImageNet	-	20%	20%
Exp-2	ImageNet → MPCS-Fixed Pair	60%	20%	20%
Exp-3	ImageNet → MPCS-Ordered Pair	60%	20%	20%
Exp-4	ImageNet → MPCS-Random Pair	60%	20%	20%
Fully Supervised Data Setting				
Exp-5	ImageNet	-	80%	20%
Exp-6	ImageNet → MPCS-Fixed Pair	60%	80%	20%
Exp-7	ImageNet → MPCS-Ordered Pair	60%	80%	20%
Exp-8	ImageNet → MPCS-Random Pair	60%	80%	20%

TABLE III: Experiment details for BACH dataset

No.	Pre-training method/weights	Fine-tuning on BACH dataset	
		Labels(%) from Train Data	Test data (%)
Exp-9	ImageNet [57]-re-implement	100% (80% train data)	20%
Exp-10	MPCS-Fixed Pair (BreakHis)	[5%, 10%, 20%, 40%, 60%, 80%, 100%]	20%
Exp-11	MPCS-Ordered Pair(BreakHis)	[5%, 10%, 20%, 40%, 60%, 80%, 100%]	20%
Exp-12	MPCS-Random Pair(BreakHis)	[5%, 10%, 20%, 40%, 60%, 80%, 100%]	20%

on proposed MPCS methods against ImageNet [15] pre-trained model to analyze the effect of learned representations. Preferred names in discussion used as MPCS-X for ImageNet → MPCS-X. Experiments(Exp-9 to Exp-12) for the second dataset BACH [2] are described in Table III based on the ResNet-50 encoder. All the images are divided into 512X512 size patches; thus, performance is measured patch-wise and image-wise (using majority voting suggested in [57]). The major objectives are 1) evaluating pre-trained models from the proposed method against the ImageNet-based transfer learning approach [2] 2) evaluating the ability to learn downstream tasks with limited labels ranging from 5% to 100% of labels from the train data portion. Finally, to evaluate the effect of learned domain-specific representations on small-scale data, a series of fine-tuning (all layers trained) and linear evaluation experiments (only fully-connected layers trainable) Exp-13 to Exp-16 are conducted Breast Cancer Cell dataset and described in Table IV. Similar to BACH dataset, images for this dataset were also divided into sizes 224X224 for training and test, and performance was measured likewise.

V. RESULTS & DISCUSSIONS

The quantitative results and qualitative analysis from extensive experiments on three datasets validate the efficiency of learned representations for the proposed self-supervised pre-training method MPCS. Preliminary investigation on data prior also supports the hypothesis of reducing human prior in self-supervised representation learning.

Results on BreakHis dataset for the experiments Exp-1 to Exp-4 in Table II in limited labels setting (20% labels only) are described in Table V. It shows that all the variants (ordered pair, random pair, and fixed pair) of MPCS pre-trained models obtain significant (*across magnification, p <0.05*) improvement (1.55 ~ 2.52)% over the ImageNet transfer learning model for all four magnifications and results are competitive with other state-of-the-art methods that have been trained in the fully supervised setting using all labels. Further, Table VI compares the performance of experiments Exp-5 to Exp-8 in Table II in the fully supervised setting (using all labels). The MPCS pre-trained models outperform several state-of-the-art methods with (3.5 ~ 8.0)% higher accuracy in malignancy

TABLE IV: Experiment details for Breast Cancer Cell Dataset

No.	Pre-training method/weights	Fine-tuning on Breast Cancer Cell Dataset		Linear-evaluation on Breast Cancer Cell Dataset	
		Train data(%)	Test data (%)	Train data(%)	Test data(%)
Exp-13	MPCS-Fixed Pair (BreakHis)	80%	20%	80%	20%
Exp-14	MPCS-Ordered Pair(BreakHis)	80%	20%	80%	20%
Exp-15	MPCS-Random Pair(BreakHis)	80%	20%	80%	20%

classification. Both encoder architectures, Efficient-net b2 and ResNet-50 perform consistently. The t-SNE visualization of the pre-trained ResNet-50 encoder in Figure 9 and class activation maps (CAM) [46] in Figure 5 show robust representation learning in the pre-tuning and fine-tuning phases, respectively. Additionally, cross-magnification evaluation is also performed as suggested in previous work [21] and [47], results support performance generalization across magnifications, and results can be found in the supplementary material. MPCS pre-

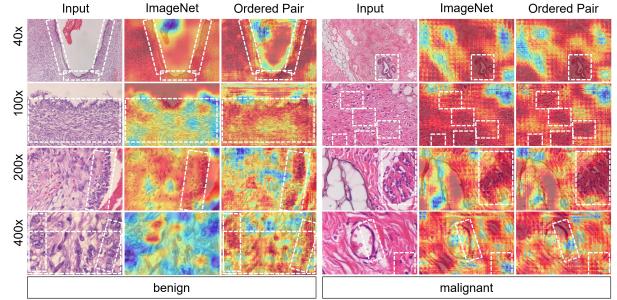


Fig. 5: Comparing CAM from ImageNet and MPCS-OP on BreakHis dataset. Only OP variant CAM is shown for space constraint. Activation - red (darker) color

trained ResNet-50 models on the BreakHis dataset are further evaluated on two additional datasets. Results on the BACH dataset for the experiments Exp-9 to Exp-12 in Table III are described in Table VII and VIII. Results in Table VII show the performance improvement on four-class classification on image-wise and patch-wise accuracy while comparing with other state-of-the-arts for the ResNet-50 encoder and other architectures. Specifically, compared with other recent work based on contrastive learning-based self-supervised method DPCL [14], the proposed method MPCS consistently outperforms over a varying range of label usage as shown in Figure 8 and obtains an improvement of 4.85% when using 100% labels. To compare with ImageNet pre-trained ResNet-50 encoder in identical settings, the method suggested in [57] is re-implemented and results are reported in Table VII for comparison and qualitative analysis through CAM is also shown for samples from the BACH dataset in Figure 6. Results on the third Breast Cell Cancer dataset for experiments Exp-13 to Exp-15 from Table IV are compared with other methods in Table IX, shows significant improvement over other methods and obtains 98.18% in fine-tuning and 96.36% in linear evaluation. CAM in Figure 7 also supports the method qualitatively. Reproducible source code for all the results reported is added in supplementary content and shall be made available on GitHub.

TABLE V: Performance evaluation of the proposed methods in limited labelled data setting when fine-tuning only on 20% labeled data. (across magnification, $p < 0.05$ for all three self-supervised methods for PLA). Ablation results for 40%, 60%, and 80% labels are provided in supplementary content.

Encoder	Metric	Method	40X	100X	200X	400X	Mean
Efficient-net b2	Image Level Accuracy	ImageNet	86.36 \pm 6.13	87.80 \pm 4.15	86.82 \pm 5.74	84.98 \pm 6.05	86.49 \pm 5.63
		FixedPair	87.26 \pm 4.46	87.45 \pm 2.35	89.38 \pm 2.18	88.12 \pm 3.84	88.05 \pm 3.20
		Ordered Pair	87.40\pm3.73	89.30\pm2.74	90.50\pm2.19	88.35\pm3.30	88.89\pm2.99
		Random Pair	86.21 \pm 4.20	89.55 \pm 2.84	89.18 \pm 4.05	87.34 \pm 3.44	88.07 \pm 3.63
	Patient Level Accuracy	ImageNet	86.13 \pm 5.15	87.76 \pm 6.03	85.79 \pm 4.10	85.51 \pm 5.27	86.30 \pm 5.14
		FixedPair	86.90 \pm 4.14	87.64 \pm 3.05	89.60 \pm 3.28	88.26 \pm 3.05	88.10 \pm 3.38
		Ordered Pair	87.19\pm3.20	88.86\pm2.58	90.20\pm3.26	88.96\pm3.22	88.80\pm3.07
		Random Pair	87.17 \pm 3.88	88.36 \pm 2.84	88.58 \pm 4.01	88.66 \pm 3.13	88.19 \pm 3.48
ResNet-50	Image Level Accuracy	ImageNet	87.40 \pm 4.88	86.22 \pm 5.71	86.02 \pm 4.74	85.30 \pm 5.95	86.24 \pm 5.32
		FixedPair	86.69 \pm 3.96	86.94 \pm 3.05	88.76 \pm 2.28	88.81\pm2.77	87.68 \pm 3.01
		Ordered Pair	87.56\pm3.48	88.60\pm3.01	89.77\pm2.19	87.61 \pm 3.48	88.38\pm3.04
		Random Pair	87.06 \pm 3.40	87.96 \pm 3.44	88.55 \pm 3.15	86.64 \pm 2.98	87.55 \pm 3.24
	Patient Level Accuracy	ImageNet	87.10 \pm 4.80	88.06 \pm 5.11	84.19 \pm 4.28	85.01 \pm 5.27	86.09 \pm 4.86
		FixedPair	87.45 \pm 3.96	86.38 \pm 3.12	88.18 \pm 3.00	88.89\pm2.98	87.72 \pm 3.27
		Ordered Pair	87.88\pm2.89	88.21\pm3.21	89.52\pm3.26	87.90 \pm 3.03	88.38\pm3.09
		Random Pair	87.17 \pm 2.98	87.96 \pm 3.02	88.76 \pm 3.55	88.06 \pm 3.00	87.99 \pm 3.14

TABLE VI: Comparison of proposed methods with state-of-the-art methods on downstream task of classifying histopathological images at different magnification factors in fully supervised setting(using 100% train set labels). RN-50 indicates ResNet-50 and Eff-net b2 indicates Efficient-net b2 encoder. Also, OP indicates Ordered Pair, FP indicates Fixed Pair, and RP indicates Random Pair.

Method	Patient Level Accuracy (RR)				Mean	Image Level Accuracy				Mean
	40X	100X	200X	400X		40X	100X	200X	400X	
Original-GLCM[51]	74.7 \pm 1.0	78.6 \pm 2.6	83.4 \pm 3.3	81.7 \pm 3.3	79.60 \pm 2.55	-	-	-	-	-
PFTAS[25]	83.80 \pm 2.0	82.10 \pm 4.9	85.10 \pm 3.1	82.30 \pm 3.8	83.33 \pm 3.45	-	-	-	-	-
MIL-NP[53]	92.1 \pm 5.9	89.1 \pm 5.2	87.2 \pm 4.3	82.7 \pm 3.0	87.77 \pm 4.6	87.8 \pm 5.6	85.6 \pm 4.3	80.8 \pm 2.8	82.9 \pm 4.1	84.28 \pm 4.20
SW[51]	88.6 \pm 5.6	84.5 \pm 2.4	85.3 \pm 3.8	81.7 \pm 4.9	85.02 \pm 4.17	89.6 \pm 6.58	85.0 \pm 4.8	84.0 \pm 3.2	80.8 \pm 3.1	84.85 \pm 4.42
MI[6]	83.08 \pm 2.08	83.17 \pm 3.51	84.63 \pm 2.72	82.10 \pm 4.42	83.25 \pm 3.18	-	-	-	-	-
Deep[50]	84.0 \pm 6.9	83.9 \pm 5.9	86.3 \pm 3.5	82.1 \pm 2.4	84.07 \pm 4.67	84.6 \pm 2.9	84.8 \pm 4.2	84.2 \pm 1.7	81.6 \pm 3.7	83.80 \pm 3.13
MILCNN[53]	86.9 \pm 5.4	85.7 \pm 4.8	85.9 \pm 3.9	83.4 \pm 5.3	85.47 \pm 4.85	86.1 \pm 4.28	83.8 \pm 3.0	80.2 \pm 2.6	80.6 \pm 4.6	82.68 \pm 3.62
GLPB[3]	84.5 \pm 4.2	83.5 \pm 2.0	89.6 \pm 5.0	88.2\pm4.0	86.45 \pm 3.8	82.1 \pm 6.4	81.4 \pm 4.8	88.4 \pm 5.0	87.2 \pm 4.5	84.78 \pm 5.18
RPDB[39]*	92.02 \pm 0.9	90.21 \pm 2.40	81.94 \pm 1.70	80.09 \pm 0.70	88.06 \pm 1.4	94.26\pm3.2	92.71 \pm 0.4	83.90 \pm 2.8	82.74 \pm 1.5	88.40 \pm 1.98
SMSE[54]	87.51 \pm 4.07	89.12 \pm 2.86	90.83 \pm 3.31	87.10 \pm 3.80	88.64 \pm 3.51	-	-	-	-	-
ImageNet (Eff-net b2)	91.91 \pm 4.25	91.93 \pm 4.20	91.46 \pm 5.17	88.10 \pm 3.88	90.85 \pm 4.36	92.12 \pm 4.18	92.66 \pm 4.20	91.83 \pm 4.55	88.35 \pm 5.21	91.24 \pm 4.54
MPCS-FP (Eff-net b2)	92.23 \pm 3.50	92.72 \pm 3.68	91.94 \pm 3.80	88.40 \pm 3.26	91.33 \pm 3.56	92.23 \pm 3.80	93.57\pm3.23	92.23 \pm 2.98	88.40 \pm 3.90	91.61 \pm 3.48
MPCS-OP (Eff-net b2)	92.45\pm3.25	93.47\pm2.98	92.44\pm3.30	89.00 \pm 3.05	91.84\pm3.15	92.67 \pm 3.36	93.63\pm3.38	92.72\pm2.80	88.74\pm3.90	91.94\pm3.36
MPCS-RP (Eff-net b2)	93.26\pm3.48	93.57\pm3.36	92.23\pm3.21	89.57\pm3.79	92.15\pm3.46	93.45\pm3.55	93.38 \pm 2.80	92.28\pm3.49	89.81\pm3.15	92.23\pm3.24
ImageNet (RN-50)	91.46 \pm 4.30	91.24 \pm 5.1	90.72 \pm 4.68	87.90 \pm 4.12	90.33 \pm 4.55	91.83 \pm 5.12	92.23 \pm 4.15	91.61 \pm 4.00	87.88 \pm 4.80	90.89 \pm 4.52
MPCS-FP (RN-50)	91.83 \pm 3.88	92.67 \pm 2.72	91.61 \pm 3.40	89.00 \pm 3.15	91.28 \pm 3.29	92.24 \pm 3.48	92.66 \pm 3.88	91.91 \pm 3.68	88.40 \pm 3.66	91.30 \pm 3.68
MPCS-OP (RN-50)	93.00\pm3.66	93.26 \pm 3.08	92.28\pm2.88	88.74\pm3.60	91.82\pm3.31	93.26\pm3.40	93.45\pm2.89	92.45\pm3.77	89.57\pm2.96	92.18\pm3.26
MPCS-RP (RN-50)	92.72 \pm 3.50	93.57\pm2.88	92.23 \pm 3.90	88.40 \pm 3.05	91.73 \pm 3.33	92.72 \pm 3.38	92.72 \pm 4.02	91.91 \pm 3.21	88.56 \pm 3.89	91.48 \pm 3.66

TABLE VII: Performance comparison of proposed method MPCS (ResNet-50 encoder) on BACH dataset with other state-of-the-arts for four class classification. RN-50 indicates ResNet-50.

Method	Image-wise accuracy		Patch-wise accuracy		Mean
	validation	test	validation	test	
PT [44]	-	90.00	-	77.40	
HN [60] (RN-50)	-	81.60	-	-	
HN [60]	-	91.30	-	82.10	
DPCL [14]	-	87.00	-	-	
ImageNet [57] re-implement	92.40 \pm 2.04	90.50 \pm 2.10	80.56 \pm 3.06	80.00 \pm 2.64	
MPCS-FP	92.50 \pm 1.90	90.55 \pm 2.05	84.25 \pm 1.88	82.79 \pm 2.05	
MPCS-OP	93.31\pm1.85	91.85\pm1.77	83.90\pm1.89	83.13\pm2.00	
MPCS-RP	93.00 \pm 1.88	91.00 \pm 2.32	83.78 \pm 2.09	82.90 \pm 2.10	

A. *Self-supervised method MPCS demonstrates label efficiency*

All three variants of the proposed method MPCS demonstrate label efficiency on downstream tasks. Results in Table V On BreakHis dataset, MPCS fine-tuned models obtain (significant improvement $p < 0.01$) proportionally bigger margins of improvement of ($2.52 \pm 0.02\%$) over ImageNet pre-trained

TABLE VIII: Performance evaluation (image-wise accuracy and f1 score) of proposed method MPCS on BACH dataset for limited label range for ResNet-50 encoder.

Label(%) from train data	F1 score (test data)			Accuracy (test data)		
	MPCS-FP	MPCS-OP	MPCS-RP	MPCS-FP	MPCS-OP	MPCS-RP
5%	0.50 \pm 0.05	0.50 \pm 0.05	0.51 \pm 0.05	51.25 \pm 5.02	50.00 \pm 4.80	53.00 \pm 5.06
10%	0.60 \pm 0.05	0.61 \pm 0.04	0.61 \pm 0.05	60.50 \pm 4.88	61.75 \pm 3.93	62.50 \pm 4.60
20%	0.65 \pm 0.03	0.70 \pm 0.04	0.68 \pm 0.02	69.25 \pm 2.92	71.00 \pm 3.90	69.00 \pm 2.89
40%	0.79 \pm 0.04	0.81 \pm 0.04	0.80 \pm 0.03	80.75 \pm 3.40	81.75 \pm 3.52	81.50 \pm 3.05
60%	0.87 \pm 0.03	0.87 \pm 0.03	0.86 \pm 0.03	87.70 \pm 3.48	87.75 \pm 3.10	86.50 \pm 3.00
80%	0.89 \pm 0.03	0.90 \pm 0.02	0.89 \pm 0.02	89.25 \pm 3.02	90.75 \pm 2.17	89.75 \pm 0.02
100%	0.90 \pm 0.02	0.91 \pm 0.02	0.90 \pm 0.02	90.55 \pm 2.05	91.85 \pm 1.77	91.00 \pm 2.32

TABLE IX: Performance measure of proposed method MPCS (ResNet-50 encoder) on Breast Cancer Cell Dataset

Method	Fine-tuned			Linear-evaluation		
	accuracy	precision	recall	accuracy	precision	recall
ST [43]	86.00 \pm 3.00	-	1.0	-	-	-
MATN [34]	91.70	-	-	-	-	-
ATN [30]	75.50 \pm 1.60	0.73 \pm 0.01	0.73 \pm 0.04	-	-	-
MPCS-FP	98.14 \pm 2.05	0.99 \pm 0.01	0.98 \pm 0.01	96.29 \pm 1.90	0.97 \pm 0.01	0.96 \pm 0.01
MPCS-OP	98.18\pm1.80	0.99\pm0.01	0.98 \pm 0.01	96.36\pm1.88	0.97\pm0.01	0.96\pm0.01
MPCS-RP	98.10 \pm 2.00	0.985 \pm 0.01	0.98 \pm 0.01	96.22 \pm 2.02	0.965 \pm 0.01	0.96 \pm 0.01

model while only 20 % labels are used for all magnification scales. These results match the performance of other state-

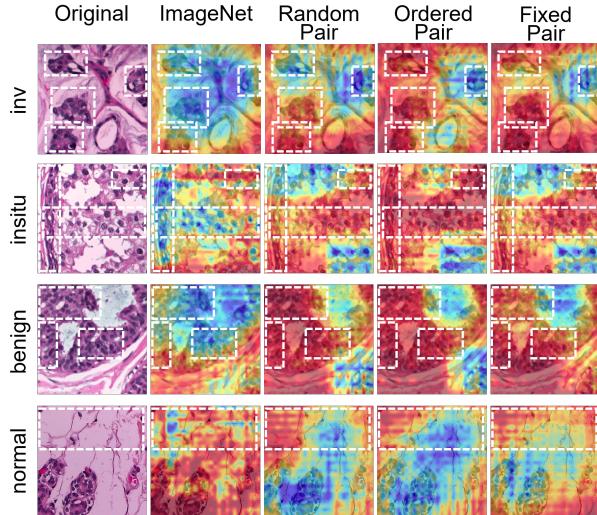


Fig. 6: CAM from ImageNet and MPCS-RP, OP, FP on BACH dataset (MPCS specifically learns to prevent activation on cells present on normal class images).

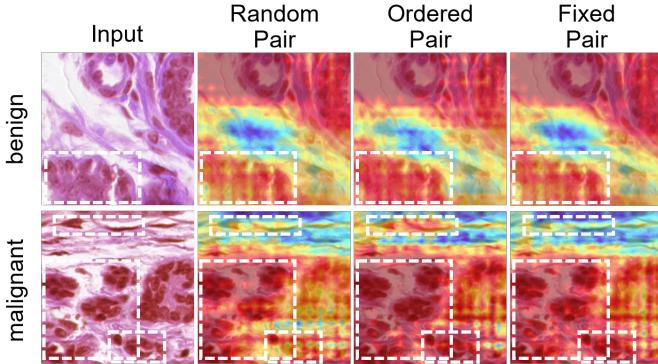


Fig. 7: Comparing CAM among MPCS-RP, OP, and FP on Breast Cancer Cell dataset. Activation - red (darker) color

of-the-art methods mentioned in Table VI which have been trained on 100% labels. Following the trend, MPCS pre-trained models consistently outperform on the BACH dataset compared with the recent contrastive learning-based method DPCL [14] in a complete range of labels from 5% to 100%, shown in Figure 8.

B. Data prior enables self-supervision on small-scale datasets

The proposed MPCS method enables self-supervised representation learning on the small-scale dataset by extensively using a data prior (supervision signal from data) e.g. magnification factors (40X, 100X, 200X, and 400x). It decreases the dependence on human-curated priors e.g. augmentation method choices during self-supervised pre-training.

C. MPCS learns robust self-supervised representations

Explicitly focusing on the robustness of learned representations, Figure 9 strongly supports the fact that MPCS pre-trained models can capture and learn discriminative features across the classes during the self-supervised pre-training phase itself without knowing about actual human-provided labels. It is worth mentioning that data points of different classes are

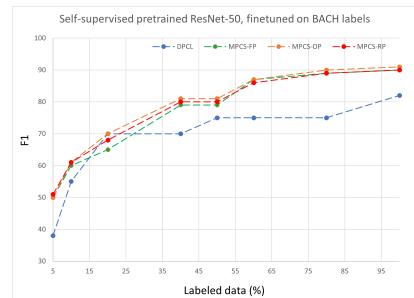


Fig. 8: Comparison with DPCL [14] for label efficiency on classification task on BACH dataset

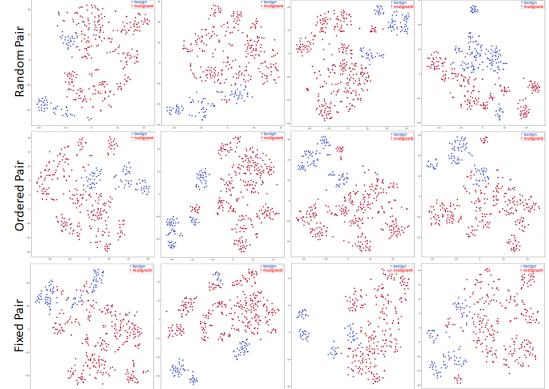


Fig. 9: tsne-visualization of the features from MPCS model (on BreakHis) without fine-tuning. Blue(benign); red(malignant).

easily separable by either linear or non-linear boundaries for all four magnifications for all variants of the MPCS method. The CAM of fine-tuned models depicted in Figures 5 and 6 also indicate that MPCS pre-trained models activate regions of interest (dark red colors show strong activation) more efficiently than the ImageNet pre-trained model for BreakHis and BACH datasets.

D. Preliminary support for the hypothesis about reducing human induced priors

The MPCS-Ordered Pair inducts weaker human-prior in pair sampling. Thus, the MPCS method obtains one DoF for randomly choosing the first input view. In comparison, the MPCS-Fixed Pair, which inducts stronger human prior by choosing both views, 200x and 400x by human-prior, gives zero DoF to the MPCS method. The MPCS-Random Pair, in which the MPCS method obtains the highest degree of freedom since the human-prior is absent. Figure 10 explains the trends when fewer labels are used, that both encoders that weaker human-prior based pair sampling tends to outperform over extreme cases of stronger human prior or the absence of it. However, it requires detailed explorations of different datasets and tasks. Similar patterns were also observed on the BACH dataset that the ordered pair outperforms over other two variants.

VI. CONCLUSIONS

The novel MPCS method enables self-supervised pre-training for comparatively small-scale breast cancer micro-

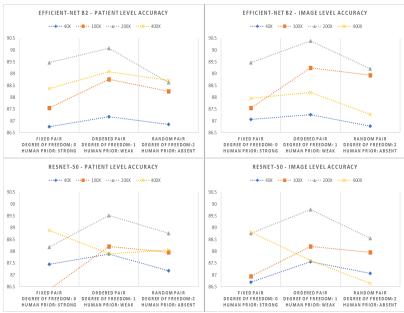


Fig. 10: Comparison for human priors - Indicates weaker human-prior(moderate DoF for method) outperforms stronger and no human-prior during limited (20% labels) data

scopic image dataset BreakHis [52] for efficient representation learning by exploiting supervision signals from data. Results on three public datasets have confirmed the excellence of the MPCS method. MPCS learns better representations by exploiting magnification priors (OP is best among all) and distinguishing among different cells in downstream tasks. It could be better to investigate on a continuous scale of magnification scales considered in this paper. In the future, we intend to investigate magnification prior to other self-supervised learning approaches focusing on redundancy reduction and Siamese networks.

REFERENCES

- [1] Vincent Andreadczyk and Paul F Whelan. Using filter banks in convolutional neural networks for texture classification. *Pattern Recognition Letters*, 84:63–69, 2016.
- [2] Guilherme Aresta, Teresa Araújo, Scotty Kwok, Sai Saketh Chennamsetty, Mohammed Safwan, Varghese Alex, Bahram Marami, Marcel Prastawa, Monica Chan, Michael Donovan, et al. Bach: Grand challenge on breast cancer histology images. *Medical image analysis*, 56:122–139, 2019.
- [3] Steve Tsham Mpinda Ataky, Jonathan de Matos, Alceu de S Britto, Luiz ES Oliveira, and Alessandro L Koerich. Data augmentation for histopathological images based on gaussian-laplacian pyramid blending. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020.
- [4] Shekoofeh Azizi, Basil Mustafa, Fiona Ryan, Zachary Beaver, Jan Freyberg, Jonathan Deaton, Aaron Loh, Alan Karthikesalingam, Simon Kornblith, Ting Chen, et al. Big self-supervised models advance medical image classification. *arXiv preprint arXiv:2101.05224*, 2021.
- [5] Adrien Bardes, Jean Ponce, and Yann LeCun. Vicreg: Variance-invariance-covariance regularization for self-supervised learning. *arXiv preprint arXiv:2105.04906*, 2021.
- [6] Neslihan Bayramoglu, Juho Kannala, and Janne Heikkilä. Deep learning for magnification independent breast cancer histopathology image classification. In *2016 23rd International conference on pattern recognition (ICPR)*, pages 2440–2445. IEEE, 2016.
- [7] Yassir Benhammou, Boujemaa Achhab, Francisco Herrera, and Siham Tabik. Breakhis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights. *Neurocomputing*, 375:9–24, 2020.
- [8] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 132–149, 2018.
- [9] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *arXiv preprint arXiv:2006.09882*, 2020.
- [10] Tianqi Chen and Carlos Guestrin. Xgboost: Reliable large-scale tree boosting system. In *Proceedings of the 22nd SIGKDD Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA*, pages 13–17, 2015.
- [11] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR, 13–18 Jul 2020.
- [12] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. *Advances in Neural Information Processing Systems*, 33:22243–22255, 2020.
- [13] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15750–15758, 2021.
- [14] Ozan Ciga, Tony Xu, and Anne Louise Martel. Self supervised contrastive learning for digital histopathology. *Machine Learning with Applications*, 7:100198, 2022.
- [15] 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. Ieee, 2009.
- [16] Erkan Deniz, Abdulkadir Şengür, Zehra Kadiroğlu, Yanhui Guo, Varun Bajaj, and Ümit Budak. Transfer learning based histopathologic image classification for breast cancer detection. *Health information science and systems*, 6(1):1–7, 2018.
- [17] Carl Doersch, Abhinav Gupta, and Alexei A Efros. Unsupervised visual representation learning by context prediction. In *Proceedings of the IEEE international conference on computer vision*, pages 1422–1430, 2015.
- [18] Elisa Drelie Gelasca, Jiyun Byun, Boguslaw Obara, and BS Manjunath. Evaluation and benchmark for biological image segmentation. In *2008 15th IEEE International Conference on Image Processing*, pages 1816–1819. IEEE, 2008.
- [19] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
- [20] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. Bootstrap Your Own Latent: A new approach to self-supervised learning. In *Neural Information Processing Systems*, Montréal, Canada, 2020.
- [21] Vibha Gupta and Arnav Bhavsar. Breast cancer histopathological image classification: is magnification important? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 17–24, 2017.
- [22] Vibha Gupta and Arnav Bhavsar. Sequential modeling of deep features for breast cancer histopathological image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2254–2261, 2018.
- [23] Vibha Gupta and Arnav Bhavsar. Partially-independent framework for breast cancer histopathological image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.
- [24] Varun Gupta, Megha Vasudev, Amit Doegar, and Nitigya Sambyal. Breast cancer detection from histopathology images using modified residual neural networks. *Biocybernetics and Biomedical Engineering*, 41(4):1272–1287, 2021.
- [25] Nicholas A Hamilton, Radosav S Pantelic, Kelly Hanson, and Rohan D Teasdale. Fast automated cell phenotype image classification. *BMC bioinformatics*, 8(1):1–8, 2007.
- [26] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [27] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [28] Xuehai He, Xingyi Yang, Shanghang Zhang, Jinyu Zhao, Yichen Zhang, Eric Xing, and Pengtao Xie. Sample-efficient deep learning for covid-19 diagnosis based on ct scans. *medrxiv*, 2020.
- [29] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*,

- pages 4700–4708, 2017.
- [30] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018.
- [31] James Keeler, David Rumelhart, and Wee Leow. Integrated segmentation and recognition of hand-printed numerals. In R. P. Lippmann, J. Moody, and D. Touretzky, editors, *Advances in Neural Information Processing Systems*, volume 3. Morgan-Kaufmann, 1991.
- [32] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [33] Daisuke Komura and Shumpei Ishikawa. Machine learning methods for histopathological image analysis. *Computational and Structural Biotechnology Journal*, 16:34–42, 2018.
- [34] Andrei V Konstantinov and Lev V Utkin. Multi-attention multiple instance learning. *Neural Computing and Applications*, pages 1–23, 2022.
- [35] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.
- [36] Hongwei Li, Fei-Fei Xue, Krishna Chaitanya, Shengda Liu, Ivan Ezhov, Benedikt Wiestler, Jianguo Zhang, and Bjoern Menze. Imbalance-aware self-supervised learning for 3d radiomic representations. *arXiv preprint arXiv:2103.04167*, 2021.
- [37] Huangjing Lin, Hao Chen, Qi Dou, Liansheng Wang, Jing Qin, and Pheng-Ann Heng. Scannet: A fast and dense scanning framework for metastatic breast cancer detection from whole-slide image. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 539–546, 2018.
- [38] Jingyu Liu, Gangming Zhao, Yu Fei, Ming Zhang, Yizhou Wang, and Yizhou Yu. Align, attend and locate: Chest x-ray diagnosis via contrast induced attention network with limited supervision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10632–10641, 2019.
- [39] Rui Man, Ping Yang, and Bowen Xu. Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks. *IEEE Access*, 8:155362–155377, 2020.
- [40] Ishan Misra and Laurens van der Maaten. Self-supervised learning of pretext-invariant representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6707–6717, 2020.
- [41] Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In *European conference on computer vision*, pages 69–84. Springer, 2016.
- [42] Abhijeet Patil, Dipesh Tamboli, Swati Meena, Deepak Anand, and Amit Sethi. Breast cancer histopathology image classification and localization using multiple instance learning. In *2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, pages 1–4. IEEE, 2019.
- [43] Hadi Rezaeilouyeh, Ali Mollahosseini, and Mohammad H Mahoor. Microscopic medical image classification framework via deep learning and shearlet transform. *Journal of Medical Imaging*, 3(4):044501, 2016.
- [44] Kaushiki Roy, Debapriya Banik, Debotosh Bhattacharjee, and Mita Nasipuri. Patch-based system for classification of breast histology images using deep learning. *Computerized Medical Imaging and Graphics*, 71:90–103, 2019.
- [45] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, pages 146–157. Springer, 2017.
- [46] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.
- [47] Milad Sikaroudi, Benyamin Ghojogh, Fakhri Karray, Mark Crowley, and Hamid R Tizhoosh. Magnification generalization for histopathology image embedding. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1864–1868. IEEE, 2021.
- [48] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.
- [49] Hari Sowrirajan, Jingbo Yang, Andrew Y Ng, and Pranav Rajpurkar. Moco pretraining improves representation and transferability of chest x-ray models. In *Medical Imaging with Deep Learning*, pages 728–744. PMLR, 2021.
- [50] Fabio A. Spanhol, Luiz S. Oliveira, Paulo R. Cavalin, Caroline Petitjean, and Laurent Heutte. Deep features for breast cancer histopathological image classification. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1868–1873, 2017.
- [51] Fabio Alexandre Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. Breast cancer histopathological image classification using convolutional neural networks. In *2016 international joint conference on neural networks (IJCNN)*, pages 2560–2567. IEEE, 2016.
- [52] 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.
- [53] PJ Sudharshan, Caroline Petitjean, Fabio Spanhol, Luiz Eduardo Oliveira, Laurent Heutte, and Paul Honeine. Multiple instance learning for histopathological breast cancer image classification. *Expert Systems with Applications*, 117:103–111, 2019.
- [54] Yibao Sun, Xingru Huang, Yaqi Wang, Huiyu Zhou, and Qianyi Zhang. Magnification-independent histopathological image classification with similarity-based multi-scale embeddings, 2021.
- [55] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.
- [56] Bastiaan S Veling, Jasper Linnmans, Jim Winkens, Taco Cohen, and Max Welling. Rotation equivariant cnns for digital pathology. In *International Conference on Medical image computing and computer-assisted intervention*, pages 210–218. Springer, 2018.
- [57] Sulaiman Vesal, Nishant Ravikumar, AmirAbbas Davari, Stephan Ellmann, and Andreas Maier. Classification of breast cancer histology images using transfer learning. In *International conference image analysis and recognition*, pages 812–819. Springer, 2018.
- [58] Yongjun Wang, Baiying Lei, Ahmed Elazab, Ee-Leng Tan, Wei Wang, Fanglin Huang, Xuehao Gong, and Tianfu Wang. Breast cancer image classification via multi-network features and dual-network orthogonal low-rank learning. *IEEE Access*, 8:27779–27792, 2020.
- [59] Jilan Xu, Junlin Hou, Yuejie Zhang, Rui Feng, Chunyang Ruan, Tao Zhang, and Weiguo Fan. Data-efficient histopathology image analysis with deformation representation learning. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 857–864, 2020.
- [60] Rui Yan, Fei Ren, Zihao Wang, Lihua Wang, Tong Zhang, Yudong Liu, Xiaosong Rao, Chunhui Zheng, and Fa Zhang. Breast cancer histopathological image classification using a hybrid deep neural network. *Methods*, 173:52–60, 2020.
- [61] Fisher Yu, Vladlen Koltun, and Thomas Funkhouser. Dilated residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 472–480, 2017.
- [62] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. *arXiv preprint arXiv:2103.03230*, 2021.
- [63] Hong-Yu Zhou, Shuang Yu, Cheng Bian, Yifan Hu, Kai Ma, and Yefeng Zheng. Comparing to learn: Surpassing imagenet pretraining on radiographs by comparing image representations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 398–407. Springer, 2020.

VII. SUPPLEMENTARY MATERIAL

A. Self-supervised methods learn magnification invariant representations

The MPCS methods not only outperform for magnification-specific tasks stated in Tables V and VI but representations learned through the proposed method also demonstrate a consistent edge in classification performance in cross-magnification evaluation over the ImageNet model. These experiments were conducted on the Efficient-net b2 encoder only to prevent additional computation usage. However, the ResNet-50 encoder can be benchmarked, if needed. Table X

TABLE X: Type 1 cross magnification performance comparison of proposed methods (Leave one magnification out on which model trained). The values represent mean performance of remaining magnifications (e.g. train on 40x and evaluated (mean) on 100x, 200x, and 400x.)

Method	Trained on 40X	Trained on 100X	Trained on 200X	Trained on 400X
	Mean Cross-Magnification Image Level Accuracy			
ImageNet	79.56±11.74	79.56±11.74	82.97±6.77	84.16±4.98
MPCS-Ordered Pair	80.99±8.91	80.99±8.91	84.40±3.81	84.20±5.58
MPCS-Random Pair	80.13±9.60	80.13±9.60	84.84±5.30	84.83±5.30
Mean Cross-Magnification Patient Level Accuracy				
ImageNet	81.01±9.59	84.38±5.78	81.45±6.89	83.31±7.31
MPCS-Ordered Pair	81.49±6.96	84.02±5.78	82.19±3.83	83.10±7.44
MPCS-Random Pair	81.20±7.09	85.05±7.15	82.97±5.84	83.76±6.63

TABLE XI: Type 2 cross magnification performance comparison of proposed methods (select one magnification in on which model was not trained). The values represent mean performance of a magnification whereas trained on other magnifications (e.g. evaluated on 40x and trained on 100x, 200x, and 400x).

Method	Evaluated on 40X	Evaluated on 100X	Evaluated on 200X	Evaluated on 400X
	Mean Cross-Magnification Image Level Accuracy			
ImageNet	84.33 ± 4.32	83.65 ± 6.07	84.31 ± 8.93	78.86 ± 10.62
MPCS-Ordered Pair	85.35 ± 4.29	84.56 ± 6.06	85.28 ± 8.20	78.98 ± 7.82
MPCS-Random Pair	86.55 ± 5.10	84.82 ± 5.54	84.99 ± 7.79	79.17 ± 9.60
Mean Cross-Magnification Patient Level Accuracy				
ImageNet	81.71 ± 6.25	83.76 ± 6.82	84.35 ± 8.40	80.33 ± 8.10
MPCS-Ordered Pair	81.97 ± 6.25	84.54 ± 6.52	84.63 ± 7.80	79.51 ± 5.20
MPCS-Random Pair	83.24 ± 6.89	84.98 ± 6.12	84.84 ± 7.04	79.92 ± 6.53

shows (type-1 mean cross magnification evaluation) mean cross-magnification accuracy where the model is evaluated on other magnifications except the magnification on which the model was trained. The MPCS-Order Pair methods outperform the ImageNet model and other methods with a mean cross-magnification ILA of 80.99% and PLA of 81.49% when the model was trained on 40x magnification and evaluated on 100x, 200x, and 400x. Whereas MPCS-Random Pair outperforms with mean cross-magnification ILA 84.84% and PLA 82.97% when trained on 200x and ILA 84.83% and PLA 83.76% when trained on 400x. For 100x, MPCS-Ordered Pair obtains ILA 80.99%, and MPCS-Random Pair obtains PLA 85.05%. Further, Table XI evaluates the mean performance of models trained on other magnifications except on which evaluation is performed (type-2 mean cross magnification

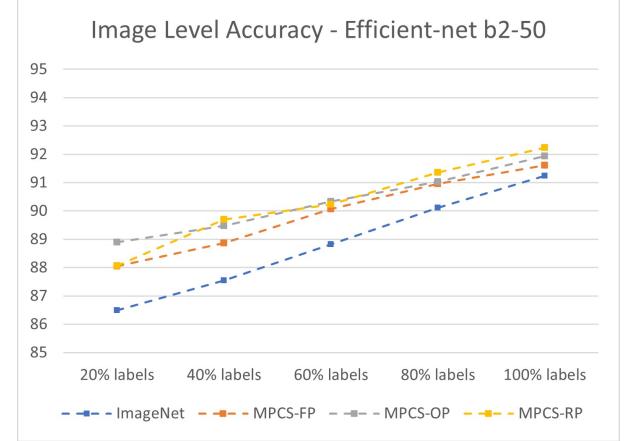


Fig. 11: Performance comparison (ILA accuracy, Efficient-net b2 model) for MPCS pre-trained models with ImageNet pre-trained model over range of labels used.

evaluation). Interestingly, type-2 cross magnification evaluation also shows similar trends except in 400x, in which the ImageNet model obtained high PLA performance. Empirical analysis on type-1 and type-2 cross magnification suggests that MPCS self-supervised pre-trained models perform better than the ImageNet model by learning magnification invariant representations.

B. Additional ablation results on using 80%, 60%, and 40% of labels of the training set on BreakHis dataset

This section describes the extended ablation study on using labels in an incremental manner. The main section of the results describes and compares all three variants of the MPCS method with ImageNet pre-trained models when fine-tuned on 20% and 100% labels of the training set are used. To continue the trend for completeness of analysis, this section adds the results for the same setting considering 40%, 60%, and 80% label utilization in fine-tuning, results described in Tables XII, XIII, and XIV, respectively. The most important observation is that MPCS methods consistently outperform the ImageNet model over the range of labels provided and specifically, the ordered pair method remains best performing in largely. Figure 11 and 12 shows the comparisons for Efficient-net b2 encoder for ILA and PLA accuracy. Similarly, Figure 13 and 14 shows the comparisons for ResNet-50 encoder for ILA and PLA accuracy. A common trend is evident that MPCS methods based models consistently performs better than ImageNet based model for entire range of labels.

It clearly shows that self-supervised learned representations improve fine-tuning task performance overall range of available labels, similar to the trend observed on the BACH dataset. Besides being able to obtain relatively higher accuracy on limited label settings, more label additions are largely beneficial to self-supervised pre-trained models than ImageNet pre-trained models.

TABLE XII: Performance evaluation of the proposed methods in limited labelled data setting when fine-tuning on 80% labels of train set.

Method	Patient Level Accuracy (RR)				Mean	Image Level Accuracy				Mean
	40X	100X	200X	400X		40X	100X	200X	400X	
ImageNet (Eff-net b2)	90.48±3.45	90.53±3.60	90.36±4.67	87.30±3.88	89.66±3.90	91.32±3.48	91.46±3.40	90.34±3.55	87.32±4.76	90.11±3.79
MPCS-FP (Eff-net b2)	91.26±3.64	91.54±3.68	90.84±2.66	88.32±3.23	90.49±3.30	91.83±3.48	92.34±3.20	91.33±2.45	88.30±3.50	90.95±3.15
MPCS-OP (Eff-net b2)	92.05±3.17	92.73±2.45	91.54±2.80	88.98±3.42	91.33±2.96	91.67±3.53	92.45±2.80	91.45±3.60	88.55±3.40	91.03±3.33
MPCS-RP (Eff-net b2)	92.16±4.18	91.51±3.18	91.13±2.81	89.16±2.43	91.00±3.15	92.25±3.61	92.48±3.20	92.01±3.49	88.71±3.15	91.36±3.79
ImageNet (RN-50)	90.16±3.40	90.01±4.1	89.32±3.54	87.00±4.45	89.12±3.87	90.53±5.12	91.03±3.23	90.31±3.59	86.48±4.92	89.58±4.21
MPCS-FP (RN-50)	90.44±3.43	91.32±2.54	90.43±3.32	88.95±2.10	90.29±2.85	90.14±3.48	91.00±3.88	90.71±3.24	88.30±2.46	90.04±3.27
MPCS-OP (RN-50)	92.00±2.42	92.16±3.08	91.15±2.88	88.30±3.30	90.90±2.92	92.16±2.80	92.15±2.89	91.05±2.43	88.50±2.45	90.00±2.64
MPCS-RP (RN-50)	91.02±3.56	91.20±3.28	91.10±3.88	88.20±3.17	90.38±2.92	91.02±2.44	91.22±3.13	90.02±3.65	87.06±3.84	89.80±3.27

TABLE XIII: Performance evaluation of the proposed methods in limited labelled data setting when fine-tuning on 60% labels of train set.

Method	Patient Level Accuracy (RR)				Mean	Image Level Accuracy				Mean
	40X	100X	200X	400X		40X	100X	200X	400X	
ImageNet (Eff-net b2)	89.28±3.56	89.03±3.40	89.22±3.12	86.10±3.32	88.40±3.35	90.02±3.54	90.16±3.53	89.11±3.21	86.00±3.52	88.82±3.45
MPCS-FP (Eff-net b2)	90.22±3.87	90.45±3.20	89.72±2.70	88.32±3.21	89.67±3.25	90.56±3.32	91.28±3.43	90.20±2.55	88.20±3.40	90.06±3.30
MPCS-OP (Eff-net b2)	91.10±3.21	91.54±2.66	90.44±2.65	88.96±3.55	90.51±3.02	90.50±3.22	91.74±2.66	90.63±3.41	88.49±3.91	90.34±3.30
MPCS-RP (Eff-net b2)	91.00±4.18	90.30±3.22	90.10±2.61	89.00±2.55	90.10±3.14	91.15±3.43	91.26±3.11	91.01±3.26	87.53±3.02	90.23±3.21
ImageNet (RN-50)	88.90±3.44	89.05±3.3	88.04±3.23	86.04±3.24	88.00±3.30	89.01±4.10	89.65±3.55	89.02±3.44	86.00±3.12	88.42±3.55
MPCS-FP (RN-50)	89.23±3.21	90.11±2.32	89.30±3.12	88.90±3.15	89.39±2.95	89.05±3.33	90.05±3.55	89.56±3.65	88.21±3.56	89.21±3.52
MPCS-OP (RN-50)	91.02±2.54	91.08±3.18	90.05±2.78	88.21±3.22	90.09±3.22	91.05±3.66	91.01±2.70	90.10±2.32	88.00±2.60	90.04±2.82
MPCS-RP (RN-50)	90.01±3.22	90.10±3.28	90.21±3.56	88.22±3.20	89.64±3.32	90.05±2.32	90.02±3.16	89.01±3.55	86.90±3.21	89.00±3.06

TABLE XIV: Performance evaluation of the proposed methods in limited labelled data setting when fine-tuning on 40% labels of train set.

Method	Patient Level Accuracy (RR)				Mean	Image Level Accuracy				Mean
	40X	100X	200X	400X		40X	100X	200X	400X	
ImageNet (Eff-net b2)	87.45±3.04	88.00±3.22	87.52±3.56	85.95±3.11	87.03±3.23	88.05±3.11	89.14±3.22	88.00±3.19	85.02±3.00	87.55±3.13
MPCS-FP (Eff-net b2)	88.66±3.62	88.90±3.10	89.66±2.03	88.30±3.12	88.88±2.98	88.02±3.72	89.28±3.04	90.00±2.16	88.15±3.06	88.86±3.00
MPCS-OP (Eff-net b2)	89.75±3.56	90.63±2.66	90.32±3.11	88.96±3.55	89.92±3.22	88.82±3.22	90.04±2.89	90.60±3.43	88.40±3.89	89.47±3.36
MPCS-RP (Eff-net b2)	89.05±3.27	90.05±3.42	89.80±2.82	88.70±2.65	89.40±3.04	90.25±3.21	91.00±3.36	90.05±3.11	87.40±3.45	89.70±2.53
ImageNet (RN-50)	88.01±3.65	88.50±3.43	87.09±3.62	85.88±3.01	87.37±3.43	88.10±3.23	88.60±3.53	88.02±3.00	85.60±3.22	87.78±3.25
MPCS-FP (RN-50)	88.20±3.61	88.17±2.62	89.00±3.81	88.90±3.56	88.58±3.4	87.90±3.25	88.10±3.21	88.50±3.61	88.10±3.56	88.15±3.41
MPCS-OP (RN-50)	89.42±2.13	90.28±3.64	89.92±2.18	88.02±3.22	89.41±2.79	89.10±3.05	90.00±2.85	89.90±2.52	87.80±2.65	89.20±2.78
MPCS-RP (RN-50)	89.01±3.32	89.10±3.82	90.00±3.46	88.15±3.26	89.06±3.47	88.10±2.35	89.30±3.20	88.90±3.05	86.85±3.22	88.29±2.96

TABLE XV: Experimentation statistics for proposed method MPCS

Dataset	Experiment Training Type	No. of Experiments
BreakHis	SSL pretrain - Efficient-net b2	15
	SSL pretrain - ResNet-50	18
	Downstream task- Efficient-net b2	400
	Downstream task- ResNet-50	400
BACH	Downstream task- ResNet-50	140
Breast Cell Cancer Dataset	Downstream task- ResNet-50	30
Total		1003

C. Experimentation statistics

An extensive experimentation strategy was designed, and experiments were performed To evaluate all the variants of the proposed self-supervised pre-training method MPCS. Specifically, 15 pre-training experiments for Efficient-net b2 and 18 pre-training experiments for ResNet-50 were performed on the BreakHis dataset. Further learned representations from pre-training models are evaluated by 800 downstream task training experiments on the BreakHis dataset covering all four magnifications (40x, 100x, 200x, and 400x), 5-cross folds, and on a wide range of labels (5% to 100% train set labels). One hundred forty downstream task training experiments were performed on the BACH dataset using BreakHis MPCS pre-trained ResNet-50 models. Finally, 30 downstream tasks experiments were performed for the Breast Cell Cancer Dataset using ResNet-50 pre-trained models covering fine-tuning and

linear evaluation. In this way, 1003 experiments are performed in the current work. Details are mentioned in Table XV.

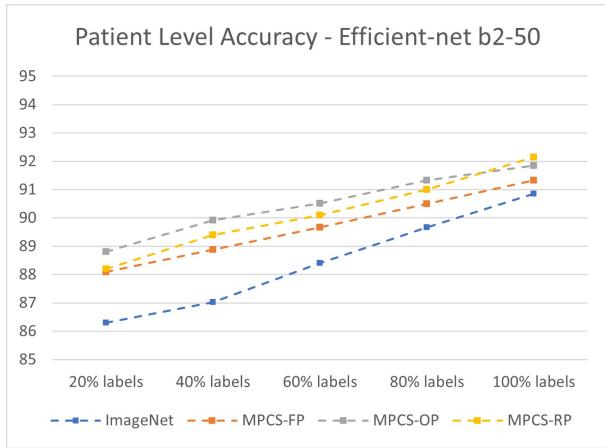


Fig. 12: Performance comparison (PLA accuracy, Efficient-net b2 model) for MPCS pre-trained models with ImageNet pre-trained model over range of labels used.

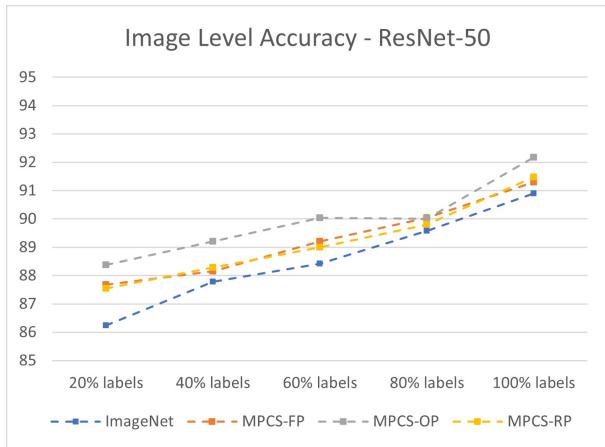


Fig. 13: Performance comparison (ILA accuracy, ResNet-50 model) for MPCS pre-trained models with ImageNet pre-trained model over range of labels used.

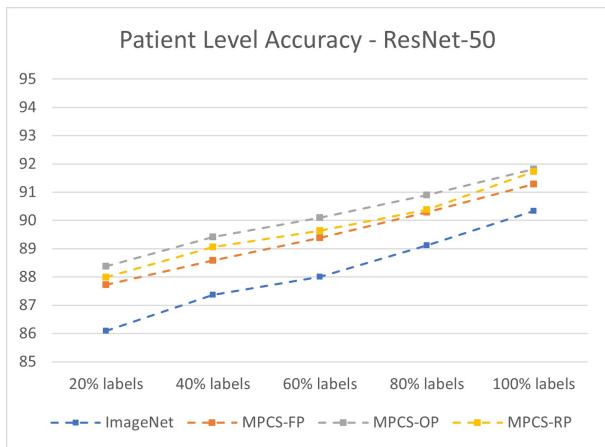


Fig. 14: Performance comparison (PLA accuracy, ResNet-50 model) for MPCS pre-trained models with ImageNet pre-trained model over range of labels used.