Hierarchical discriminative learning improves visual representations of biomedical microscopy

Cheng Jiang[1*]   Xinhai Hou[1*]   Akhil Kondepudi[1]   Asadur Chowdury[1]
Christian W. Freudiger[2]   Daniel A. Orringer[3]   Honglak Lee[1,4]   Todd C. Hollon[1]

[1]University of Michigan   [2]Invenio Imaging   [3]New York University
[4]LG AI Research   [*]Equal Contribution

{chengjia, xinhaih, tocho}@umich.edu   https://hidisc.mlins.org/

Abstract

Learning high-quality, self-supervised, visual representations is essential to advance the role of computer vision in biomedical microscopy and clinical medicine. Previous work has focused on self-supervised representation learning (SSL) methods developed for instance discrimination and applied them directly to image patches, or fields-of-view, sampled from gigapixel whole-slide images (WSIs) used for cancer diagnosis. However, this strategy is limited because it (1) assumes patches from the same patient are independent, (2) neglects the patient-slide-patch hierarchy of clinical biomedical microscopy, and (3) requires strong data augmentations that can degrade downstream performance. Importantly, sampled patches from WSIs of a patient's tumor are a diverse set of image examples that capture the same underlying cancer diagnosis. This motivated HiDisc, a data-driven method that leverages the inherent patient-slide-patch hierarchy of clinical biomedical microscopy to define a <u>hi</u>erarchical <u>disc</u>riminative learning task that implicitly learns features of the underlying diagnosis. HiDisc uses a self-supervised contrastive learning framework in which positive patch pairs are defined based on a common ancestry in the data hierarchy, and a unified patch, slide, and patient discriminative learning objective is used for visual SSL. We benchmark HiDisc visual representations on two vision tasks using two biomedical microscopy datasets, and demonstrate that (1) HiDisc pretraining outperforms current state-of-the-art self-supervised pretraining methods for cancer diagnosis and genetic mutation prediction, and (2) HiDisc learns high-quality visual representations using natural patch diversity without strong data augmentations.
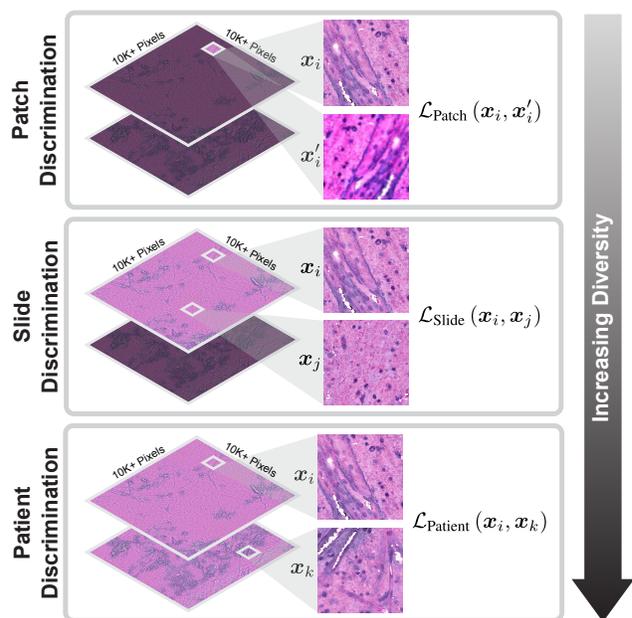
Figure 1. **Hierarchical self-supervised discriminative learning for visual representations**. Clinical biomedical microscopy has a hierarchical patch-slide-patient data structure. HiDisc combines patch, slide, and patient discrimination into a unified self-supervised learning task.

1. Introduction

Biomedical microscopy is an essential imaging method and diagnostic modality in biomedical research and clinical medicine. The rise of digital pathology and whole-slide images (WSIs) has increased the role of computer vision and machine learning-based approaches for analyzing microscopy data [51]. Improving the quality of visual representation learning of biomedical microscopy is critical to

introducing decision support systems and automated diagnostic tools into clinical and laboratory medicine.

Biomedical microscopy and WSIs present several unique computer vision challenges, including that image resolutions can be large ($10K \times 10K$ pixels) and annotations are often limited to weak slide-level or patient-level labels. Moreover, even weak annotations are challenging to obtain in order to protect patient health information and ensure patient privacy [54]. Additionally, data that predates newly developed or future clinical testing methods, such as genomic or methylation assays, also lack associated weak annotations. Because of large WSI sizes and weak annotations, the majority of computer vision research in biomedical microscopy has focused on WSI classification using a weakly supervised, patch-based, multiple instance learning (MIL) framework [2, 7, 20, 37, 38, 48]. Patches are arbitrarily defined fields-of-view (e.g., 256×256 pixels) that can be used for model input. The classification tasks include identifying the presence of cancerous tissue, such as breast cancer metastases in lymph node biopsies [13], differentiating specific cancer types [7, 11, 18], predicting genetic mutations [11, 26, 32], and patient prognostication [8, 29]. A limitation of end-to-end MIL frameworks for WSI classification is the reliance on weak annotations to train a patch feature extractor and achieve high-quality patch-level representation learning. This limitation, combined with the challenge of obtaining fully annotated, high-quality WSIs, necessitates better methods for self-supervised representation learning (SSL) of biomedical microscopy.

To date, research into improving the quality and efficiency of patch-level representation learning with*out* annotations has been limited. Previous studies have focused on using known SSL methods, such as contrastive learning [35, 47, 50], and applying them directly to WSI patches for visual pretraining. These SSL methods are not optimal because the majority use instance (i.e., patch) discrimination as the pretext learning task [5, 9, 10, 15, 55]. Patches belonging to the same slide or patient are correlated, which can decrease the learning efficiency. Instance discrimination alone does not account for patches from a common slide or patient being different and diverse views of the same underlying pathology. Moreover, previous SSL methods neglect the inherent patient-slide-patch data hierarchy of clinical biomedical microscopy as shown in Figure 1. This hierarchical data structure is not used to improve representation learning when training via a standard SSL objective. Lastly, most SSL methods require strong data augmentations for instance discrimination tasks [9]. However, strong and domain-agnostic augmentations can worsen representation learning in microscopy images by corrupting semantically important and discriminative features [21, 50].

Here, we introduce a method that leverages the inherent patient-slide-patch hierarchy of clinical biomedical microscopy to define a self-supervised hierarchical discriminative learning task, called HiDisc. HiDisc uses a self-supervised contrastive learning framework such that positive patch pairs are defined based on a common ancestry in the data hierarchy, and a combined patch, slide, and patient discriminative learning objective is used for visual SSL. By sampling patches across the data hierarchy, we introduce increased diversity between the positive examples, allowing for better visual representation learning and bypassing the need for strong, out-of-domain data augmentations. While we examine the HiDisc learning objective in the context of contrastive learning, it can be generalized to any siamese representation learning method [10].

We benchmark HiDisc self-supervised pretraining on two computer vision tasks using two diverse biomedical microscopy datasets: (1) multiclass histopathologic cancer diagnosis using stimulated Raman scattering microscopy [41] and (2) molecular genetic mutation prediction using light microscopy of hematoxylin and eosin (H&E)-stained cancer specimens [30]. These tasks are selected because of their clinical importance and they represent examples of how deep learning-based computer vision methods can push the limits of what is achievable through biomedical microscopy [18, 24, 26, 31]. We benchmark HiDisc in comparison to several state-of-the-art SSL methods, including SimCLR [9], BYOL [15], and VICReg [1]. We demonstrate that HiDisc has superior performance compared to other SSL methods across both datasets and computer vision tasks. Our results demonstrate how hierarchical discriminative learning can improve self-supervised visual representations of biomedical microscopy.

2. Related Work

Biomedical microscopy and computational pathology Biomedical microscopy refers to a diverse set of microscopy methods used in both clinical medicine and biomedical research. The most common clinical use of biomedical microscopy is light microscopy combined with H&E staining of clinical tissue specimens, such as tissue biopsies for cancer diagnosis [34]. The introduction of fast and efficient whole-slide digitization resulted in a rapid increase in the availability of WSIs and accelerated the field of computational pathology [42, 51]. Computational pathology aims to discover and characterize histopathologic features within biomedical microscopy data for cancer diagnosis, prognostication, and to estimate response to treatment.

The introduction of deep learning to WSI has resulted in clinical-grade computational pathology with diagnostic performance on par with board-certified pathologists [2, 18, 20]. See [43] for a comprehensive survey of deep learning-based methods in computational pathology. Ilse et al. presented MIL framework using an attention-based global pooling operation for slide-level inference [22]. Campanella et al.
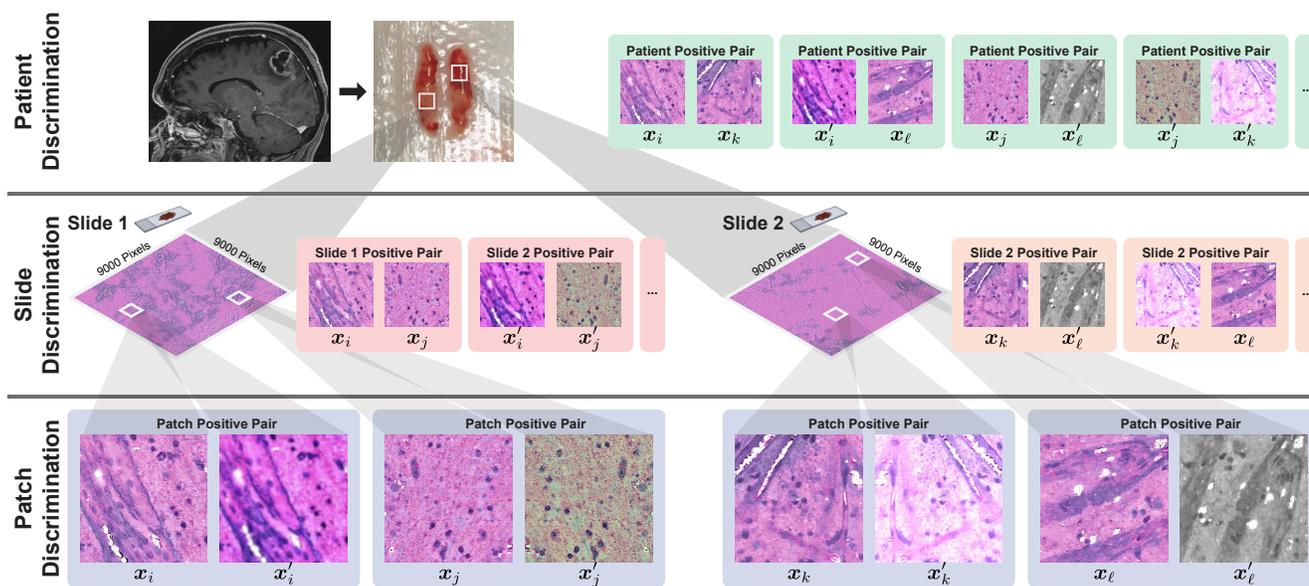
Figure 2. **HiDisc Overview.** Motivated by the patient-slide-patch data hierarchy of clinical biomedical microscopy, HiDisc defines a patient, slide, and patch discriminative learning objective to improve visual representations. Because WSI and microscopy data are inherently hierarchical, defining a unified hierarchical loss function does not require additional annotations or supervision. Positive patch pairs are defined based on a common ancestry in the data hierarchy. A major advantage of HiDisc is the ability to define positive pairs without the need to sample from or learn a set of strong image augmentations, such as random erasing, shears, color inversion, etc. Because each field-of-view in a WSI is a different view of a patient's underlying cancer diagnosis, HiDisc implicitly learns image features that predict that diagnosis.

extended the strategy of a trainable aggregation operation using a recurrent neural network for gigapixel WSI classification tasks [2]. Lu et al. updated the attention-based MIL method to allow for better interpretability and efficient weakly supervised training using transformers [38]. HiDisc is complementary to any MIL framework and can be used as an effective self-supervised pretraining strategy.

Other biomedical microscopy methods have an increasing role in clinical medicine. Electron [39], fluorescence [44, 46, 53], and stimulated Raman scattering microscopy [14, 41] are a few examples of imaging methods that generate microscopy images used for patient diagnosis. Several studies have applied deep learning-based methods to these modalities for image analysis [18, 19, 23, 24, 44, 53].

Self-supervised pretraining in biomedical microscopy
Self-supervised pretraining has been used in computational pathology to improve patch-level representation learning [7, 28, 43, 45, 57]. Generally, a two-stage approach is used where first an SSL method is applied for patch-level feature learning using instance discrimination, and then the patch-level features are aggregated for slide- or patient-level diagnosis. SSL patch pretraining can reduce the amount of data needed compared to end-to-end MIL training [43]. Contrastive predictive coding [36], SimCLR [28], MoCo [45],

VQ-VAE [6], and VAE-GAN [57] are examples of deep self-supervised visual representation learning methods applied to biomedical microscopy images [43].

Chen et al. presented a study using vision transformers and self-supervised pretraining at different image scales within individual WSIs [7]. They aim to represent the hierarchical structure of visual histopathologic tokens (e.g., cellular features, supracellular structures, etc.) at varying image resolutions using a transformer pyramid, resulting in a single slide-level representation. Knowledge distillation was used for SSL at each image resolution [5]. HiDisc is complementary to their method and can be used for SSL at any image resolution or, more generally, field-of-view pretraining for any MIL method for slide-level representations.

3. Methods

3.1. The Patient-Slide-Patch Hierarchy

The motivation for HiDisc is that fields-of-view from clinical WSIs, sampled from within a patient's tumor, are a diverse set of image examples that capture the same underlying cancer diagnosis. Our approach focuses on how to use these diverse fields-of-view in the context of the known clinical patient-slide-patch hierarchical structure to improve visual representation learning. Most patients included in

public cancer histopathology datasets, including The Cancer Genome Atlas (TCGA) [4] and OpenSRH [24], contain multiple WSIs as part of their clinical cancer diagnosis. These WSIs may be sampled from different locations in the patient's tumor, or different regions within the same tumor specimen. Both histopathologic and molecular heterogeneity has been well described within human cancers, encouraging clinicians to obtain multiple specimens/samples/views of the patient's tumor [49]. To leverage the hierarchy shown in Figure 1, we create positive pairs at the patch-, slide-, and patient-level to define different discriminative learning tasks with a corresponding increase in visual feature diversity:

- **Patch discrimination:** Positive pairs are created from different random augmentations of the same patch. This strategy is similar to existing work on SSL via instance discrimination [1, 9, 10].

- **Slide discrimination:** Positive pairs are created from different augmented patches sampled from the same WSI. These pairs capture local feature diversity within the same specimen. Regional differences in cytologic and histoarchitectural features can be captured at this hierarchical level.

- **Patient discrimination:** Positive pairs are created from different WSIs from the same patient. Patches from different WSIs have the same underlying cancer diagnosis, but can have the greatest degree of feature diversity due to spatially separated tumor specimens. Additionally, diversity in specimen quality, processing, and staining, etc., is captured at this level.

An overview of the hierarchical discrimination tasks is shown in Figure 2.

3.2. Hierarchical Discrimination (HiDisc)

The formulation of the HiDisc loss function is based on NT-Xent [9] and inspired by [27, 56] for the purpose of multiple positive pairs. The fundamental difference is that no class labels are used during training with a HiDisc loss.

Discrimination level	Number of samples treated as independent	Number of Positive pairs
Patch	$n \cdot n_s \cdot n_p$	n_a
Slide	$n \cdot n_s$	$n_p \cdot n_a$
Patient	n	$n_s \cdot n_p \cdot n_a$

Table 1. **Batch composition at all discrimination levels.** The number of samples in the batch treated as independent and the number of positive pairs at each discrimination level. n, number of patients in the batch, n_s, number of slides sampled per patient, n_p, number of patches sampled per slide, n_a, number of augmentations performed on each patch.

Algorithm 1 HiDisc Pseudocode in PyTorch style

```
# f: backbone + projection mlp
# n: number of patients, batch size
# ns: number of slides sampled per patient
# np: number of patches sampled per slide
# na: number of augmentations per patch
# w: width of each patch
# h: height of each patch
# L: self supervised loss function
# lambdas: weight coefficients for each loss term

for x in loader:
    # Load minibatch x with n patients
    # x.shape is (n, ns, np, na, w, h)

    # Forward pass, z.shape is (n * ns * np * na, d)
    z = f(x.reshape(n * ns * np * na, w, h))

    # Compute HiDisc loss for patch,
    # slide and patient level discrimination
    loss_patch = L(z.reshape(n * ns * np, na, d))
    loss_slide = L(z.reshape(n * ns, np * na, d))
    loss_patient = L(z.reshape(n, ns * np * na, d))
    loss = dot(lambdas, # lambdas.shape is (1, 3)
        [loss_patch, loss_slide, loss_patient])

    # Back-propagate and update network
    loss.backward()
    optimizer.step()
```

HiDisc utilizes the natural hierarchy inherent to biomedical microscopy images to improve visual SSLf.

We randomly sample a minibatch of n patients, n_s slides from each patient, n_p patches from each slide, and we augment each patch n_a times. For patients with less than n_s slides, the slides are repeated and sampled. We assume $n_p \ll$ numbers of patches available for WSI sampling. Note that if a patient has only one WSI, then patient discrimination degenerates to slide discrimination.

The HiDisc loss consists of the sum of three losses, each of which corresponds to a discrimination task at a different level of the patch-slide-patient hierarchy. Similar to the supervised contrastive learning loss [27], the component HiDisc losses are designed to fit more than one pair of positives within each level of the hierarchy. We define the HiDisc loss at the level ℓ to be:

$$L_{\text{HiDisc}}^{\ell} = \sum_{i \in \mathcal{I}} \frac{-1}{|\mathcal{P}_\ell(i)|} \sum_{p \in \mathcal{P}_\ell(i)} \log \frac{\exp(z_i \cdot z_p / \tau)}{\sum_{a \in \mathcal{A}(i)} \exp(z_i \cdot z_a / \tau)},$$
(1)

where $\ell \in \{\text{Patch, Slide, Patient}\}$ is the level of discrimination, and $\mathcal{I}$ is the set of all images in the minibatch. $\mathcal{A}_\ell(i)$ is the set of all images in $\mathcal{I}$ except for the anchor image i,

$$\mathcal{A}(i) = \mathcal{I} \setminus \{i\},$$
(2)

and $\mathcal{P}_\ell(i)$ is a set of images that are positive pairs of i at the ℓ-level,

$$\mathcal{P}_\ell(i) = \{p \in \mathcal{A}_\ell(i) : \text{ancestry}_\ell(p) = \text{ancestry}_\ell(i)\}, \quad (3)$$

where ancestry$_\ell(\cdot)$ is the ℓ-level ancestry for an augmented patch in the batch. For example, patches $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ from the same patient would have the same patient ancestry, i.e., ancestry$_{\text{Patient}}(\boldsymbol{x}_i) = $ ancestry$_{\text{Patient}}(\boldsymbol{x}_j)$.

Patch-, slide-, and patient-level HiDisc losses share the same overall contrastive objective, but capture positive pairs at different levels in the hierarchy. Each loss has a different number of positive pairs within a minibatch. Details about this relationship are shown in Table 1.

Finally, the complete HiDisc loss is the sum of the patch-, slide-, and patient-level discrimination losses defined above:

$$\mathcal{L}_{\text{HiDisc}} = \sum_{\ell \in \{\text{Patch, Slide, Patient}\}} \lambda_\ell L^\ell_{\text{HiDisc}}, \qquad (4)$$

where λ_ℓ is a weighting hyperparameter for level ℓ in the total loss. The pseudocode in PyTorch-style detailing the training process is shown in Algorithm 1.

4. Experiments

We evaluate HiDisc using two different computational histopathology tasks: multiclass histopathologic cancer diagnosis and molecular genetic mutation prediction. We present quantitative classification performance metrics, as well as qualitative evaluation with tSNE visualizations [52] of the learned patch-level representations.

4.1. Datasets

Stimulated Raman histology (SRH) We validate HiDisc on a multiclass histopathological cancer diagnosis task using an SRH dataset. Stimulated Raman histology is an optical microscopy method that provides rapid, label-free, sub-micron resolution images of unprocessed biological tissues [14, 41]. The SRH dataset includes specimens from patients who underwent brain tumor biopsy or tumor resection. Patients were consecutively and prospectively enrolled at the University of Michigan for intraoperative SRH imaging, and this study was approved by the Institutional Review Board (HUM00083059). Informed consent was obtained for each patient prior to SRH imaging and approved the use of tumor specimens for research and development. The SRH dataset consists of 852K patches from 3560 slide images from 896 patients with classes consisting of normal brain tissue and 6 different brain tumor diagnoses: high-grade glioma (HGG), low-grade glioma (LGG), meningioma, pituitary adenoma, schwannoma, and metastatic tumor. All slides are divided into 300×300 patches, and they are preprocessed to exclude the empty or non-diagnostic regions using a segmentation model [18].

TCGA diffuse gliomas We additionally validate HiDisc using WSIs from The Cancer Genome Atlas (TCGA) dataset. We focus on WSIs from brain tumor patients diagnosed with diffuse gliomas, the most common and deadly primary brain tumor [34]. Molecular genetic mutation classification is used as the evaluation task. The most important genetic mutation that defines lower grade versus high grade diffuse gliomas is isocitrate dehydrogenase-1/2 (IDH) mutational status [3]. IDH-mutant tumors are known to have a better prognosis and overall survival (median survival ∼10 years) compared to IDH-wildtype tumors (∼1.5 years). The classification task is to predict IDH mutational status using formalin-fixed, paraffin-embedded H&E-stained WSI images at 20× magnification from the TCGA dataset. Predicting IDH mutational status from WSIs is currently not feasible for board-certified neuropathologists [3, 12]; genetic mutation prediction from WSIs could avoid time-consuming and expensive laboratory techniques like genetic sequencing. WSIs are divided into 300×300 pixel fields-of-view, and blank patches are excluded. The rest of the patches are stain normalized using the Macenko algorithm [40]. The TCGA data we included consists of a total of 879 patients and 1703 slides, and 11.3M patches.

4.2. Implementation details

We train HiDisc using ResNet-50 [17] as the backbone feature extractor and a one-layer MLP projection head to project the embedding to 128-dimensional latent space. We use an AdamW [33] optimizer with a learning rate of 0.001 and a cosine decay scheduler after warmup in the first 10% of the iterations. For a fair comparison, we control the total number of images in one minibatch ($n \cdot n_a \cdot n_s \cdot n_p$) as 512 and 448 for SRH and TCGA data, respectively. We train HiDisc with patient discrimination by setting $n_a = n_s = n_p = 2$, slide discrimination by setting $n_a = n_s = 2, n_p = 1$, and patch discrimination by setting $n_a = 2, n_s = n_p = 1$. The number of patients n sampled from each batch is adjusted accordingly. λ_ℓ for each level of discriminating loss is set to 1, and temperature τ is set as 0.7 for all experiments. We define weak augmentation as random horizontal and vertical flipping. The strong augmentations are similar as [9], including random erasing, color jittering, and affine transformation (for details, see Appendix A). We train HiDisc till convergence for both datasets (100K and 60K iterations for SRH and TCGA, respectively) with three random seeds. Training details for all baselines, including SimCLR [9], SimSiam [10], BYOL [15], VICReg [1], and SupCon [27], are similar to HiDisc.

4.3. Evaluation protocols

***k*NN evaluation** Standard protocols to evaluate self-supervised representation learning include linear and fine-tuning evaluation. However, both methods are sensitive to hyperparameters, such as learning rate [5]. Therefore, we use the k nearest neighbor (kNN) classifier for quantitative

	SRH - Patch		SRH - Patient		TCGA - Patch		TCGA - Patient	
	Accuracy	MCA	Accuracy	MCA	Accuracy	AUROC	Accuracy	AUROC
SimCLR [9]	81.0 (0.1)	73.9 (0.2)	83.1 (0.7)	78.4 (0.6)	77.8 (0.0)	85.2 (0.1)	80.7 (0.6)	88.9 (0.3)
SimSiam [10]	80.3 (1.9)	73.6 (2.7)	82.3 (1.7)	77.0 (4.0)	68.4 (0.3)	74.1 (0.3)	76.6 (0.6)	82.4 (0.3)
BYOL [15]	83.5 (0.1)	78.2 (0.2)	84.8 (1.0)	82.7 (1.0)	80.0 (0.1)	87.5 (0.1)	83.1 (0.6)	89.8 (0.2)
VICReg [1]	82.1 (0.3)	76.0 (0.4)	82.1 (0.7)	78.7 (1.9)	75.5 (0.1)	82.9 (0.1)	77.0 (0.5)	86.0 (0.3)
HiDisc-Patch	80.8 (0.0)	73.5 (0.1)	82.6 (0.3)	77.9 (0.3)	77.2 (0.1)	84.7 (0.1)	81.0 (0.5)	88.1 (0.2)
HiDisc-Slide	86.9 (0.2)	**83.2 (0.2)**	**87.6 (0.5)**	**87.0 (1.4)**	82.7 (0.2)	89.3 (0.2)	**84.3 (0.3)**	**92.3 (0.3)**
HiDisc-Patient	**87.4 (0.1)**	**83.5 (0.2)**	**87.9 (0.5)**	86.4 (0.6)	**83.1 (0.1)**	**90.1 (0.1)**	83.6 (0.3)	91.8 (0.2)
Supervised	88.9 (0.3)	86.3 (0.3)	88.5 (0.5)	89.1 (0.5)	85.1 (0.3)	91.7 (0.2)	88.3 (0.4)	95.2 (0.2)

Table 2. **Main results.** We compare HiDisc to state-of-the-art SSL. Supervised refers to models trained with supervision from class labels. Standard deviations are reported in parentheses. HiDisc-Slide and HiDisc-Patient outperform all baseline methods in all metrics for both tasks. Our HiDisc benchmark outperforms previously reported fully *supervised* baselines from the existing literature on the same genetic mutation classification task on TCGA [30]. MCA, mean class accuracy, AUROC, area under the receiver operating characteristic curve.
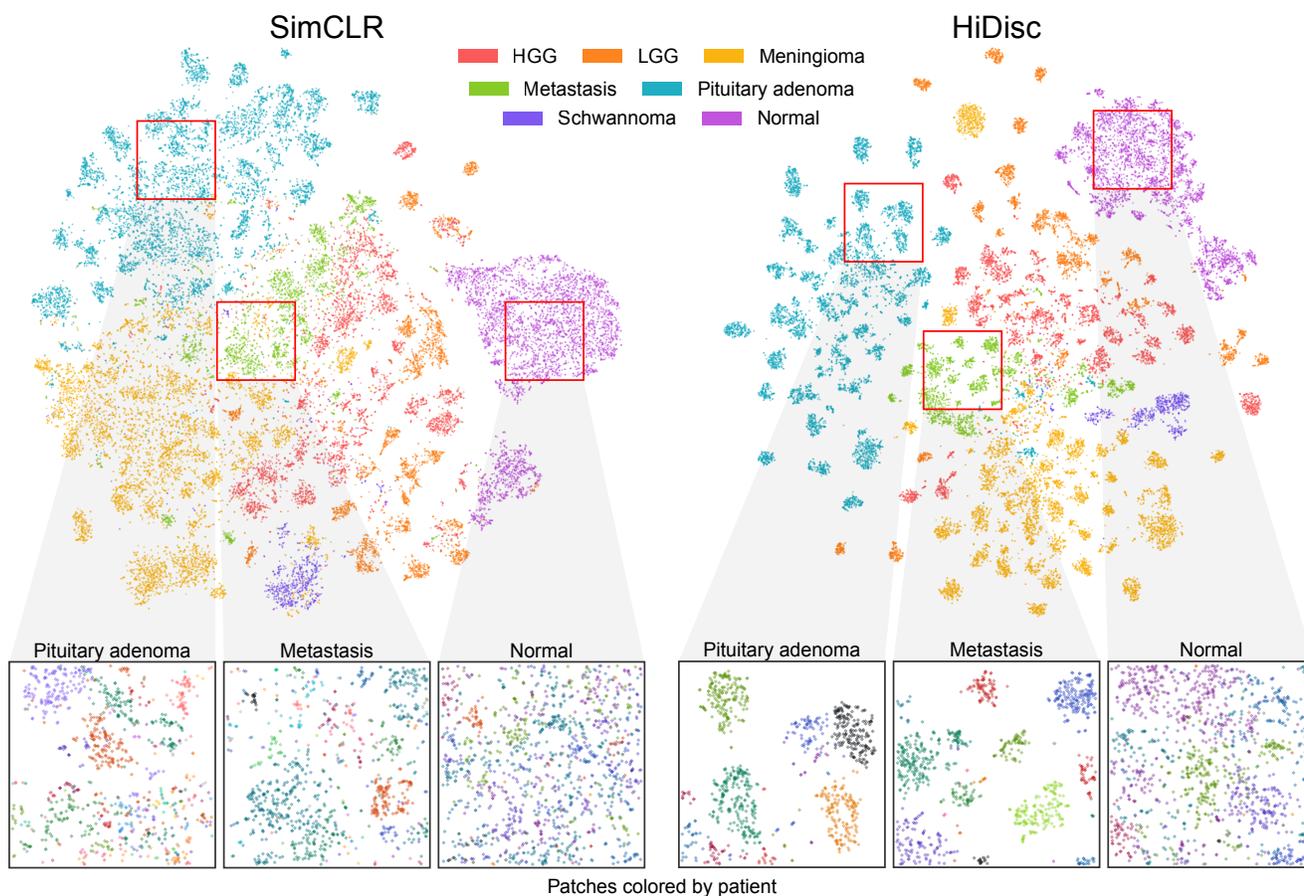

Figure 3. **Visualization of learned SRH representations using SimCLR and HiDisc. Top.** Randomly sampled patch representations are visualized after SimCLR versus HiDisc pretraining using tSNE [52]. Representations are colored based on brain tumor *diagnosis*. HiDisc qualitatively achieves higher quality feature learning and class separation compared to SimCLR. Expectedly, HiDisc shows within-diagnosis clustering that corresponds to patient discrimination. **Bottom.** Magnified cropped regions of the above visualizations show subclusters that correspond to individual patients. Patch representations in magnified crops are colored according to *patient* membership. We see patient discrimination within the different tumor diagnoses. Importantly, we do not see patient discrimination within normal brain tissue because there are minimal-to-no differentiating microscopic features between patients. This demonstrates that in the absence of discriminative features at the slide- or patient-level, HiDisc can achieve good feature learning using patch discrimination without overfitting the other discrimination tasks. HGG, high grade glioma; LGG, low grade glioma; Normal, normal brain tissue.

evaluation. We freeze the pretrained backbone to compute the representation vectors for both training and testing data, and the nearest neighbor classifier is used to match each patch in the testing dataset to the most similar k patches in the training set based on cosine similarity. It also outputs a prediction score measured by the cosine similarity between each test image and its k nearest neighbors. Due to the size of the TCGA dataset, we randomly sample 400 patches from each slide for evaluation across three different random seeds. Using the kNN classifier, we can compute accuracy (ACC), mean class accuracy (MCA), and area under receiver operating characteristic curve (AUROC) for patch metrics. We use MCA for SRH dataset because it is a multiclass classification problem and the classes are imbalanced. The AUROC is used for the TCGA dataset since it is a balanced binary classification task.

Slide and patient metrics   In contrast to patch-level evaluation, slide and patient predictions are more practical for cancer diagnosis and other clinical uses [2]. After getting patch prediction scores by kNN evaluation, we use average pooling over the scores within each WSI or patient to obtain an aggregated prediction score. Compared to most MIL methods, this non-parametric method directly evaluates representation learning without additional training.

5. Results

5.1. Quantitative metrics

In this section, we evaluate the representations learned by self-supervised HiDisc pretraining using the training and evaluation protocols described in Section 4. We compare HiDisc and SSL baselines on the testing set of both datasets in Table 2. Since HiDisc-Patch is most similar to SimCLR, we observe similar performances. After we add slide and patient discrimination in HiDisc-Slide and HiDisc-Patient, we observe a significant boost in patch accuracy (+6.1 and +6.6 on SRH, +5.5 and +5.9 on TCGA), and a similar increase in other performance metrics as well. Among all methods, HiDisc-Patient has the best performance and outperforms the best baseline, BYOL, with an improvement of 3.9% and 3.1% in classification accuracy for SRH and TCGA, respectively. The GPU wall time needed to train BYOL is roughly 1.5x longer than HiDisc because it requires updating the weights of the target network using exponential moving average. Appendix B shows additional model evaluation metrics.

5.2. Qualitative evaluation

We also qualitatively evaluate the learned patch representations with tSNE [52] colored by class and patient label for both SimCLR and HiDisc. Figures 3 and 4 show the learned representations for the SRH and TCGA datasets,

respectively. Here, we randomly sample patches from the validation set and plot them by class membership (tumor or molecular subtype). We observe that HiDisc learns better representations for both classification tasks, with more discernible clusters for each class. We also observe better patient clusters within each tumor class in the representations learned by HiDisc. Furthermore, patient clusters are not observed in normal brain tissues for both SimCLR and HiDisc-Patch, as there are minimal differentiating microscopic features between patients.

5.3. Ablation Studies

Weak Augmentation   We demonstrate that HiDisc is capable of preserving excellent performance without the use of strong, domain-agnostic data augmentations as shown in Table 3. SimCLR suffers from dimensional collapse with weak augmentations [25]. Similar to SimCLR, HiDisc-Patch collapses because it is limited to the diversity from augmentations alone. HiDisc-Slide and HiDisc-Patient performance remain high across both datasets and tasks. HiDisc-Patient outperforms HiDisc-Slide, especially when evaluating at the patch level. We hypothesize that this is a result of additional diversity between positive pairs contributed by patient-level discrimination. We also provide supervised contrastive learning (SupCon) baselines [27] as an upper performance bound. Figure 5 shows SimCLR fails to learn semantically meaningful features while HiDisc achieves high-quality representations. Overall, we observe that HiDisc, especially HiDisc-Patient, performs well regardless of whether strong augmentations are used.
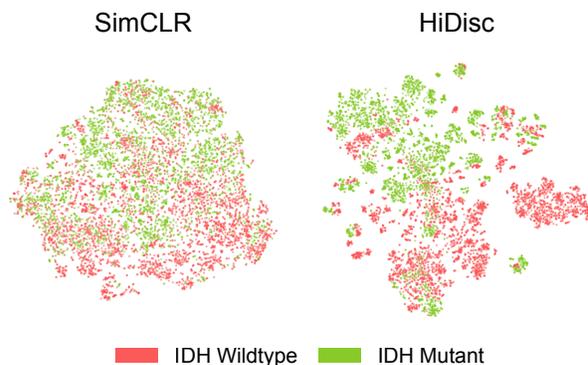

Figure 4. **Visualization of learned TCGA representations using SimCLR and HiDisc.** We randomly sample patches from the validation set, and visualize these representations using tSNE [52]. Representations on the plots are colored by IDH mutational status. Qualitatively, we can observe that HiDisc forms better representations compared to SimCLR, with clusters within each mutation that corresponds to patient membership. This observation is consistent with the visualizations for the SRH dataset in Figure 3.

	SRH - Patch		SRH - Patient		TCGA - Patch		TCGA - Patient	
	Accuracy	MCA	Accuracy	MCA	Accuracy	AUROC	Accuracy	AUROC
SimCLR [9]	31.5 (2.3)	23.1 (1.9)	40.2 (6.9)	28.9 (4.5)	57.1 (1.1)	58.4 (2.2)	58.1 (1.1)	72.8 (3.2)
HiDisc-Patch	31.3 (0.6)	22.2 (0.5)	47.4 (2.1)	33.1 (1.6)	59.0 (0.8)	61.5 (1.3)	61.2 (4.2)	75.8 (2.5)
HiDisc-Slide	82.8 (0.2)	77.4 (0.3)	**84.2 (0.5)**	**82.3 (0.4)**	79.6 (0.1)	86.3 (0.2)	77.7 (0.4)	85.3 (0.3)
HiDisc-Patient	**84.9 (0.2)**	**78.9 (0.1)**	84.7 (0.5)	80.9 (1.4)	**82.9 (0.2)**	**89.6 (0.2)**	**82.3 (0.3)**	**90.3 (0.3)**
SupCon [27]	90.0 (0.2)	87.4 (0.3)	90.0 (0.5)	90.3 (0.4)	85.4 (0.4)	92.0 (0.2)	88.4 (0.8)	95.2 (0.4)

Table 3. **Results for weak augmentation.** Comparison of HiDisc with the SimCLR and SupCon on SRH and TCGA datasets with weak augmentation (random flipping). Both SimCLR and HiDisc-Patch patch collapse since weak augmentation will make the pretext task trivial. SupCon, on the other hand, is not sensitive to data augmentations since its positive pairs are defined by class labels. HiDisc-Slide and HiDisc-Patient are only slightly affected by the augmentation and achieve performance close to the supervised method.
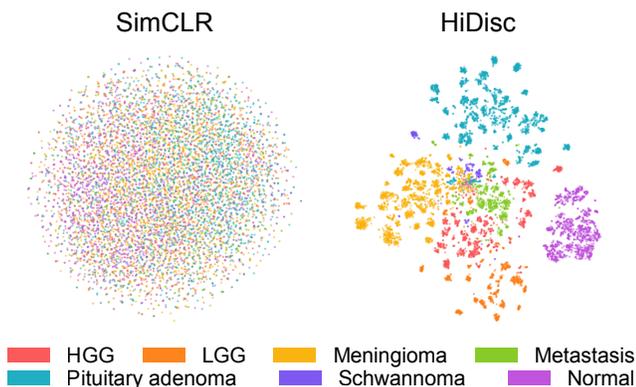

Figure 5. **Hierarchical self-supervised discriminative learning without strong data augmentations.** Randomly sampled patch representations are shown after SimCLR and HiDisc-Patient pretraining without the use of strong, domain-agnostic data augmentations on SRH dataset. HiDisc achieves high-quality representation learning, while SimCLR is unable to learn semantically meaningful features via instance discrimination alone. HGG, high grade glioma; LGG, low grade glioma; Normal, normal brain tissue

Other Ablation studies Additional experiments are included in Appendix C. We perform ablation studies on the **weighting factor** λ_ℓ for each level of discrimination. We also train HiDisc with different **number of iterations**, **learning rate** and **batch size**. Some settings may marginally improve model performance but cost a significant amount of computation resources.

6. Conclusion

We present HiDisc, a unified, hierarchical, self-supervised representation learning method for biomedical microscopy. HiDisc is able to outperform other state-of-the-art SSL methods for visual representation learning. The performance increase is driven by leveraging the inherent patient-slide-patch hierarchy of clinical WSIs. The inherent data hierarchy provides image diversity by defining positive patch pairs based on a common ancestry in the hierarchy and does not require strong, domain-agnostic augmentations. By combining patch, slide, and patient discrimination into a single learning objective, HiDisc implicitly learns image features of the patient's underlying diagnosis without the need for patient-level annotations or supervision.

Limitations Like other WSI classification methods, HiDisc representation learning is currently limited to single-resolution fields-of-view that are arbitrarily defined. Expanding HiDisc to include multiple image resolutions could improve its ability to capture multiscale image features of the patient's underlying diagnosis. Also, we have limited the evaluation to a contrastive learning framework and HiDisc can also be evaluated using other siamese learning frameworks [1, 5, 10].

Broader Impacts We have limited this investigation to biomedical microscopy and WSIs. However, many imaging medical modalities, such as magnetic resonance imaging and fundoscopy [16], have a clinical hierarchical data structure that could benefit from a similar hierarchical representation learning framework. We hope that hierarchical discriminative learning will extend beyond microscopy to other medical and non-medical imaging domains.

Acknowledgements and Competing Interests

References

[1] Adrien Bardes, Jean Ponce, and Yann LeCun. VI-CReg: Variance-Invariance-Covariance regularization for Self-Supervised learning. Feb. 2022. 2, 4, 5, 6, 8

[2] Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.*, 25(8):1301–1309, Aug. 2019. 2, 3, 7

[3] Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse Lower-Grade gliomas. *N. Engl. J. Med.*, 372(26):2481–2498, June 2015. 5

[4] Cancer Genome Atlas Research Network, John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas Pan-Cancer analysis project. *Nat. Genet.*, 45(10):1113–1120, Oct. 2013. 4

[5] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. pages 9650–9660, Apr. 2021. 2, 3, 5, 8

[6] Chengkuan Chen, Ming Y Lu, Drew F K Williamson, Tiffany Y Chen, Andrew J Schaumberg, and Faisal Mahmood. Fast and scalable search of whole-slide images via self-supervised deep learning. *Nat Biomed Eng*, Oct. 2022. 3

[7] Richard J Chen, Chengkuan Chen, Yicong Li, Tiffany Y Chen, Andrew D Trister, Rahul G Krishnan, and Faisal Mahmood. Scaling vision transformers to gigapixel images via hierarchical Self-Supervised learning. June 2022. 2, 3

[8] Richard J Chen, Ming Y Lu, Drew F K Williamson, Tiffany Y Chen, Jana Lipkova, Zahra Noor, Muhammad Shaban, Maha Shady, Mane Williams, Bumjin Joo, and Faisal Mahmood. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell*, 40(8):865–878.e6, Aug. 2022. 2

[9] 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, 2020. 2, 4, 5, 6, 8, 12, 13

[10] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. Nov. 2020. 2, 4, 5, 6, 8

[11] Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyö, Andre L Moreira, Narges Razavian, and Aristotelis Tsirigos. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nat. Med.*, 24(10):1559–1567, Oct. 2018. 2

[12] Jeanette E Eckel-Passow, Daniel H Lachance, Annette M Molinaro, Kyle M Walsh, Paul A Decker, Hugues Sicotte, Melike Pekmezci, Terri Rice, Matt L Kosel, Ivan V Smirnov, Gobinda Sarkar, Alissa A Caron, Thomas M Kollmeyer, Corinne E Praska, Anisha R Chada, Chandralekha Halder, Helen M Hansen, Lucie S McCoy, Paige M Bracci, Roxanne Marshall, Shichun Zheng, Gerald F Reis, Alexander R Pico, Brian P O'Neill, Jan C Buckner, Caterina Giannini, Jason T Huse, Arie Perry, Tarik Tihan, Mitchell S Berger, Susan M Chang, Michael D Prados, Joseph Wiemels, John K Wiencke, Margaret R Wrensch, and Robert B Jenkins. Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. *N. Engl. J. Med.*, 372(26):2499–2508, June 2015. 5

[13] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes van Diest, Bram van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen A W M van der Laak, the CAMELYON16 Consortium, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, Oscar Geessink, Nikolaos Stathonikos, Marcory Crf van Dijk, Peter Bult, Francisco Beca, Andrew H Beck, Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, Aoxiao Zhong, Qi Dou, Quanzheng Li, Hao Chen, Huang-Jing Lin, Pheng-Ann Heng, Christian Haß, Elia Bruni, Quincy Wong, Ugur Halici, Mustafa Ümit Öner, Rengul Cetin-Atalay, Matt Berseth, Vitali Khvatkov, Alexei Vylegzhanin, Oren Kraus, Muhammad Shaban, Nasir Rajpoot, Ruqayya Awan, Korsuk Sirinukunwattana, Talha Qaiser, Yee-Wah Tsang, David Tellez, Jonas Annuscheit, Peter Hufnagl, Mira Valkonen, Kimmo Kartasalo, Leena Latonen, Pekka Ruusuvuori, Kaisa Liimatainen, Shadi Albarqouni, Bharti Mungal, Ami George, Stefanie Demirci, Nassir Navab, Seiryo Watanabe, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda, Hady Ahmady Phoulady, Vassili Kovalev, Alexander Kalinovsky, Vitali Liauchuk, Gloria Bueno, M Milagro Fernandez-Carrobles, Ismael Serrano, Oscar Deniz, Daniel Racoceanu, and Rui Venâncio. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*, 318(22):2199–2210, Dec. 2017. 2

[14] Christian W Freudiger, Wei Min, Brian G Saar, Sijia Lu, Gary R Holtom, Chengwei He, Jason C Tsai, Jing X Kang, and X Sunney Xie. Label-free biomedical imaging with high sensitivity by stimulated raman scattering microscopy. *Science*, 322(5909):1857–1861, Dec. 2008. 3, 5

[15] 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. June 2020. 2, 5, 6

[16] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, Ramasamy Kim, Rajiv Raman, Philip C Nelson, Jessica L Mega, and Dale R Webster. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22):2402–2410, Dec. 2016. 8

[17] K He, X Zhang, S Ren, and J Sun. Deep residual learning for image recognition. *Proc. IAPR Int. Conf. Pattern Recogn.*, 2016. 5

[18] Todd C Hollon, Balaji Pandian, Arjun R Adapa, Esteban Urias, Akshay V Save, Siri Sahib S Khalsa, Daniel G Eichberg, Randy S D'Amico, Zia U Farooq, Spencer Lewis, Petros D Petridis, Tamara Marie, Ashish H Shah, Hugh J L Garton, Cormac O Maher, Jason A Heth, Erin L McKean, Stephen E Sullivan, Shawn L Hervey-Jumper, Parag G Patil, B Gregory Thompson, Oren Sagher, Guy M McKhann, 2nd, Ricardo J Komotar, Michael E Ivan, Matija Snuderl, Marc L Otten, Timothy D Johnson, Michael B Sisti, Jeffrey N Bruce, Karin M Muraszko, Jay Trautman, Christian W Freudiger, Peter Canoll, Honglak Lee, Sandra Camelo-Piragua, and Daniel A Orringer. Near real-time intraoperative brain tumor diagnosis using stimulated raman histology and deep neural networks. *Nat. Med.*, Jan. 2020. 2, 3, 5, 12

[19] Todd C Hollon, Balaji Pandian, Esteban Urias, Akshay V Save, Arjun R Adapa, Sudharsan Srinivasan, Neil K Jairath, Zia Farooq, Tamara Marie, Wajd N Al-Holou, Karen Eddy, Jason A Heth, Siri Sahib S Khalsa, Kyle Conway, Oren Sagher, Jeffrey N Bruce, Peter Canoll, Christian W Freudiger, Sandra Camelo-Piragua, Honglak Lee, and Daniel A Orringer. Rapid, label-free detection of diffuse glioma recurrence using intraoperative stimulated raman histology and deep neural networks. *Neuro. Oncol.*, July 2020. 3

[20] L Hou, D Samaras, T M Kurc, Y Gao, J E Davis, and J H Saltz. Patch-Based convolutional neural network for whole slide tissue image classification. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2424–2433, June 2016. 2

[21] Zeshan Hussain, Francisco Gimenez, Darvin Yi, and Daniel Rubin. Differential data augmentation techniques for medical imaging classification tasks. *AMIA Annu. Symp. Proc.*, 2017:979–984, 2017. 2

[22] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2127–2136. PMLR, 2018. 2

[23] Cheng Jiang, Abhishek Bhattacharya, Joseph R Linzey, Rushikesh S Joshi, Sung Jik Cha, Sudharsan Srinivasan, Daniel Alber, Akhil Kondepudi, Esteban Urias, Balaji Pandian, Wajd N Al-Holou, Stephen E Sullivan, B Gregory Thompson, Jason A Heth, Christian W Freudiger, Siri Sahib S Khalsa, Donato R Pacione, John G Golfinos, Sandra Camelo-Piragua, Daniel A Orringer, Honglak Lee, and Todd C Hollon. Rapid automated analysis of skull base tumor specimens using intraoperative optical imaging and artificial intelligence. *Neurosurgery*, Mar. 2022. 3

[24] Cheng Jiang, Asadur Chowdury, Xinhai Hou, Akhil Kondepudi, Christian Freudiger, Kyle Conway, Sandra Camelo-Piragua, Daniel Orringer, Honglak Lee, and Todd Hollon. OpenSRH: optimizing brain tumor surgery using intraoperative stimulated raman histology. In *Advances in Neural Information Processing Systems*, volume 35, pages 28502–28516, 2022. 2, 3, 4

[25] Li Jing, Pascal Vincent, Yann LeCun, and Yuandong Tian. Understanding dimensional collapse in contrastive self-supervised learning. Oct. 2021. 7

[26] Jakob Nikolas Kather, Lara R Heij, Heike I Grabsch, Chiara Loeffler, Amelie Echle, Hannah Sophie Muti, Jeremias Krause, Jan M Niehues, Kai A J Sommer, Peter Bankhead, Loes F S Kooreman, Jefree J Schulte, Nicole A Cipriani, Roman D Buelow, Peter Boor, Nadi-Na Ortiz-Brüchle, Andrew M Hanby, Valerie Speirs, Sara Kochanny, Akash Patnaik, Andrew Srisuwananukorn, Hermann Brenner, Michael Hoffmeister, Piet A van den Brandt, Dirk Jäger, Christian Trautwein, Alexander T Pearson, and Tom Luedde. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat Cancer*, 1(8):789–799, Aug. 2020. 2

[27] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. Apr. 2020. 4, 5, 7, 8

[28] Bin Li, Yin Li, and Kevin W Eliceiri. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. *Conf. Comput. Vis. Pattern Recognit. Workshops*, 2021:14318–14328, June 2021. 3

[29] Jana Lipkova, Tiffany Y Chen, Ming Y Lu, Richard J Chen, Maha Shady, Mane Williams, Jingwen Wang, Zahra Noor, Richard N Mitchell, Mehmet Turan, Gulfize Coskun, Funda Yilmaz, Derya Demir, Deniz Nart, Kayhan Basak, Nesrin Turhan, Selvinaz Ozkara, Yara Banz, Katja E Odening, and Faisal Mahmood. Deep learning-enabled assessment of cardiac allograft rejection from endomyocardial biopsies. *Nat. Med.*, 28(3):575–582, Mar. 2022. 2

[30] Sidong Liu, Zubair Shah, Aydin Sav, Carlo Russo, Shlomo Berkovsky, Yi Qian, Enrico Coiera, and Antonio Di Ieva. Isocitrate dehydrogenase (IDH) status prediction in histopathology images of gliomas using deep learning. *Sci. Rep.*, 10(1):1–11, May 2020. 2, 6

[31] Xiaoxuan Liu, Samantha Cruz Rivera, David Moher, Melanie J Calvert, and Alastair K Denniston. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat. Med.*, 26(9):1364–1374, Sept. 2020. 2

[32] Zhijie Liu, Wei Su, Jianpeng Ao, Min Wang, Qiuli Jiang, Jie He, Hua Gao, Shu Lei, Jinshan Nie, Xuefeng Yan, Xiaojing Guo, Pinghong Zhou, Hao Hu, and Minbiao Ji. Instant diagnosis of gastroscopic biopsy via deep-learned single-shot femtosecond stimulated raman histology. *Nat. Commun.*, 13(1):4050, July 2022. 2

[33] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. Nov. 2017. 5

[34] David N Louis, Arie Perry, Pieter Wesseling, Daniel J Brat, Ian A Cree, Dominique Figarella-Branger, Cynthia Hawkins, H K Ng, Stefan M Pfister, Guido Reifenberger, Riccardo Soffietti, Andreas von Deimling, and David W Ellison. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro. Oncol.*, June 2021. 2, 5

[35] Ming Y Lu, Richard J Chen, Jingwen Wang, Debora Dillon, and Faisal Mahmood. Semi-Supervised histology classification using deep multiple instance learning and contrastive predictive coding. Oct. 2019. 2

[36] Ming Y Lu, Richard J Chen, Jingwen Wang, Debora Dillon, and Faisal Mahmood. Semi-Supervised histology classifi-

cation using deep multiple instance learning and contrastive predictive coding. *ArXiv*, 2019. 3

[37] Ming Y Lu, Tiffany Y Chen, Drew F K Williamson, Melissa Zhao, Maha Shady, Jana Lipkova, and Faisal Mahmood. AI-based pathology predicts origins for cancers of unknown primary. *Nature*, 594(7861):106–110, June 2021. 2

[38] Ming Y Lu, Drew F K Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng*, 5(6):555–570, June 2021. 2, 3

[39] S A Luse. Electron microscopic studies of brain tumors. *Neurology*, 10:881–905, Oct. 1960. 3

[40] Marc Macenko, Marc Niethammer, J S Marron, David Borland, John T Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E Thomas. A method for normalizing histology slides for quantitative analysis. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1107–1110, June 2009. 5, 12

[41] Daniel A Orringer, Balaji Pandian, Yashar S Niknafs, Todd C Hollon, Julianne Boyle, Spencer Lewis, Mia Garrard, Shawn L Hervey-Jumper, Hugh J L Garton, Cormac O Maher, Jason A Heth, Oren Sagher, D Andrew Wilkinson, Matija Snuderl, Sriram Venneti, Shakti H Ramkissoon, Kathryn A McFadden, Amanda Fisher-Hubbard, Andrew P Lieberman, Timothy D Johnson, X Sunney Xie, Jay K Trautman, Christian W Freudiger, and Sandra Camelo-Piragua. Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated raman scattering microscopy. *Nat Biomed Eng*, 1, Feb. 2017. 2, 3, 5, 12

[42] Anil V Parwani. *Whole Slide Imaging*. Springer International Publishing. 2

[43] Linhao Qu, Siyu Liu, Xiaoyu Liu, Manning Wang, and Zhijian Song. Towards label-efficient automatic diagnosis and analysis: a comprehensive survey of advanced deep learning-based weakly-supervised, semi-supervised and self-supervised techniques in histopathological image analysis. *Phys. Med. Biol.*, 67(20), Oct. 2022. 2, 3

[44] Yair Rivenson, Hongda Wang, Zhensong Wei, Kevin de Haan, Yibo Zhang, Yichen Wu, Harun Günaydın, Jonathan E Zuckerman, Thomas Chong, Anthony E Sisk, Lindsey M Westbrook, W Dean Wallace, and Aydogan Ozcan. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat Biomed Eng*, 3(6):466–477, June 2019. 3

[45] Charlie Saillard, Olivier Dehaene, Tanguy Marchand, Olivier Moindrot, Aurélie Kamoun, Benoit Schmauch, and Simon Jegou. Self supervised learning improves dMMR/MSI detection from histology slides across multiple cancers. Sept. 2021. 3

[46] Nader Sanai, Laura A Snyder, Norissa J Honea, Stephen W Coons, Jennifer M Eschbacher, Kris A Smith, and Robert F Spetzler. Intraoperative confocal microscopy in the visualization of 5-aminolevulinic acid fluorescence in low-grade gliomas. *J. Neurosurg.*, 115(4):740–748, Oct. 2011. 3

[47] Yoni Schirris, Efstratios Gavves, Iris Nederlof, Hugo Mark Horlings, and Jonas Teuwen. DeepSMILE: Contrastive self-

supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer. *Med. Image Anal.*, 79:102464, July 2022. 2

[48] Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, and Yongbing Zhang. TransMIL: Transformer based correlated multiple instance learning for whole slide image classification. May 2021. 2

[49] Andrea Sottoriva, Inmaculada Spiteri, Sara G M Piccirillo, Anestis Touloumis, V Peter Collins, John C Marioni, Christina Curtis, Colin Watts, and Simon Tavaré. Intra-tumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci. U. S. A.*, 110(10):4009–4014, Mar. 2013. 4

[50] Karin Stacke, Jonas Unger, Claes Lundström, and Gabriel Eilertsen. Learning representations with contrastive Self-Supervised learning for histopathology applications. *Machine Learning for Biomedical Imaging*, 1(August 2022):1–10, Aug. 2022. 2

[51] Hamid Reza Tizhoosh and Liron Pantanowitz. Artificial intelligence and digital pathology: Challenges and opportunities. *J. Pathol. Inform.*, 9:38, Nov. 2018. 1, 2

[52] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *J. Mach. Learn. Res.*, 9(86):2579–2605, 2008. 5, 6, 7

[53] Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, Mauricio Rocha-Martins, Fabián Segovia-Miranda, Caren Norden, Ricardo Henriques, Marino Zerial, Michele Solimena, Jochen Rink, Pavel Tomancak, Loic Royer, Florian Jug, and Eugene W Myers. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods*, 15(12):1090–1097, Dec. 2018. 3

[54] Jenna Wiens, Suchi Saria, Mark Sendak, Marzyeh Ghassemi, Vincent X Liu, Finale Doshi-Velez, Kenneth Jung, Katherine Heller, David Kale, Mohammed Saeed, Pilar N Ossorio, Sonoo Thadaney-Israni, and Anna Goldenberg. Do no harm: a roadmap for responsible machine learning for health care. *Nat. Med.*, 25(9):1337–1340, Sept. 2019. 2

[55] Wu, Xiong, Yu, and Lin. Unsupervised feature learning via non-parametric instance discrimination. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 0, pages 3733–3742, June 2018. 2

[56] Shu Zhang, Ran Xu, Caiming Xiong, and Chetan Ramaiah. Use all the labels: A hierarchical Multi-Label contrastive learning framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16660–16669, 2022. 4

[57] Yu Zhao, Fan Yang, Yuqi Fang, Hailing Liu, Niyun Zhou, Jun Zhang, Jiarui Sun, Sen Yang, Bjoern Menze, Xinjuan Fan, and Jianhua Yao. Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4836–4845, June 2020. 3