Contextual unsupervised deep clustering in digital pathology

Mariia Sidulova MARIIA.SIDULOVA@FDA.HHS.GOV
U.S. Food and Drug Administration, USA

Seyed Kahaki SEYED.KAHAKI@FDA.HHS.GOV
U.S. Food and Drug Administration, USA

Ian Hagemann HAGEMANI@WUSTL.EDU
Washington Univ. School of Medicine, USA,

Alexej Gossmann ALEXEJ.GOSSMANN@FDA.HHS.GOV
U.S. Food and Drug Administration, USA

Abstract

Clustering can be used in medical imaging research to identify different domains within a specific dataset, aiding in a better understanding of subgroups or strata that may not have been annotated. Moreover, in digital pathology, clustering can be used to effectively sample image patches from whole slide images (WSI). In this work, we conduct a comparative analysis of three deep clustering algorithms – a simple two-step approach applying K-means onto a learned feature space, an end-to-end deep clustering method (DEC), and a Graph Convolutional Network (GCN) based method – in application to a digital pathology dataset of endometrial biopsy WSIs. For consistency, all methods use the same Autoencoder (AE) architecture backbone that extracts features from image patches. The GCN-based model, specifically, stands out as a deep clustering algorithm that considers spatial contextual information in predicting clusters. Our study highlights the computation of graphs for WSIs and emphasizes the impact of these graphs on the formation of clusters. The main finding of our research indicates that GCN-based deep clustering demonstrates heightened spatial awareness compared to the other methods, resulting in higher cluster agreement with previous clinical annotations of WSIs.

Data and Code Availability The implementation of all considered models is available in the form of a Python package with detailed documentation at https://github.com/DIDSR/DomId. All presented experiments are documented in the same repository. The WSI dataset is not publicly available at the time of writing.

Institutional Review Board (IRB) In this study, de-identified endometrial biopsy slides and clinical annotations were retrieved from Washington University School of Medicine in St. Louis (WUSM) with approval of the Institutional Review Board (IRB) and under a waiver of HIPAA consent.

1. Introduction

Clustering algorithms have proven to be a valuable tool in identifying unannotated patterns and grouping similar entities within complex underlying distributions. By understanding the previously unrecognized or hidden structures in the data, one can develop further machine learning (ML) or deep learning (DL) models that are more accurate, reliable, interpretable, and fair. Researchers have used unsupervised learning in application to digital pathology datasets. The majority of these studies have predominantly relied on unsupervised learning approaches to cluster image patches by extracting features from the patches and subsequently applying unsupervised learning algorithms to group the patches into clusters, e.g., Yao et al. (2020); Lu et al. (2021). While these methods have yielded promising results, their drawback is that cluster assignments are primarily predicated on features extracted from individual patches, and all patches are treated independently, thus neglecting spatial and contextual information from neighboring or related patches. Furthermore, it is a common drawback in previous studies to train the clustering model separately from the feature extraction model. Dividing the process in such a manner into two independently performed steps can result in a feature space that is likely not op-

timal for the clustering task. As part of this study, we demonstrate the inferior clustering performance of such two-step procedures in comparison to end-to-end deep clustering approaches. The main emphasis of this study is, however, on the utilization of Graph Convolutional Networks (GCN) as an integral component within our unsupervised deep clustering algorithm for WSI data. Our objective is to gain a better understanding of whether contextual information about the relationships between WSI patches contributes substantially to the formation of more robust and meaningful clusters within digital pathology data.

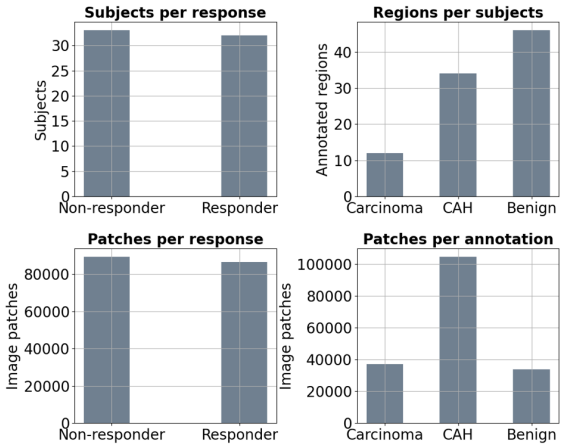


Figure 1: Summary of the data label distributions. Response to hormonal treatment is available at subject-level. Regions within WSIs were annotated as CAH, Carcinoma, and Benign.

2. Related Works

2.1. Unsupervised Learning in Digital Pathology

Unsupervised learning has been applied to pathology data from the granular pixel-level analysis to the comprehensive examination of whole-slide images. At the pixel level, unsupervised learning techniques are employed for tasks of tissue segmentation. For example, in the study Landini et al. (2019), a K-means clustering algorithm was applied at the pixel level to identify lung cancer areas in human tissue with the goal of determining the histological cancer subtype. Cheng et al. (2018) utilized a deep autoencoder (AE) to cluster cell patches into different types, and extracted features were used to characterize distributions of different types of cells in regions of interest (ROIs).

Clustering of image patches in digital pathology is commonly used for patch extraction and/or novel sampling schemes, which improve over the commonly used approaches such as random or non-overlap sampling. For example, in Yao et al. (2020), patch-level K-means clustering is applied to select important patch clusters, which subsequently are used within the method for final prediction. In Lu et al. (2021), an attention-based network is combined with clustering layers to constrain and refine the feature space with the overall goal of identifying ROIs and interpreting the important morphology used for diagnosis. Sidulova et al. (2023) proposed a conditional clustering algorithm for clustering digital pathology patches with the goal of uncovering hidden subgroups in the data. The issue of scalability of deep clustering methods in digital pathology has been highlighted in these works.

2.2. Graph-based Neural Networks in Digital Pathology

Exploring spatial relationships within digital histopathology images has shown to be useful for investigating microenvironment heterogeneity, which could potentially have important clinical implications. Graph neural networks have demonstrated desirable performance characteristics in digital pathology applications. As discussed in the paper by Ahmedt-Aristizabal et al. (2022), graph-based studies can generally be divided into three main categories based on the main objective: classification (Jaume et al., 2021), segmentation (Anklin et al., 2021) and ROI retrieval (Ozen et al., 2021). Depending on the aim of the study, researchers construct graph representations differently. Among common graph calculation methods, the Cell-Graph (CG) (Sureka et al., 2020) and Patch-Graph (PG) (Ye et al., 2019; Zhao et al., 2020) approaches have been mainly employed in classification studies; the Tissue-Graph (TG) (Anklin et al., 2021) has shown to be successful in enhancing segmentation accuracy.

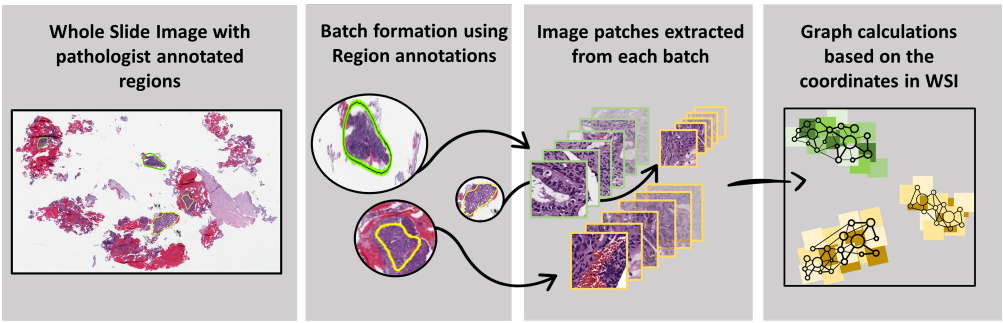


Figure 2: Summary of the patch extraction process and graph calculations for the extracted patches.

3. Methods

3.1. Structural Deep Clustering Network (SDCN), adjusted training process and batching for WSI data

SDCN (Bo et al., 2020) is a deep neural network model that combines GCN and AE architectures for the purpose of unsupervised clustering. However, in its original formulation, the algorithm exhibits substantial scalability issues preventing its application to digital pathology, where WSI sizes are in the gigapixel range or larger. This is primarily due to its requirement for a graph construction on all available data and the necessity to process all data in a single batch during training. In order to circumvent this issue, we modify the training process of SDCN and introduce a new batching approach suitable for WSI data, as described in the following.

In this work, the input to the model is a batch of patches X and the graph structure G connecting the patches in the batch. In the presented experiments, each batch contains 900 patches randomly selected from the same WSI. The methods for patch extraction and sampling on this dataset are described in Kahaki et al. (2023). In our study, each training batch contains information about one subject only. Within each batch, only three regions have been presented, each containing 300 patches. However, in order to introduce stochasticity enhancing the model's learning, new batches are sampled in each epoch of training. Specifically, for each epoch, we randomly select three regions from the pool of annotated regions available for each subject, which exceeds the number used per batch. As shown in Figure 2, a graph is constructed for each batch for the GCN layers.

To construct the graph, we connect the patches based on the patch location coordinates within the WSI, whereby each patch is connected to the 8 patches closest to it. A constraint imposed by GCN is that there cannot be any graph connections between patches from different batches (mini-batches), which are needed for optimization methods such as stochastic gradient descent or Adam. Therefore, an appropriate batching approach has to be specified based on domain knowledge – different pathology slides constitute different batches in our experiments.

We designed the GCN specifically to leverage spatial relationships between individual patches within each WSI. Given that patches from different WSIs are effectively separated by an infinite spatial distance, each WSI is treated as a distinct batch, isolating its patches as a complete unit for training. This approach results in the graph utilized by our SDCN being a collection of disjoint subgraphs, each representing a separate WSI and thus treated as an individual batch. Our definition of inter-relationship information, confined to the spatial coordinates within a single WSI, necessitates this structure.

Thus, we have proposed a modified SDCN algorithm, adapted for the digital pathology domain, and we refer to it as *SDCN + WSI batching*.

Each batch of WSI patches is passed through a convolutional neural network (CNN) based AE. The encoder of this AE has three convolutional layers with 32, 64, and 128 filters, respectively, followed by a fully connected layer. The decoder consists of transposed convolutional layers with 128, 64, and 32 filters. Additionally, ReLU activation functions and batch normalization are used within both architectures. Following Bo et al. (2020), the activations H^i of each AE encoding layer ($i = 1, 2, 3$) are then passed into the GCN layers as node features. The GCN consists of three hidden graph convolutional layers. The combination of the adjacency matrix A from the graph

G of the batch and the features H^i forms the input to each hidden layer of the GCN. The node features from the last hidden layer of the GCN are passed for cluster assignments Z to an output layer with a Softmax activation function.

The cluster assignment is achieved by the "dual self-supervised module" (Bo et al., 2020), which refers to measuring the similarity q_{ij} of the ith sample's latent representation and the jth cluster's centroid for all i, j by using Student's t-distribution as a kernel, which is then followed by calculating the target distribution p_{ij} for self-supervised training. $Q = \{q_{ij}\}_{i,j}$ and $P = \{p_{ij}\}_{i,j}$ can be viewed as distributions of the cluster assignments for all samples, and the KL-divergence $\mathcal{L}_{\text{clu}} = \text{KL}(P\|Q)$ is used as the clustering loss. In order to train the GCN, P is also used to supervise the GCN output Z via the loss term $\mathcal{L}_{\text{gcn}} = \text{KL}(P\|Z)$. The AE is trained with a conventional reconstruction loss $\mathcal{L}_{\text{rec}}$ (mean squared error). Thus, the full objective of SDCN is to minimize

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \alpha\mathcal{L}_{\text{clu}} + \beta\mathcal{L}_{\text{gcn}}.$$

The primary goal of this study is to evaluate the potential utility of GCNs in discerning meaningful clusters within digital pathology datasets. The identified clusters could subsequently aid in discerning previously unannotated subgroups or hidden subpopulations in the data. Our hypothesis posits that the integration of contextual information facilitated by GCNs may yield discernible patterns or structures within the dataset that hold substantive biological or pathological relevance. In terms of practical applications, we believe that this methodology could serve as a valuable tool for uncovering latent biases embedded within a given dataset, and thus provide valuable information for subsequent training and testing of clinical DL-based tools (e.g., diagnostic or prognostic models) on the dataset.

3.2. Dataset

In this study, we utilized a set of endometrial biopsy whole-slide images obtained from patients with complex atypical hyperplasia (CAH) and/or endometrioid endometrial adenocarcinoma (EC) from Washington University School of Medicine in St. Louis. All patients in this dataset received progestin for non-surgical treatment of endometrial disease for between 3 and 15 months, followed by a second biopsy at which the response to treatment was ascertained by routine pathology.

Each slide was scanned at 40x using an Aperio AT2 scanner to obtain a WSI. Subsequently, these WSIs were reviewed by an expert pathologist and non-exhaustively annotated to highlight CAH, carcinoma, and benign regions. Additionally, a consensus of a second pathologist has been obtained for the annotations. Moreover, for this study, we have also extracted patches from some areas outside of the regions of annotation. The patches from the outside areas were presumed to be benign in our analyses. A random sample of the outside patches was visually reviewed by a pathologist, and it was verified that the vast majority of them were benign. The data contained clinical information, including age, race, BMI, and response to hormonal treatment. Thus, every patch is assigned either a response or non-response label, depending on the treatment outcomes for the subject.

Figure 1 presents a summary of the dataset. It encompasses data from 65 subjects with one WSI per subject. Patches of size 256×256 pixels were extracted from 9 to 50 regions per WSI, including regions of annotation and regions outside, as described above, and coordinates for the exact patch location within the WSI have been recorded. The total number of patches that have been used for training is equal to $175,500$. Detailed information about the label distribution in the extracted patches can be found in Figure 1. Because the intended purpose is finding subgroups in the given dataset only (by an unsupervised method), a separate test set is not used.

3.3. Experimental Setup

In this paper, we compare three clustering algorithms: K-means in a trained deep representation space of an AE (*AE + K-means*), Deep Embedding Clustering (*DEC*) (Xie et al., 2016), and the proposed GCN based approach (*SDCN + WSI batching*). For consistency and fair comparison, all of the models use the same pre-trained AE model as their backbone (see Section 3.1), which is responsible for extracting feature representations from each patch. For K-means, the model's objective is to minimize the sum of squared distances between these features and assigned cluster centroids in the latent space. DEC (Xie et al., 2016) is a popular deep unsupervised clustering algorithm that, in an end-to-end fashion, simultaneously learns feature representations that are optimized for clustering and cluster assignments. However, DEC does not take any re-

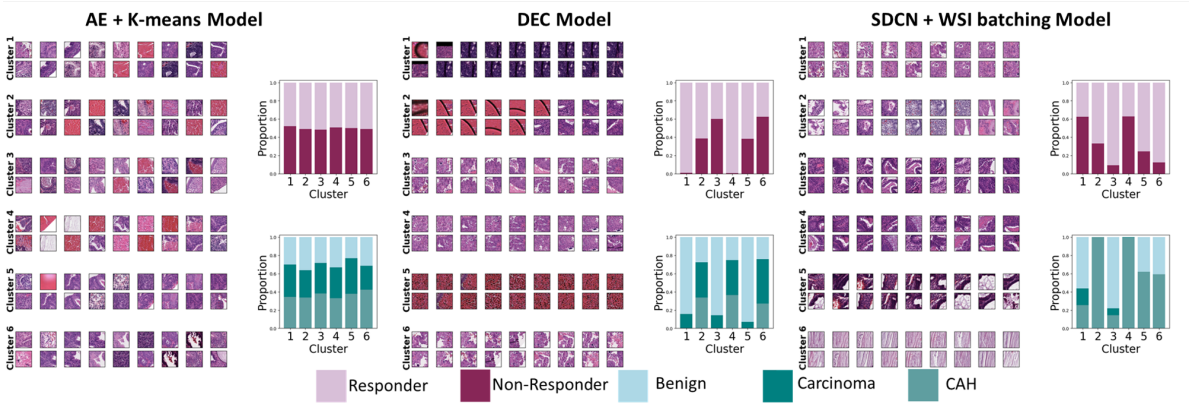


Figure 3: Summary of the patch distributions per identified cluster for the compared models. Samples of patches in each of the identified clusters and per-cluster label distributions are shown.

lationships between data samples into account (i.e., patches extracted from WSIs in this work). Finally, we compare these approaches to the modified SDCN algorithm, which is detailed in Section 3.1. While K-means and DEC rely on the direct extraction of features from individual patches for cluster formation, SDCN takes contextual information from neighboring patches into consideration to make the final cluster assignment prediction. We chose to identify 6 clusters with each model, as we have 6 combinations of response labels and annotation labels.

Three evaluation metrics were employed to evaluate the resulting cluster assignments: proportion of agreement with the annotation label, denoted as p_A(region annotation label), proportion of agreement with the response label, denoted as p_A(response label), and a proposed metric we refer to as the Contextual Patch Alignment Index (CPAI) for brevity. CPAI estimates the probability that two neighboring patches in the graph are assigned to the same cluster, and it is computed as an empirical proportion of the observations that satisfy that property. Proportions of agreement are calculated using the Hungarian algorithm to determine the assignment between the identified clusters and the ground truth labels. Due to visual similarity of the patches within annotated regions and the biological prior knowledge about the tissue samples, we hypothesised that assigned clusters should also form contiguous regions, therefore we developed the CPAI metric to measure the degree to which formed clusters are contiguous.

4. Results

In Figure 3, one can see sample patches for each of the identified clusters for each model. For a more quantitative description of the results, we have included label distributions for each of the identified clusters. The top bar graph shows the proportion of responder and nonresponder samples in each cluster; the bottom shows the proportion of different annotations – carcinoma, CAH, and benign. It could be observed that for K-Means, identified clusters do not have a strong association with either response or annotation labels. For the DEC model, clusters 1 and 4 have a greater presence of patches from the subjects who have responded to hormonal treatment in CAH/EC patients. Clusters 1, 3, and 5 have a larger presence of Benign samples. For SDCN, clusters 1 and 4 strongly associate with non-responder labels, and clusters 2 and 4 with CAH labels.

To quantify the degree of association, we measured percent agreement with region annotation labels and percent agreement with response labels, which is shown in Table 1. It could be seen that SDCN has the highest percent agreement with both annotation and response labels.

In Figures 4(a) and 4(b), one can see example cluster masks on an annotated region, and cluster masks for two randomly chosen WSIs. Evidently, the SDCN's contextual understanding is reflected by a proclivity to assign neighboring image patches to the same cluster, while the other clustering methods yield scattered or patchy patterns within areas of the WSI. This indicates that the domain-specific knowl-

	SDCN	DEC	k-means
CPAI	0.88	0.47	0.34
p_A(region annotation label)	0.60	0.29	0.20
p_A(response label)	0.51	0.30	0.20

Table 1: Quantitative assessment of clustering performance per model. *Notation:* CPAI: Contextual Patch Assignment Index (see Section 3.3), p_A: proportion agreement.

edge encoding relationships between individual image patches are effectively encapsulated within the GCN-based model. While in this example, only the spatial information is integrated into the SDCN through the graph relationships, other known interrelationships or associations between locations or areas of a WSI could be encoded in the same fashion. To investigate the impact of contextual (spatial) information on cluster assignment, we estimated the probability of two neighboring patches belonging to the same domain, denoted CPAI (see above), and it is reported in Table 1 alongside other metrics. As anticipated, this proportion is notably highest for the SDCN model, showcasing its superior spatial awareness.

5. Conclusions

In this paper, we proposed a GCN-based deep clustering approach for WSI data, utilizing the SDCN model (Bo et al., 2020) at its core. We compare it to multiple other DL-based clustering models when applied to digital pathology data at the patch level. The results reveal that the DEC and SDCN-based approaches can identify clusters that track with region annotations provided by trained pathologists, as well as treatment response labels. While the two-step approach of K-means and AE lacks strong correlations, both end-to-end DL clustering algorithms DEC and SDCN exhibit varying but pronounced relationships between specific clusters and response and annotation labels. Nevertheless, the proposed modified SDCN demonstrates superior spatial awareness in clustering neighboring image patches, resulting in an overall better clustering performance than DEC, which demonstrates the role of graph-based deep learning methods for effective cluster assignments for digital pathology applications. Thus, the presented experiments and results suggest that incorporation of the spatial local context to the model could be helpful in further identification of larger-scale features, such as carcinoma.

From a practical standpoint, our methodology holds promise as a tool for detecting and understanding hidden biases or unrecognized subgroups in a given WSI dataset, thereby offering important information for the refinement of clinical deep learning tools, such as diagnostic and prognostic models, on that data during their training or testing phases.

While our current methodology primarily leverages spatial data within individual WSIs, integration of other contextual variables, such as detailed clinical variables, additional finer-grained pathology annotations and patient outcomes as well as demographic information could be explored in future studies. We also acknowledge that adopting different types of inter-patch relationship information could alter the graph's structure, potentially leading to non-disjoint subgraphs that would challenge our batching approach. This underscores the importance of incorporating both domain knowledge and the specifics of our methodology when defining batching criteria.

Acknowledgments

The authors would like to thank Dr. Weijie Chen for helpful discussion. This project was supported in part by an appointment to the Research Participation Program at the U.S. Food and Drug Administration administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.

References

David Ahmedt-Aristizabal, Mohammad Ali Armin, Simon Denman, Clinton Fookes, and Lars Petersson. A survey on graph-based deep learning for computational histopathology. *Computerized Medical Imaging and Graphics*, 95:102027, 2022.

Valentin Anklin, Pushpak Pati, Guillaume Jaume, Behzad Bozorgtabar, Antonio Foncubierta-

Rodriguez, Jean-Philippe Thiran, Mathilde Sibony, Maria Gabrani, and Orcun Goksel. Learning Whole-Slide Segmentation from Inexact and Incomplete Labels Using Tissue Graphs. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, pages 636–646. Springer, 2021. doi: 10.1007/978-3-030-87196-3_59.

Deyu Bo, Xiao Wang, Chuan Shi, Meiqi Zhu, Emiao Lu, and Peng Cui. Structural deep clustering network. In *Proceedings of the web conference 2020*, pages 1400–1410, 2020.

Jun Cheng, Xiaokui Mo, Xusheng Wang, Anil Parwani, Qianjin Feng, and Kun Huang. Identification of topological features in renal tumor microenvironment associated with patient survival. *Bioinformatics*, 34(6):1024–1030, 2018.

Guillaume Jaume, Pushpak Pati, Behzad Bozorgtabar, Antonio Foncubierta, Anna Maria Anniciello, Florinda Feroce, Tilman Rau, Jean-Philippe Thiran, Maria Gabrani, and Orcun Goksel. Quantifying explainers of graph neural networks in computational pathology. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8106–8116, 2021.

Seyed M. M. Kahaki, Ian S. Hagemann, Kenny Cha, Christopher J. Trindade, Nicholas Petrick, Nicolas Kostelecky, and Weijie Chen. Weakly supervised deep learning for predicting the response to hormonal treatment of women with atypical endometrial hyperplasia: A feasibility study. In *Medical Imaging 2023: Digital and Computational Pathology*, page 31. SPIE, 2023. doi: 10.1117/12.2652912.

Gabriel Landini, Shereen Fouad, David Randell, and Hisham Mehanna. Epithelium and stroma segmentation using multiscale superpixel clustering. *Journal of Pathology Informatics*, 10, 2019.

Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021.

Yigit Ozen, Selim Aksoy, Kemal Kösemehmetoğlu, Sevgen Önder, and Ayşegül Üner. Self-supervised learning with graph neural networks for region of interest retrieval in histopathology. In *2020 25th International conference on pattern recognition (ICPR)*, pages 6329–6334. IEEE, 2021.

Mariia Sidulova, Xudong Sun, and Alexej Gossmann. Deep unsupervised clustering for conditional identification of subgroups within a digital pathology image set. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pages 666–675. Springer, 2023.

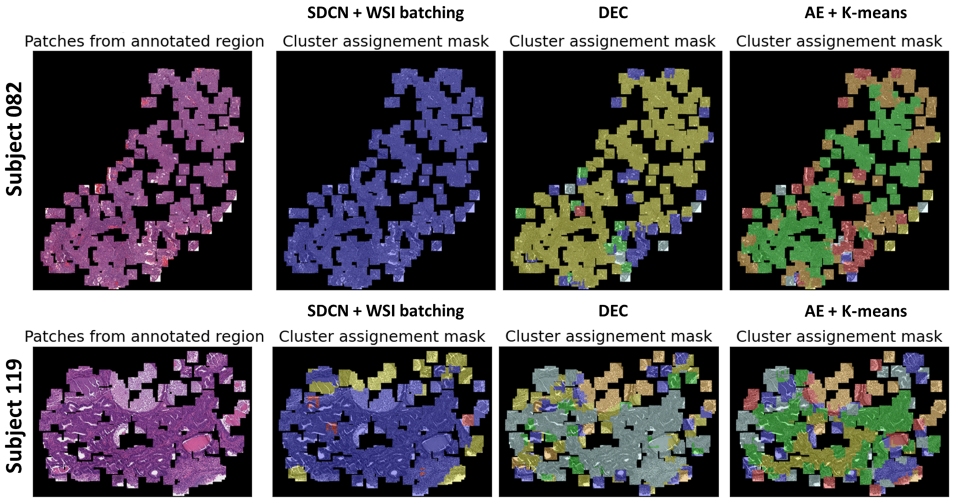
Mookund Sureka, Abhijeet Patil, Deepak Anand, and Amit Sethi. Visualization for histopathology images using graph convolutional neural networks. In *2020 IEEE 20th international conference on bioinformatics and bioengineering (BIBE)*, pages 331–335. IEEE, 2020.

Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, pages 478–487. PMLR, 2016.

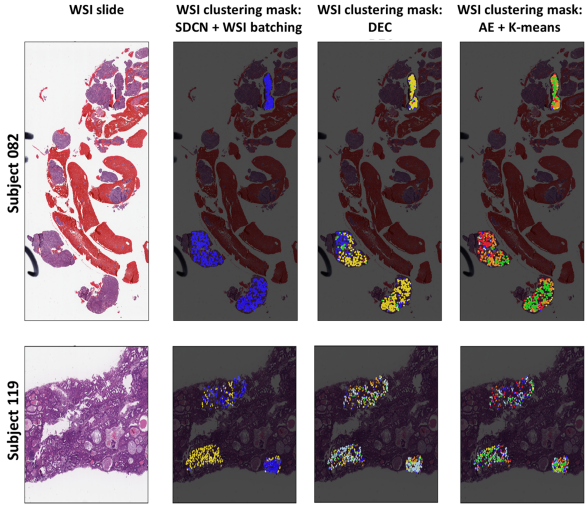
Jiawen Yao, Xinliang Zhu, Jitendra Jonnagaddala, Nicholas Hawkins, and Junzhou Huang. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. *Medical Image Analysis*, 65:101789, 2020.

Haili Ye, Da-Han Wang, Jianmin Li, Shunzhi Zhu, and Chenyan Zhu. Improving histopathological image segmentation and classification using graph convolution network. In *Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition*, pages 192–198, 2019.

Yu Zhao, Fan Yang, Yuqi Fang, Hailing Liu, Niyun Zhou, Jun Zhang, Jiarui Sun, Sen Yang, Bjoern Menze, Xinjuan Fan, et al. Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4837–4846, 2020.



(*a*) Example cluster assignment masks on WSI regions.



(*b*) Example WSIs with cluster prediction masks.

Figure 4: Sample cluster prediction masks for the studied methods overlaid on randomly chosen WSI regions (as outlined by a pathologist) and on the full WSIs. Each patch within the selected region of the WSI is categorized into clusters, with each cluster represented by a different color on the map. **(a)** The leftmost panels feature two randomly selected regions from two randomly selected WSIs. Sample regions for subject 082 and subject 119 were annotated by a pathologist as CAH, and 300 patches were extracted from each region at random locations for computational analysis. The remaining three sets of panels display maps depicting various cluster assignments by the studied models. **(b)** Two sample WSIs with cluster assignment masks. Both (a) and (b) demonstrate SDCN's capability to effectively encode domain-specific knowledge relating to the individual image patches within a WSI through the use of a graph. In this example, SDCN incorporates spatial information to yield clustering results that manifest as contiguous regions within the WSI, enhancing interpretability and potential biological or clinical relevance. In contrast, alternative deep clustering methods often result in fragmented or patchy patterns within the WSI, as observed in both sub-figures (a) and (b).