

“D3.6”

Graph representation of the histopathology knowledge

Version: 2.0

Last Update: 17/12/21

Distribution Level: *PU*

Distribution level

PU = Public,

RE = Restricted to a group of the specified Consortium,

PP = Restricted to other program participants (including Commission Services),

CO= Confidential, only for members of the ExaMode Consortium (including the Commission Services)



The ExaMode Project Consortium groups the following Organizations:

Partner Name	Short name	Country
HAUTE ECOLE SPECIALISEE DE SUISSE OCCIDENTALE	HES-SO	Switzerland
UNIVERSITA DEGLI STUDI DI PADOVA	UNIPD	Italy
SIRMA AI	SIRMA AI	Bulgaria
STICHTING RADBOUD UNIVERSITAIR MEDISCH CENTRUM - RADBOUDUMC	RUMC	Netherlands
MICROSCOPEIT SP ZOO	MICROSCOPEIT	Poland
AZIENDA OSPEDALIERA PER L'EMERGENZA CANNIZZARO	AOEC	Italy
SURF	SURFSARA BV	Netherlands

Document Identity

Creation Date:	24/11/2021
Last Update:	17/12/2021

Revision History

Version	Edition	Author(s)	Date
0	1	Niccolò Marini	24/11/2021
Comments:	First draft of the deliverable		
0	2	Marek Wodzinski	29/11/2021
Comments:	Completion of the first draft of the deliverable		
0	3	Niccolò Marini	1/12/2021
Comments:	Final version of the deliverable		
1	0	Manfredo Atzori	1/12/2021
Comments:	Proof check of the content of the deliverable		

1	1	Todor Primov	5/12/2021
Comments:	Internal review comments		
1	2	Svetla Boytcheva	9/12/2021
Comments:	Internal review comments		
2	0	Niccolò Marini	17/12/21
Comments:	Revision		

Executive summary

Histopathology involves the analysis of tissue samples in order to identify diseases, such as cancer. Computational pathology is a domain aiming to develop algorithms to automatically analyze gigapixel digitized histopathology images, called whole slide images (WSI). WSIs are high-resolution images (in the order of $100'000^2$ pixels per image) produced by scanning stained tissue samples. Convolutional neural networks (CNNs) are the state-of-the-art algorithm for most of the tasks involving the automatic analysis of WSIs. Even though they allow to reach high performance for several tasks, such as classification or segmentation, the training of CNNs is hindered by several factors, such as the variability of WSI staining across different medical centres; the lack of large datasets with local annotations for training and testing; and the high variability in the morphology of tissues. These challenges limit the development of CNN models that can generalize well to data acquired in diverse medical centres.

The objective of this deliverable is to develop a graph representation of the histopathology visual knowledge, focusing on the first ExaMode use case (colon cancer).

A morphology-based representation of the visual knowledge of colon WSIs is obtained by training an H&E-adversarial with classes linked to high-level colon morphologies (malignant, dysplasia, glands, non-informative tissue).

The morphology-based representation is used to build a graph-representation of the visual histopathology knowledge, that guarantees separability between the classes and shows the relationship among them. The graph representation of the visual knowledge is obtained by computing the adjacency matrix between the nodes obtained after hierarchical clustering. The histopathology visual knowledge graph represents the relationships between the color invariant content of the histopathology images. A visual representation that is stain-invariant and that corresponds to the most relevant findings detected by pathologists may help to learn robust features that can be re-used to solve specific tasks in the computational pathology domain. The importance of the graph representation is fundamental to allow linking with other datasets in future applications and to allow creating in Deliverable 3.7 a link with text-based ontologies developed in WP2.

The morphology-based representation of the visual knowledge of colon WSIs is as well applied to two concrete problems in order to verify its capability to improve algorithms that are often developed in the domain of digital pathology. First, it is used to automatically annotate specific classes in unlabelled WSIs and, second, to remove info-informative patches during the training of a Multiple Instance Learning (MIL) CNN. The methods are evaluated using colon cancer WSI, provided by the proprietary hospitals and publicly available datasets.

Table of Contents

1	Introduction	6
2	Methods	10
2.1	Dataset	10
2.2	Morphology-based representation	12
2.3	Graph representation of visual histopathology knowledge	13
2.4	Morphology-based representation applications	14
2.4.1	Application 1: Pseudo-labeling of colorectal WSIs for patch classification	14
2.4.2	Application 2: Filtering of WSIs non-informative patches for colorectal WSIs classification	14
3	Results	15
3.1	Graph representation of visual histopathology knowledge	15
3.2	Morphology-based representation applications	31
3.2.1	Application 1: Pseudo-labeling of colorectal WSIs for patch classification	31
3.2.2	Application 2: Filtering of WSIs non-informative patches for colorectal WSIs classification	31
4	Discussion & Conclusion	32
5	References	33
6	Annexes	37
6.1	Semi-supervision	37
6.2	Multiple Instance Learning	37
6.3	Metrics	38
6.4	Hyperparameters and implementation	38
6.5	Graph representation of visual histopathology knowledge with 25 clusters	38

List of abbreviations

CNN	Convolutional Neural Network
WSI	Whole Slide Image
MIL	Multiple Instance Learning
HGD	High-grade dysplasia
LGD	Low-grade dysplasia

1 Introduction

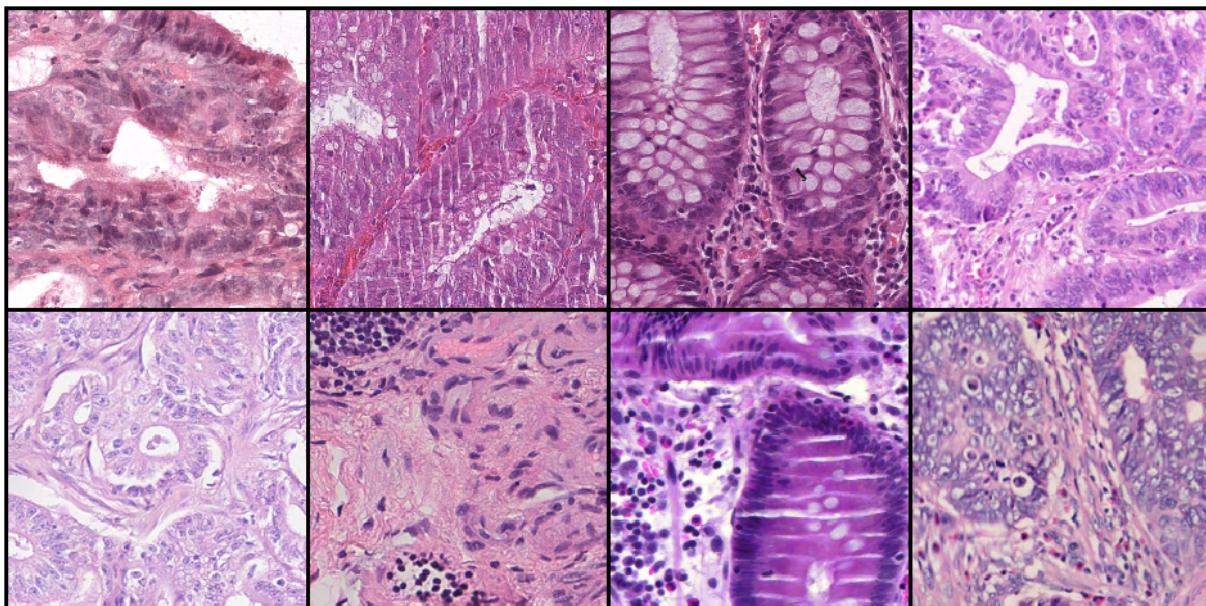
The digitalization of histopathology workflows, along with the advancements of deep learning, is paving the way to the development of algorithms that can learn from clinical data. Even though several challenges still remain.

Histopathology represents the gold standard domain for the diagnosis of several diseases, such as cancer¹. The domain involves the analysis of tissue specimens collected through biopsies or tissue resections and analyzed by a medical expert (pathologist) in order to identify manifestations of possible diseases. The analysis of samples is time-consuming (it can also last a few hours per image²) due to the variability of tissue morphologies, the experience, the subjective evaluation of findings made by pathologists³ and the regions selected to be more carefully evaluated. Furthermore, the inter-pathologist agreement is generally low⁴⁻⁷. Digital pathology involves the acquisition and the management of digitized tissue samples, defined as whole slide images (WSI). The images are acquired via machinery called whole slide scanner, with an optical magnification usually in the order of 20-40x (spatial high-resolution of 0.25-0.5 µm per pixel)^{8,9}. During the acquisition, several magnification levels of the images are stored to zoom from the lowest to the highest levels and to allow pathologists to detect and analyze different details of the images.

Computational pathology is a relatively recent domain that aims to develop computer-based tools for the automatic analysis of WSIs. In the last years, Convolutional Neural Networks (CNNs) reached high performance in several tasks, such as histopathology images classification and segmentation, emerging as the state-of-the-art method to solve several computational pathology tasks. However, despite an increasing number of applications and methods yearly developed, the training of CNNs is challenging due to several factors: the WSI colour variability; the lack of large datasets with local annotations; the high variability in tissue morphology.

The colour (stain) variability involves a lack of standardization in the preparation and acquisition procedures across different medical centers¹⁰⁻¹², particularly the reagents to stain the images, the model of scanner adopted, and tissue thickness. Stain heterogeneity leads to low model generalization on data acquired in different contexts from those used to train the models¹³⁻¹⁷. The most common staining used in digital pathology is based on Eosin and Hematoxylin (H&E)¹⁸, even though standardization processes are being developed there is no standard procedure for the concentration of the H&E solution¹⁹. The chemicals are applied for making the cells and their parts visible. The Eosin stains the cytoplasm and extracellular structures with shades of pink, while the Hematoxylin stains the cellular nuclei with shades of blue. Therefore, using different concentrations, the nuclei are stained with heterogeneous shades of blue and the cytoplasm with heterogeneous shades and saturations of pink. It is possible to observe the colour heterogeneity in Figure 3. A homogeneous representation of colour is essential for algorithms that depend on colour information²⁰. An example of stain variability is shown in Figure 1.

Colon



Prostate

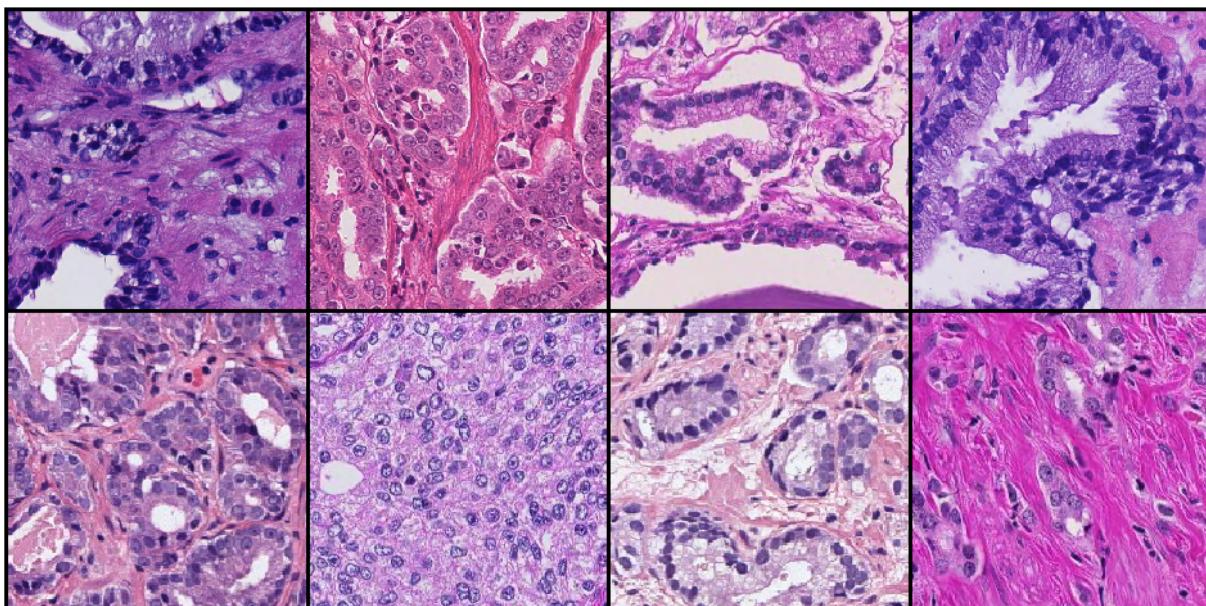


Figure 1. Example of colour variations in WSIs. The patches come from both proprietary and publicly available repositories. The above part of the Figure includes colon tissue, while the bottom one is prostate tissue.

Local annotations (also known as strong labels) are annotations made by a pathologist at pixel- or patch-level^{21,22}. Since a medical expert is required, local annotations production is time-consuming and expensive^{23,24}. Large datasets with locally-annotated images are needed to train CNNs in a fully-supervised fashion²⁵. Currently, fully-supervision shows the highest performance. WSIs are gigapixel images; therefore, having manually annotated data available can improve the learning process, focusing the analysis only on a portion of the image.

The high variability in tissue morphology involves the heterogeneous appearance of the tissue structures, such as the gland's shape or the cellular shapes. The heterogeneous appearance leads to a low inter-pathologist agreement⁴⁻⁷. For example, in Arvaniti et al.⁴, two pathologists reach Cohen's Kappa score $\kappa = 0.67$ in the diagnosis of prostate Gleason grading (four classes) and $\kappa = 0.67$ in the diagnosis of prostate Gleason scoring (six classes), in Vennalaganti et al.⁶ seven pathologists reach $\kappa = 0.43$ in the diagnosis of low-grade dysplasia in patients with Barret's esophagus (three classes), in Turner et al.⁷ 28 pathologists reach $\kappa = 0.45$ in the diagnosis of colon polyp type (five classes) and $\kappa = 0.67$ in the diagnosis of dysplasia grade (two classes); in Costantini et al.⁵ 4 pathologists reach $\kappa = 0.34$ in the diagnosis of colon adenoma histologic type (3 classes) and $\kappa = 0.54$ in the diagnosis of dysplasia grade (two classes). The fact that the inter-pathologist agreement is low also creates semantic problems in the image annotation procedure. For example, in publicly available colon datasets, such as GlaS²⁶, CRC²⁷, UNITO²⁸, Xu²⁹, the class "normal" defines concepts such as normal mucosa glands, stroma, normal submucosa. Even though these classes do not include any dangerous conditions for the patient, the morphologies they refer to vary (the stroma usually does not include glands).

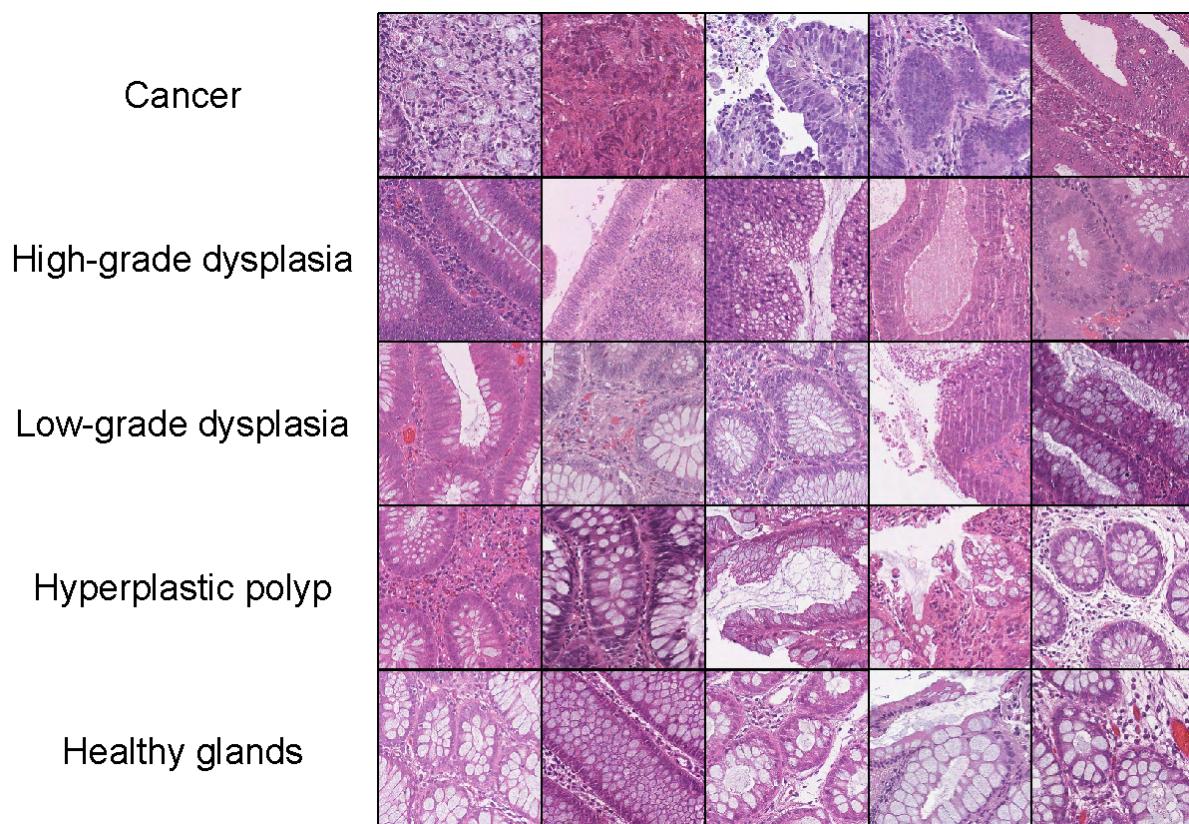


Figure 2. An example of the high variability of tissue morphology for five colon tissue conditions (cancer, high-grade dysplasia, low-grade dysplasia, hyperplastic polyp, normal glands). For each condition, five examples are shown.

This deliverable aims to present a graph representation of the visual histopathology knowledge for the first ExaMode use case, i.e. colon cancer, aiming at alleviating the challenges presented in CNN training and allowing linking with text-based ontologies in future deliverables. The visual representation is stain-invariant, based on the most informative morphologies examined by pathologists during the tissue analysis and obtained by combining heterogeneous datasets to target the image variability.

The morphology-based representation is learnt using an H&E-adversarial CNN¹² trained in a fully-supervised fashion, using patches from several datasets, re-annotated with classes corresponding to the

morphologies analyzed by a pathologist. The patches include colon tissue, from two proprietary datasets - i.e. Azienda Ospedaliera Cannizzaro, Catania, Italy (AOEC) and Radboud Medical University Center, Nijmegen, The Netherlands (Radboudumc) - and from AIDA³⁰.

The morphology-based representation is visualized through a graph that allows verifying the property of the representation, such as the separability between the classes and the relationship between different classes. The graph-representation includes nodes (classes) and edges (relationship between classes). The nodes include patches with similar characteristics and belonging to the same class, while the edges are weighted with a similarity metric, showing a stronger relationship to classes sharing similar characteristics, such as high-grade and low-grade dysplasia.

In order to show how the learned representation can be used in different tasks, the morphology-based representation is applied to create pseudo-annotations and filter the non-informative patches during a Multiple Instance Learning CNN training. In both cases, the methods are compared with the performance of a CNN trained with Multiple Instance Learning.

Colon was selected as the first ExaMode use case study because of the difficulty to diagnose it and its impact on society. Colon is one of the most commonly diagnosed cancer in the world³¹. A 75% increase in diagnosis is predicted by 2040 for both genders and an extensive age range of ages³². The diagnosis of colon cancer involves the identification of malignant polyps³¹ (i.e. agglomerations of cells in the colon surface). It can include many classes, such as *adenocarcinoma*, *high-grade dysplasia* (HGD), *low-grade dysplasia* (LGD), *hyperplastic polyp* and *normal glands*. Data are collected by two private hospitals and publicly available datasets.

2 Methods

2.1 Dataset

The data used in this work are collected with three objectives. Firstly, data with pixel-wise annotations are provided, in order to train a CNN to learn relevant tissue morphologies which are used to create the morphology-based representation and the graph-representation of the visual histopathology knowledge. Secondly, weakly-annotated data are provided, in order to evaluate the morphology-based representation application to pseudo-labeling (including weak annotations), combining predicted morphologies and weak annotations. Thirdly, weakly-annotated data are used as well in a MIL CNN training, using the learnt morphologies to filter non-informative patches.

Source	Class					
	Cancer	High-grade dysplasia	Low-grade dysplasia	Hyperplastic polyp	Normal glands	Total images
AOEC	362	454	716	186	34	1543
Radboudumc	42	30	211	131	518	915
Total	352	440	873	294	552	2310
AIDA ³⁰	31	0	4	1	65	101
CRC ²⁷	69	0	0	0	71	140
UNITO ²⁸	0	1370	5804	545	950	8669
Xu ²⁹	355	0	0	0	362	717
Public dataset	454	1370	5808	546	1448	9627

Table 1. **Overview of the dataset composition.** The dataset includes colon images and reports from digital pathology workflows (AOEC and Radboudumc) and publicly available datasets. The table reports the number of WSI for each class (corresponding to the weak annotations).

Correspondingly to the objectives, data includes three main sets: the first consisting of strongly-annotated patches, coming from private and publicly available datasets including AOEC, Radboudumc and AIDA, which were re-annotated using macro classes linked to tissue morphologies (Table 1); the

second including pseudo-labeled patches, coming from private data (Table 2); and the third including weakly-annotated WSIs (Table 3).

Dataset	Malignant	Dysplasia	Polyps	Non Informative	Total
Training set					
AOEC	4682	39841	7763	9548	61834
Radboudumc	2333	2049	1199	2717	8298
AIDA	7881	3065	28691	72050	111687
Validation set					
AOEC	393	5952	1496	979	8820
Radboudumc	180	741	175	227	1323
AIDA	0	231	3168	8504	11903

Table 2. Overview of the strongly-annotated patches used to train the CNN. The annotations are mapped to four classes: malignant, including carcinomas, dysplasia, including high-grade and low-grade dysplasia, polyps, including normal glands and hyperplastic polyps, non-informative, including stroma, subserosa and submucosa.

The first set (Table 1) represents the morphologies searched by pathologists during colon WSIs analysis. Data are collected from private hospitals (AOEC and Radboudumc) and AIDA³⁰, a publicly available dataset. The strong-annotations, that include several classes, are mapped to macro classes: “malignant” includes different types of carcinomas, “dysplasia” includes high-grade and low-grade dysplasia, “polyps” includes well-shaped polyps such as hyperplastic polyps and normal glands, “non-informative” includes stroma, submucosa, subserosa and all the morphologies not linked to dangerous diseases. The data

The second set (Table 2) includes pseudo-labels, combined with the CNN trained predictions to classify morphologies and the weak annotations. The WSIs from AOEC and Radboudumc are pseudo-labeled.

The third set (Table 3) includes the WSI coming with weakly-annotations. The images coming from AOEC (except for a subset) and Radboudumc are used to train the MIL CNN, while the images coming from publicly available datasets are used to test the CNN (in combination with the AOEC test set).

Dataset	Cancer	HGD	LGD	Hyperplastic	Normal glands	Total
Training set						
AOEC	31638	30550	37627	9163	4049	113027
Radboudumc	3362	1119	10704	2525	20962	38672
Validation set						
AOEC	4537	3057	4763	459	40	12856
Radoudumc	463	759	3421	481	1210	6334

Table 3. Overview of the pseudo-labeled patches, coming from AOEC and Radboudumc.

2.2 Morphology-based representation

The morphology-based representation of visual colon histopathology knowledge is based on the procedure performed by pathologists during the analysis of samples. The analysis involves the identification of polyps, e.g. lesions of the colon surface. According to the characteristics of the lesion, it is possible to differentiate between different conditions, leading to different diseases and diagnosis³³.

The procedure's steps are:

1. The identification of the presence of polyps.
2. The identification of the type and the number of polyps. The two most important types of polyps are adenoma-serrated polyps and malignant polyps.
3. The identification of dysplasia and its grade (low-, medium- and high-grade) and the annotation of details about pathologic abnormalities.
4. The identification of malignant polyps and their features (tumour type, histological tumour grade, lymphovascular invasion and margin involvement)

Considering the steps in the procedure, the most relevant morphologies identified by pathologists are cancer, dysplasia, polyps. Therefore, to build a robust image representation is necessary to have features that allow distinguishing between those morphologies.

In order to achieve this goal, the strongly-annotated patches from AOEC, Radboudumc and AIDA, are re-annotated using the following macro-classes: malignant (including carcinomas and

adenocarcinomas), dysplasia (including high- and low-grade dysplasia), glands (including hyperplastic polyps and healthy glands) and non-informative (including stroma, submucosa, subserosa). The patches numbers are reported in Table 1. The re-annotated patches are used to train a CNN in a fully-supervised fashion. The image representation learnt is evaluated on two applications, to analyze if it is possible to use it to solve different pathologist tasks. Furthermore, the CNN is trained through an H&E-adversarial method¹², to learn stain-invariant features.

2.3 Graph representation of visual histopathology knowledge

Obtaining a graph representation of the visual histopathology knowledge is the aim of the deliverable. This result can be beneficial to better understand how the CNNs perform decisions and how to create an ontology of the visual features, in order to link it with the text-based ontology.

The graph representation allows verifying the presence of representation properties, such as the separability between different classes and the relationships between them, modelling the nodes as the classes (in this case cancer, high-grade dysplasia, low-grade dysplasia, hyperplastic polyp and normal glands) and the edges as the weighted relationship between them.

The five classes are chosen considering the availability of pixel-wise annotations, allowing to assign a concept to the node of the graphs, through the comparison between the tissue included in the nodes and the annotations. The graph representation is generated at patch-level, allowing to identify relevant morphologies, allowing to identify the relationships between the components included within a WSI. These components are usually findings leading to a global diagnosis, described in the reports, and used to generate the textual ontology. The patch-level graph representation allows to identify the different components included within a WSI (intra-WSI), that can be separated according to their morphology. The representation may be further extended at case-level (inter-WSI), allowing to link the global concepts identified within the reports with the WSIs. The extension of the representation at image-level may directly link the textual ontologies from the reports to the images, allowing to query images from the textual concepts and then exploring the graph at the patch level. Thus, a heterogeneous graph can be created to link the image-based ontology to the text-based ontology, provided that the text and image latent spaces are directly related.

In order to create the graph representation the following steps are performed: (i) the embedding calculation of each strongly annotated patch, (ii) the hierarchical clustering of the embeddings with a predefined number of clusters, (iii) the calculation of each cluster distribution with respect to the strong annotations, (iv) calculation of the distance and the centroids of each cluster, (v) the calculation of distance between each cluster centroid to create the adjacency matrix and the graph representation, (vi) the dimensionality reduction of the latent space using t-SNE (t-Distributed Stochastic Neighbor embedding) and PCA (Principal Component Analysis).

The model used for calculating the embeddings is the encoder of Resnet34, trained to classify colorectal morphologies (as described in Paragraph 2.2). The hierarchical clustering is performed directly in the latent space by agglomerative clustering with a predefined number of clusters (5 and 25) using the Euclidean distance and the ward linkage. The distribution within a given cluster is calculated with respect to the strong annotations of the patches. To address the class imbalance, the distributions are normalized with respect to the total number of patches of a given class. The centroid for each cluster is calculated in order to calculate the distances between the clusters. Then, the adjacency matrix is calculated and the graph representation is created. In order to enable a visual representation of the latent space the latent space dimensionality is reduced to two dimensions. The dimensionality reduction is calculated using t-SNE (2 dimensions) initialized by the PCA results (100 dimensions) and the perplexity set to the square root of the number of patches. t-SNE is a technique used for non-linear dimensionality reduction (2 dimension) of high-dimensional data, such as the embeddigs generated in

this work (128 elements). T-SNE aggregates similar samples, identifying similar examples and projecting in a new low-dimensional space (usually 2 or 3 dimensions). The projection in the new space involves the minimization of the Kullback-Leibler divergence of the joint probabilities among the low- and the high-dimensional data embeddings. The PCA algorithm is used to create an intermediate representation (embeddings with 100 elements) before the application of the method, in order to reduce the resources needed to generate the new space.

2.4 Morphology-based representation applications

2.4.1 Application 1: Pseudo-labeling of colorectal WSIs for patch classification

The first application where the image representation is applied is a semi-supervision task, using the CNN trained to classify morphologies to pseudo-label weakly-annotated WSIs (from AOEC and Radboudumc).

The annotation task involves five classes: cancer, high-grade dysplasia, low-grade dysplasia, hyperplastic polyp and healthy glands, while the CNN predicts four classes (malignant, dysplasia, polyps and non-informative) linked to high-level morphologies.

The weak labels of the WSIs are used to fill the semantic gap between the morphology labels (high-level) and the five classes (low-level), through the following rules:

1. A patch is pseudo-labeled as cancer if 1) the CNN classifies the patch as malignant and the patch come from a WSI weakly-annotated with cancer
2. A patch is pseudo-labeled as high-grade dysplasia if 1) the CNN classifies the patch as dysplasia and comes from a WSI weakly-annotated with high-grade dysplasia. It is discarded if a patch comes from a WSI weakly-annotated with both high-grade and low-grade dysplasia.
3. A patch is pseudo-labeled as low-grade dysplasia if 1) the CNN classifies the patch as dysplasia and comes from a WSI weakly-annotated with low-grade dysplasia. It is discarded if a patch comes from a WSI weakly-annotated with both high-grade and low-grade dysplasia.
4. A patch is pseudo-labeled as a hyperplastic polyp if 1) the CNN classifies the patch as a polyp and comes from a WSI weakly-annotated with a hyperplastic polyp. It is discarded if a patch comes from a WSI weakly-annotated with both hyperplastic polyp and healthy glands.
5. A patch is pseudo-labeled as healthy glands if 1) the CNN classifies the patch as a polyp and the patch comes from a WSI weakly-annotated with healthy glands. It is discarded if a patch comes from a WSI weakly-annotated with both hyperplastic polyp and healthy glands.

The pseudo-labeled data are used to train another CNN at patch-level.

2.4.2 Application 2: Filtering of WSIs non-informative patches for colorectal WSIs classification

The second application where the image representation is tested is a Multiple Instance Learning task, using the CNN trained to classify morphologies to filter the weakly-annotated WSIs, both at training and test time. The morphology-based CNN is used to remove clusters of patches labeled as non-informative from the WSI bag, reducing the noise introduced by the weak annotations.

3 Results

3.1 Graph representation of visual histopathology knowledge

The visualization of the latent space is shown in Figure 3. It presents the exemplary patches in the 2-D latent space after the dimensionality reduction using PCA followed by T-SNE. In Figure 4 the patches are presented as points with the color representing the corresponding strong annotation. It can be observed that different types of patches are being aggregated in different regions of the latent space.

The clustering results for five clusters in the 2D latent space are shown in Figure 5. The clusters are encoded in color representing the most influential class, apart from a single, central cluster for which the strong annotations are not well separated. The corresponding dendrogram is shown in Figure 6. Exemplary images together with the cluster distributions are shown in Figures 7-16. They show 25 randomly chosen images within each cluster and the percentage of cases from a given class (with respect to all occurrences of a given class) in the selected cluster. Finally, the graph representation of visual features is shown in Figure 17.



Figure 3. Visualization of the WSI patches distributed in the latent space with dimensionality reduced using the T-SNE. Note that similar types of patches are close to each other.

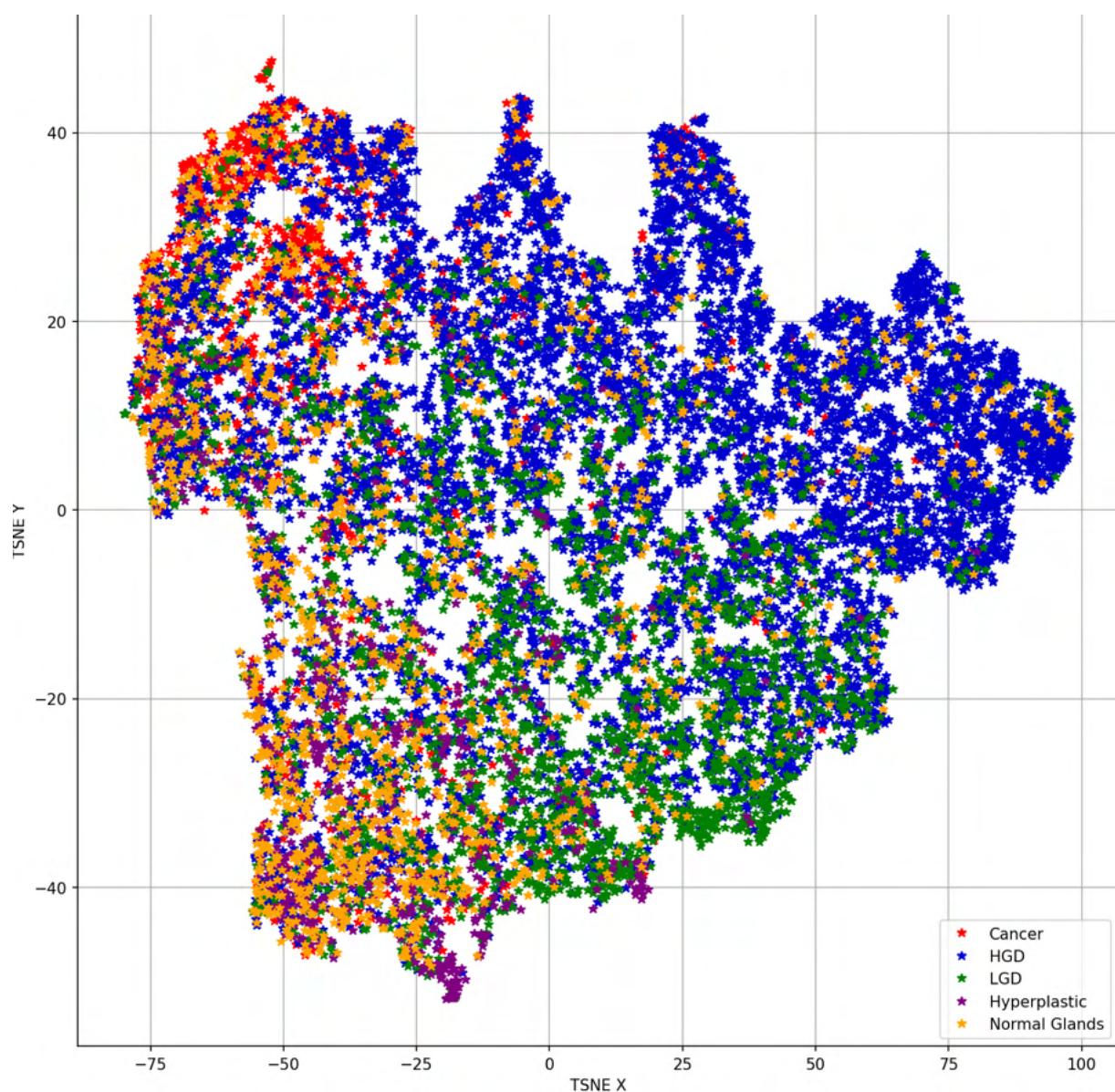


Figure 4. Distribution of the strong annotations within the latent space with dimensionality reduced using T-SNE.

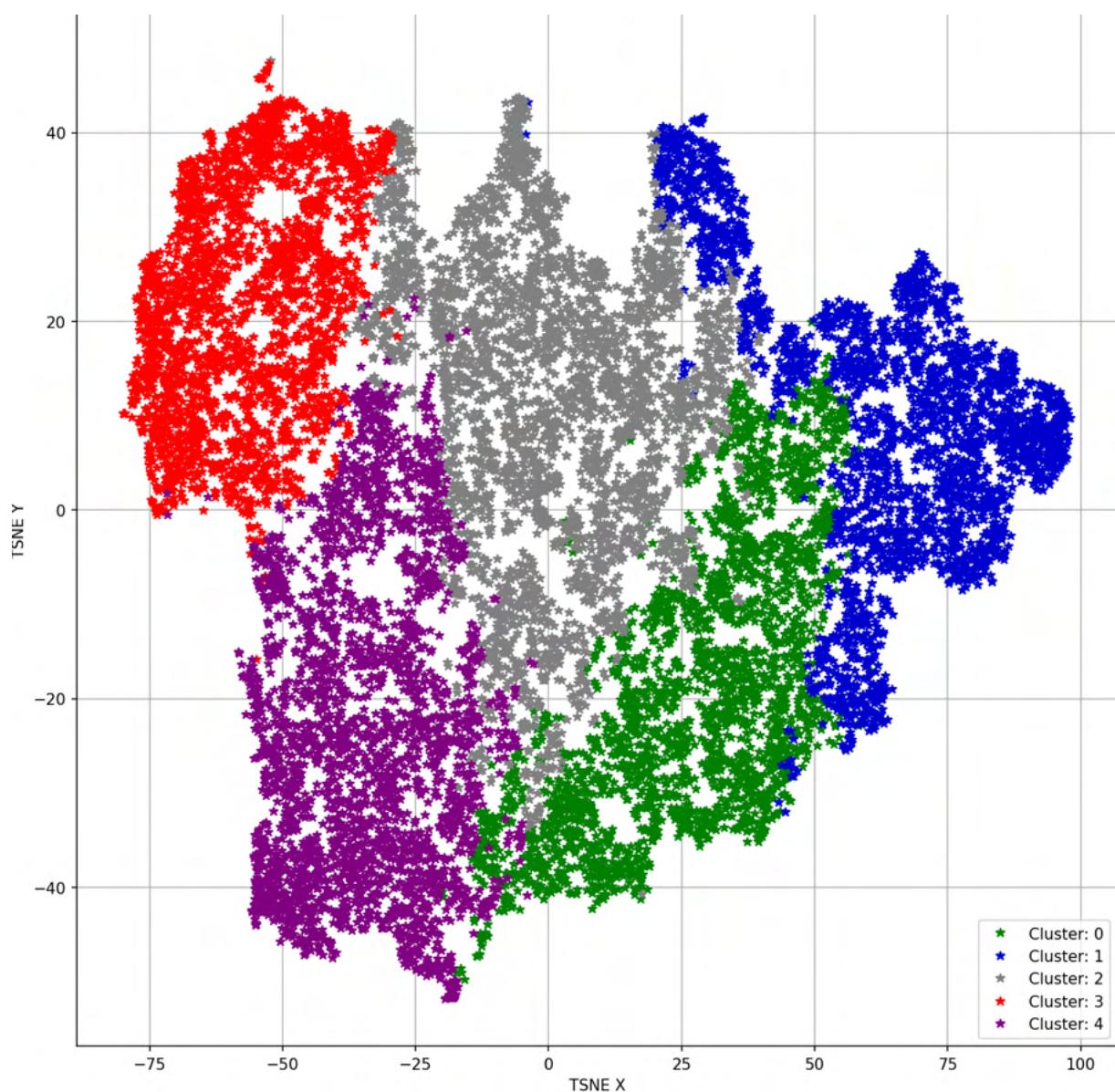


Figure 5. Visualization of the cluster distribution (hierarchical clustering, ward linkage, Euclidean distance) in the latent space. Please note that the clusters were calculated before the dimensionality reduction. in red: cluster 3, cancer; in purple, cluster 4, hyperplastic polyps and normal glands; in grey, cluster 2, cases from all the classes; in green, cluster 0, low-grade dysplasia; in blue, cluster 1, high-grade dysplasia.

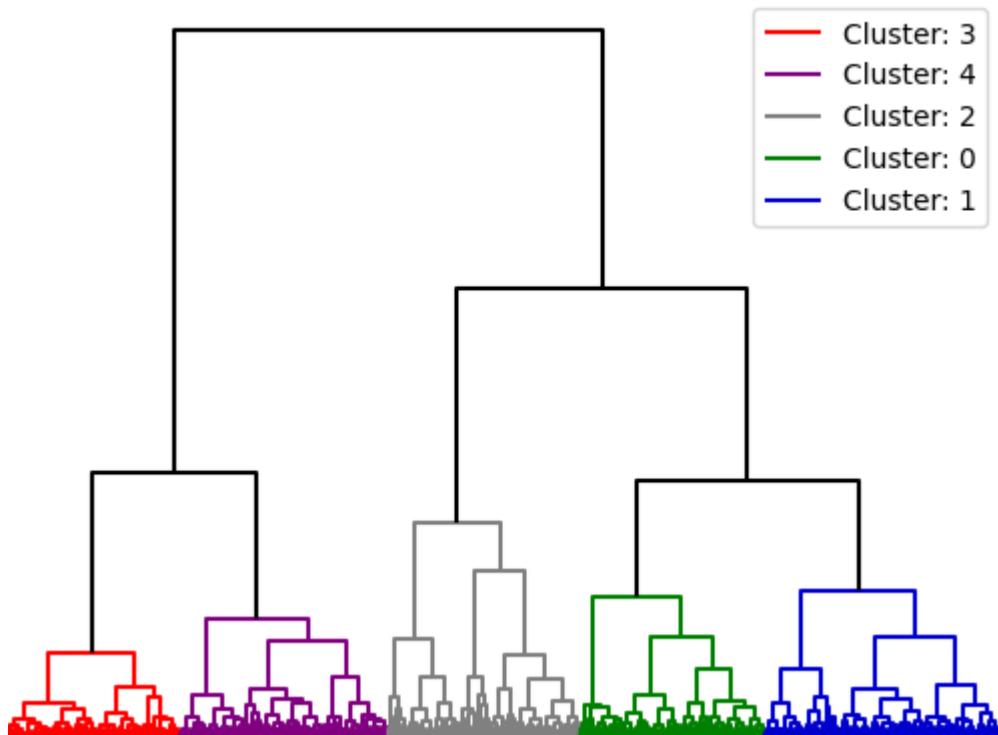


Figure 6. Dendrogram representing the structure of the latent space. From the left to the right: in red: cluster 3, cancer; in purple, cluster 4, hyperplastic polyps and normal glands; in grey, cluster 2, cases from all the classes; in green, cluster 0, low-grade dysplasia; in blue, cluster 1, high-grade dysplasia.

Cluster: 0

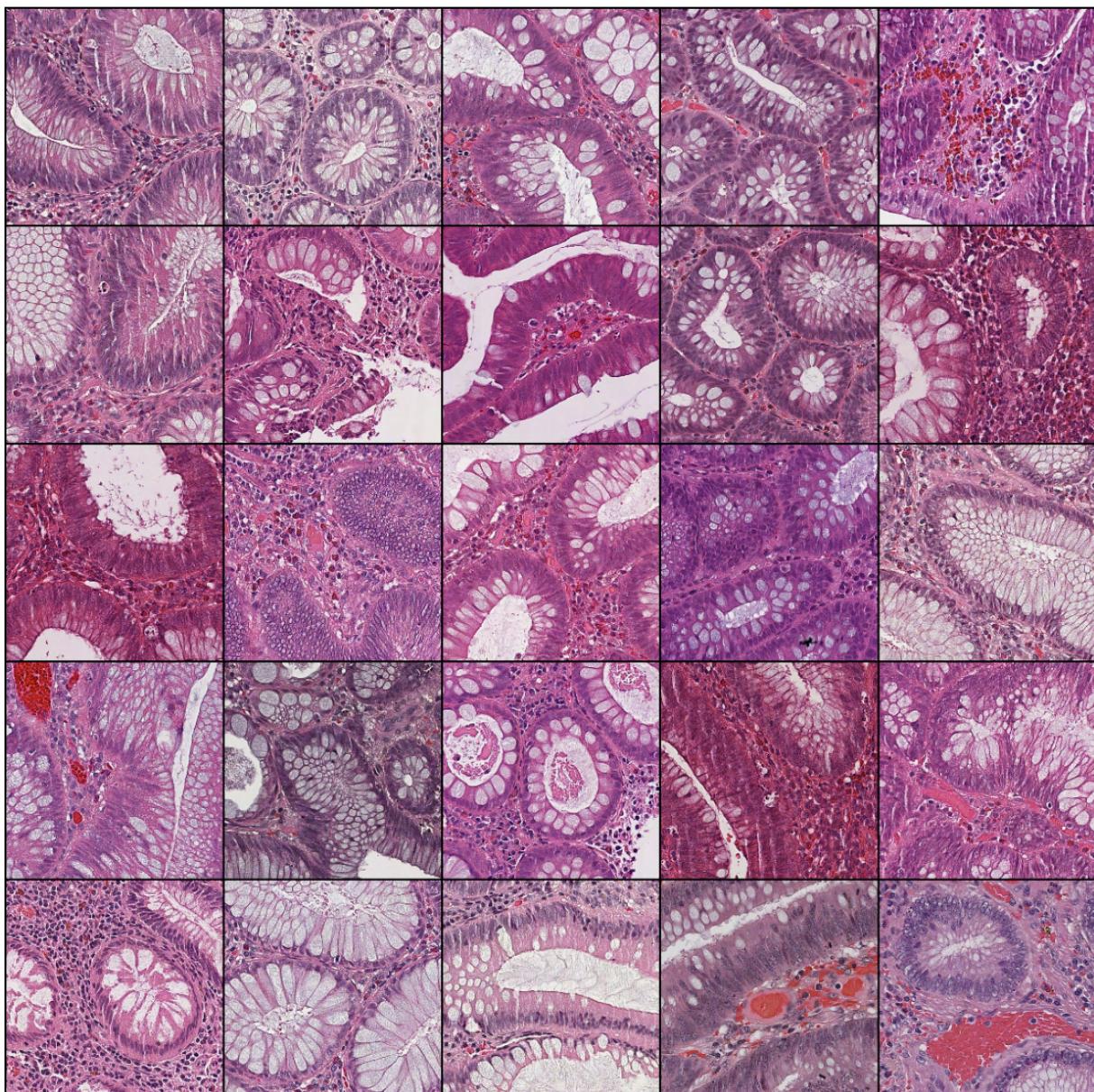


Figure 7. Exemplary patches from Cluster 0 (represented mostly by LGD).

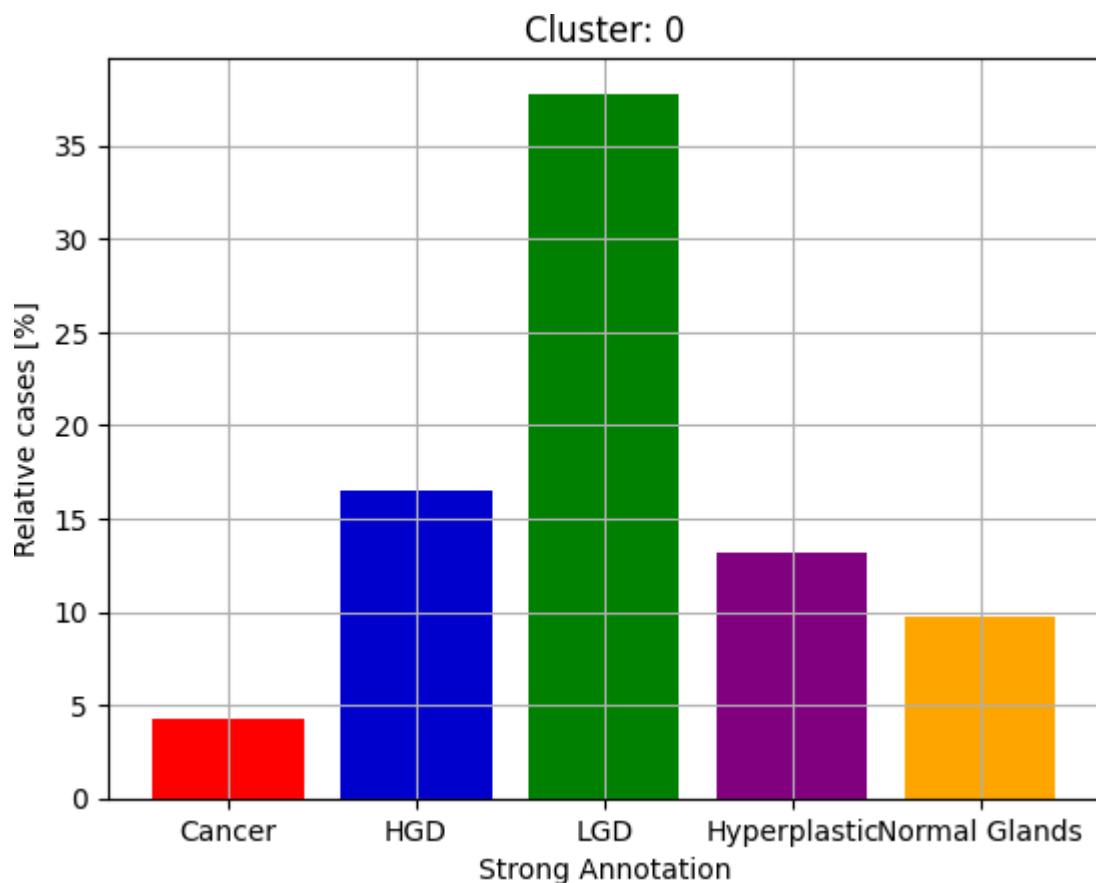


Figure 8. Distribution of the patches in Cluster 0. Please note that the number of relative cases is calculated with respect to all cases of a given class to account for class imbalance.

Cluster: 1

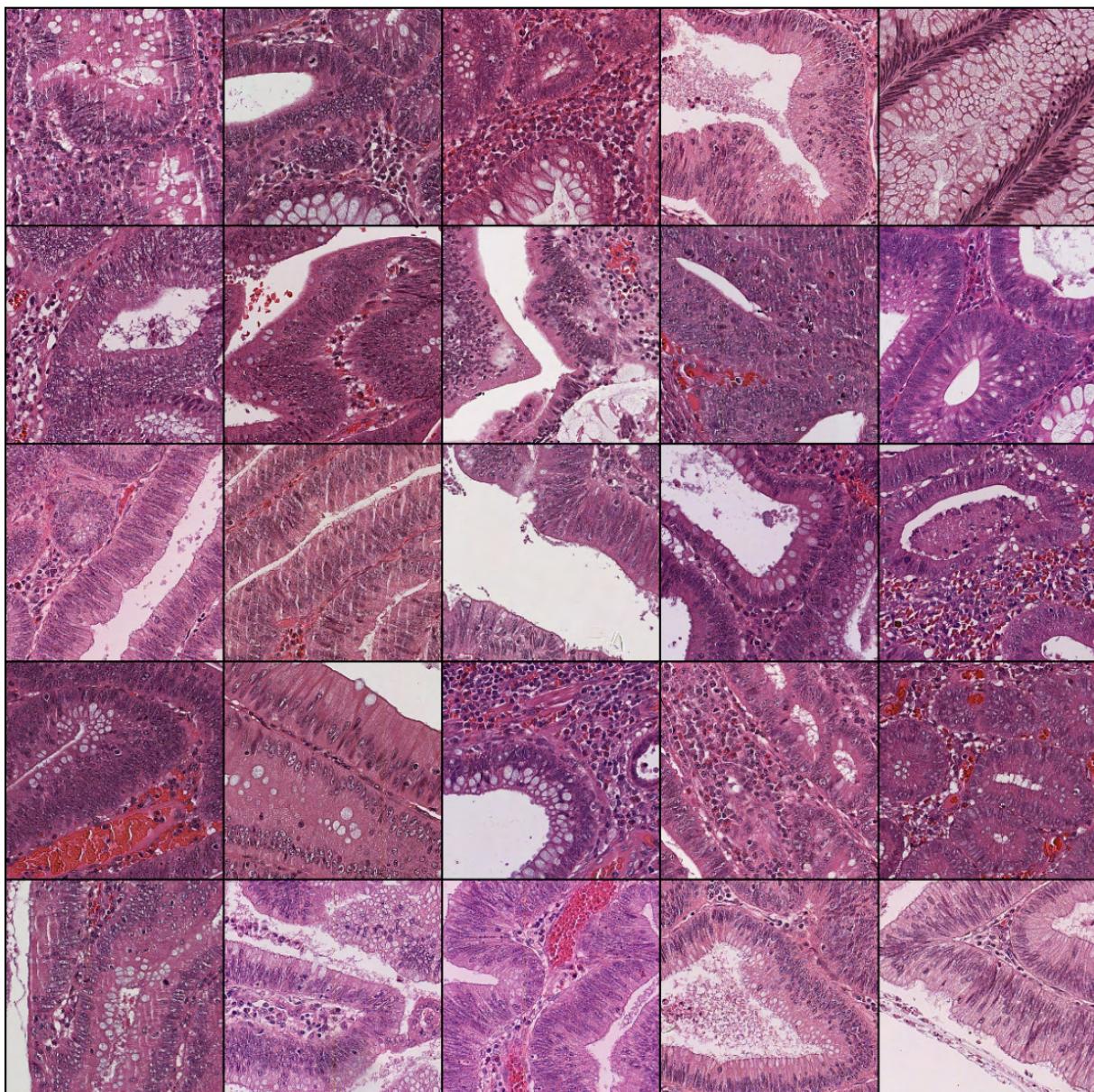


Figure 9. Exemplary patches from Cluster 1 (represented mostly by HGD).

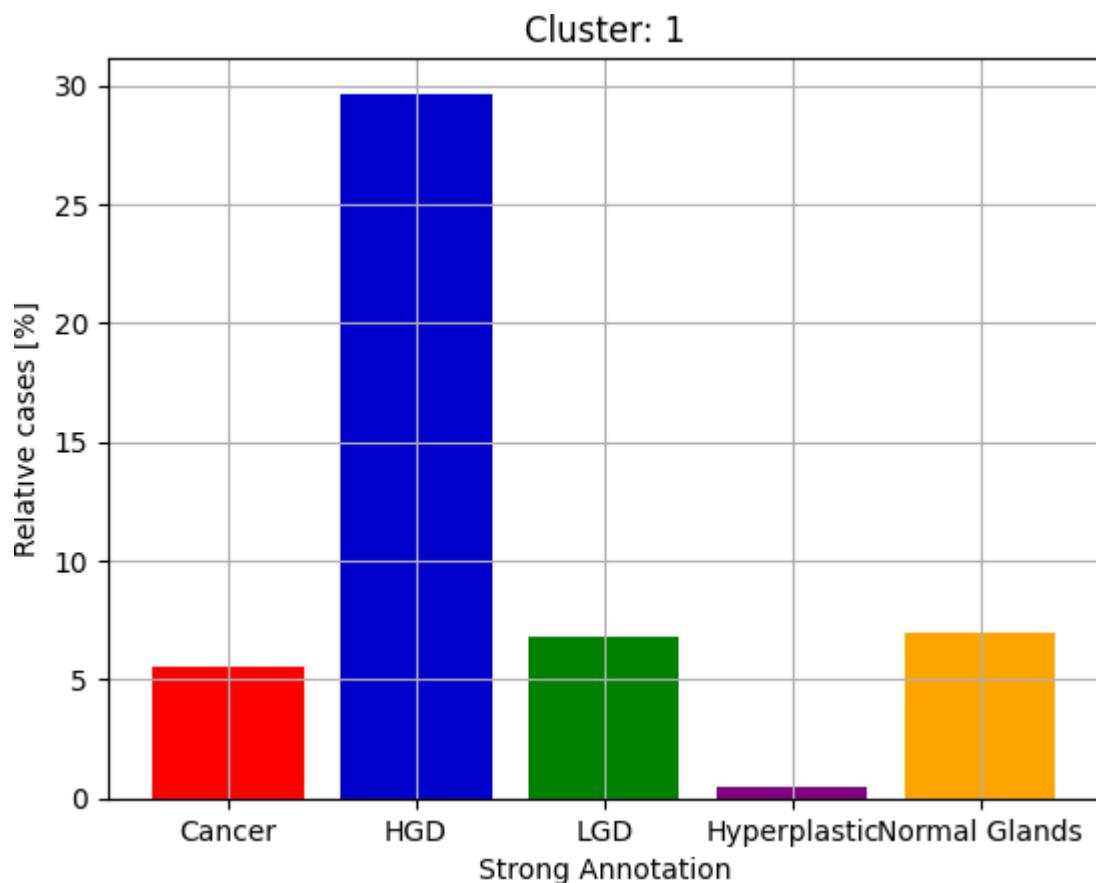


Figure 10. Distribution of the patches in Cluster 1. Please note that the number of relative cases is calculated with respect to all cases of a given class to account for class imbalance.

Cluster: 2

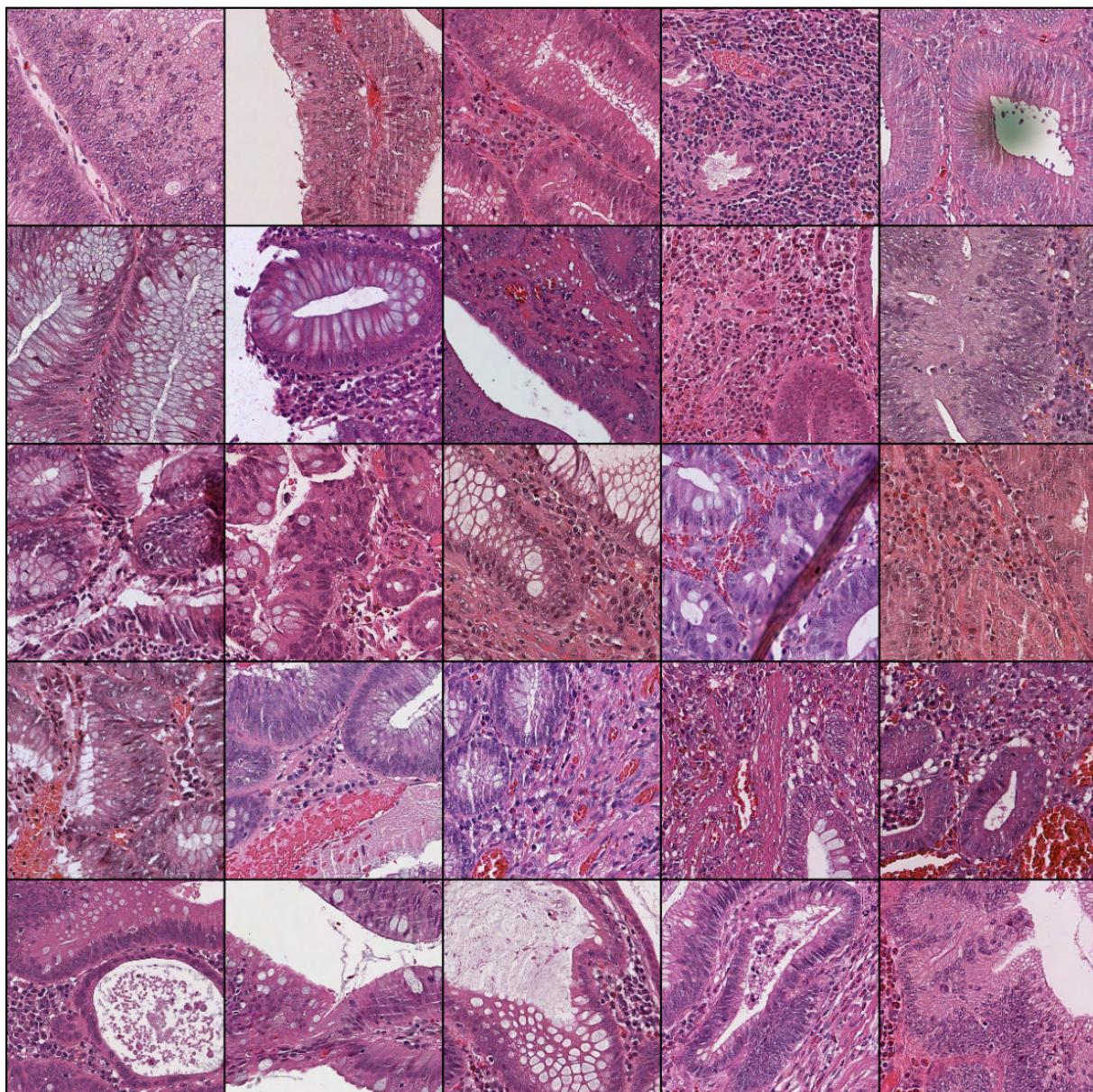


Figure 11. Exemplary patches from Cluster 2 (cluster containing cases from all the classes).

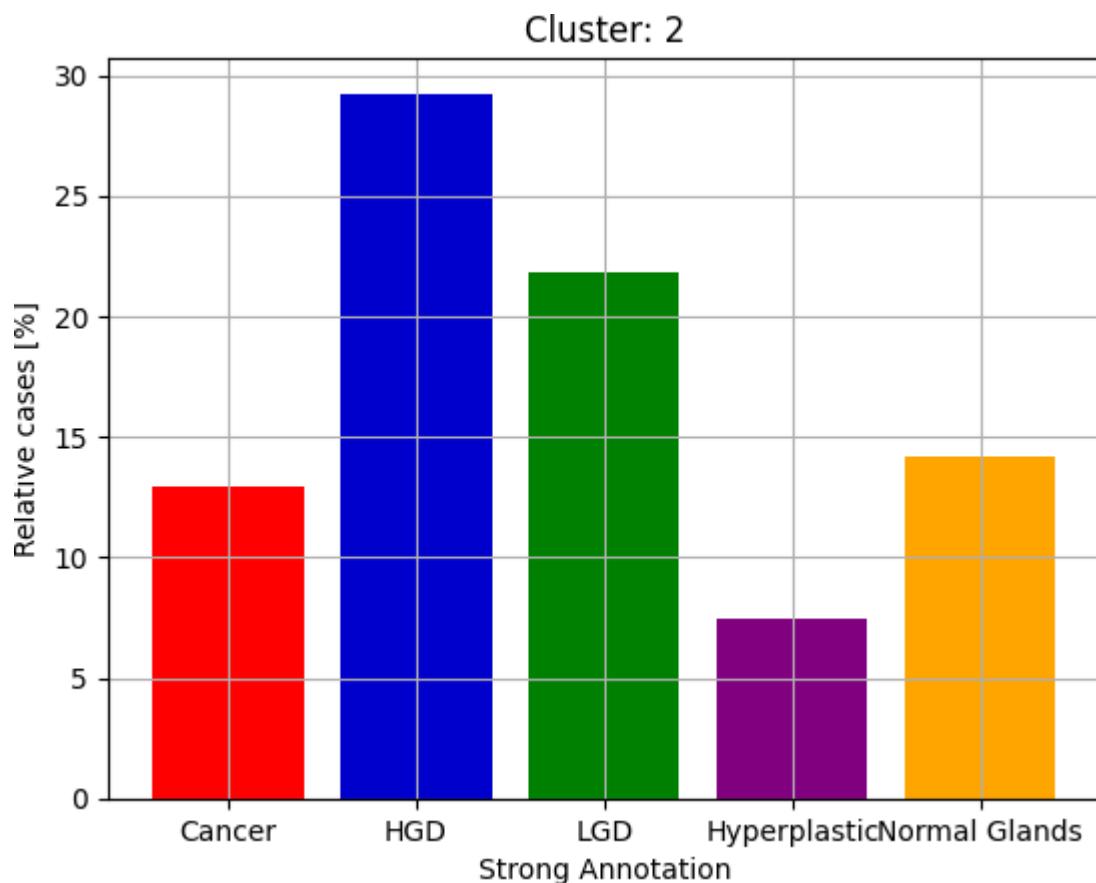


Figure 12. Distribution of the patches in Cluster 2. Please note that the number of relative cases is calculated with respect to all cases of a given class to account for class imbalance.

Cluster: 3

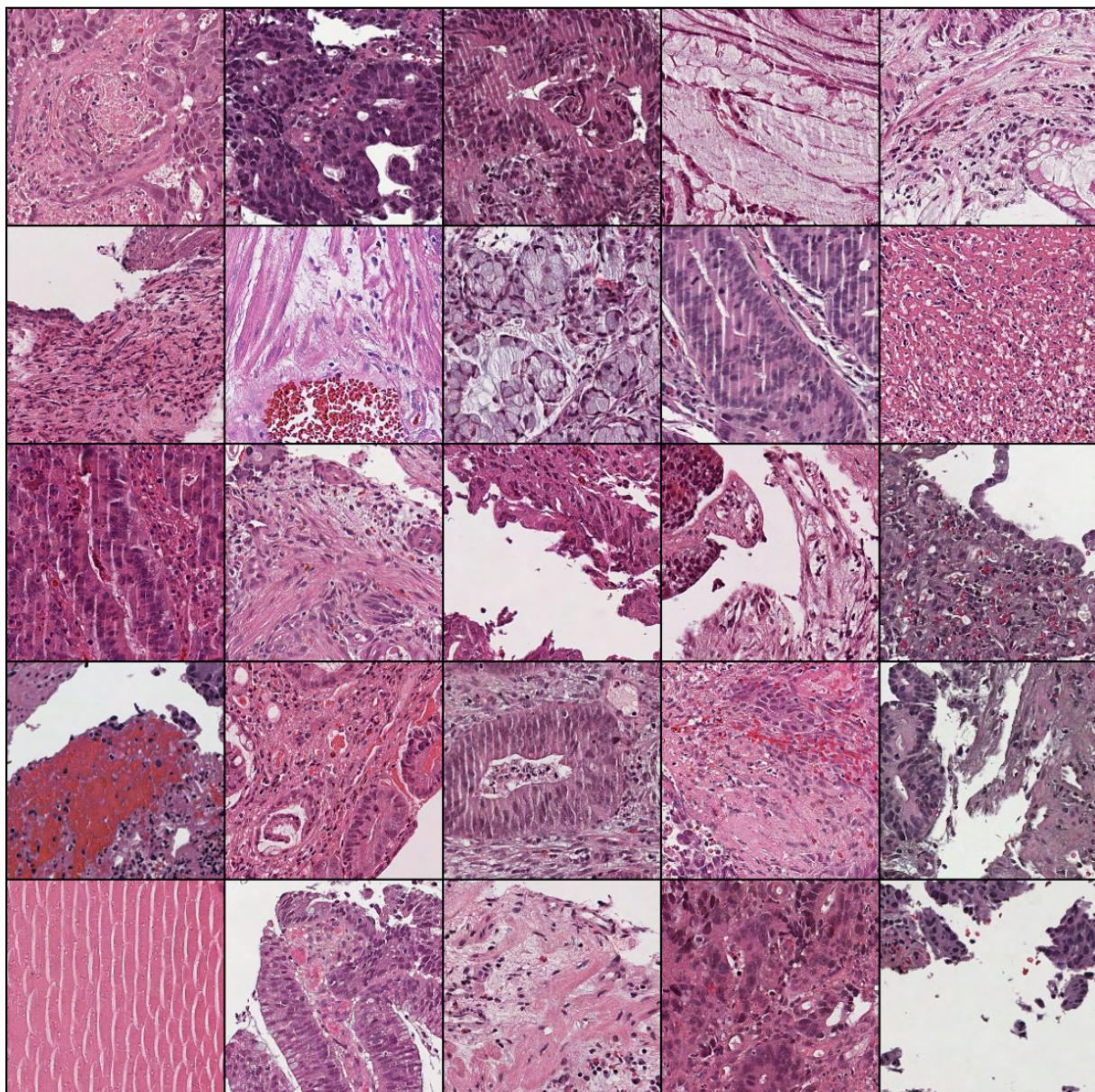


Figure 13. Exemplary patches from Cluster 3 (represented mostly by cancer).

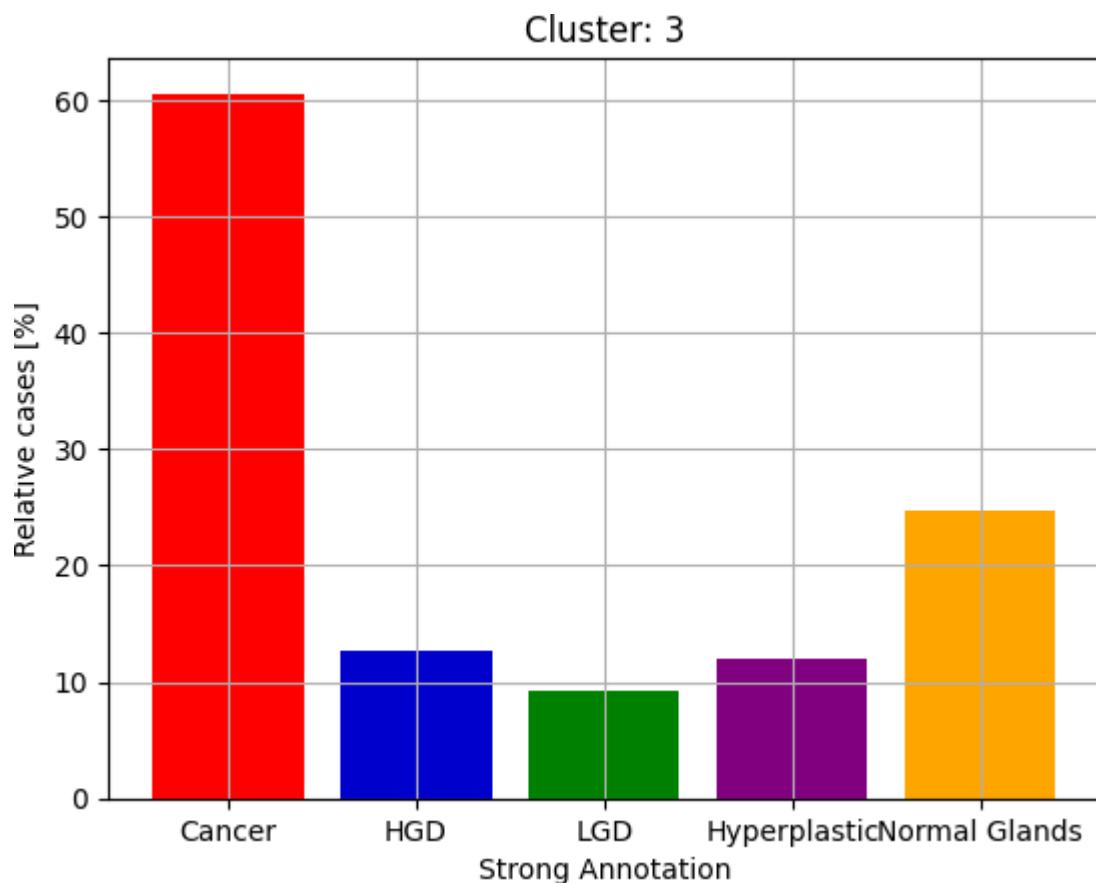


Figure 14. Distribution of the patches in Cluster 3. Please note that the number of relative cases is calculated with respect to all cases of a given class to account for class imbalance.

Cluster: 4

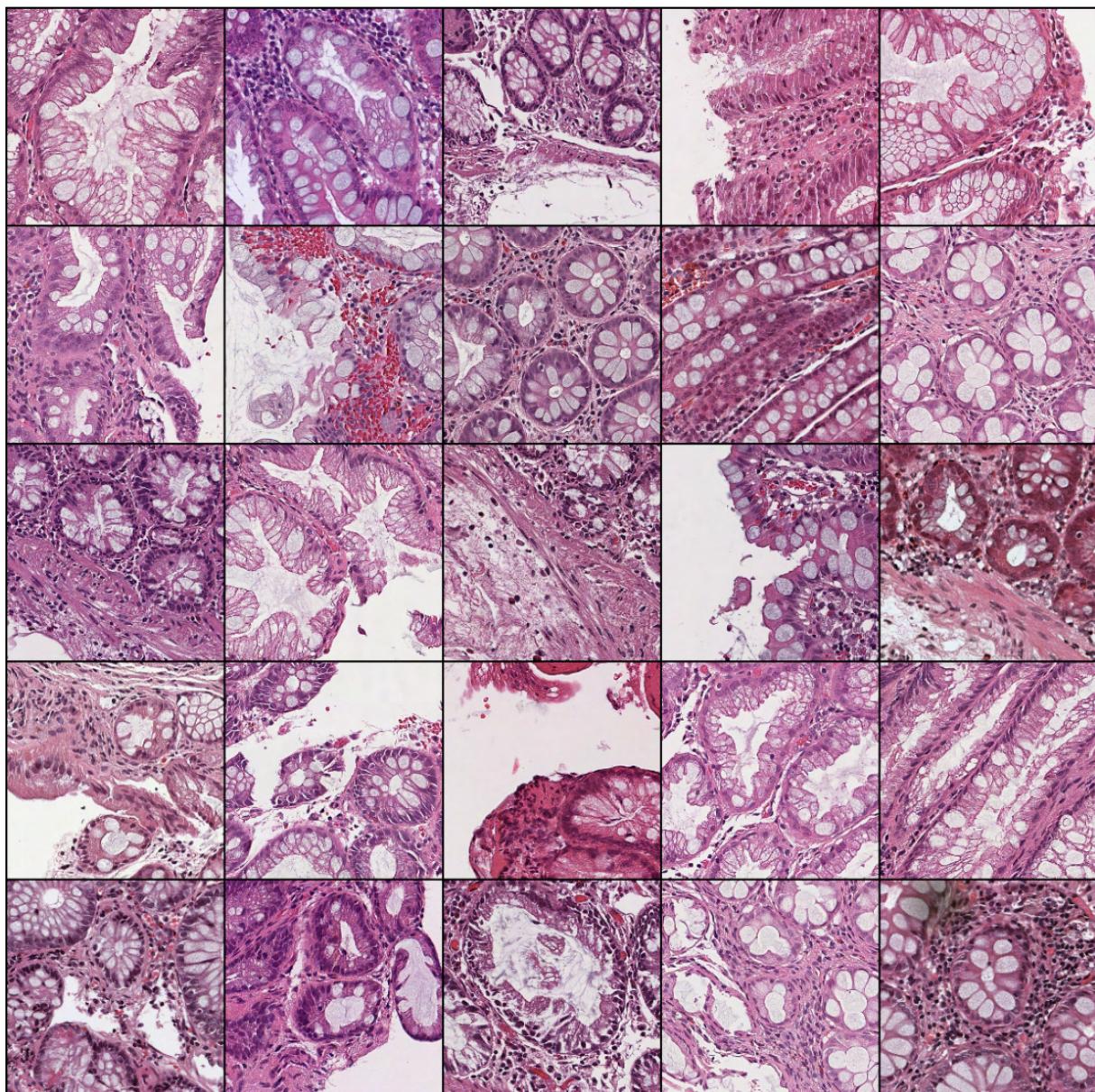


Figure 15. Exemplary patches from Cluster 4 (represented mostly by hyperplastic polyps and normal glands).

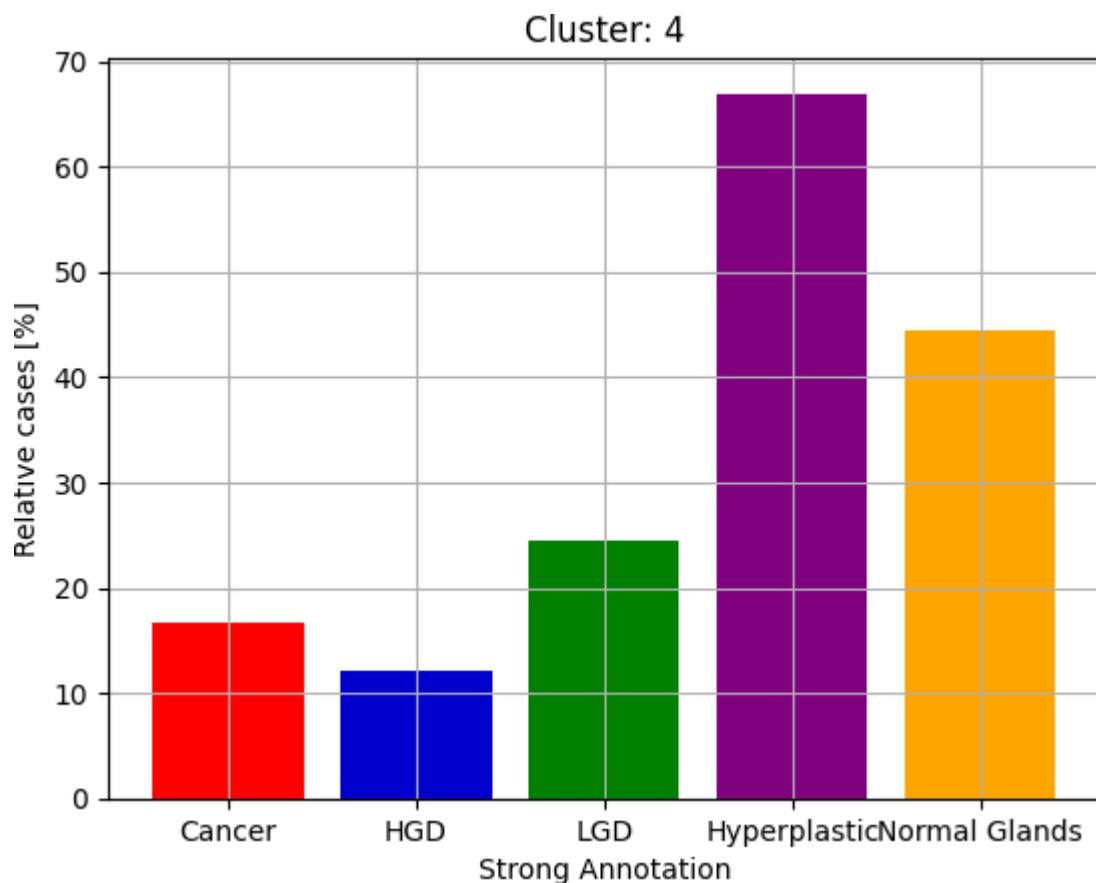


Figure 16. Distribution of the patches in Cluster 4. Please note that the number of relative cases is calculated with respect to all cases of a given class to account for class imbalance.

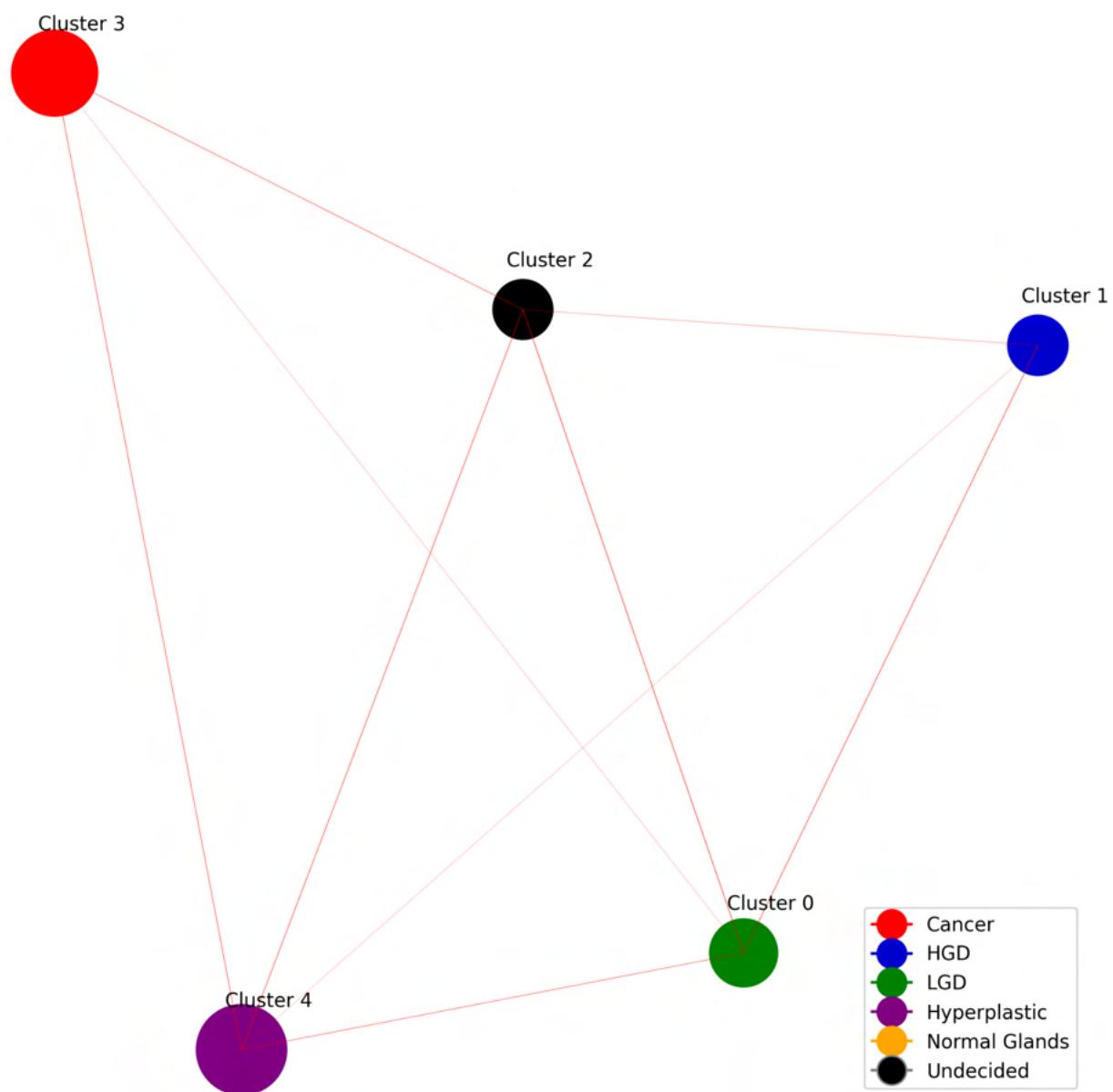


Figure 17. The graph representation of visual features is represented by the cluster centroids and the relative distance between given classes.

3.2 Morphology-based representation applications

3.2.1 Application 1: Pseudo-labeling of colorectal WSIs for patch classification

The CNN trained with pseudo-labels outperforms the same CNN, trained with MIL, at patch-level classification, in three datasets (AOEC, Radboudumc and AIDA).

The models are evaluated using Cohen's κ -score and the comparison is made considering the average and the standard deviation of ten models.

In AOEC dataset, the semi-supervised CNN reaches κ -score = 0.586 ± 0.028 , while the MIL reaches κ -score = 0.563 ± 0.018 . In Radboudumc dataset, the semi-supervised CNN reaches κ -score = 0.591 ± 0.069 , while the MIL reaches κ -score = 0.556 ± 0.068 . In AIDA dataset, the semi-supervised CNN reaches κ -score = 0.748 ± 0.025 (while the MIL reaches κ -score = 0.560 ± 0.022).

Dataset	MIL: patch-level	Semi-supervision
AOEC	0.563 ± 0.018	0.586 ± 0.028
Radboudumc	0.556 ± 0.068	0.591 ± 0.069
AIDA	0.560 ± 0.022	0.748 ± 0.025

Table 4. Results for the performance of the CNN trained with pseudo-labeled patches (semi-supervision), compared with a MIL CNN. The performance is evaluated at the patch-level with Cohen's κ -score. The average and the standard deviation (of the models involved in the k-fold cross-validation) are reported for each dataset.

3.2.2 Application 2: Filtering of WSIs non-informative patches for colorectal WSIs classification

The CNN trained with MIL and non-informative patches filtering, using the morphology CNN, outperforms the same CNN, trained with MIL, at image-level classification, in publicly available datasets, while it reaches comparable results in the AOEC dataset (same source used to train the network).

The models are evaluated using micro-accuracy and the comparison is made considering the average and the standard deviation of ten models.

In the AOEC dataset, the MIL CNN combined with the patches filtering reaches micro-accuracy = 0.878 ± 0.009 , comparable with the MIL without filtering, which reaches micro-accuracy = 0.879 ± 0.010 . In the CRC dataset, the MIL CNN combined with the patches filtering reaches micro-accuracy = 0.786 ± 0.023 , comparable with the MIL without filtering, which reaches micro-accuracy = 0.784 ± 0.025 . In the UNITO dataset, the MIL CNN combined with the patches filtering reaches micro-accuracy = 0.798 ± 0.019 , comparable with the MIL without filtering, which reaches micro-accuracy = 0.793 ± 0.033 . In the Xu dataset, the MIL CNN combined with the patches filtering reaches micro-accuracy = $0.796 \pm$

0.029, outperforming the MIL without filtering, which reaches micro-accuracy = 0.759 ± 0.025 . In the AIDA dataset, the MIL CNN combined with the patches filtering reaches micro-accuracy = 0.877 ± 0.018 , outperforming the MIL without filtering, which reaches micro-accuracy = 0.782 ± 0.014 .

Dataset	MIL	MIL + clustering
AOEC	0.879 ± 0.010	0.878 ± 0.009
CRC ²⁷	0.784 ± 0.025	0.786 ± 0.023
UNITO ²⁸	0.793 ± 0.033	0.798 ± 0.019
Xu ²⁹	0.759 ± 0.025	0.796 ± 0.029
AIDA ³⁰	0.782 ± 0.014	0.877 ± 0.018

Table 5. Results for the performance of the CNN trained with MIL and non-informative patches filtering through morphology CNN, compared with a MIL CNN, without filtering. The performance is evaluated at the image-level with micro-accuracy. The average and the standard deviation (of the models involved in the k-fold cross-validation) are reported for each dataset.

4 Discussion & Conclusion

The representation presented in this deliverable highlights the relationship between different morphologies in WSIs and can help to solve computational pathology tasks, guaranteeing higher performance than other approaches.

The representation is obtained by training an H&E-CNN in a fully supervised fashion to classify different morphologies included in colorectal cancer WSIs: cancer, dysplasias, normal glands and non-informative. This training setup guarantees two essential properties: the separability between classes and the stain-invariance.

The classes selected to evaluate the separability in the latent space are: cancer, high-grade dysplasia, low-grade dysplasia, hyperplastic polyps and normal glands. The classification of the patches corresponding to those classes is not a simple task, as shown in Figure 3-4. In particular, Figure 4 shows how the decision border for a particular class is not well-defined. Even though it is possible to identify macro-regions with most samples belonging to a class (such as blue in the top right for HGD, green bottom right for LGD), this is not possible for all the classes. For example, it is not possible to distinguish between cancer and images annotated as normal glands in the top left corner between images annotated as a hyperplastic polyp and normal glands (sometimes they are grouped in the same class). While the latter can be expected (hyperplastic polyps and normal glands show very similar morphology), the first case can be explained by noting the following points: 1) normal glands patches include regions with stroma (due to a not fine-grained annotation) that have a morphology similar to the images annotated as cancer 2) the variability of cancer can include morphologies where the lumen of the patches can be

confused with a normal image, as shown in Xu et al. Therefore, dealing with the mentioned classes is not simple.

However, the representation presented in the deliverable allows to better separate the classes. This fact is shown in the five clusters presented in Figures 7-16. Each cluster includes patches mostly-annotated with the same class. For each cluster, the distribution of the patches (in terms of annotations) and 25 randomly selected patches are presented. It is the case of cluster 0 (lgd), cluster 1 (hgd), cluster 3 (cancer). The other two clusters are particular: cluster 2 includes patches from all the classes, while cluster 4 from both hyperplastic polyps and normal glands. The examples from cluster 2 present high variability in terms of tissue morphology and in terms of classes. However, the morphologies can't be easily linked to a class. A hypothesis to explain this cluster is that it includes the patches that are 1) mislabeled and 2) that lie in the decision border of the classes. CCluster 4 presents the problem already mentioned regarding the differences between hyperplastic polyp and normal glands. This cluster, even including patches from two classes, is acceptable in terms of morphologies. Indeed, the examples are particularly clear: both include well-defined glands with a lumen. Furthermore, the hyperplastic polyp is linked to a benign condition, such as the normal glands. Another point to stress involves the relationship between the clusters, shown by the graph in Figure 17. The graph's nodes are the clusters, while the edges are the relationship between the clusters (in this case the similarity between the clusters). The larger the node, the better the separability between

The thicker the edge between two nodes, the more similar are the patches within two clusters. Among the different relationships, two examples are important to be analyzed. The first example involves clusters 0 and 1, including patches with respectively LGD and HGD. The similarity is high, meaning that the clusters include similar morphologies. This is the desired property, since they have similar tissue morphologies and sometimes even the pathologists cannot distinguish between them. The second example involves clusters 3-0 (LGD and cancer) and 1-4 (HGD and hyperplastic/normal glands). In this case, the similarity between the clusters is lower (the edges are thinner).

Considering the examples included in the clusters, it is possible to visualize the stain-invariance property. In fact, the patches in the clusters present several colour variations, meaning that the features learnt during the training refer to different morphologies and are not influenced by the Hematoxylin & Eosin component of the images.

The representation is used in two computational pathology tasks (semi-supervision and MIL), to check if it can improve the performance of a CNN in terms of classification, compared with a CNN trained via MIL. In the first case, the representation is combined with weak annotations to pseudo-label over 1'000 WSIs. The pseudo-labels are used to train a CNN in patch-level classification, outperforming the same CNN trained with MIL, in three datasets (AOEC, Radboudumc, AIDA). In the second case, the representation is used to filter non-informative patches during MIL training, allowing the model to use only meaningful patches. Also in this case, the CNN obtains higher performance (comparable for some datasets) with the same network trained without the filtering.

5 References

1. Gurcan, M. N. *et al.* Histopathological Image Analysis: A Review. *IEEE Reviews in Biomedical Engineering* (2009) doi:10.1109/RBME.2009.2034865.
2. Krupinski, E. A., Graham, A. R. & Weinstein, R. S. Characterizing the development of visual search expertise in pathology residents viewing whole slide images. *Human Pathology* **44**, 357–364 (2013).
3. Wei, J. W. *et al.* Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Scientific Reports* **9**, 1–8 (2019).

4. Arvaniti, E. *et al.* Automated Gleason grading of prostate cancer tissue microarrays via deep learning. *Scientific Reports* **8**, 12054 (2018).
5. Costantini, M. *et al.* Interobserver agreement in the histologic diagnosis of colorectal polyps: The experience of the multicenter adenoma colorectal study (SMAC). *Journal of Clinical Epidemiology* **56**, 209–214 (2003).
6. Vennalaganti, P. *et al.* Discordance Among Pathologists in the United States and Europe in Diagnosis of Low-Grade Dysplasia for Patients With Barrett's Esophagus. *Gastroenterology* **152**, 564-570.e4 (2017).
7. Turner, J. K., Williams, G. T., Morgan, M., Wright, M. & Dolwani, S. Interobserver agreement in the reporting of colorectal polyp pathology among bowel cancer screening pathologists in Wales. *Histopathology* **62**, 916–924 (2013).
8. Marini, N. *et al.* Multi-Scale Task Multiple Instance Learning for the Classification of Digital Pathology Images with Global Annotations. in *Proceedings of the MICCAI Workshop on Computational Pathology* (eds. Atzori, M. *et al.*) vol. 156 170–181 (PMLR, 2021).
9. Marini, N. *et al.* Multi_Scale_Tools: A Python Library to Exploit Multi-Scale Whole Slide Images. *Frontiers in Computer Science* **0**, 68 (2021).
10. Tellez, D. *et al.* Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology.
11. Ren, J., Hacihaliloglu, I., Singer, E. A., Foran, D. J. & Qi, X. Unsupervised Domain Adaptation for Classification of Histopathology Whole-Slide Images. *Frontiers in Bioengineering and Biotechnology* **7**, 102 (2019).
12. Marini, N., Atzori, M., Otálora, S., Marchand-Maillet, S. & Müller, H. H\&E-Adversarial Network: A Convolutional Neural Network To Learn Stain-Invariant Features Through Hematoxylin \& Eosin Regression. in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops* 601–610 (2021).
13. Otálora, S., Atzori, M., Andrearczyk, V., Khan, A. & Müller, H. Staining invariant features for improving generalization of deep convolutional neural networks in computational pathology. *Frontiers in Bioengineering and Biotechnology* **7**, 198 (2019).
14. Ren, J., Hacihaliloglu, I., Singer, E. A., Foran, D. J. & Qi, X. Unsupervised Domain Adaptation for Classification of Histopathology Whole-Slide Images. *Frontiers in Bioengineering and Biotechnology* **7**, 102 (2019).
15. Tellez, D. *et al.* Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical Image Analysis* **58**, 101544 (2019).
16. Khan, A., Atzori, M., Otálora, S., Andrearczyk, V. & Müller, H. Generalizing convolution neural networks on stain color heterogeneous data for computational pathology. <https://doi.org/10.1111/12.2549718> **11320**, 173–186 (2020).
17. Stacke, K., Eilertsen, G., Unger, J. & Lundstrom, C. Measuring Domain Shift for Deep Learning in Histopathology. *IEEE journal of biomedical and health informatics* **25**, 325–336 (2021).
18. Fischer, A. H., Jacobson, K. A., Rose, J. & Zeller, R. Hematoxylin and eosin staining of tissue and cell sections. *Cold Spring Harbor Protocols* **3**, (2008).
19. LARSON, K., HO, H. H., ANUMOLU, P. L. & CHEN, M. T. Hematoxylin and Eosin Tissue Stain in Mohs Micrographic Surgery: A Review. *Dermatologic Surgery* **37**, 1089–1099 (2011).

20. Baldi, P. Deep Learning in Biomedical Data Science. *Annual Review of Biomedical Data Science* **1**, 181–205 (2018).
21. Campanella, G. *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine* **25**, 1301–1309 (2019).
22. Hegde, N. *et al.* Similar image search for histopathology: SMILY. *npj Digital Medicine* **2**, (2019).
23. Campanella, G. *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine* **25**, 1301–1309 (2019).
24. Marini, N., Otálora, S., Müller, H. & Atzori, M. Semi-supervised training of deep convolutional neural networks with heterogeneous data and few local annotations: An experiment on prostate histopathology image classification. *Medical Image Analysis* **73**, 102165 (2021).
25. Karimi, D., Dou, H., Warfield, S., Analysis, A. G.-M. I. & 2020, undefined. Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. *Elsevier*.
26. Sirinukunwattana, K., Snead, D. R. J. & Rajpoot, N. M. A Stochastic Polygons Model for Glandular Structures in Colon Histology Images. *IEEE Transactions on Medical Imaging* **34**, 2366–2378 (2015).
27. Awan, R. *et al.* Glandular Morphometrics for Objective Grading of Colorectal Adenocarcinoma Histology Images. *Scientific Reports* **7**, 2220–2243 (2017).
28. UNITOPATHO | IEEE DataPort. <https://ieee-dataport.org/open-access/unitopatho>.
29. Xu, Y. *et al.* Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. *BMC Bioinformatics* (2017) doi:10.1186/s12859-017-1685-x.
30. Stadler, C. B. *et al.* Proactive Construction of an Annotated Imaging Database for Artificial Intelligence Training. *Journal of Digital Imaging* 2020 **34:1** 34, 105–115 (2020).
31. Benson, A. B. *et al.* NCCN Guidelines ® Insights Colon Cancer, Version 2.2018 Featured Updates to the NCCN Guidelines. *JNCCN Journal of the National Comprehensive Cancer Network* vol. 16 359–369 (2018).
32. Rahib, L., Wehner, M. R., Matrisian, L. M. & Nead, K. T. Estimated Projection of US Cancer Incidence and Death to 2040. *JAMA Network Open* **4**, 214708 (2021).
33. Zauber, A. G. *et al.* Colonoscopic Polypectomy and Long-Term Prevention of Colorectal-Cancer Deaths. *New England Journal of Medicine* **366**, 687–696 (2012).
34. Foucart, A., Debeir, O. & Decaestecker, C. SNOW: Semi-supervised, noisy and/or weak data for deep learning in digital pathology. *Proceedings - International Symposium on Biomedical Imaging 2019-April*, 1869–1872 (2019).
35. Zeki, I. *et al.* Billion-scale semi-supervised learning for image classification. (2019).
36. Shaw, S., Pajak, M., Lisowska, A., Tsafaris, S. A. & O’Neil, A. Q. Teacher-Student chain for efficient semi-supervised histology image classification. (2020).
37. Tolkach, Y., Dohmgörden, T., Toma, M. & Kristiansen, G. High-accuracy prostate cancer pathology using deep learning. *Nature Machine Intelligence* 2020 **2:7** 2, 411–418 (2020).
38. Bulten, W. *et al.* Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *The Lancet Oncology* **21**, 233–241 (2020).
39. Ilse, M., Tomczak, J. M. & Welling, M. *Attention-based Deep Multiple Instance Learning*. (2018).

40. Hashimoto, N. *et al.* Multi-scale Domain-adversarial Multiple-instance CNN for Cancer Subtype Classification with Unannotated Histopathological Images. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 3852–3861 (2020).
41. Lu, M. Y. *et al.* Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering* 1–16 (2021) doi:10.1038/s41551-020-00682-w.
42. Chikontwe, P., Kim, M., Nam, S. J., Go, H. & Park, S. H. Multiple Instance Learning with Center Embeddings for Histopathology Classification. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 12265 LNCS 519–528 (Springer Science and Business Media Deutschland GmbH, 2020).
43. Carbonneau, M. A., Cheplygina, V., Granger, E. & Gagnon, G. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition* (2018) doi:10.1016/j.patcog.2017.10.009.
44. Foulds, J. & Frank, E. A review of multi-instance learning assumptions. *The Knowledge Engineering Review* 25, 1–25 (2010).
45. Wang, Y., Li, J. & Metze, F. A Comparison of Five Multiple Instance Learning Pooling Functions for Sound Event Detection with Weak Labeling. in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* vols. 2019-May 31–35 (Institute of Electrical and Electronics Engineers Inc., 2019).
46. Chicco, D. Ten quick tips for machine learning in computational biology. *BioData Mining* vol. 10 1–17 (2017).

6 Annexes

6.1 Semi-supervision

Semi-supervised learning is a technique that lies between unsupervised learning and fully-supervised training.

Recent work highlighted that unlabeled data can be used to improve the generalization of CNN, when used with a small amount of labeled data. In particular, that is true in digital pathology, where large datasets without local annotations are available, while it is hard to collect large datasets with local annotations³⁴.

The technique involves the pseudo-labeling (annotations made by automatic algorithms) of data (in digital pathology of patches) used to train the same model or another model³⁵. In the latter case, the technique is also known as the teacher/student paradigm.

Examples of the paradigm in digital pathology^{24,34,36-38}, show that CNNs trained with the combination of pseudo-labeled data and strongly-annotated data reach higher results, compared with CNNs trained only with strongly-annotated data, but also that these models can better generalize on heterogeneous data, since they can be trained with images coming from different medical sources.

6.2 Multiple Instance Learning

The CNN is trained using a Multiple Instance Learning (MIL) framework, trained at the instance-level.

MIL^{23,39-43} is a weakly-supervised framework that allows facing problems where data are organized as a bag of instances⁴³ and the information available on the data regards the entire bag, without any local information about the instances. The framework is based on the MIL assumption⁴⁴, which coordinates the relationship between the bag and the instances. The original MIL assumption asserts that a bag is positive if it includes at least one positive instance, while it is negative if it does not include any positive instance. The assumption can be relaxed to be adopted in problems where the bag is identified by the distribution of its patches, such as multiclass problems. A MIL problem can be formulated at two levels⁴³: the bag-level and the instance-level. In both levels, the component that aggregates the instance features or the instances predictions is called the pooling algorithm. Histopathology image classification can be formulated as a MIL problem, where a WSI represents a bag X_n that includes p patches and the information available on the data regards the entire WSI, as in the MIL CNN presented in this paper. The MIL CNN presented in this paper produces predictions for the single instances.

The model includes several components: a convolutional backbone (ResNet34) pre-trained on ImageNet and frozen, an intermediate fully-connected layer, a classifier and an attention network. The convolutional backbone is pre-trained on ImageNet and it is frozen during the training. The intermediate fully-connected layer produces smaller feature vectors, called embeddings from the ResNet features, for each patch within a WSI. The classifier produces predictions from the patch embeddings. The attention network is a pooling layer (state-of-the-art algorithm), that aggregates the predictions for the single patches to have a global prediction for the WSI^{39,45}. The attention network gives weight to each patch, depending on their importance in the global predictions. The sum of the weight is 1. The choice of

training a MIL instance-based framework show advantages and disadvantages. The most important advantage is that this kind of CNN can be used in clinical settings, showing the pathologist the model's predictions for each patch. Furthermore, the attention pooling layer may underline some regions of interest for the pathologist. The most crucial disadvantage is that the model's performance at WSI (bag) level is lower than MIL embedding-based frameworks⁴³.

6.3 Metrics

The performance of the model is evaluated at patch-level and WSI-level.

At patch-level, the classification is a multiclass problem and the model is evaluated using Cohen's κ -score. Cohen's κ -score measures the agreement between raters. It is usually adopted in scientific literature to evaluate the agreement between pathologists. In this case, it measures the agreement between the model predictions and the ground truth. The metric varies between -1 (complete disagreement) and 1 (complete agreement). Cohen's κ -score = 0 means agreement by chance.

At WSI-level and image-level, the classification is a multilabel problem and the model is evaluated using the micro-average of accuracy. Accuracy is the fraction of correct predictions (true positives + true negatives) on the total number of the predictions and varies between 0 (total wrong predictions) and 1 (perfect predictions). The accuracy metric is averaged using micro-accuracy, working with the single true positives, etc.

6.4 Hyperparameters and implementation

The hyperparameters used to train the model are optimized using a grid search algorithm⁴⁶.

The grid search algorithm aims to identify the optimal configuration of CNN hyperparameters.

In this case, the optimal configuration allows the CNN to reach the lowest loss function on WSI classification, on the validation partition.

The grid search involves several hyperparameters: the number of epochs for training the model (15, after this number of epochs the model does not reach a lower loss function), the optimized (Adam; SGD and Adam were tested), the learning rate (10^{-3} ; 10^{-2} , 10^{-3} , 10^{-4} , 10^{-5} were tested), the decay rate (10^{-3} ; 10^{-2} , 10^{-3} , 10^{-4} , 10^{-5} were tested) and the number of nodes within the embedding layer before the classifier (128; 32, 64, 128, 256 were tested).

The whole pipeline is implemented with several Python libraries.

Pytorch 1.1.0 is used to model, train and test the CNNs. Openslide 3.4.1 and ASAP 1.9 are used to access the WSIs. Scikit-learn 0.23.1 is used to evaluate the performance metrics of the models.

Albumentations 1.8 is used for implementing the data augmentation pipeline.

All the experiments are executed on a Tesla V100 GPU.

6.5 Graph representation of visual histopathology knowledge with 25 clusters

The visual representation of the latent space is also created for 25 clusters to show a more grained distribution of the latent space. The process is the same as for the 5 clusters. The cluster separation in latent space is shown in Figure 18. The graph visualization is presented in Figure 19. The exemplary images and distributions of chosen clusters are shown in Figures 20-33.

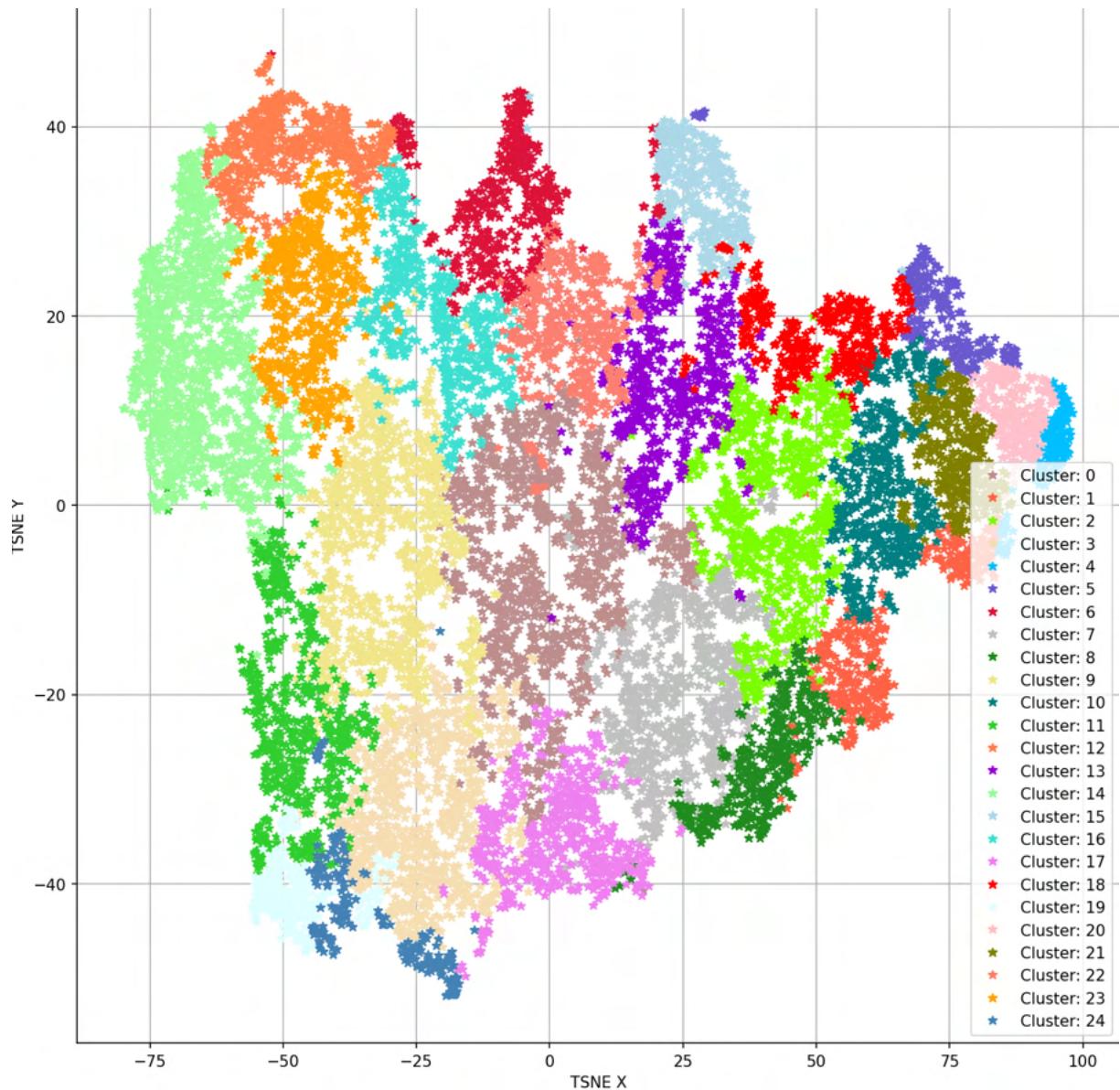


Figure 18. Visualization of the clusters in the latent space (hierarchical clustering, ward linkage, Euclidean distance, 25 clusters).

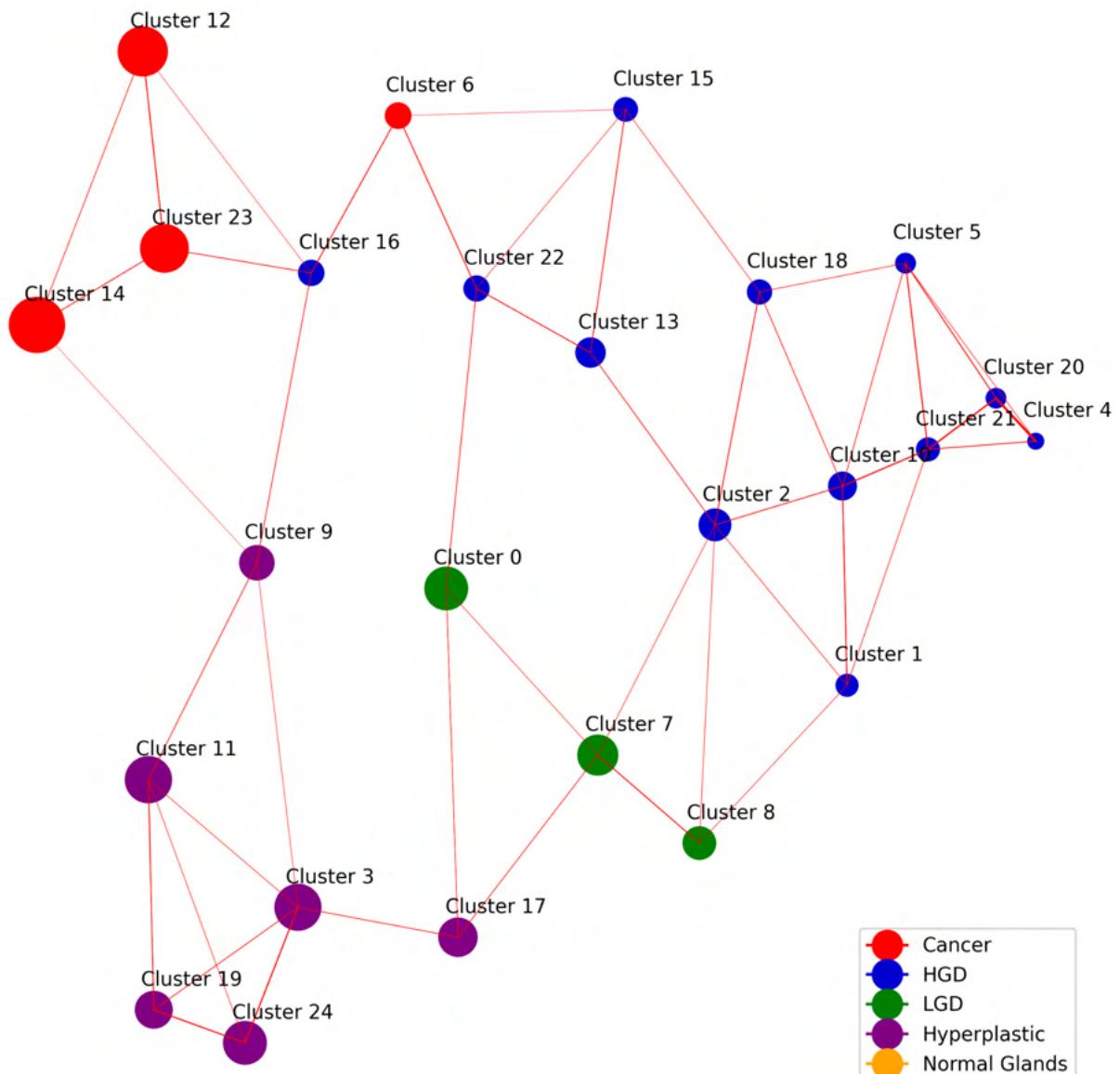


Figure 19. The graph representation of the latent space clusters. The cluster color denotes the most frequently occurring class within a given cluster.

Cancer Clusters

Cluster: 12

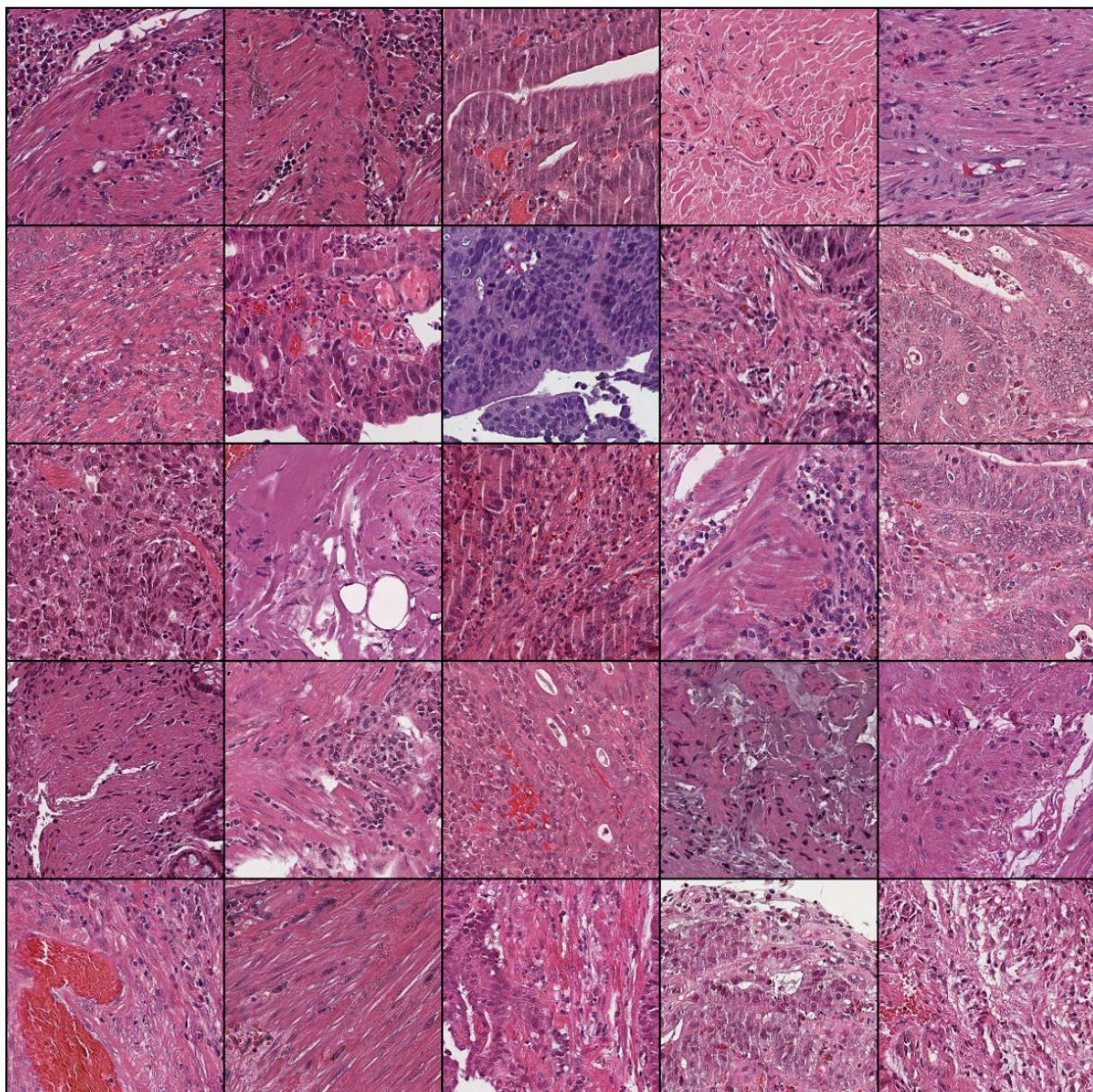


Figure 20. Exemplary patches from Cluster 12.

Cluster: 12

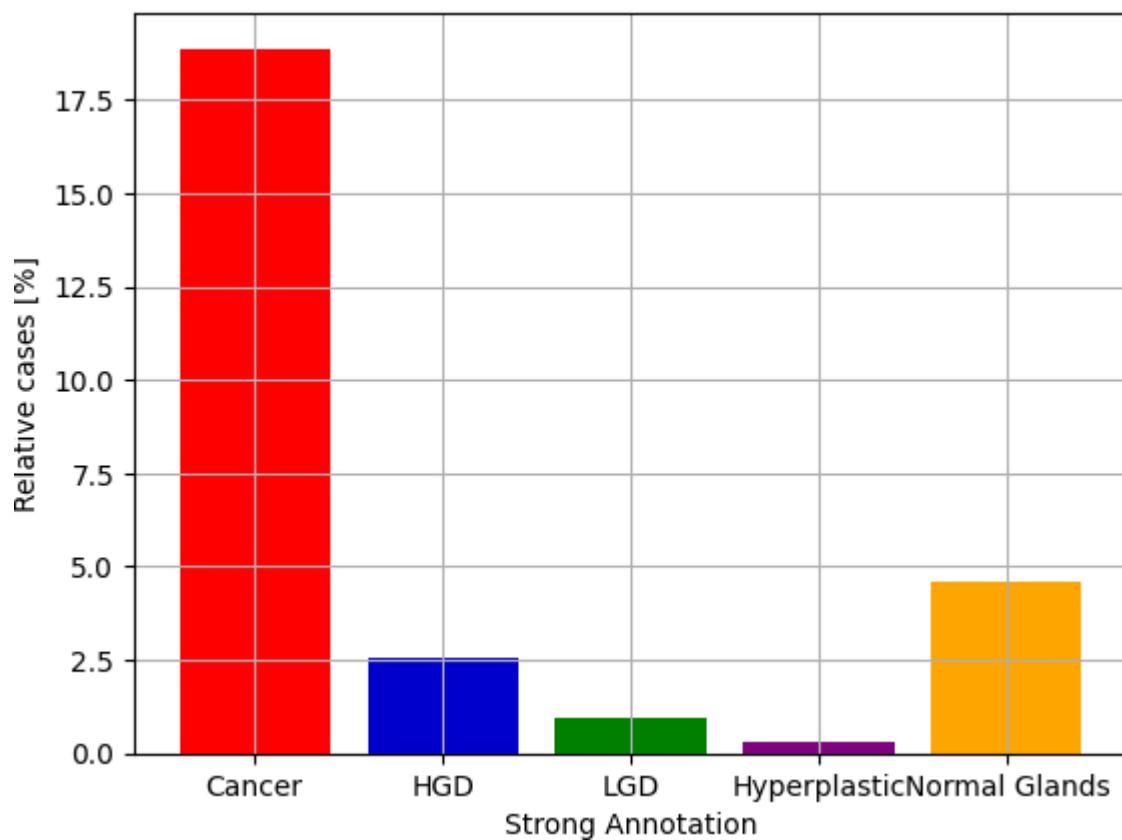


Figure 21. Distribution of the patches in Cluster 12. Please note that the number of relative cases is calculated with respect to all cases of a given class to account for class imbalance.

Cluster: 23

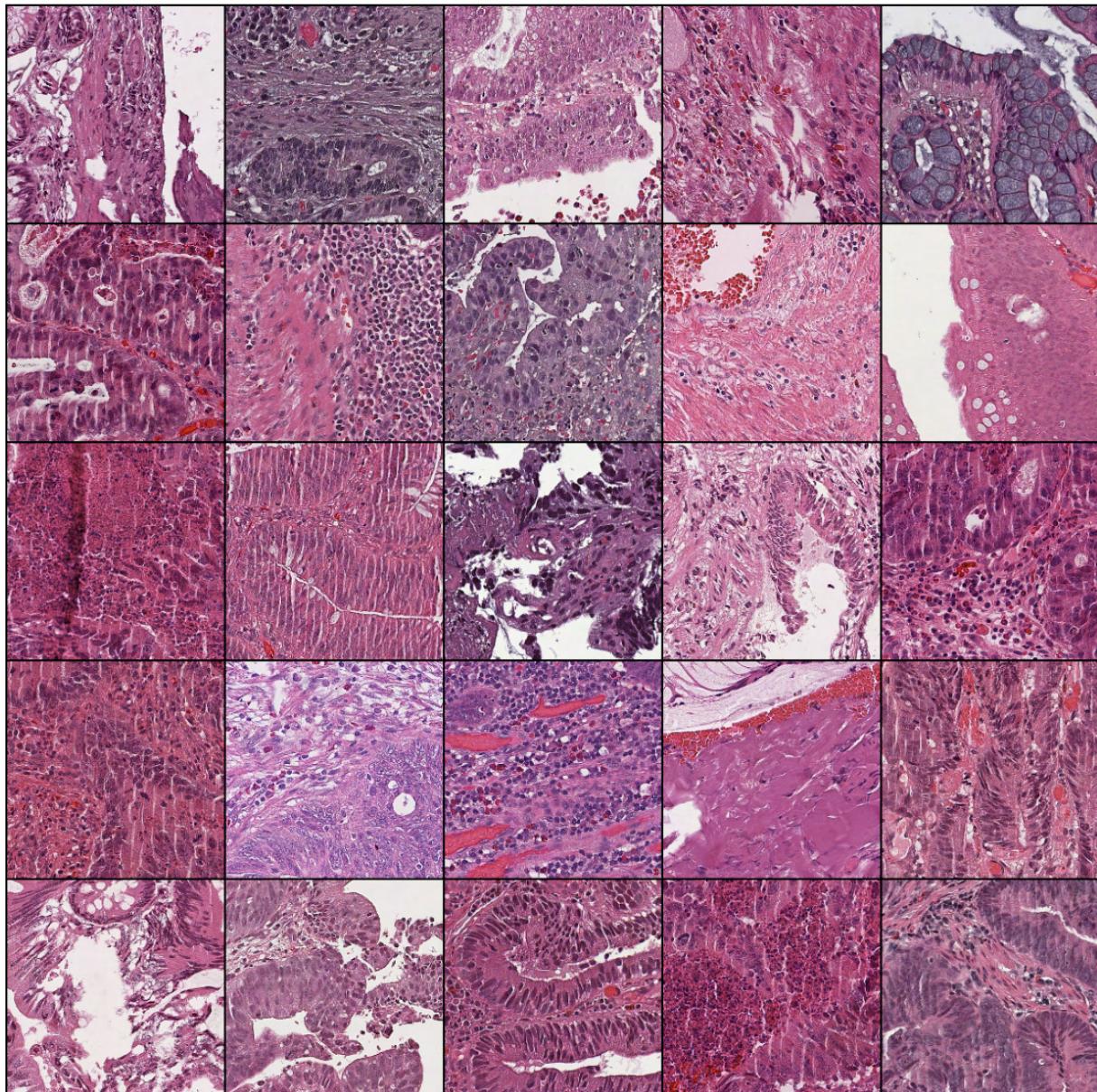


Figure 22. Exemplary patches from Cluster 23.

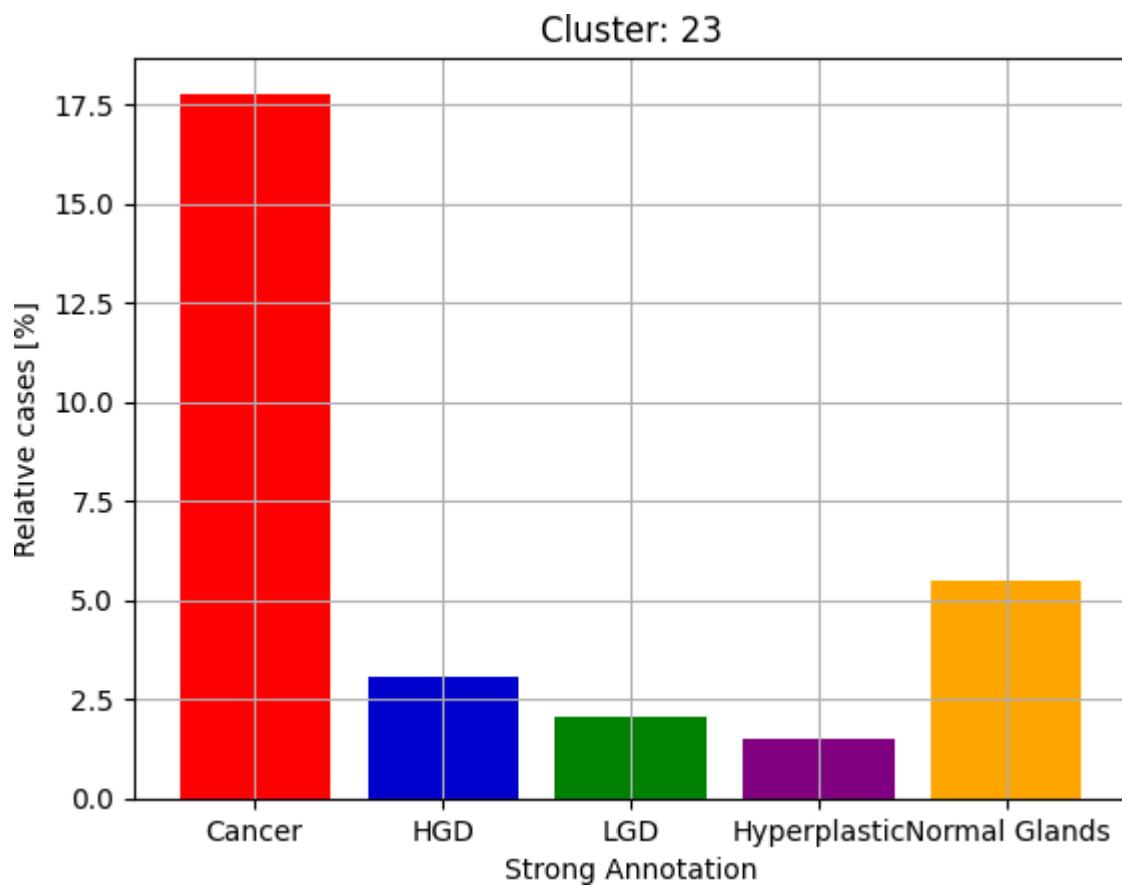


Figure 23. Distribution of the patches in Cluster 23. Please note that the number of relative cases is calculated with respect to all cases of a given class to account for class imbalance.

Cluster: 14

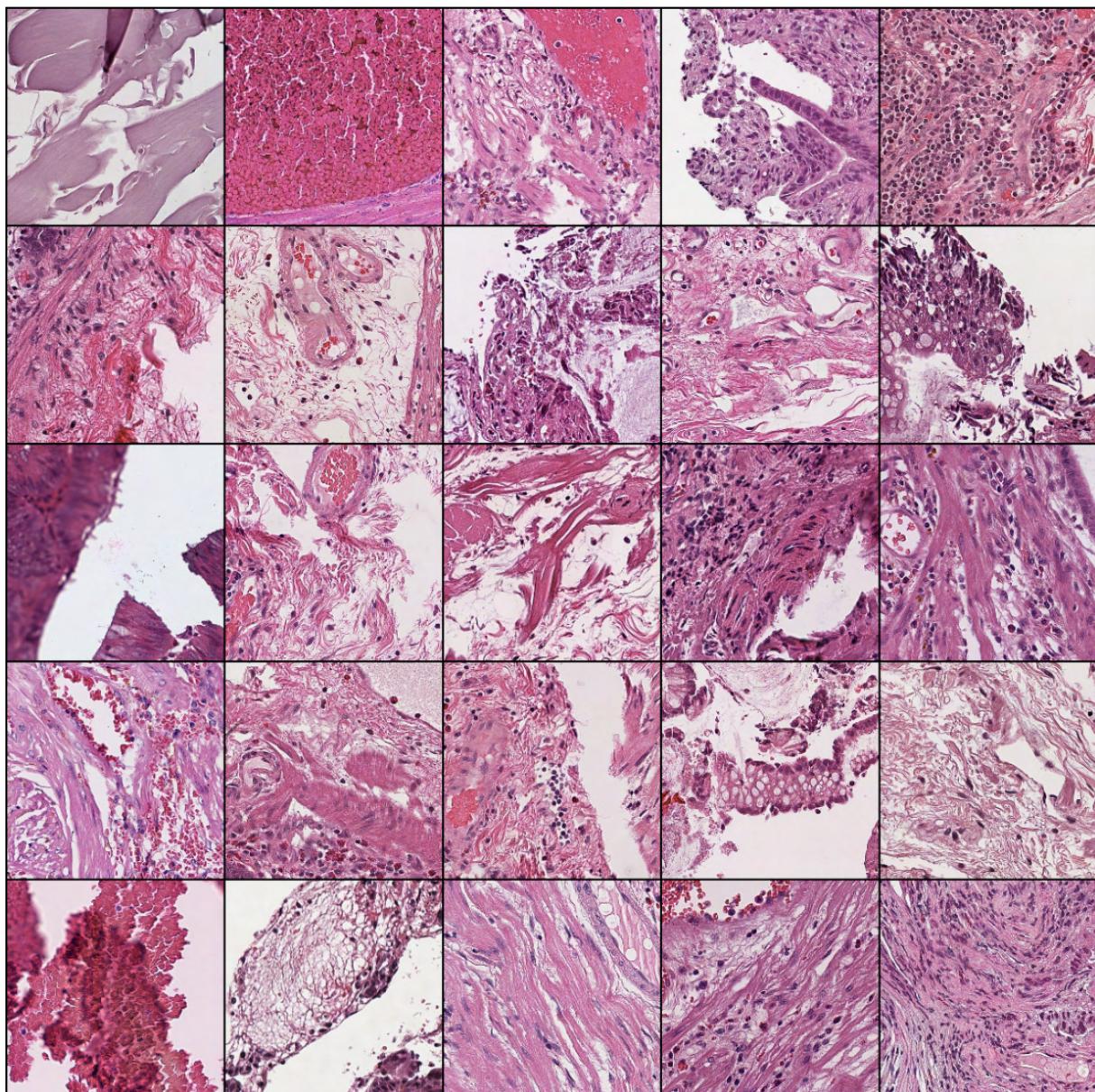


Figure 24. Exemplary patches from Cluster 14.

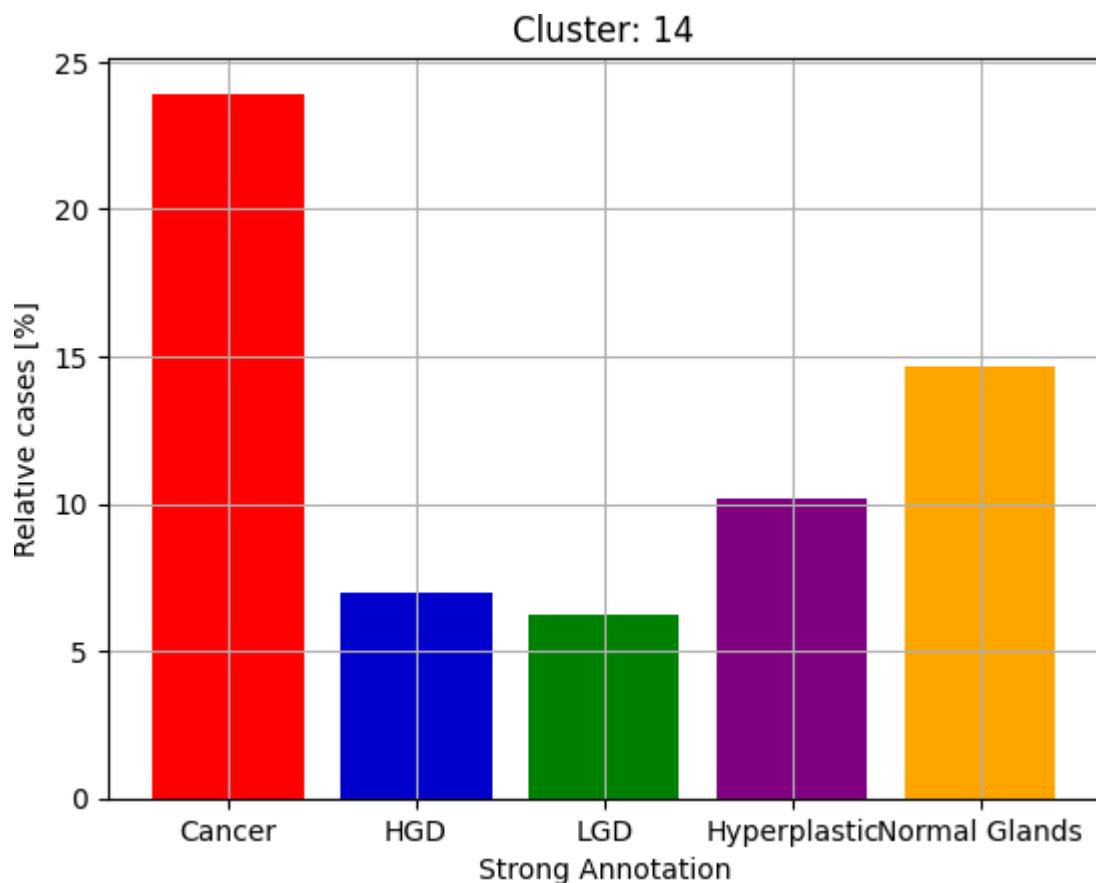


Figure 25. Distribution of the patches in Cluster 14. Please note that the number of relative cases is calculated with respect to all cases of a given class to account for class imbalance.

Cluster: 6

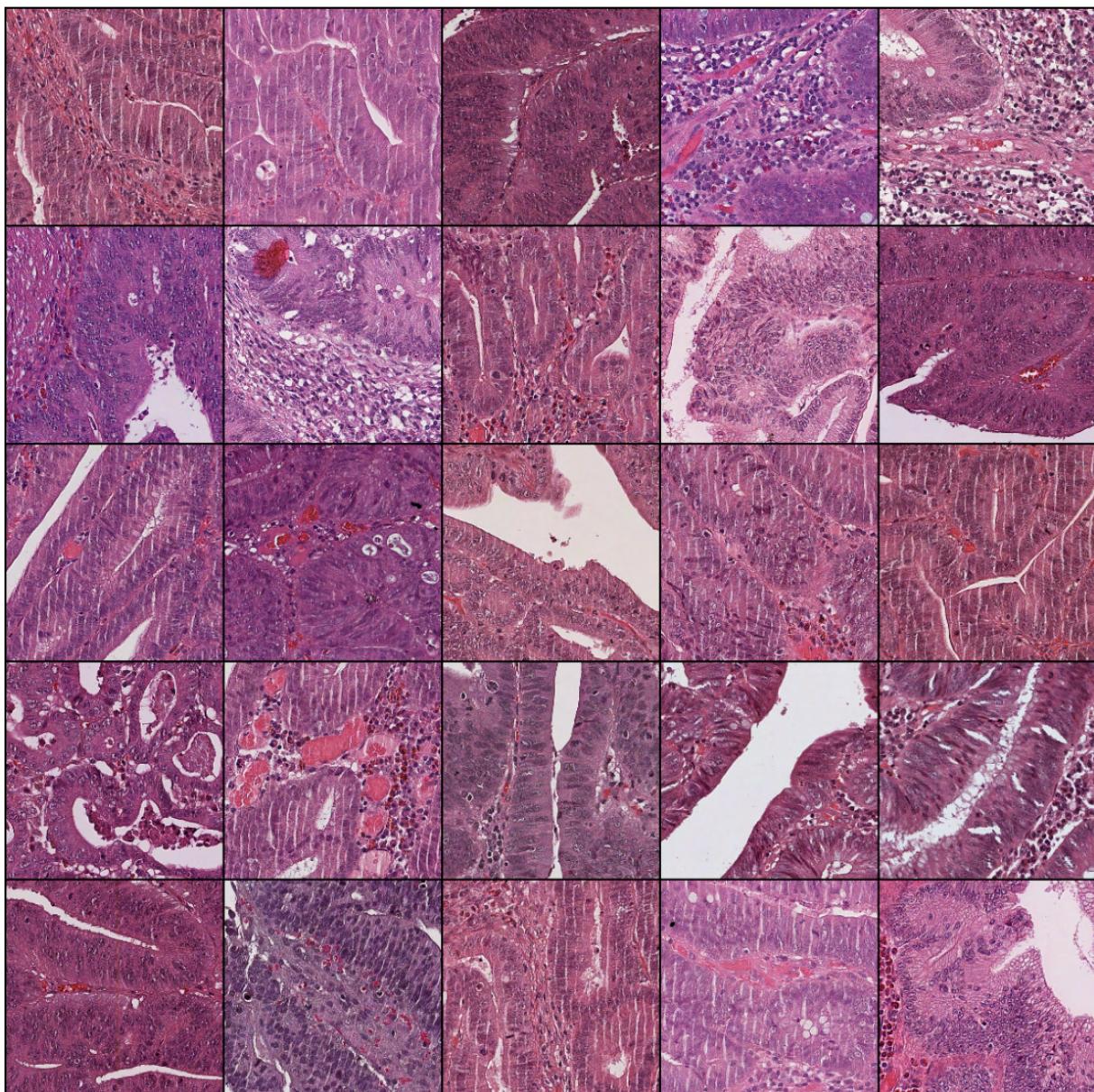


Figure 26. Exemplary patches from Cluster 6.

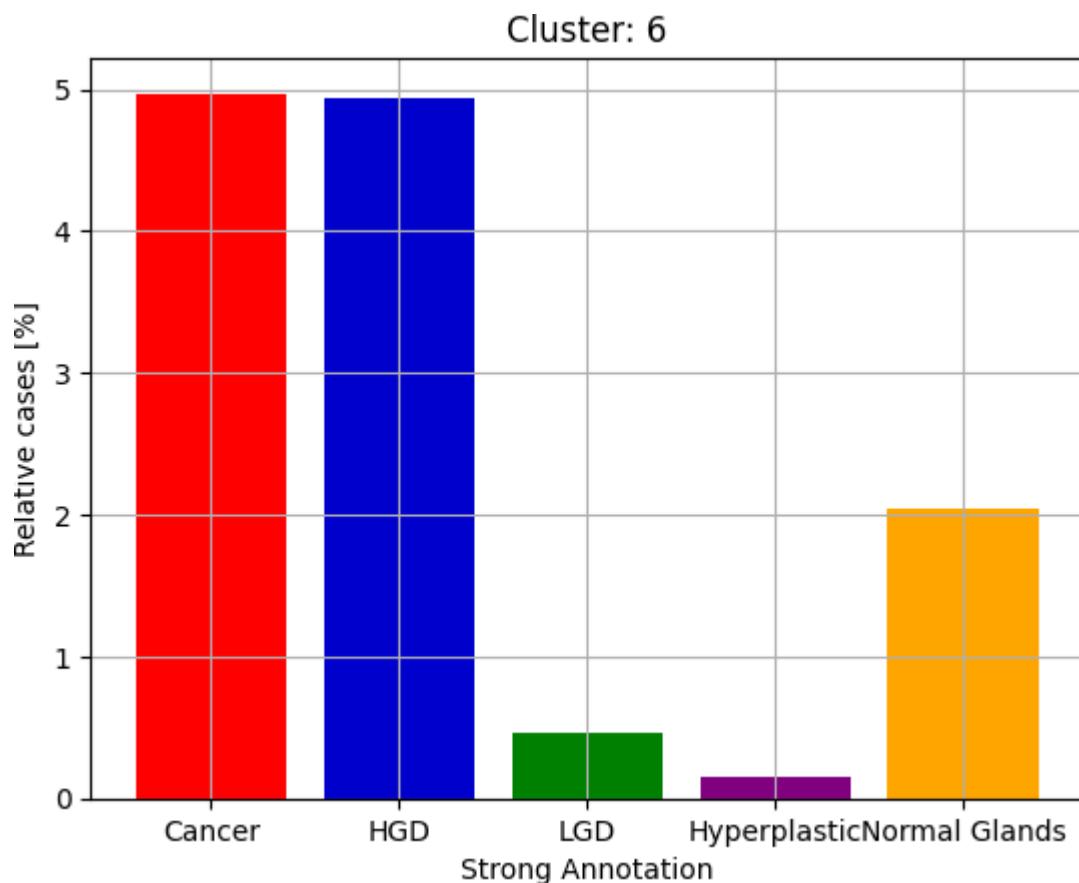


Figure 27. Distribution of the patches in Cluster 6. Please note that the number of relative cases is calculated with respect to all cases of a given class to account for class imbalance.

Hyperplastic

Cluster: 24

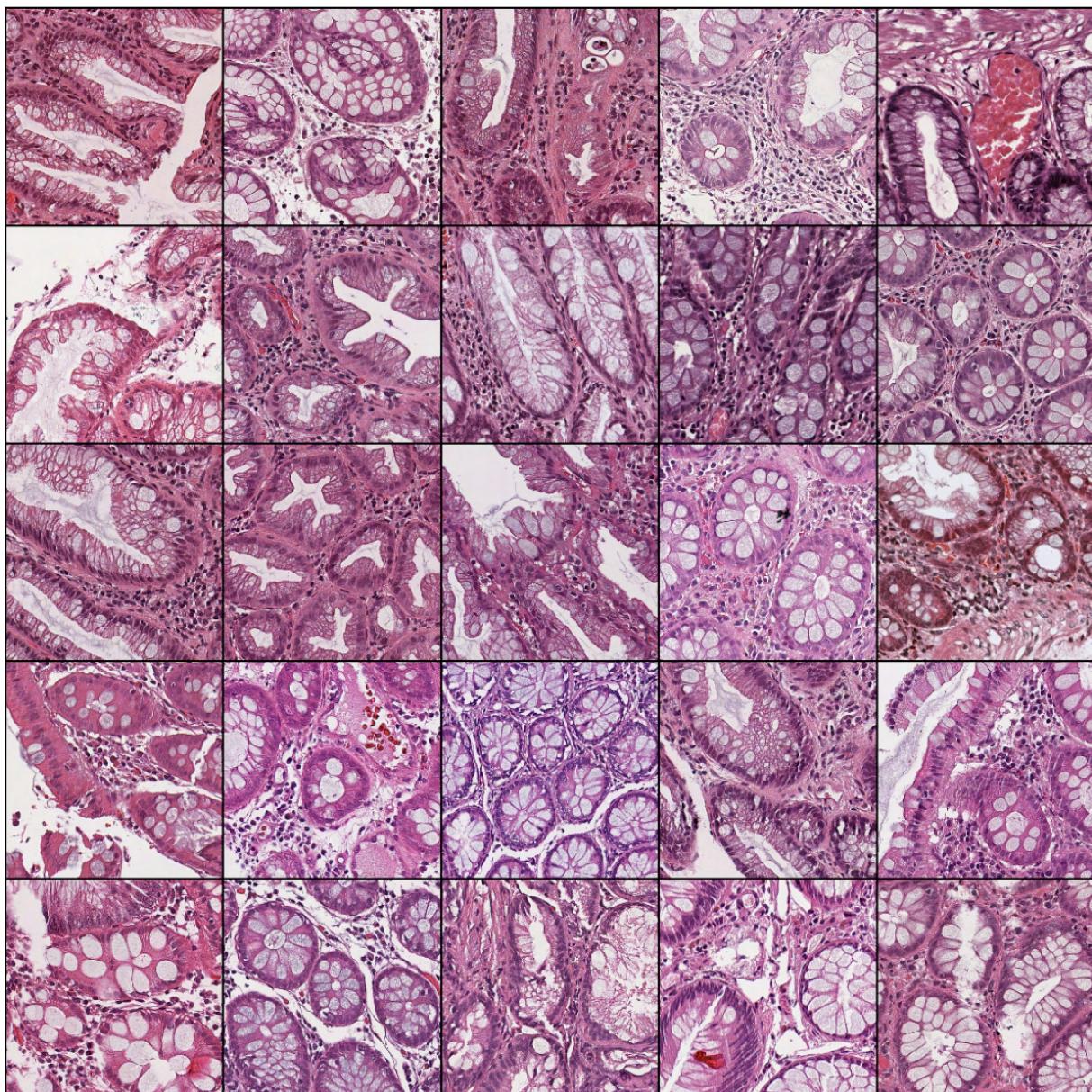


Figure 28. Exemplary patches from Cluster 24.

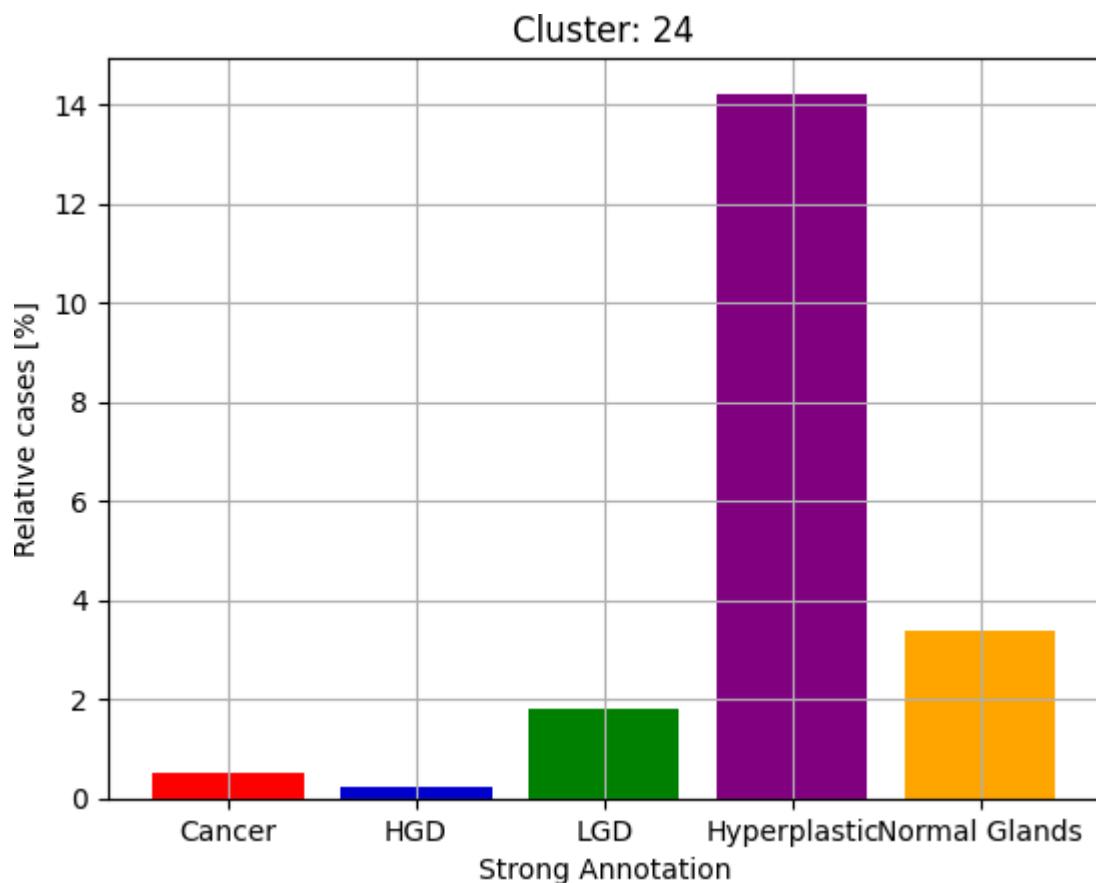


Figure 29. Distribution of the patches in Cluster 24. Please note that the number of relative cases is calculated with respect to all cases of a given class to account for class imbalance.

LGD

Cluster: 8

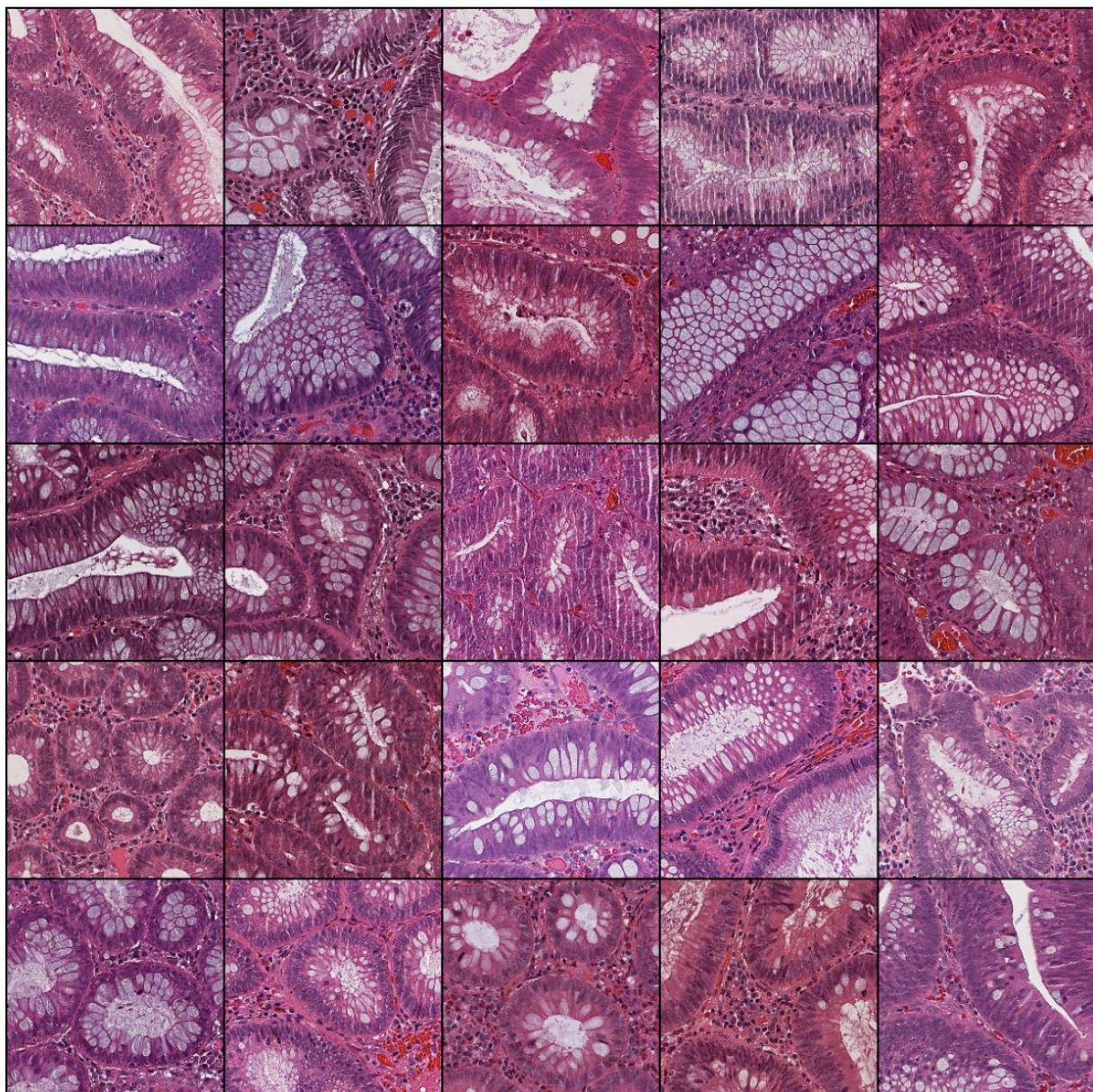


Figure 30. Exemplary patches from Cluster 8.

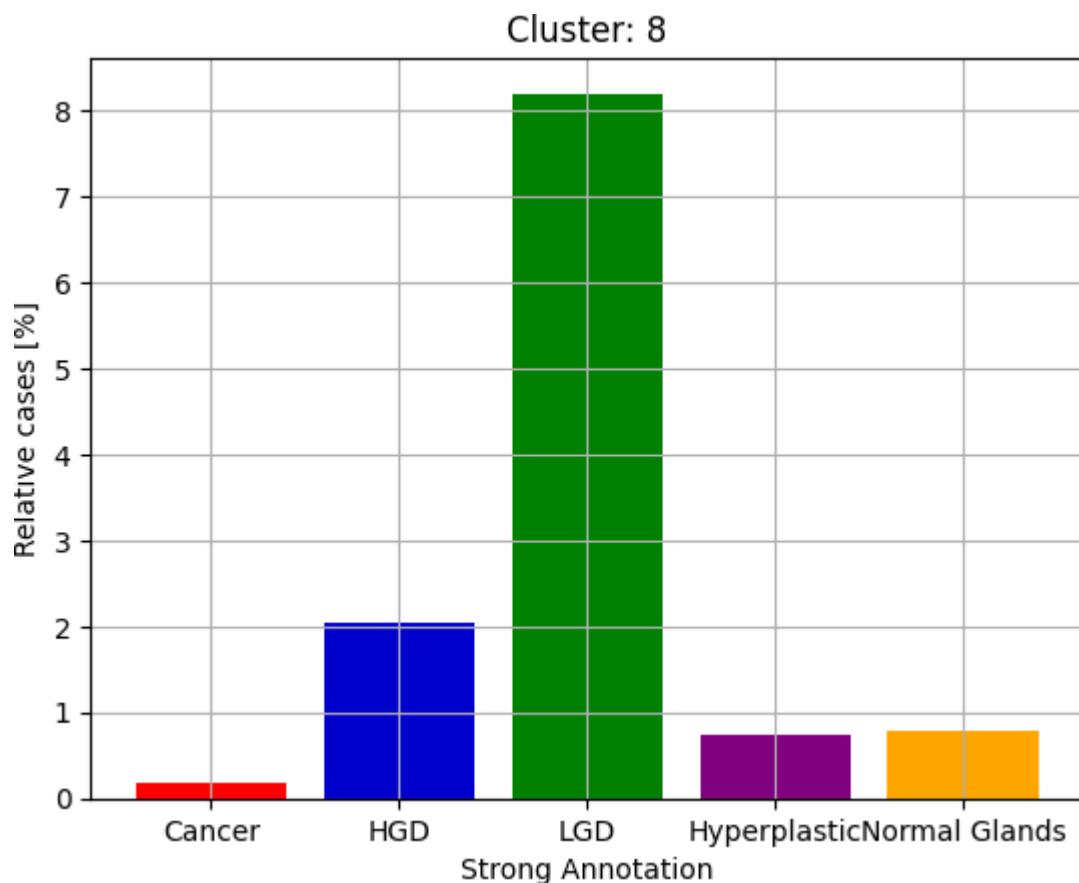


Figure 31. Distribution of the patches in Cluster 8. Please note that the number of relative cases is calculated with respect to all cases of a given class to account for class imbalance.

HGD

Cluster: 10

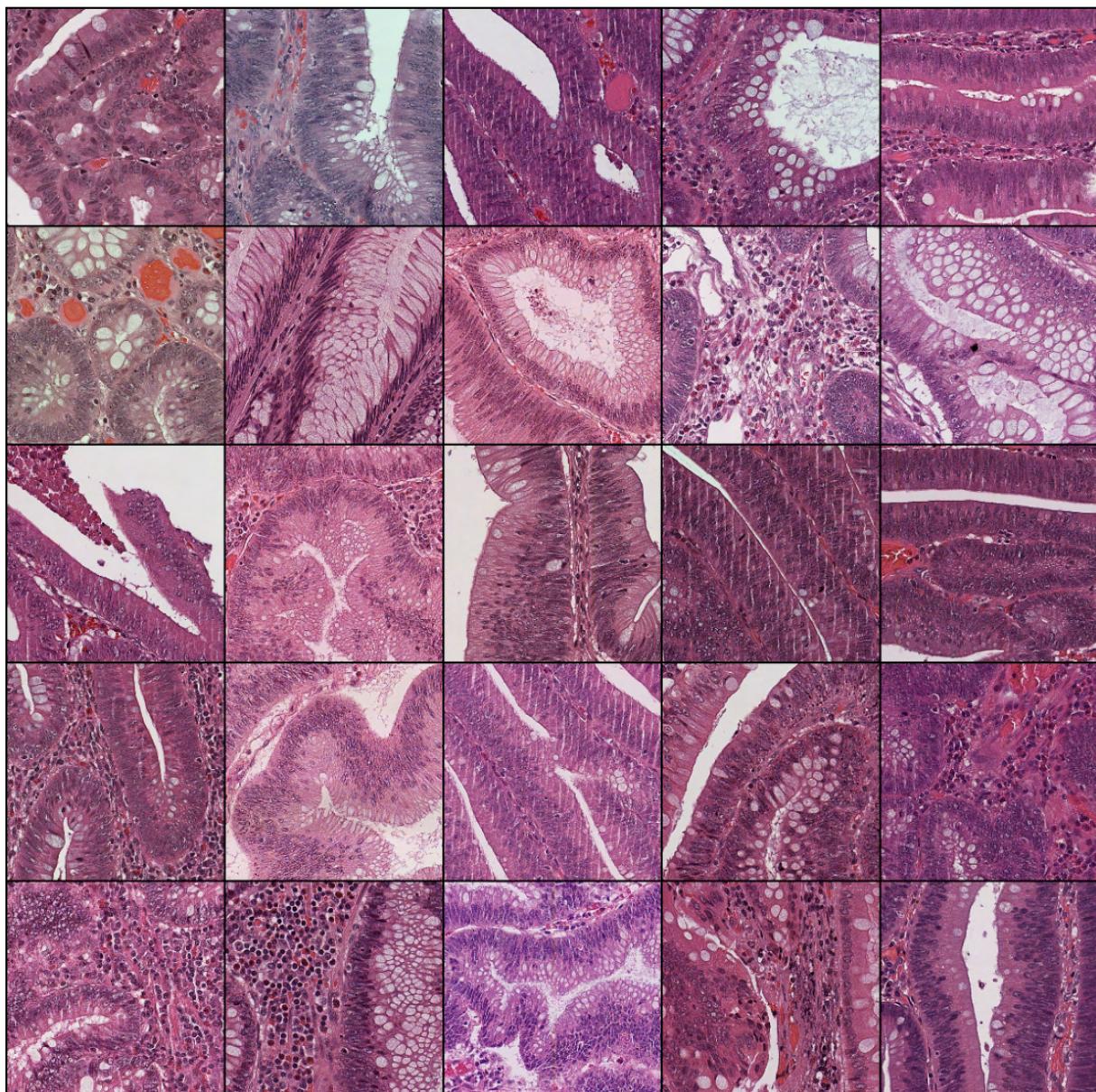


Figure 32. Exemplary patches from Cluster 10.

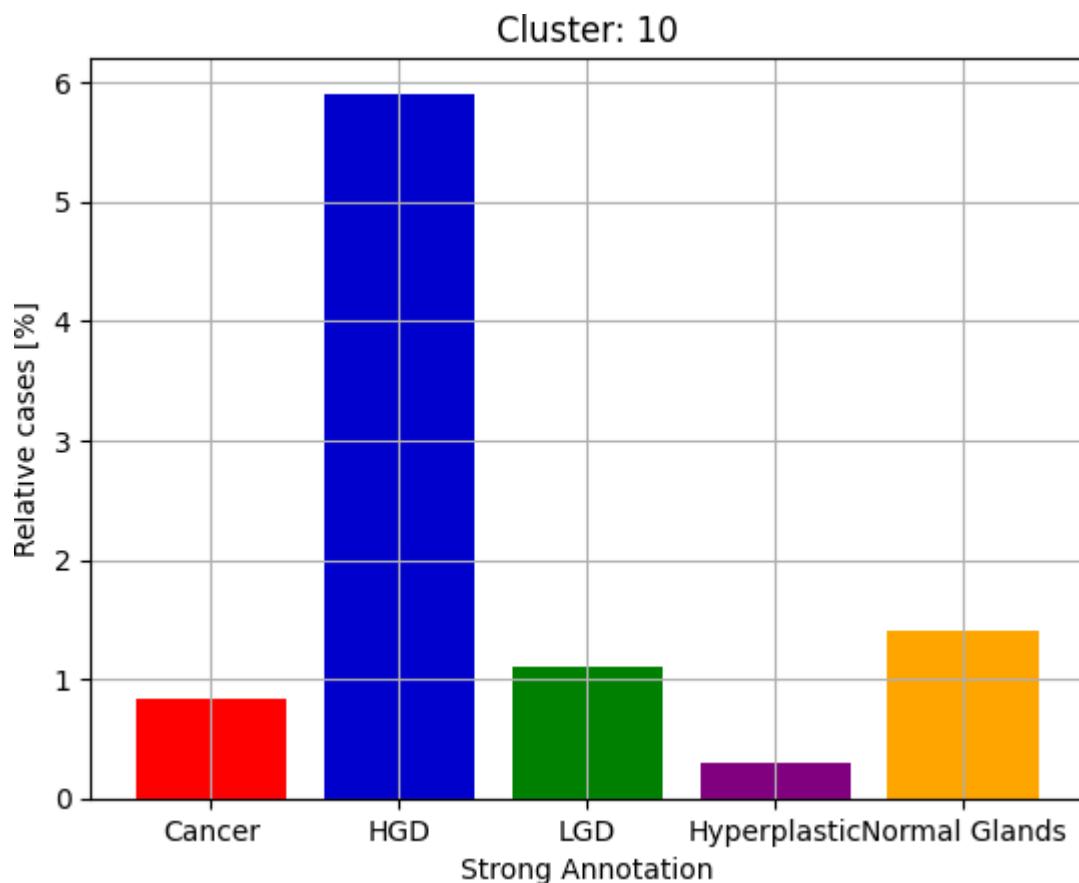


Figure 33. Distribution of the patches in Cluster 10. Please note that the number of relative cases is calculated with respect to all cases of a given class to account for class imbalance.