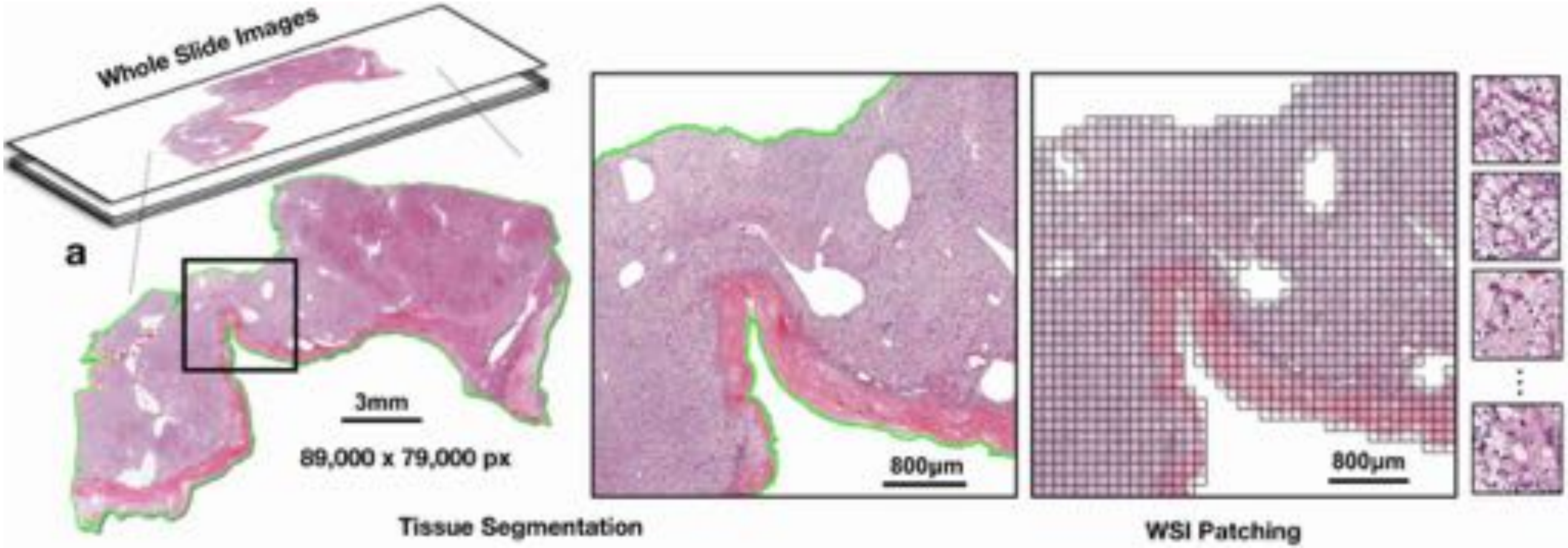
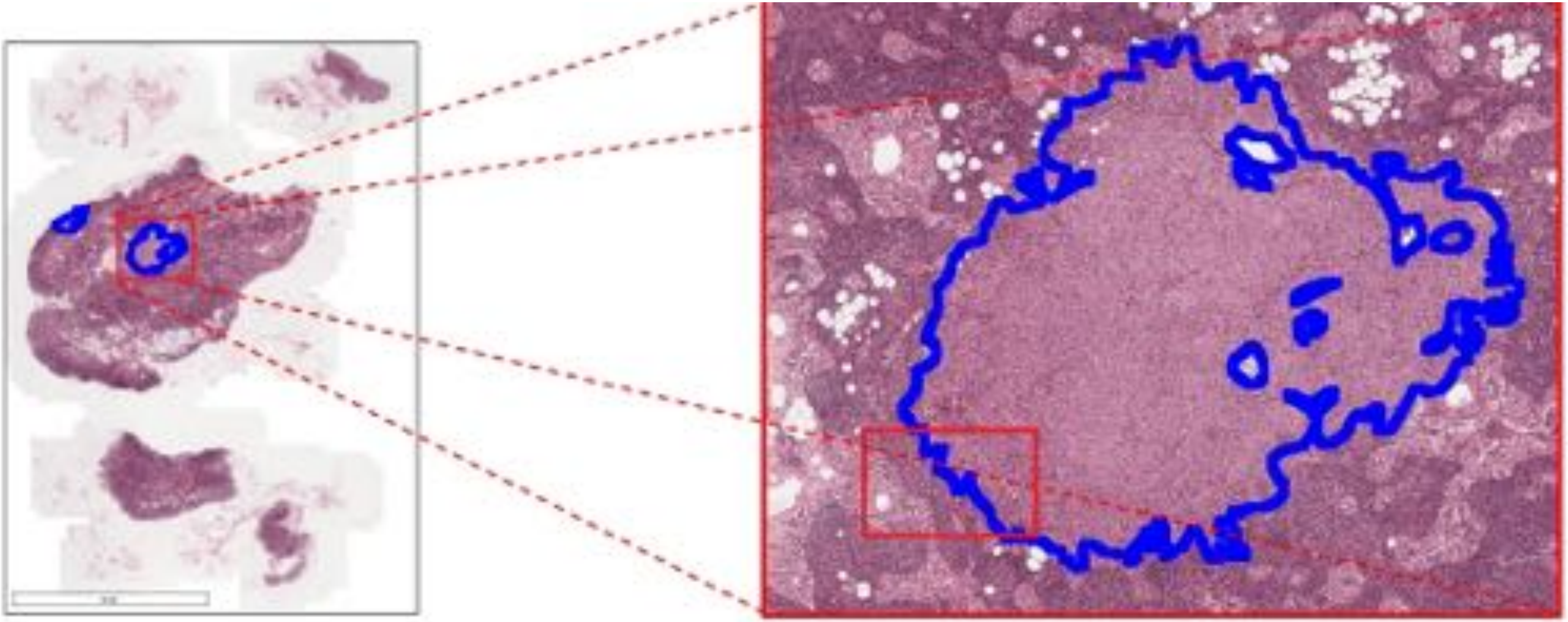
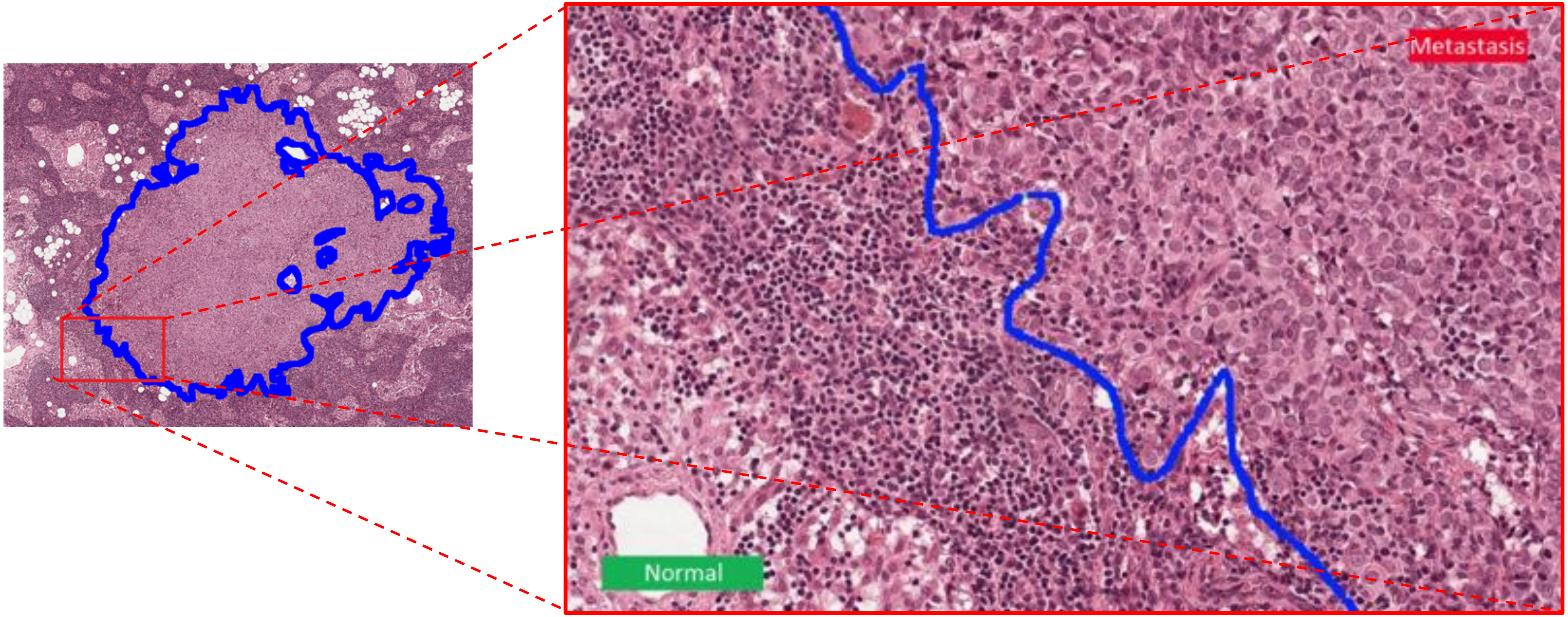
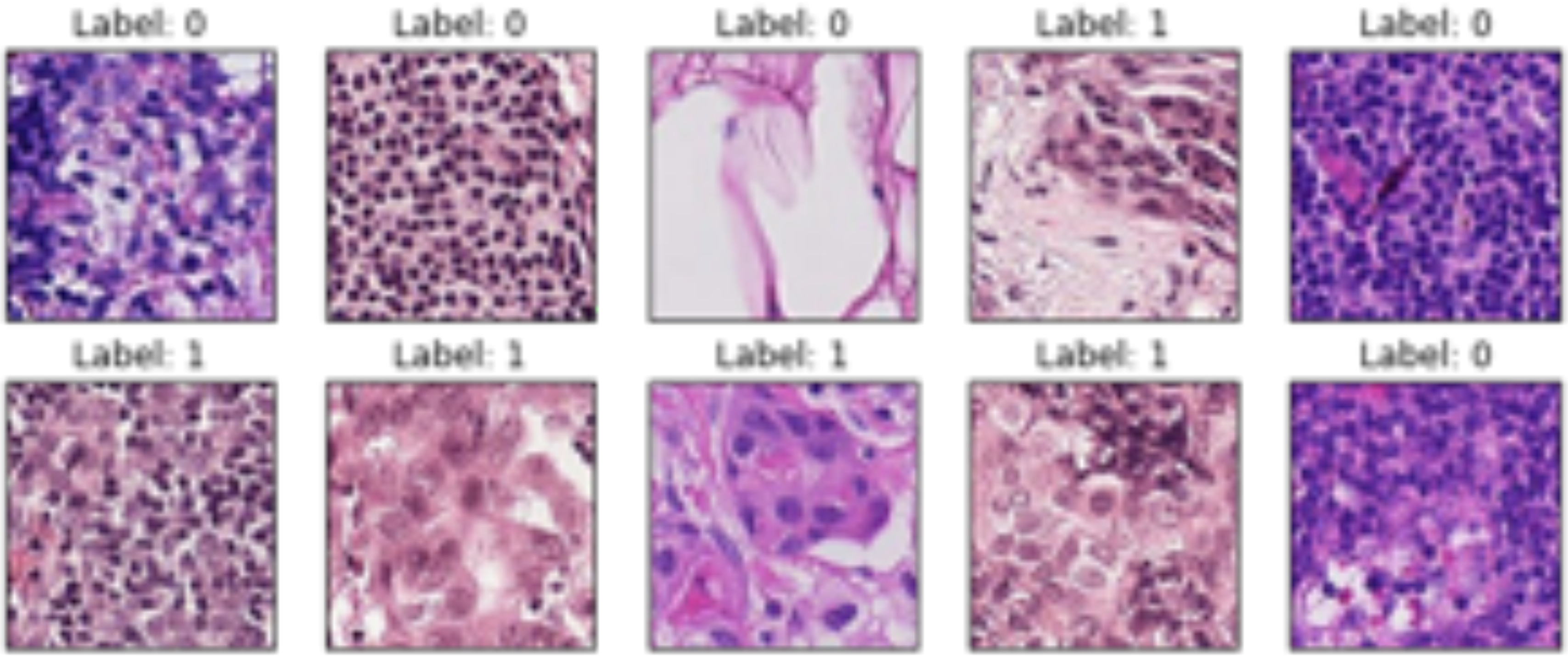

Whole Slide Images
a
3mm
89,000 x 79,000 px
800µm
800µm
Tissue Segmentation
WSI Patching



The data in this challenge contains a total of 400 whole-slide images (WSIs)

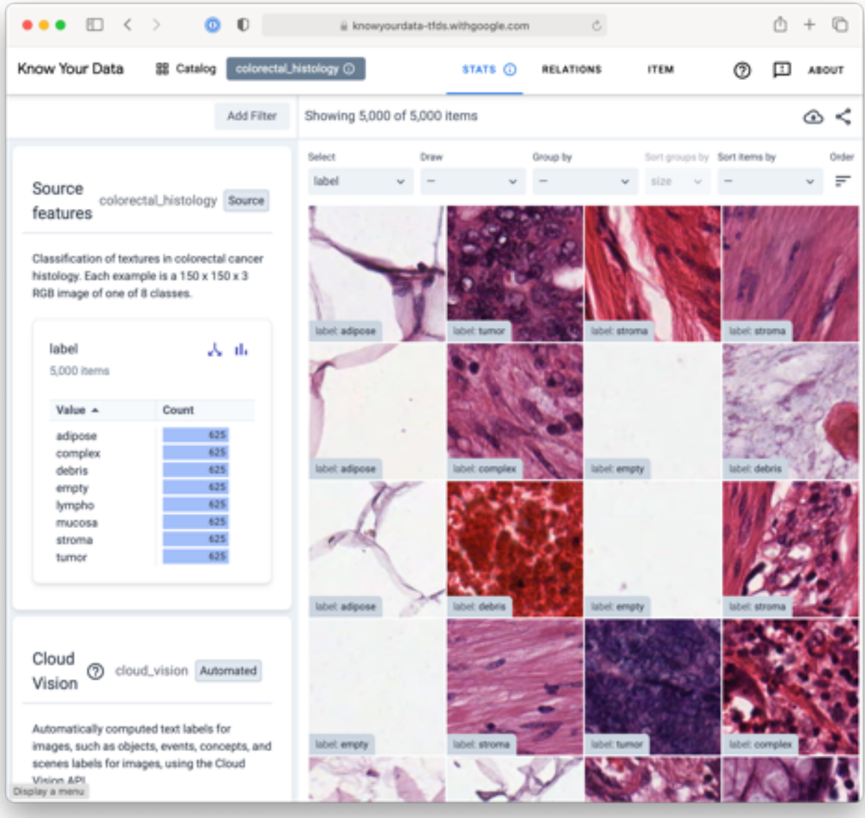



Label: 0
Label: 0
Label: 0
Label: 1
Label: 0
Label: 1
Label: 1
Label: 1
Label: 1
Label: 0

Colorectal histology dataset:

(The dataset for the final project)

Classification of textures in colorectal cancer histology.
Each example is a 150 x 150 x 3 RGB image of one of 8 classes



https://knowyourdata-tfds.withgoogle.com/#tab=STATS&dataset=colorectal_histology

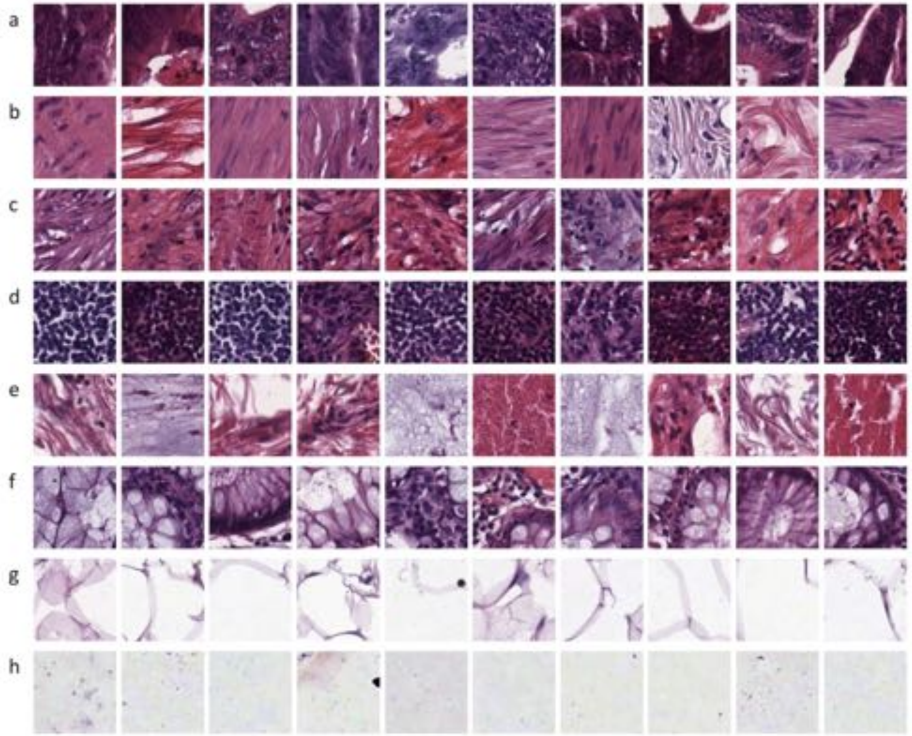


Figure 1. Representative images from our dataset. Here, the first 10 images of every tissue class in our dataset are shown. They represent the wide variation of illumination, stain intensity and tissue textures present in routine histopathological images. Images were extracted from 10 independent samples of colorectal cancer (CRC) primary tumours. (**a**) tumour epithelium, (**b**) simple stroma, (**c**) complex stroma (stroma that contains single tumour cells and/or single immune cells), (**d**) immune cell conglomerates, (**e**) debris and mucus, (**f**) mucosal glands, (**g**) adipose tissue, (**h**) background.

SCIENTIFIC REPORTS

OPEN

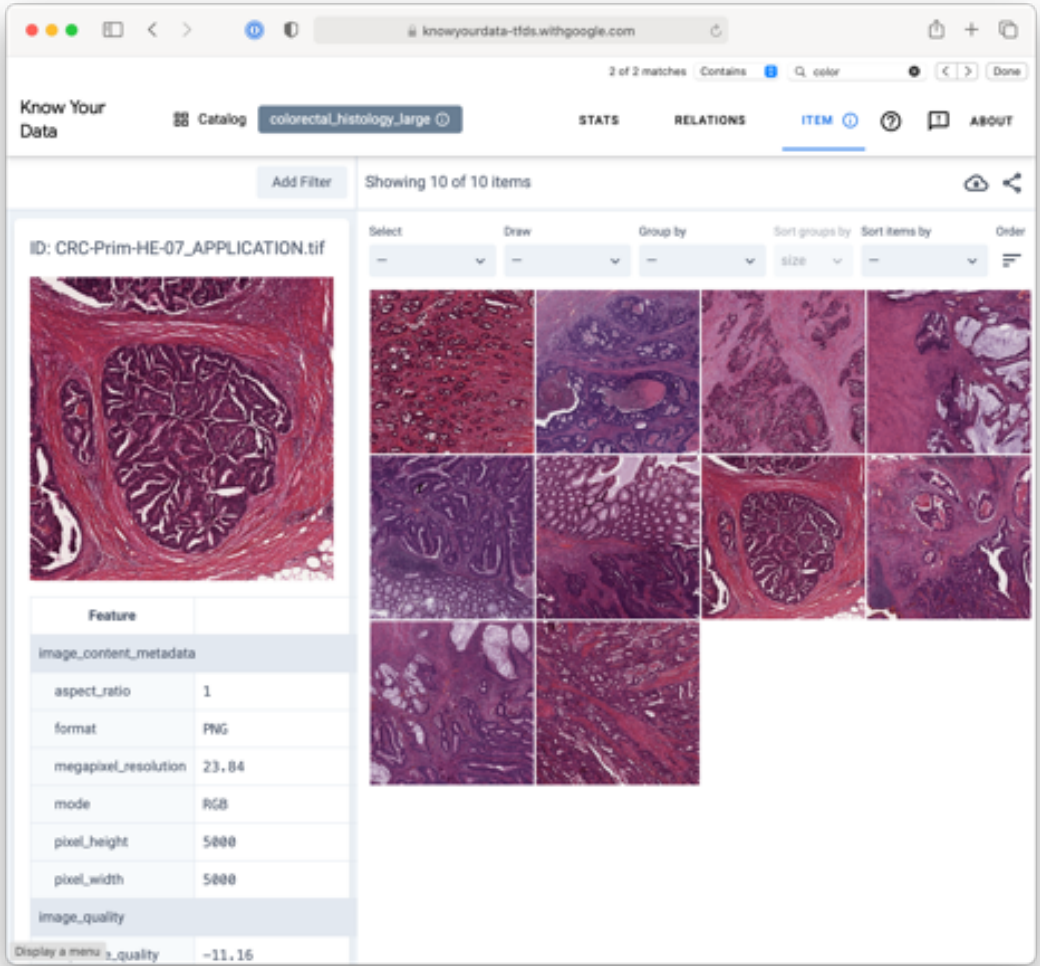
Multi-class texture analysis in colorectal cancer histology

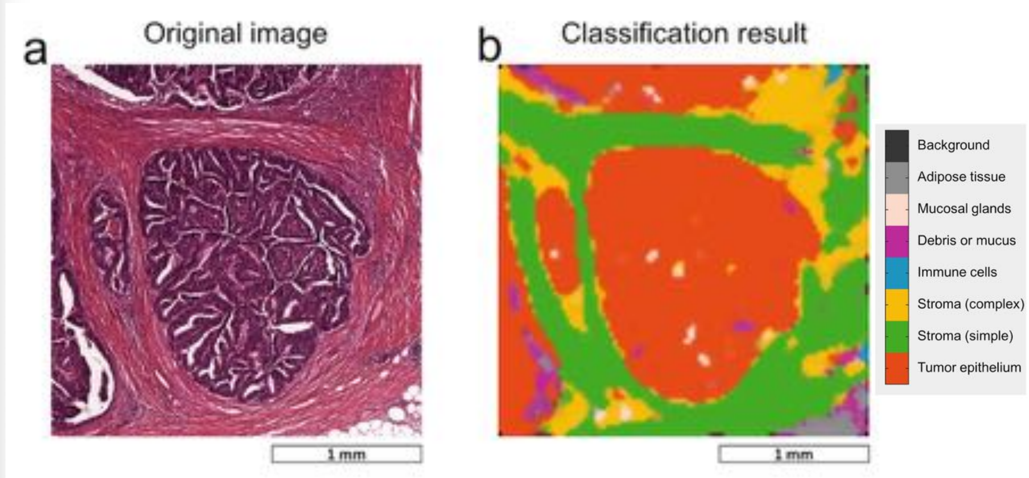
Jakob Nikolas Kather[1,2], Cleo-Aron Weis[1], Francesco Bianconi[3], Susanne M. Melchers[4], Lothar R. Schad[2], Timo Gaiser[1], Alexander Marx[1] & Frank Gerrit Zöllner[2]

Received: 02 March 2016
Accepted: 25 May 2016
Published: 16 June 2016

Automatic recognition of different tissue types in histological images is an essential part in the digital pathology toolbox. Texture analysis is commonly used to address this problem; mainly in the context of estimating the tumour/stroma ratio on histological samples. However, although histological images typically contain more than two tissue types, only few studies have addressed the multi-class problem. For colorectal cancer, one of the most prevalent tumour types, there are in fact no published results on multiclass texture separation. In this paper we present a new dataset of 5,000 histological images of human colorectal cancer including eight different types of tissue. We used this set to assess the classification performance of a wide range of texture descriptors and classifiers. As a result, we found an optimal classification strategy that markedly outperformed traditional methods, improving the state of the art for tumour-stroma separation from 96.9% to 98.6% accuracy and setting a new standard for multiclass tissue separation (87.4% accuracy for eight classes). We make our dataset of histological images publicly available under a Creative Commons license and encourage other researchers to use it as a benchmark for their studies.

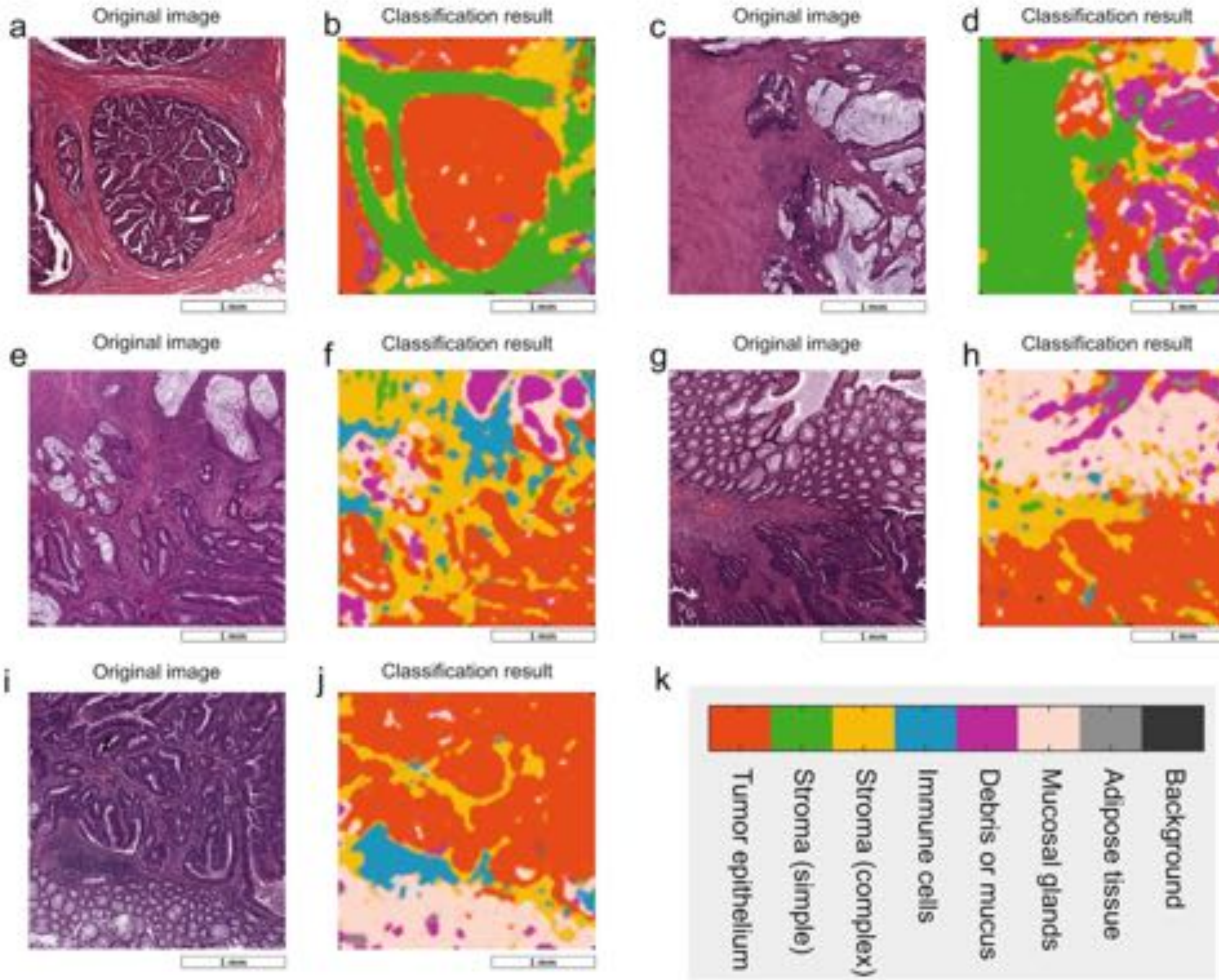
https://www.researchgate.net/publication/303998214_Multi-class_texture_analysis_in_colorectal_cancer_histology/link/5762810e08ae0c8c2553b63b/download





(5000x5000 pixels image)

https://knowyourdata-tfds.withgoogle.com/#tab=STATS&dataset=colorectal_histology_large



a
Original image
b
Classification result
c
Original image
d
Classification result
e
Original image
f
Classification result
g
Original image
h
Classification result
i
Original image
j
Classification result
k
Tumor epithelium
Stroma (simple)
Stroma (complex)
Immune cells
Debris or mucus
Mucosal glands
Adipose tissue
Background

Project goals

- Train a ConvNet classifier
  (either design your own or use a pretrained network + transfer-learning/fine-tuning)

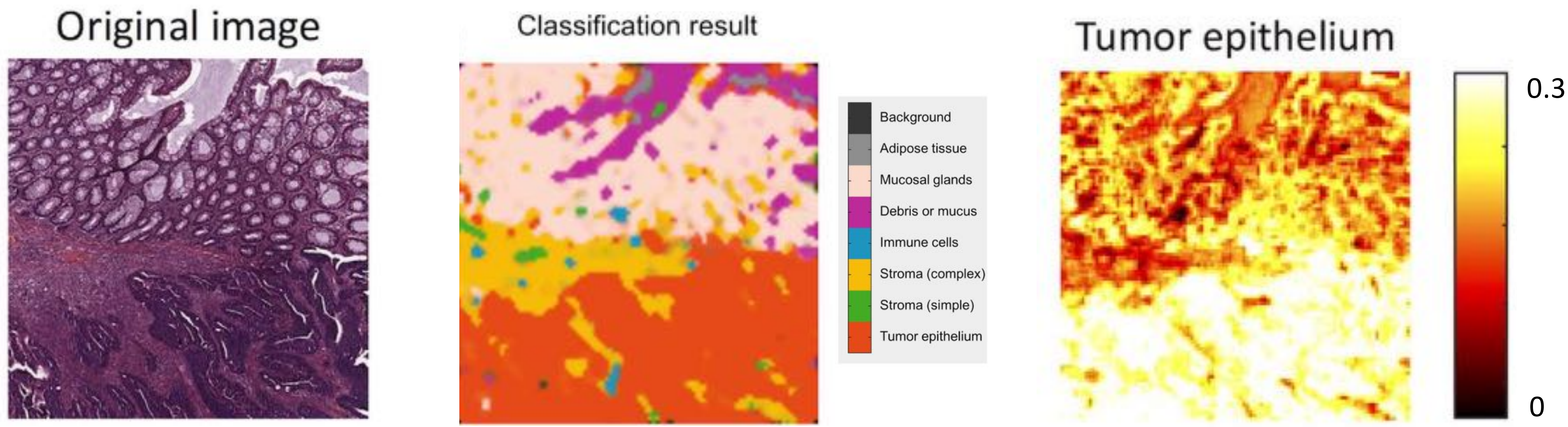
- Use train/validation protocols
  (e.g., 90% data for training, 10% for validation)

- Try several schemes, and compare your results in a table
  (e.g., different architectures, optimizers, preprocessing & data-augmentation, …)

Project goals

- Using your best model:
  - Plot its confusion matrix.
  - Use its *one-before-last* layer (its "features vector" layer), and **PCA** or ***t*-SNE** to visualize the labelled-data distribution in **2-D**.
  - Pick each of the large 5000x5000 pixels images, slice them into smaller 150x150 patches, and use your model to classify each patch. Visualize the classification results (on-top or side-by-side of the large 5000x5000 images) in two ways:
    - Color each patch (in the large image) using the predicted class label (8 colors for 8 classes).
    - Color each patch using the probability value of the "tumor ep." class (a 'heatmap' result).

Visualization examples:



5000x5000 pixels

Hint: you can use some overlap (e.g., 50%) when dividing the large image into patches, to get a nicer results.