

The Cancer Genome Atlas Kidney Clear Cell Carcinoma ECE 4783 - Module 1 and 2 Feedback

Alex Faché and Sam Lovejoy, UG 2



Tuesday March 10, 2020

Content

- Clinical Problem Statement
- Literature Critique
- Flow Chart
- Module 1
 - Preprocessing
- Module 2
 - Feature Selection
 - Feature Extraction
- Project Timeline and Contribution
- References

Clinical Problem Statement

- to develop a decision support model for diagnosis and prognosis of cancer
- follow image processing pipeline for clinical decision making
- 1 - Image Processing
 - using "quality controlled" images, apply normalization and data augmentation
- 2 - Image Feature Extraction
 - extract color, texture, and morphological features from images
- 3 - Image Classification for Clinical Decision Making
 - develop models for three classifier techniques, apply two cross-validation schemes, analyze four performance metrics for each predictive model

Literature Critique: Alex

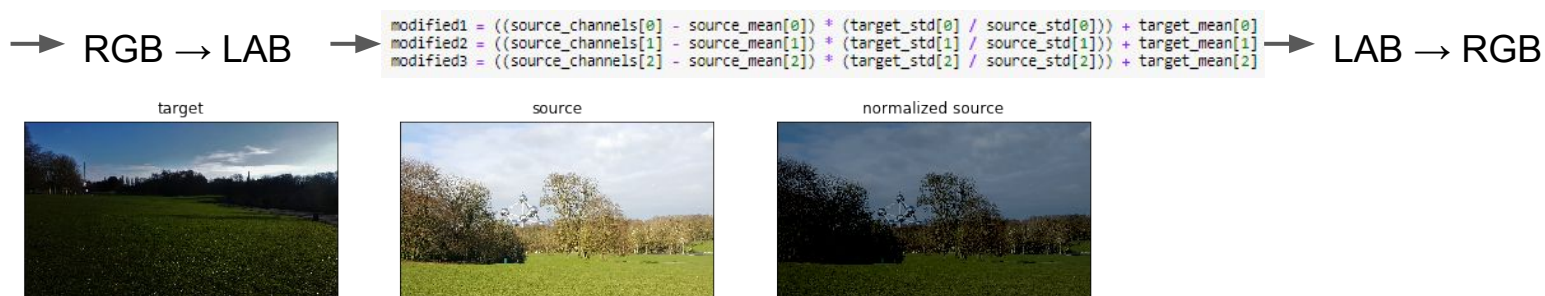
Automatic Batch-Invariant Color Segmentation of Histological Cancer Images ¹	Automatic Classification for Pathological Prostate Images Based on Fractal Analysis ²	Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features ³
color normalization	texture features	morphological features
Sonal Kothari, John H. Phan, Richard A. Moffitt, Todd H. Stokes, Shelby E. Hassberger, Kaiser Chaudry, Andrew N. Young, May D. Wang	Po-Whei Huang, Cheng-Hsiung Lee	Haibo Wang, Angel Cruz-Roa, Ajay Basavanahally, Hannah Gilmore, Natalie Shih, Mike Feldman, John Tomaszewski, Fabio Gonzalez, Anant Madabhushia
2011	2009	2014
Conf Proc IEEE Int Symp Biomed Imaging	IEEE Transactions on Medical Imaging, Vol 28, No 7, July	Journal of Medical Imaging
<ul style="list-style-type: none"> normalize over color map instead of all pixels selects unique colors does not include frequency of colors 	<ul style="list-style-type: none"> fractal based feature set is more accurate than feature sets from multiwavelets, Gabor filters, GLCM 	<ul style="list-style-type: none"> cascaded strategy of CNN with handcrafted features faster approach and far less computationally expensive
<ul style="list-style-type: none"> all pixel normalization tends to distort colors more in the normalized image 	<ul style="list-style-type: none"> Gleason grading is based on texture properties apply fractal dimension to analyze texture is well suited 	<ul style="list-style-type: none"> light CNN that is less computationally expensive difficult to detect mitotic nuclei due to highly variable size and shape
<ul style="list-style-type: none"> use expert domain knowledge to increase overall performance of color segmentation 	<ul style="list-style-type: none"> fractal dimension feature set of small size that effectively grades prostate images according to Gleason grading methods used are differential box-counting and entropy-based fractal dimension estimation 	<ul style="list-style-type: none"> mitosis detection is one of 3 key factors in breast cancer grading existing methods used either handcrafted features or CNN

Literature Critique: Sam

Object-and spatial-level quantitative analysis of multispectral histopathology images for detection and characterization of cancer	Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles	Histological image feature mining reveals emergent diagnostic properties for renal cancer
All three, but mostly morphological	Texture features	Multiple
L. Boucheron	J. Barker, A. Hoogi, A. Depeursinge, D. L. Rubin	S. Kothari, J. H. Phan, A. N. Young, and M. D. Wang
2008	2016	2011
PhD Thesis UC Santa Barbara	Medical Image Analysis	IEEE Bioinformatics Biomedicine
Used multispectral analysis over RGB analysis for object recognition, attempts to classify breast cancer histopathology images using a known dataset from Yale University	Examines coarseness versus fineness in localized regions of brain tumor images for differentiation among cancer subtypes.	Color features come from 28 unique Gabor filters each with energy E and entropy H for square cropped subregions of renal cancer images.
This thesis did not implement feedback loops, there was no recursion in pixel-level analysis for object recognition.	Uses color segmentation, texture, and morphological features in "tiles", localized regions of whole slide (very large) images.	Texture features come from histograms using the eight fractal method. Eight statistics are used for each of 31 tiles plus the object length for a total of 249 texture features.
Non-nuclear features for feature extraction provided the most reliable results, though the algorithms were computationally expensive. No increases in accuracy.	The authors achieved an accuracy rating of between 95.5% and 100% for a dataset of 302 images using an elastic net classifier.	Shape and Topology features come from elliptical centers around the center nuclei and segmented stain objects. Other algorithms such as Voronoi diagram are used for topology features.

Module 1: Flow Chart

Reinhard's Method
(normalization)



Cropping (224 x 224) → `current_image = normalized_images[i, r:r+crop_size[0], c:c+crop_size[1], :]` → perform 5 crops per image

Rotating (90°, 180°, 270°) →

```
if k == 0: #rotate 90 CCW
    temp = np.transpose(current_image, axes=(1, 0, 2))
    modified_image[::-1, :, :] = temp
elif k == 1: #rotate 180 CCW
    temp[:, ::-1, :] = current_image
    modified_image[:, ::-1, :] = temp
elif k == 2: #rotate 270 CCW
    temp = np.transpose(current_image, axes=(1, 0, 2))
    modified_image[:, ::-1, :] = temp
```

 → perform 1 of each

Flipping (Vertical, Horizontal) →

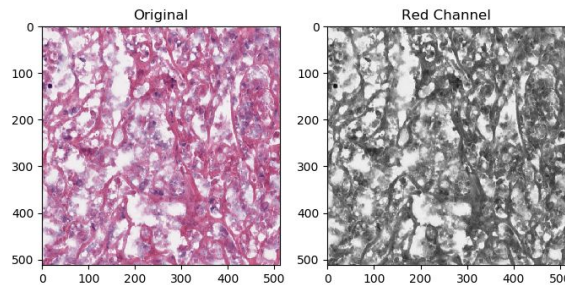
```
elif k == 3: #flip vertically
    modified_image[::-1, :, :] = current_image
elif k == 4: #flip horizontally
    modified_image[:, ::-1, :] = current_image
```

 → perform 1 of each

Module 2: Flow Chart

100 features
(color)

→ Red Channel →



→ Red
Statistics

```
Average: 0.578278  
Standard Deviation: 0.20219949  
Variance: 0.040884633  
Max: 1.0  
Min: 0.03137255
```

Shannon Entropy

100 features (texture) →

$$H = - \sum p(x) \log p(x)$$

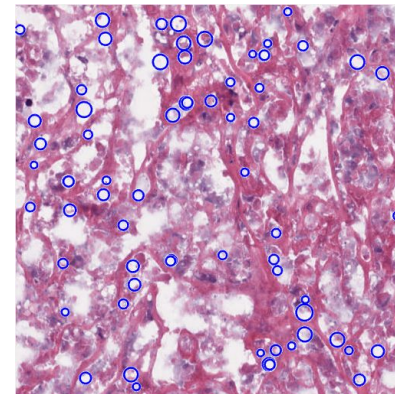
→ 6.182

Module 2: Flow Chart

100 features
(morphology)

HCT

```
hct.m x +
1 - img = imread('necrosis.png');
2 - [centers, radii] = imfindcircles(img, [4,18]);
3 - imshow(img)
4 - hold on
5 - viscircles(centers, radii, 'EdgeColor', 'b');
```



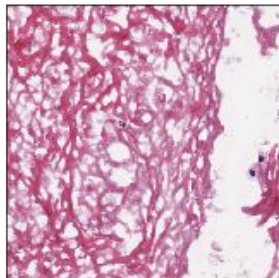
PCA

Data Reduction
Algorithm

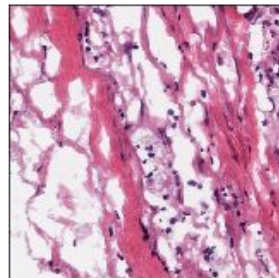
```
hct.m x pca_script.m x +
1 - I = double(imread('necrosis.png'));
2 - X = reshape(I, size(I, 1)*size(I, 2), 3); % Reshape into 1-dimension (3-bit depth)
3 - coeff = pca(X); % get the pca coefficients
4 - Itransformed = X*coeff; % apply them
5 - Ipc1 = reshape(Itransformed(:, 1), size(I, 1), size(I, 2)); % build red
6 - Ipc2 = reshape(Itransformed(:, 2), size(I, 1), size(I, 2)); % build blue
7 - Ipc3 = reshape(Itransformed(:, 3), size(I, 1), size(I, 2)); % build green
8 - % show stuffs
9 - figure, imshow(Ipc1, []);
10 - figure, imshow(Ipc2, []);
11 - figure, imshow(Ipc3, []);
```


Original Data

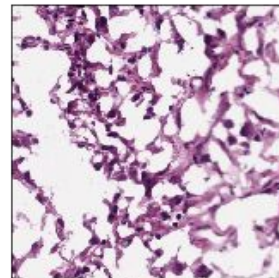
- Necrosis
 - premature death of cells
- Stroma
 - healthy connective tissue cells
- Tumor
 - abnormal cell divides uncontrollably destroying body tissue



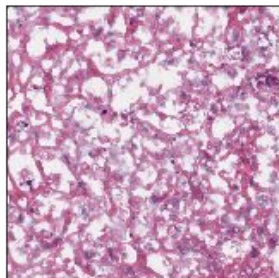
Necrosis



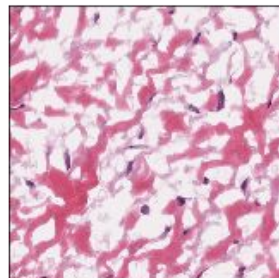
Stroma



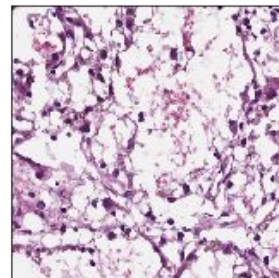
Tumor



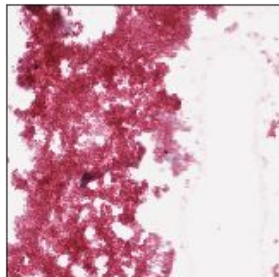
Necrosis



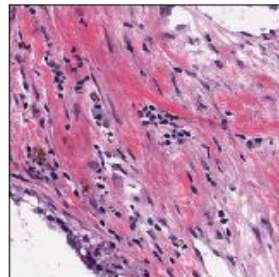
Stroma



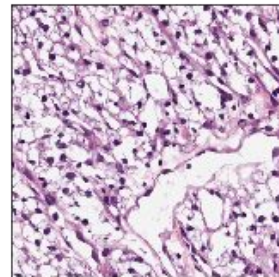
Tumor



Necrosis

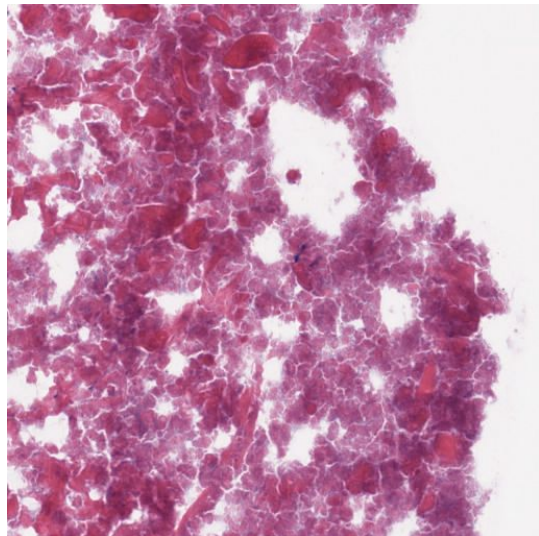


Stroma



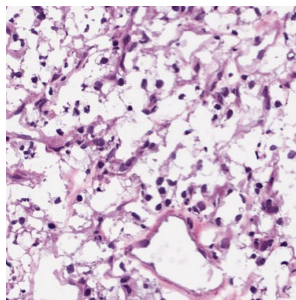
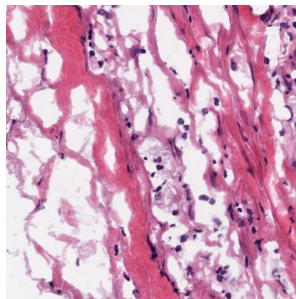
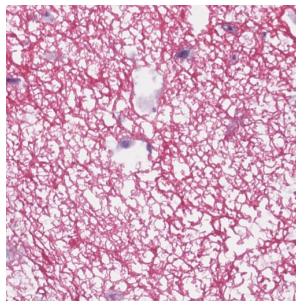
Tumor

Preprocessing: Necrosis Target

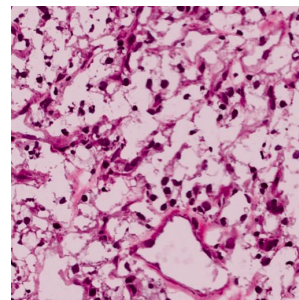
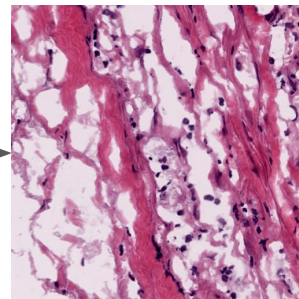
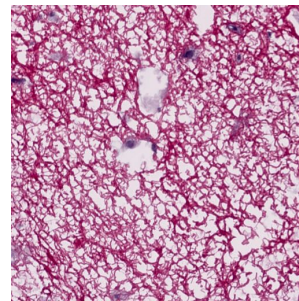


Target

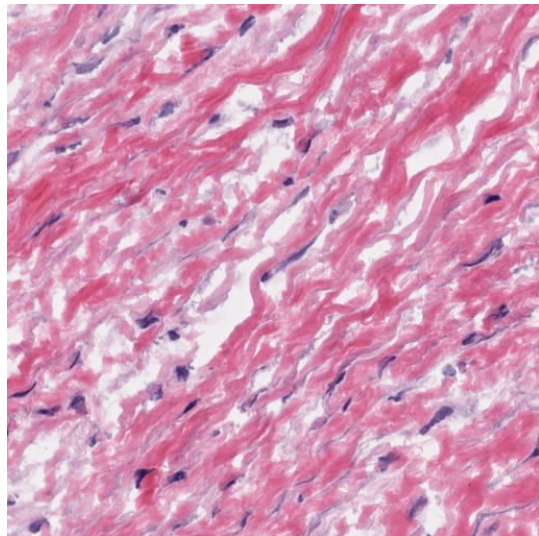
Source



Normalized
Source

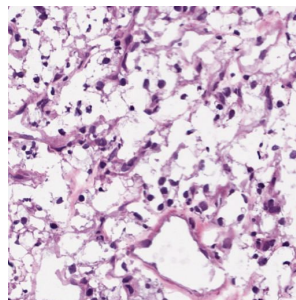
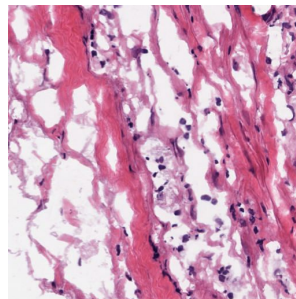
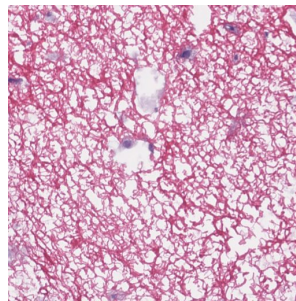


Preprocessing: Stroma Target

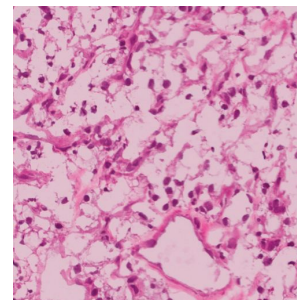
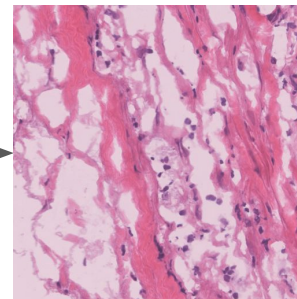
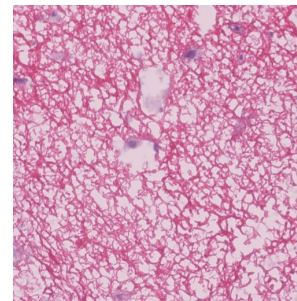


Target

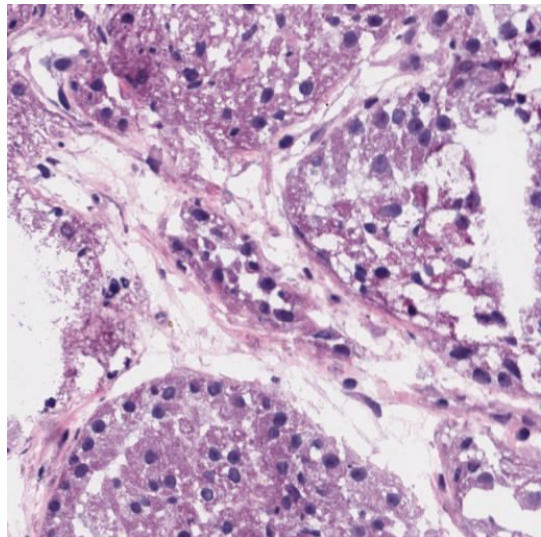
Source



Normalized
Source

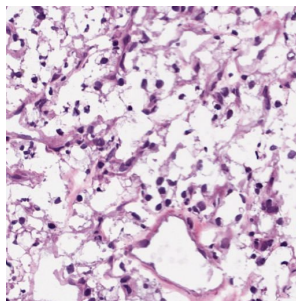
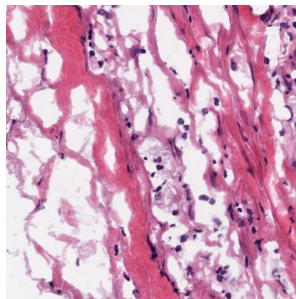
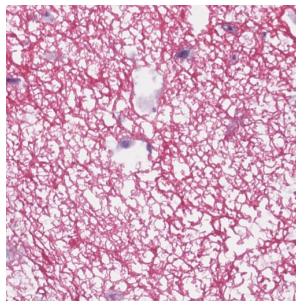


Preprocessing: Tumor Target

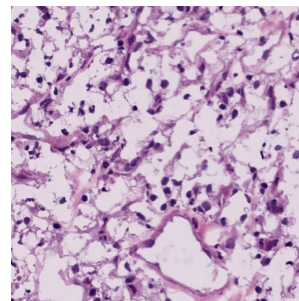
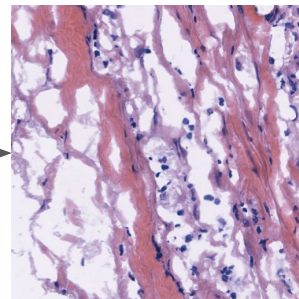
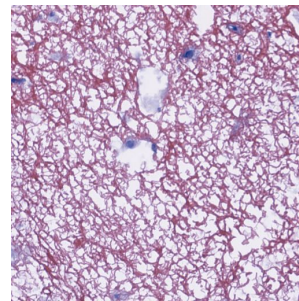


Target

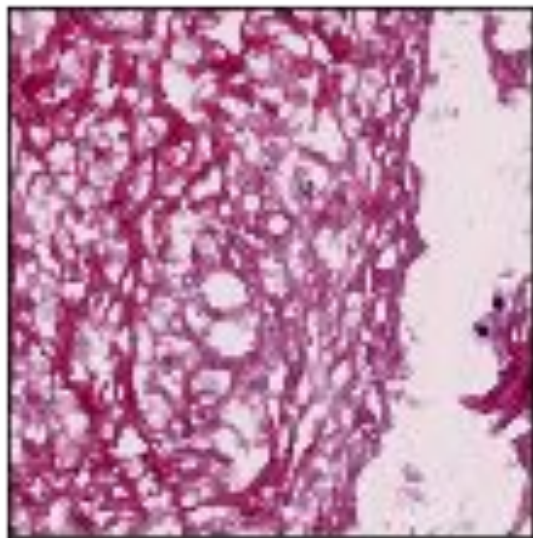
Source



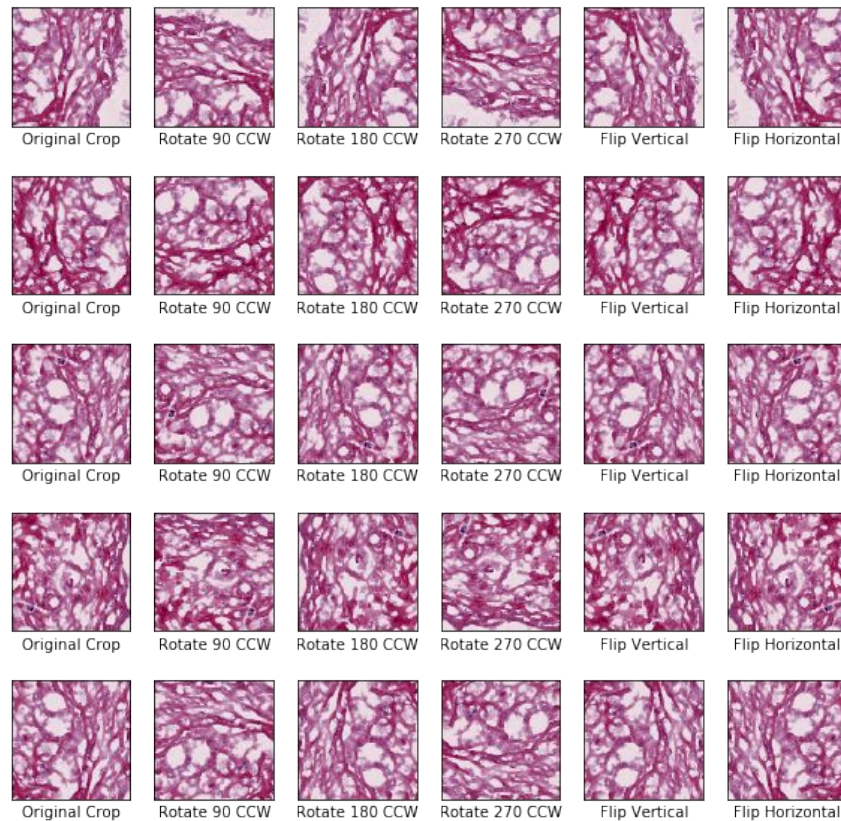
Normalized
Source



Preprocessing: Cropping, Flipping, Rotating

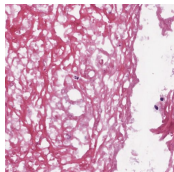


Necrosis

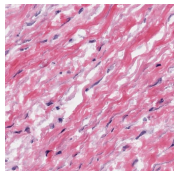


Feature Selection and Extraction - Color HSV

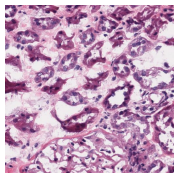
```
norm_h = normalize(h, norm='l2')  
norm_s = normalize(s, norm='l2')  
norm_v = normalize(v, norm='l2')  
  
return np.array([np.mean(norm_h), np.mean(norm_s), np.mean(norm_v), np.std(norm_h), np.std(norm_s), np.std(norm_v)], dtype=np.float32)
```



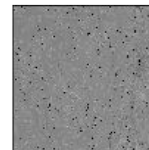
[0.04386097 0.03393478 0.04388783 0.00541669 0.02831175 0.00519455]



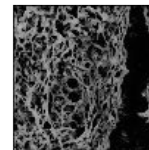
[0.04407357 0.03931944 0.04402243 0.00326273 0.02017689 0.00389242]



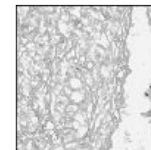
[0.04395115 0.03608433 0.0431277 0.00462836 0.0255156 0.00965021]



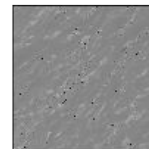
H



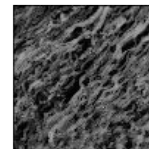
S



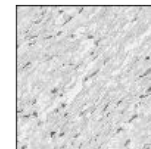
V



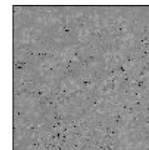
H



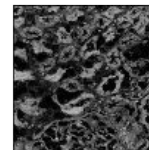
S



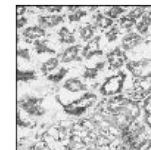
V



H



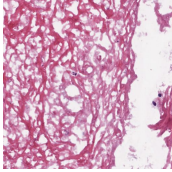
S



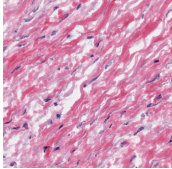
V

Feature Selection and Extraction - Texture GLCM

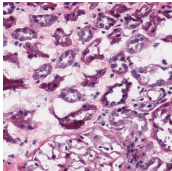
```
contrast = feature.greycomprops(glcm, prop='contrast')  
homogeneity = feature.greycomprops(glcm, prop='homogeneity')  
energy = feature.greycomprops(glcm, prop='energy')  
correlation = feature.greycomprops(glcm, prop='correlation')
```



[0.33699545 0.34934375 0.35286507 0.35348353 0.10693018 0.05439614 0.02205111 0.00702876]



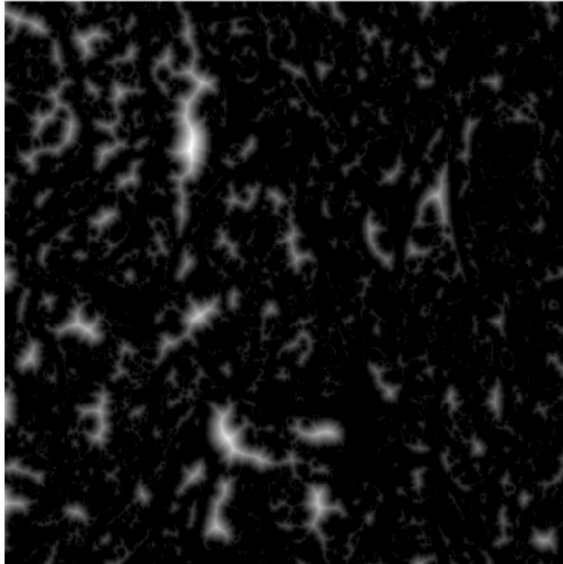
[0.32264647 0.348883 0.35213572 0.35340744 0.14456578 0.05727689 0.03162951 0.01015728]



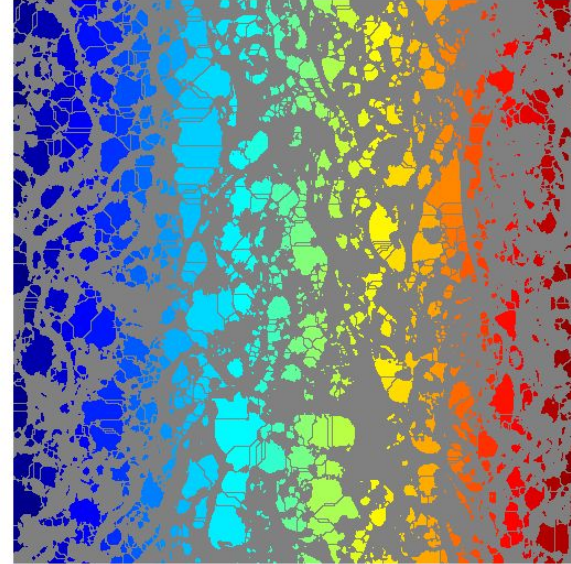
[0.3376464 0.3502538 0.35295105 0.35349736 0.10485661 0.04819005 0.02062924 0.00629445]

Feature Selection and Extraction - Morphological

Distance Transform of Necrosis 31



Watershed Transform



Feature Selection and Extraction - Summary

Color	Texture	Morphological	Dimensionality Reduction
RGB (μ , σ)	entropy	symmetry	Principal Component Analysis
HSI (μ , σ)	contrast, homogeneity, energy, correlation	erosion	Random Forest sequential forward selection
LAB (μ , σ)	nucleus density for subregions (μ , σ)	dilation	
	nucleus shape (minor axis, major axis)	Watershed, Canny edge detection	
	DCT (μ , σ)	SNR	
	DFT (μ , σ)	noise mask	
	histogram analysis	edge eccentricity	
	Jaccard similarity index		
	Sørensen–Dice coefficient		
	Autoencoder		

Project Timeline and Contribution

1. Module 1 and 2 Feedback
 - a. March 10, 2020
2. Module 2 Literature Review
 - a. March 12, 2020
3. Final Presentation and Report
 - a. April 6, 2020

- Alex
 - Module 1
 - feature extraction
- Sam
 - Module 2
 - 100+ Feature extraction
 - PCA reduction and Image Compression
 - Autoencoder implementation

References

1. S. Kothari, J. H. Phan, R. A. Moffitt, T. H. Stokes, S. E. Hassberger, Q. Chaudry, et al., "Automatic batch-invariant color segmentation of histological cancer images," Conf Proc IEEE Int Symp Biomed Imaging, ISBI, pp. 657-660, 2011.
2. P. Huang and C. Lee, "Automatic Classification for Pathological Prostate Images Based on Fractal Analysis," in IEEE Transactions on Medical Imaging, vol. 28, no. 7, pp. 1037-1050, July 2009.
3. H. Wang, A.C. Roa et al, "Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features". Journal of Medical Imaging, 1(3), 034003 (2014).
4. L. Boucheron, "Object-and spatial-level quantitative analysis of multispectral histopathology images for detection and characterization of cancer," PhD thesis, University of California, Santa Barbara, 2008.
5. J. Barker, A. Hoogi, A. Depeursinge, D. L. Rubin, "Automated classification of brain tumor type in whole-slide digital pathology images using local representative tiles," Medical Image Analysis, vol. 30, pp. 60-71, 2016.
6. S. Kothari, J. H. Phan, A. N. Young, and M. D. Wang, "Histological image feature mining reveals emergent diagnostic properties for renal cancer," Conf Proc IEEE Bioinform Biomed, BIBM, 2011.