HISTOPATHOLOGY PATCH CLASSIFICATION USING MACHINE LEARNING

Anindya Shaha, Prem Prasad

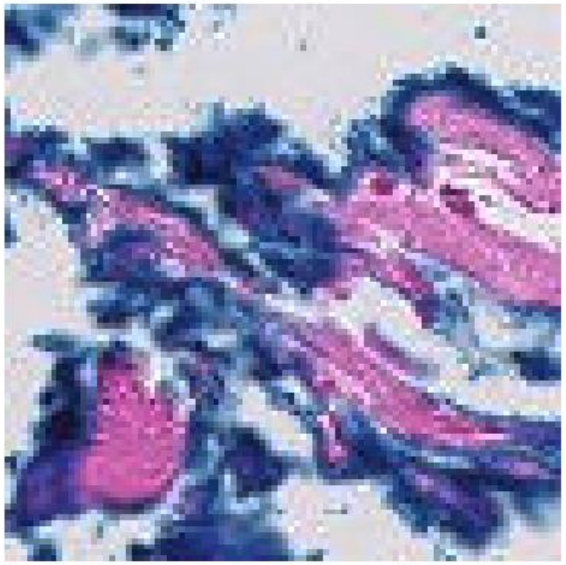
The *Classical* Approach, Computer-Aided Diagnosis

Universitat de Girona

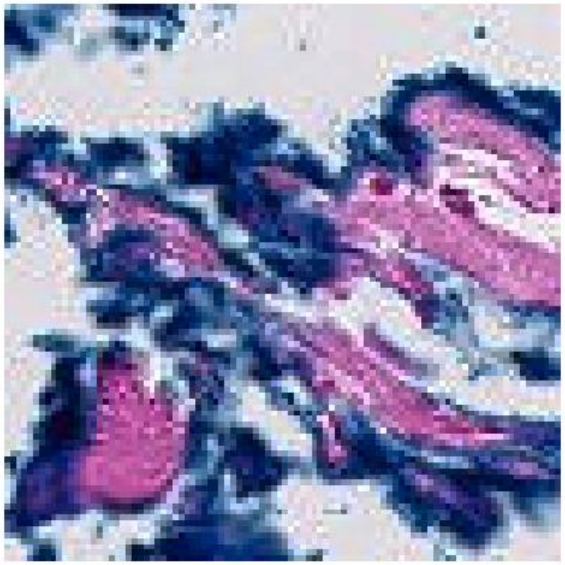
MAIA

DENOISING & CLAHE

Original RGB Image



CLAHE Enhanced Image



Denoised Image

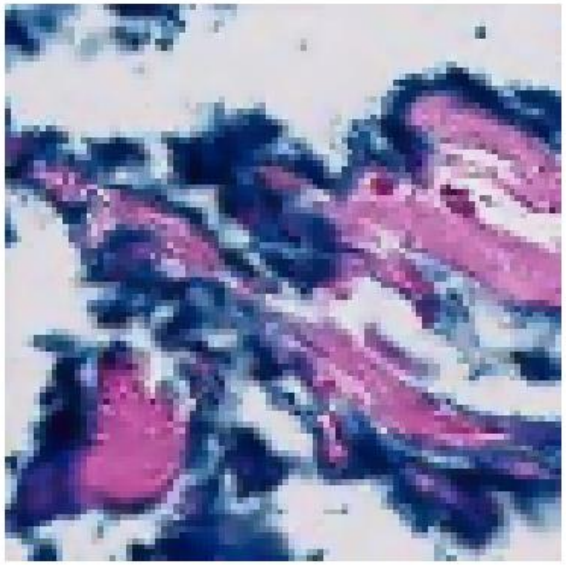


Figure 1: Original *RGB* image (left) after CLAHE enhancement at clip limit 3.0 (middle), which inadvertently boosts the color noise throughout the image. Hence the image is denoised using partial derivative equations based on the ROF model (right) with a weight of 10 to minimize prior distortions.

M.N. Gurcan et al. (2009), "Histopathological Image Analysis: A Review", IEEE Rev. Biomed. Engg.
L. Rudin, S. Osher, E. Fatemi (1992), "Nonlinear Total Variation Based Noise Removal Algorithms", Physica D.
K. Zuiderveld (1994), "Contrast Limited Adaptive Histogram Equalization", Graphic Gems IV.

COLOR SPACES

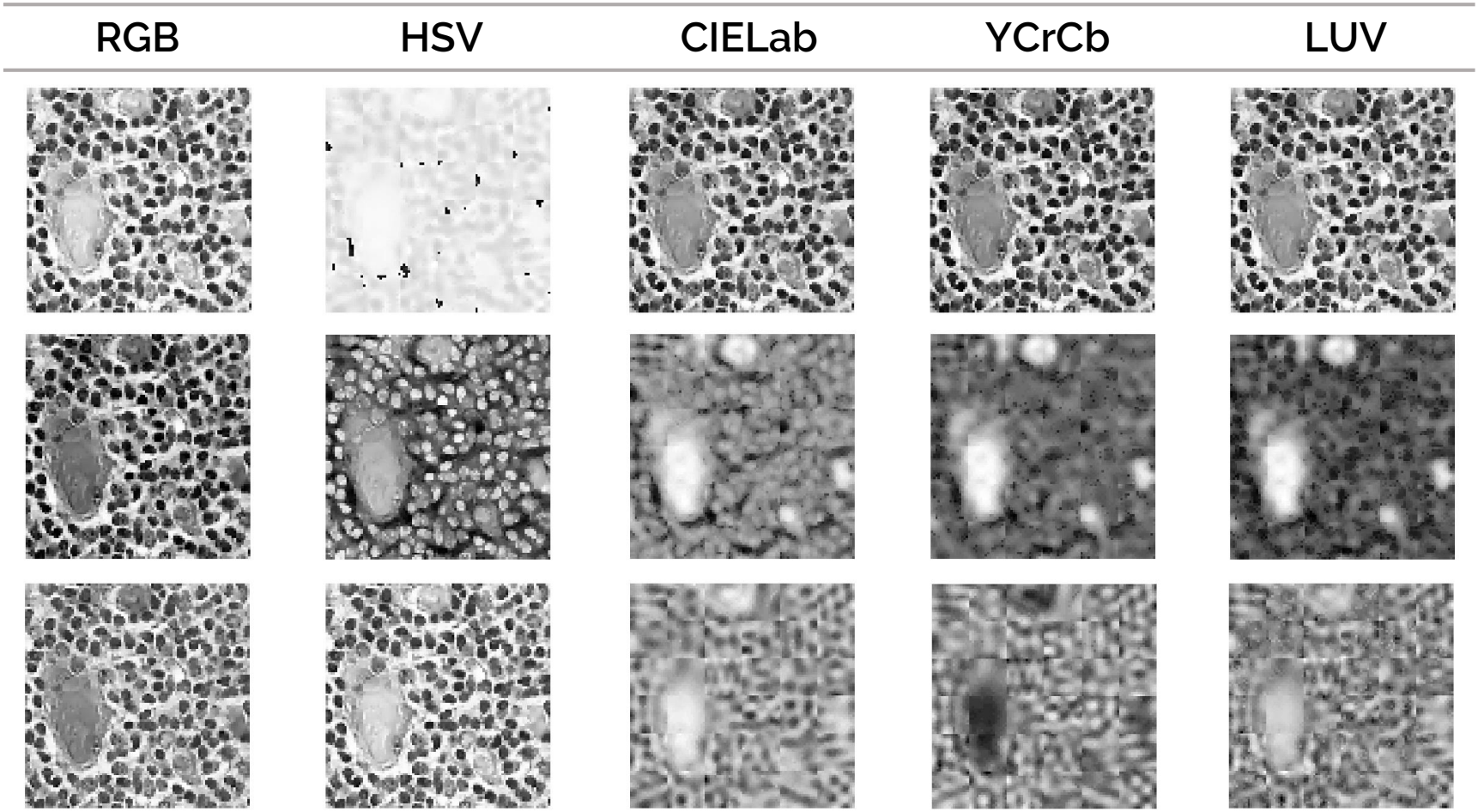


Figure 2: Color spaces used during computation of color moments and multi-color space texture features

M.N. Gurcan et al. (2009), "Histopathological Image Analysis: A Review", IEEE Rev. Biomed. Engg.
J. Yang et al. (2018), "Detection of Breast Cancer on Digital Histopathology Images", Elsevier IMU.

MULTI-COLOR SPACE FEATURES

- **Color Moments**
  Mean, Standard Deviation, Skew, Kurtosis
- **Gray-Level Co-occurrence Matrix (GLCM)**
  Contrast, Dissimilarity, Homogeneity, Correlation, Entropy, ASM
- **Local Binary Patterns (LBP)**
  Points, P = 8; Radius, R = 2; Bins = 10
- **Basic Texture Features**
  Entropy, Smoothness, Uniformity

4 Features × 3 Channels × 5 Color Spaces × 2 Color Constancy Modes = **120 Features**

6 Features × 3 Channels × 5 Color Spaces × 2 Color Constancy Modes = **180 Features**

10 Features × 3 Channels × 5 Color Spaces × 2 Color Constancy Modes = **300 Features**

3 Features × 3 Channels × 5 Color Spaces × 2 Color Constancy Modes = **90 Features**

O. Simon et al. (2018), "Multi-Radial LBP Features as a Tool for Rapid Glomerular Detection and Assessment in Whole Slide Histopathology Images", Nature Scientific Reports
M.N. Gurcan et al. (2009), "Histopathological Image Analysis: A Review", IEEE Rev. Biomed. Engg.
J. Yang et al. (2018), "Detection of Breast Cancer on Digital Histopathology Images", Elsevier IMU.

GABOR WAVELETS

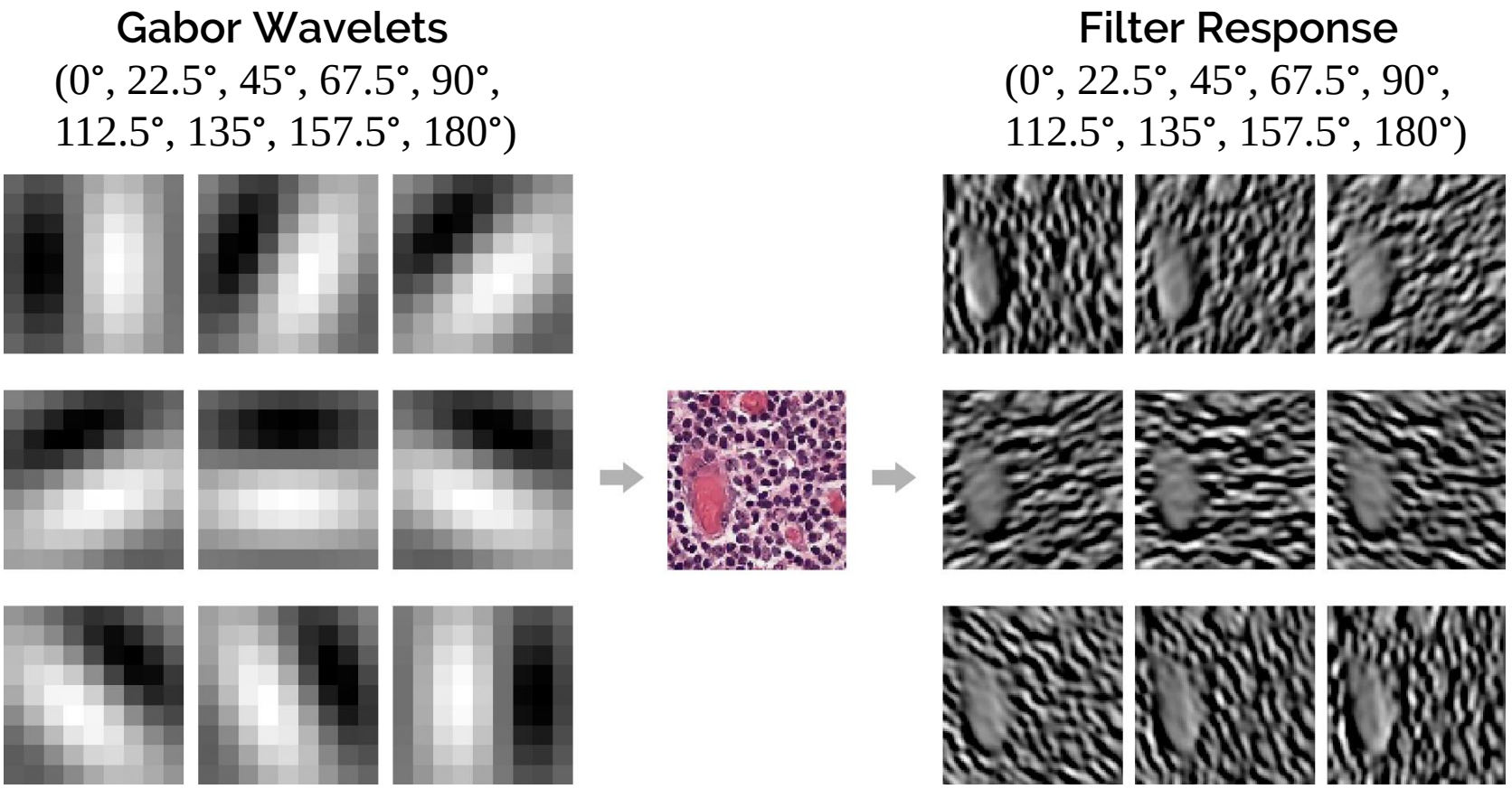


Figure 3: ***Gabor Filter*** response images and their respective local energy values for a *Gray-Level* image (converted from the original *RGB* image) constitute as 9 feature values to be used for texture classification.

M.N. Gurcan et al. (2009), "Histopathological Image Analysis: A Review", IEEE Rev. Biomed. Engg.
B. Ginley et al. (2017), "Unsupervised Labeling of Glomerular Boundaries using Gabor Filters and Statistical Testing in Renal Histology", IEEE JMI

HISTOGRAM OF GRADIENTS

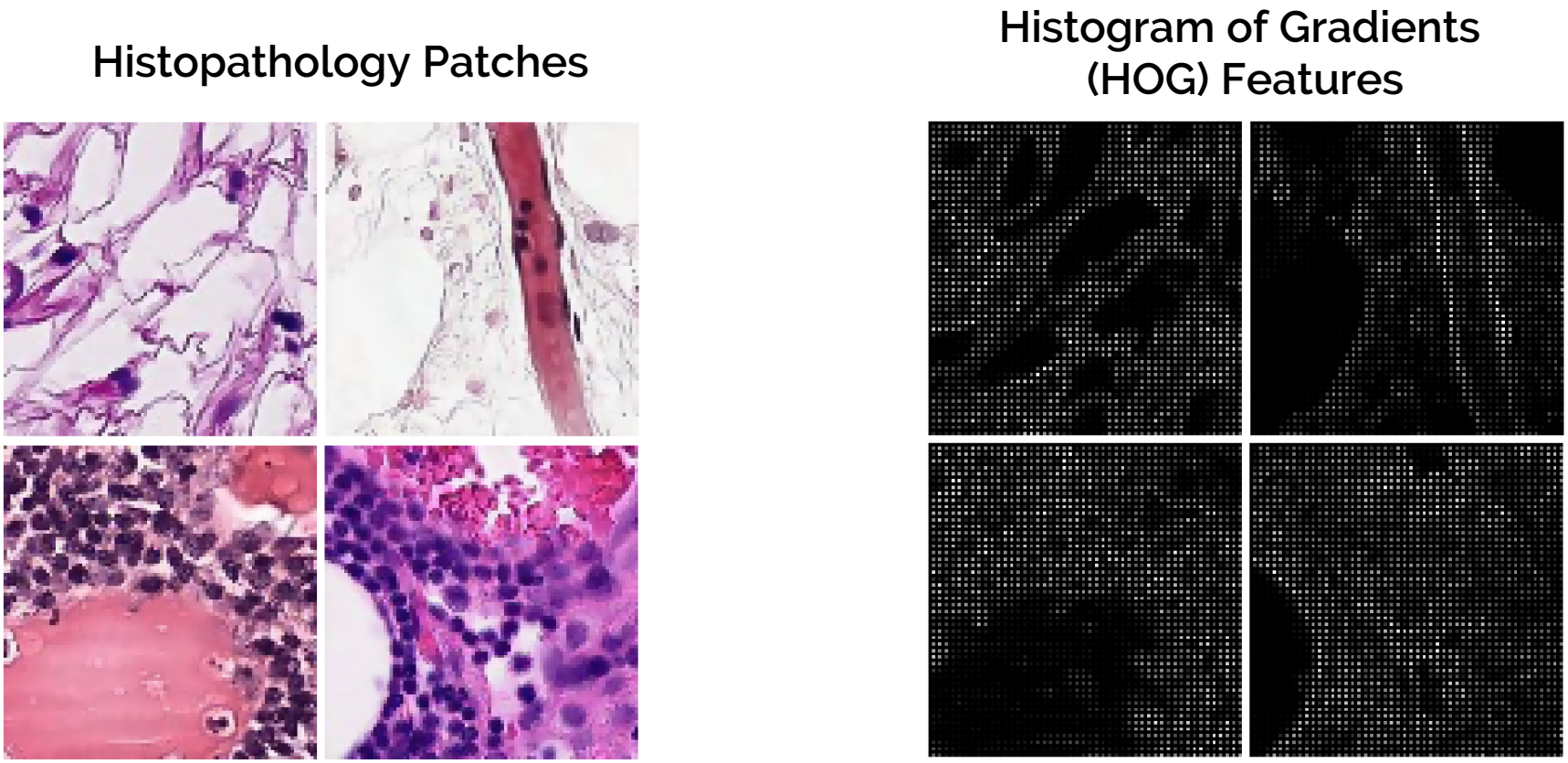


Figure 4: Multi-channel ***HOG Features*** (right) and their corresponding original *RGB* images (left) at 8 orientations, (2,2) pixels per cell and (1,1) cells per block.

M.N. Gurcan et al. (2009), "Histopathological Image Analysis: A Review", IEEE Rev. Biomed. Engg.
T.J. Alhindi (2018), "Comparing LBP, HOG and Deep Features for Classification of Histopathology Images", IJCNN

VORONOI & DELAUNAY FEATURES

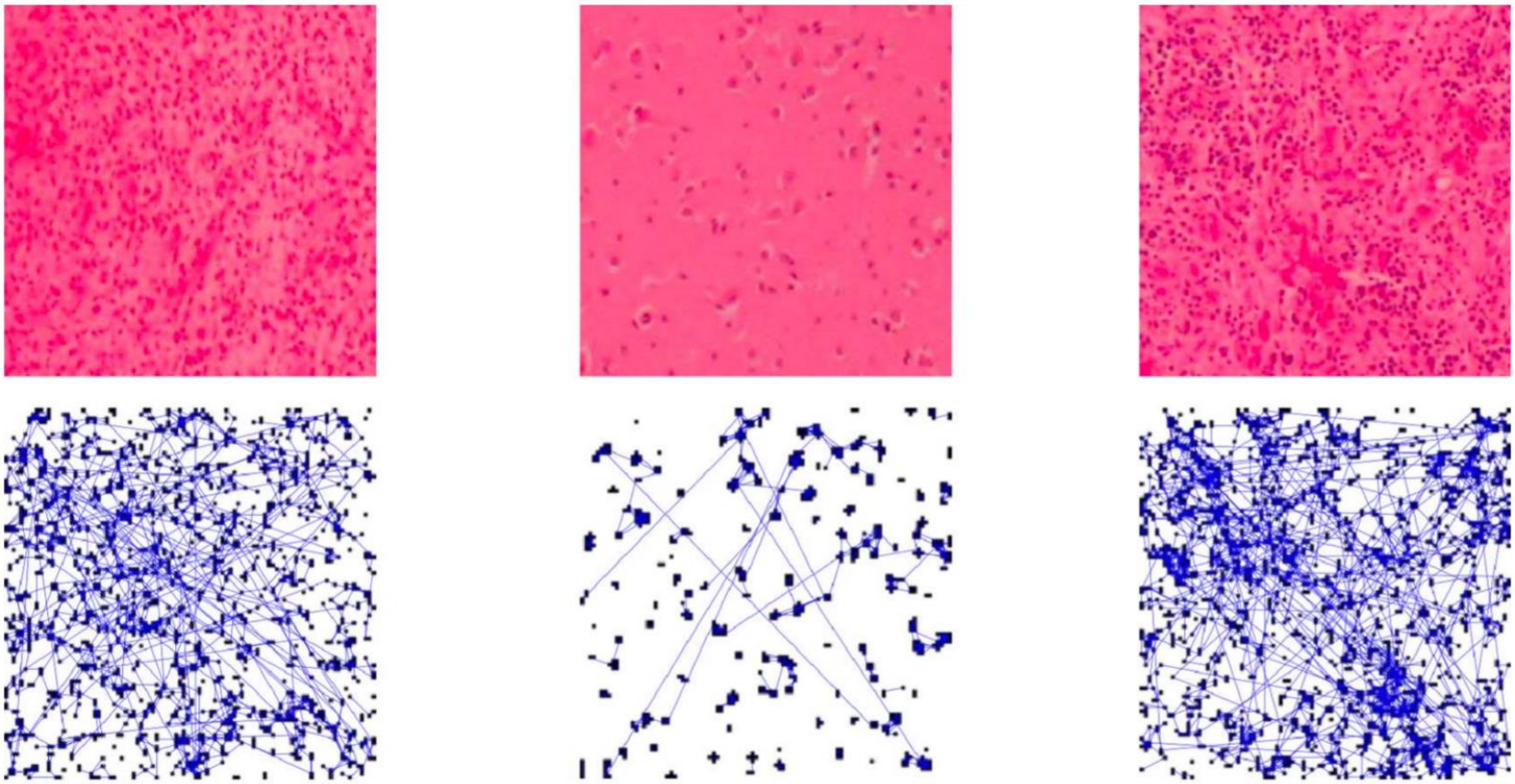


Figure 5: Histopathology images (top row) and cell graphs (bottom row) for cancerous (left), healthy (center) and inflamed (right) brain tissues [C. Gunduz et al.]

M.N. Gurcan et al. (2009), "Histopathological Image Analysis: A Review", IEEE Rev. Biomed. Engg.
C. Gunduz et al. (2004), "The Cell Graphs of Cancer," Bioinformatics.
S. Doyle (2008), "Automated Grading of Breast Cancer Histopathology using Spectral Clustering with Textural and Architectural Image Features", IEEE ISBI

VORONOI & DELAUNAY FEATURES

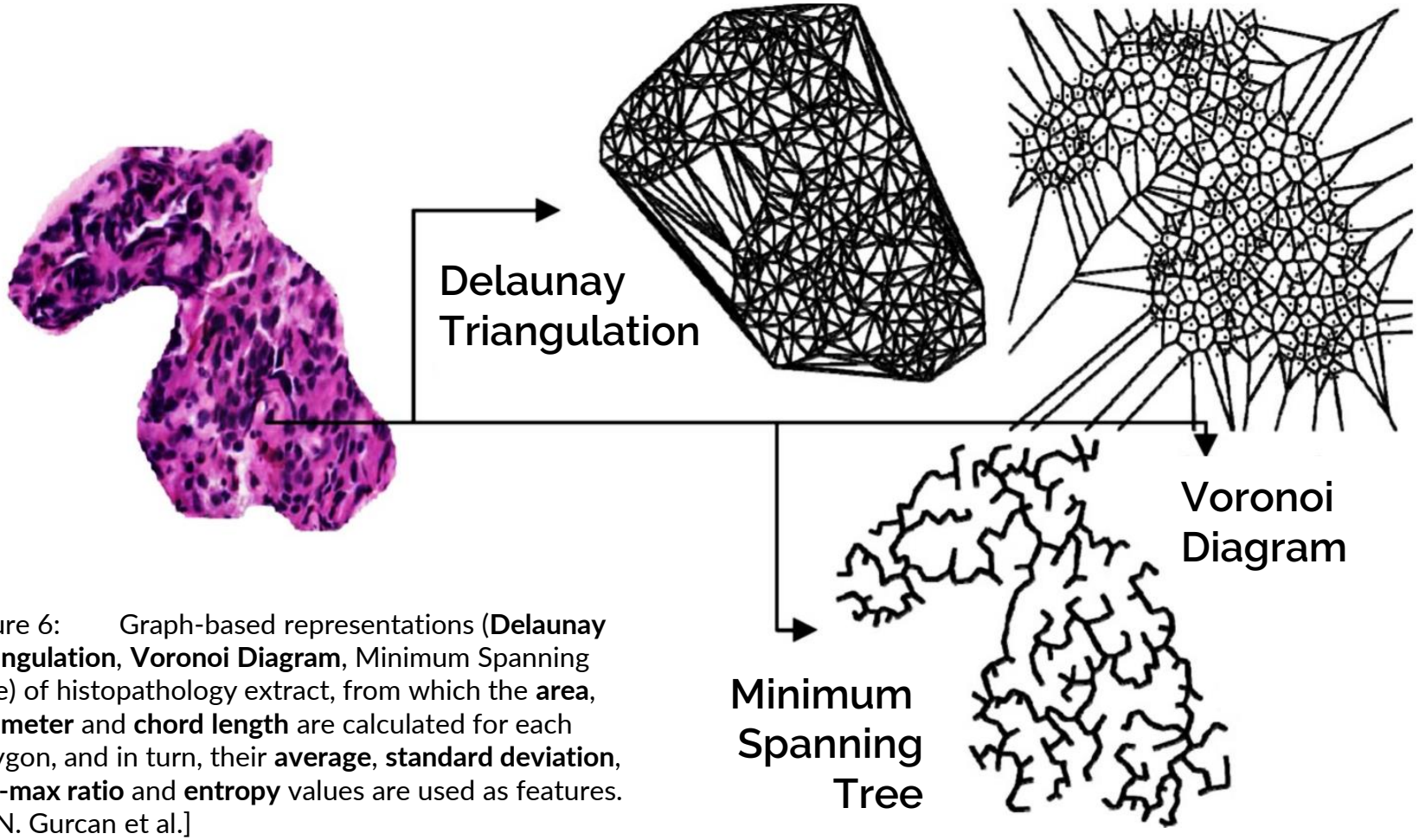


Figure 6: Graph-based representations (**Delaunay Triangulation**, **Voronoi Diagram**, Minimum Spanning Tree) of histopathology extract, from which the **area**, **perimeter** and **chord length** are calculated for each polygon, and in turn, their **average**, **standard deviation**, **min-max ratio** and **entropy** values are used as features. [M.N. Gurcan et al.]

M.N. Gurcan et al. (2009), "Histopathological Image Analysis: A Review", IEEE Rev. Biomed. Engg.
T.R. Jones (2005), "Voronoi-Based Segmentation of Cells on Image Manifolds", Springer CVBIA
H. Sharma (2015), "A Review of Graph-Based Methods for Image Analysis in Digital Histopathology", Diagnostic Pathology

FEATURE SELECTION

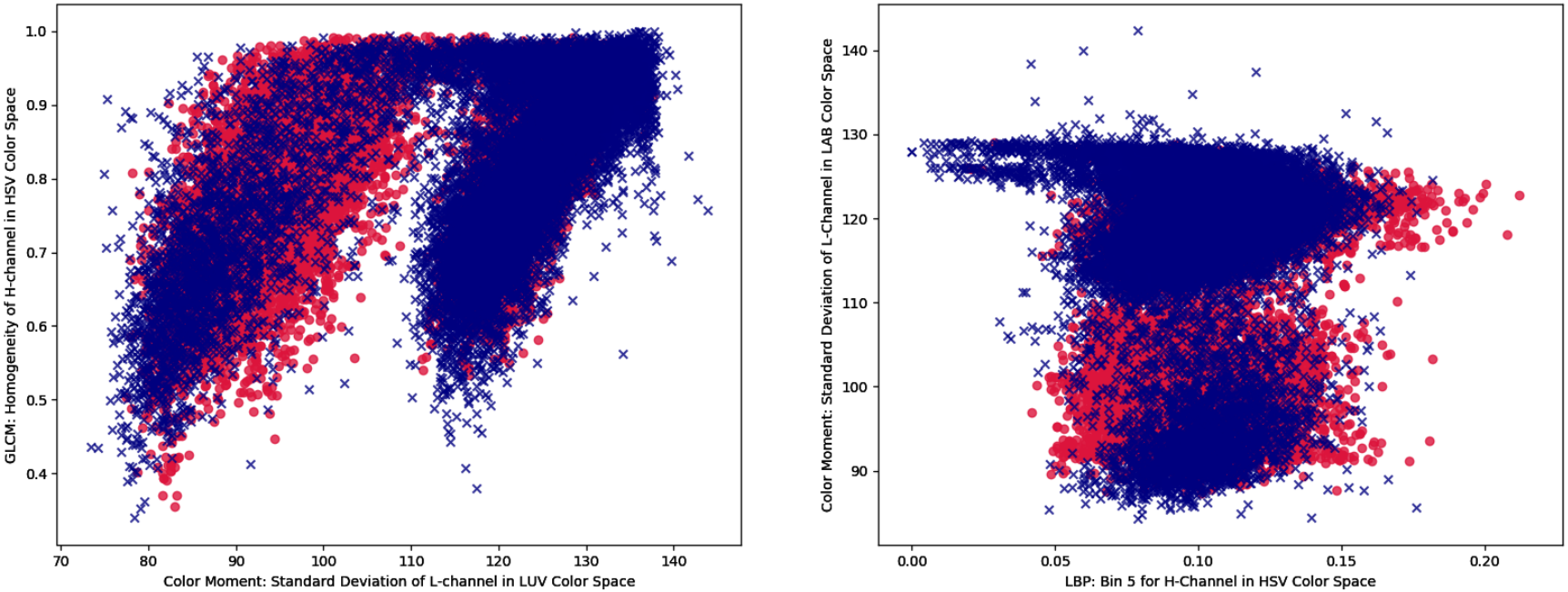


Figure 7: Scatter plot of all M0 samples (marked in red) and all B0 samples (marked in blue) in a 2-D feature space is used to assess the discriminative properties of any given pair of features. Highly redundant features are discarded in the final pipeline to streamline the training process and limit the potential risk of overfitting.

FINAL MODEL & OPTIMIZATION

PDE-ROF Denoising
~~Gray World CC~~
~~White-Balance (LAB)~~
~~Gray-Edge CC~~
~~Max-RGB CC~~
~~Occlusion Removal~~
CLAHE

Color Moments
GLCM Features
LBP Features
~~Color Markers~~
Basic Texture Features
~~HOG Features~~
Gabor Wavelets
Voronoi Features
Delaunay Features

SVM
~~Random Forests~~

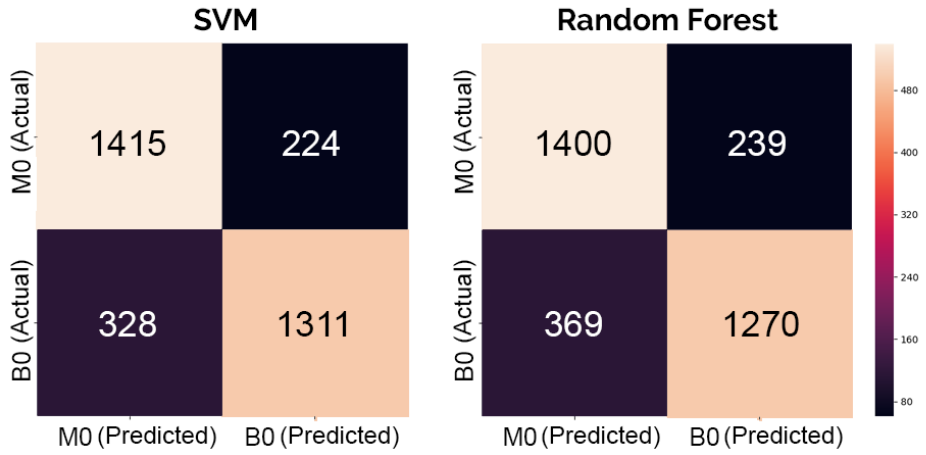
Higher Data Expression
Lower Color Noise
Higher Accuracy
(≈+5%)

Stronger Class Representations
Less Underfitting
Higher Accuracy
(≈+15%)

Less Overfitting
Higher Accuracy
(≈+2%)

EXPERIMENTAL RESULTS

Metric	SVM	Random Forest
Accuracy	0.8316	0.8145
Precision	0.8329	0.8165
Recall	0.8316	0.8145
F1 Score	0.8314	0.8142
Sensitivity	0.8000	0.7750
Specificity	0.8630	0.8540
AUC	0.9181	0.9079



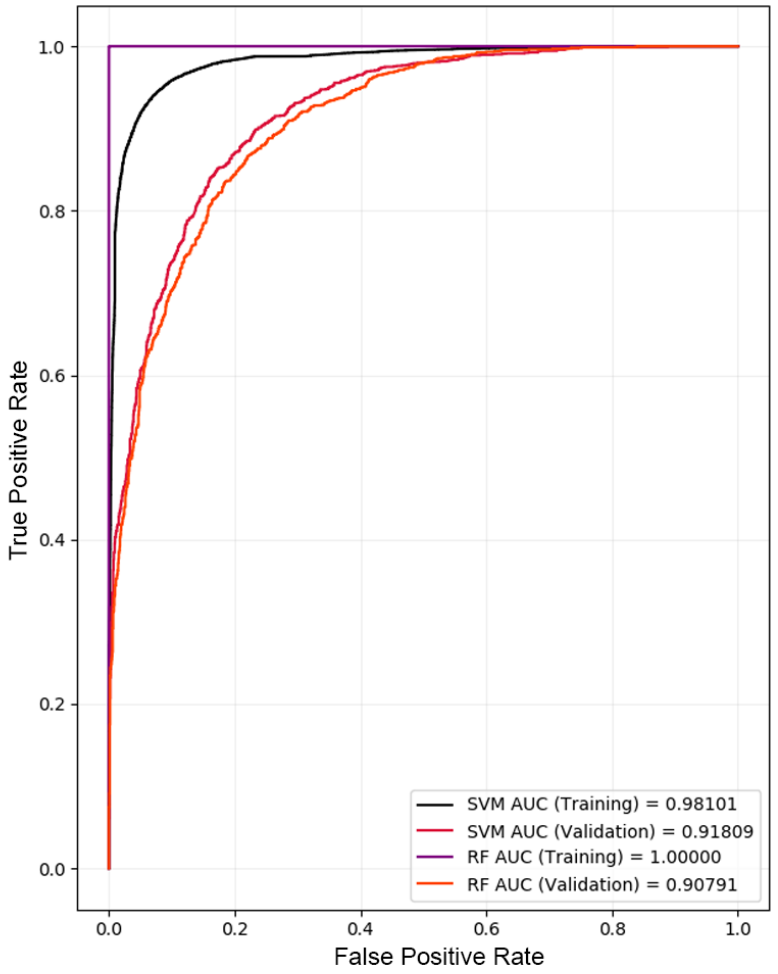


Figure 8: Evaluation metrics, confusion matrices and Receiver Operating Characteristic (ROC) curve for SVM and Random Forest on the validation set, after performing grid search to tune each classifier.

FUTURE WORK

Challenges

- Limiting the usage of **features whose pre-processing steps cause heavy degradation to the image quality** of small histopathology patches.
- Exploring **alternatives to OpenCV's color space transformations**, which cause severe pixelation at the scale of small histopathology patches.
- Incorporating **adaptive, noise invariant nuclei extraction steps** as the pre-requisite to cell graph-based feature computation.
- Advanced **classification strategies to optimize computation time and accuracy** (e.g. cascaded SVM with weak features in the starting layer, and increasingly more discriminating features through the subsequent layers, ensuring that every feature is not calculated for every sample).