

Evaluating the Effect of Image Representation on Digital Pathology Patch Classification

Justin T. Pontalba

Faculty of Biomedical Engineering

Yeats School of Graduate Studies, Ryerson University

Toronto, Canada

jportalb@ryerson.ca

Abstract—This paper evaluates the effect of image representation on support vector machine (SVM) classification for digital hematoxylin and eosin based images. From the red, green, and blue space, the images are transformed to the luminance, hematoxylin, and eosin image spaces. Several SVM models are separately trained on each image representation and testing patches are classified. The three models, performed the same achieving an accuracy score of 0.83. However, the hematoxylin-based model achieved the highest average cross validation accuracy of 0.750.

Index Terms—digital pathology, stain deconvolution, support vector machine

I. INTRODUCTION

Analyzing slide-mounted tissue sections under a microscope is a laboratory-based method for determining the prognoses and diagnoses of cancer. However, the transition of tissue-mounted slides to the digitization of whole-slide images (WSI) have augmented the practice of pathologist departments as they started to utilize digital imaging for simple tasks. To digitize the WSIs, tissue biopsies are excised from the tissue of interest, stained with contrast stains (hematoxylin and Eosin (H&E)) or immunohistochemical stains, and digitized with a whole-slide scanner. Upon, digitization, the digital WSI can be viewed in special software [1]. Other than visualization, the digitization of slides proposes the opportunity to use image processing, computer vision, machine learning - broadly *computational pathology*, to perform qualitative and quantitative tasks to improve pathology workflow [1]. Morphological analysis of histologic structures, automated grading of tumors, and automated selection of desired regions of interest are common applications proposed in literature. In particular, tasks such automated classification of tissue regions has emerged as computational power and the popularity of artificial intelligence increased as well.

This paper evaluates the performance of *support vector machine (SVM)* classifiers on various representations of digital H&E images. The trained models determine whether a digital pathology image patch is *benign* or *malignant*. This paper employs the feature extraction method by [2] and performs a comparative analysis between images of three different representations. [2] uses filter banks comprised of gaussian, laplacian of gaussian, anisotropic edge, and bar filters. In a similar way, this paper utilizes filter banks of multi-scale *Gabor* filters which have been previously used to extract features in similar

classification tasks [3] [4]. The first analysis was performed using *luminance* images taken from the $l\alpha\beta$ colour space. In addition, the second and third analyses were performed using *hematoxylin* and *eosin* stain images respectively, which were obtained after performing *stain deconvolution*. Since, [2] demonstrated success in classifying breast images using just the luminance channel, we propose to extend the experiment by developing classification models on just the hematoxylin and eosin images. Using these types of images could indicate if nuclei or stromal features alone are useful in differentiating benign or malignant tissue [5] [6].

This paper is organized as follows. *Materials and Methods* are described in Section II, *Results* are presented in Section III, and *Discussions and Conclusions* are given in the last section, Section IV.

A. Related Works

Classification of digital pathology images is a subject that has been researched, where many pipelines involve the extraction of important features for discerning the diagnoses of tissues. In previous works, [2] [5] [6], first-order statistic descriptors, mean, median, standard deviation, skewness, and kurtosis, were used in an attempt to capture the disorder of nuclear shape and texture. Cancerous nuclei exhibit properties of *apoptosis* and *necrosis* which describes programmed cell death and the degradation of tissue respectively. During these processes nuclei pleomorphism - variations of any nuclear size, shape or chromatin appearance variations, are observed [7]

Stromal characteristics differ between healthy and cancerous tissue. As a tumour develops, the formation of blood vessels add to the vascular density of the stroma. Furthermore, stromal characteristics can exhibit scanty or excessive connective tissue and numerous inflammatory cells such as lymphocytes, plasma cells and macrophages as the body attempts to destroy the tumour.

In this paper, both nuclei and stromal images are considered individually in discerning benign and cancerous tissue. Both structures are compared to luminance images which depict a combination of both tissue types.

II. MATERIALS AND METHODS

A. Data

The images used in this paper are of unknown tissue type, where 26 region of interest images (ROIs) are known to be malignant and 32 ROIs are known to be benign. The tissue is stained with hematoxylin and eosin (H&E) contrast staining, and the ROIs are of size 896 pixels by 768 pixels.

B. Experimental Design

Figure 1 depicts the experimental design used to address the effect of image representation on ROI classification. First, during the *pre-processing* the input images are transformed to the luminance, hematoxylin, and eosin representations. Next, features are extracted by first performing convolution on each image of the dataset with each filter of the filter bank. First order statistics are evaluated per output patch, which are then grouped to combat redundancy. The *Bag-Of-Words* technique is used for this task and can be seen as a feature *reduction* or *selection* method. Furthermore, *Support Vector Machine* classifiers are developed by using a *Grid Search* methodology for hyperparameter tuning, followed by K-fold cross validation. The classification results are then quantified using *Sensitivity*, *Specificity*, *f1-score*, and *average cross-validation accuracy*.

C. Pre-processing

1) *Colour Transformation*: To transform the image dataset into various representations, the red, green, and blue (RGB) images are transformed to the $l\alpha\beta$, hematoxylin, and eosin spaces.

The $l\alpha\beta$ colour space separates colour information in the α and β -channels, from the *luminance* or *lightness* in the l -channel. The l -channel represents much of the intensity information in the image, which is useful for feature extraction through performing convolution with each input image and each Gabor filter. Because the Gabor filtering method utilized in this paper requires a gray-scale input, the l -channel was chosen as oppose to the α, β -channels because the l -channel contains more information where nuclei and stromal structures are more discernible, as depicted in **Figure 2**. In order to convert an RGB image to the $l\alpha\beta$ space, an intermediate conversion from the RGB space to the XYZ space must be done first. The equation below depicts this conversion [8].

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_r & X_g & X_b \\ Y_r & Y_g & Y_b \\ Z_r & Z_g & Z_b \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

Once in the XYZ space, the image is converted to the $l\alpha\beta$ space using the following equations [8]:

$$L* = \begin{cases} 116(\frac{Y}{Y_n})^{\frac{1}{3}} - 16 & \text{if } \frac{Y}{Y_n} > 0.008856 \\ 903.3(\frac{Y}{Y_n}) & \text{if } \frac{Y}{Y_n} < 0.008856 \end{cases} \quad (2)$$

$$a* = 500 * (f(\frac{X}{X_n}) - f(\frac{Y}{Y_n})) \quad (3)$$

$$b* = 200 * (f(\frac{Y}{Y_n}) - f(\frac{Z}{Z_n})) \quad (4)$$

$$L* = \begin{cases} t^{\frac{1}{3}} & \text{if } t > 0.008856 \\ 7.787 * t + \frac{16}{116} & \text{if } t < 0.008856 \end{cases} \quad (5)$$

For this conversion, *Scikit-learn* libraries are used, which uses the 2-standard observer and illuminant D65. Reference values of $X/X_n = 95.047$, $Y/Y_n = 100$, and $Z/Z_n = 108.883$ are used in *Scikit-learn's* implementation [9].

2) *Stain Deconvolution*: The hematoxylin and eosin stain images are generated through a transformation called *stain deconvolution*. The stain deconvolution transformation converts an RGB image to an image comprised of the stains contained in the RGB image. For H&E stained images, the resulting image would be of hematoxylin, eosin, and an error channel that represents remaining colour if the transformation was not applied accurately. Stain deconvolution is made possible by the *Lambert-Beer's Law*. This law describes the relationship between the attenuation of light through a substance with respect to the properties of that substance [10]. In this paper, the *substances* are the combination of contrast stains and the tissue. The equation below describes Lambert-Beer's law [10]:

$$I_C = I_0 * e^{-A_{c_C}}, \quad (6)$$

where I_C is the source image, I_0 is the background bright-field, and A_{c_C} is the absorption coefficient. Of interest, are the concentration of stains contained in the source image. Therefore, for multiple stains, the concentrations can be models as:

$$OD = -\log(\frac{I_C}{I_0}) = \sum_{n=1}^3 A_{C_1} + A_{C_2} + A_{C_3}, \quad (7)$$

where A_{C_n} represents the RGB image multiplied by a stain matrix:

$$\begin{bmatrix} R_{11} & G_{12} & B_{13} \\ R_{21} & G_{22} & B_{23} \\ R_{31} & G_{32} & B_{33} \end{bmatrix} \quad (8)$$

In this paper, the stain matrix is estimated in the HSV space and then converted to the RGB space to perform stain deconvolution. The HSV space is used because this representation isolates colour information in the hue channel. As such, the hematoxylin and eosin hues are first estimated from a kernel density estimate (KDE) of the hue image weighted by the saturation image. The stain hues are defined as the minimum and maximum range where the majority of the KDE distribution are found. To find this range, a gaussian mixture model methodology is utilized. The modes corresponding to the hematoxylin and eosin stains are found. The min and max values of the range are defined by equations **9** and **10** respectively, where $mode_n$ represents the mode of each stain and $\sigma_H(h)$ is the standard deviation of the hue channel of the source image.

$$\text{hematoxylin} = mode_{hem} - \sigma_H(h) \quad (9)$$

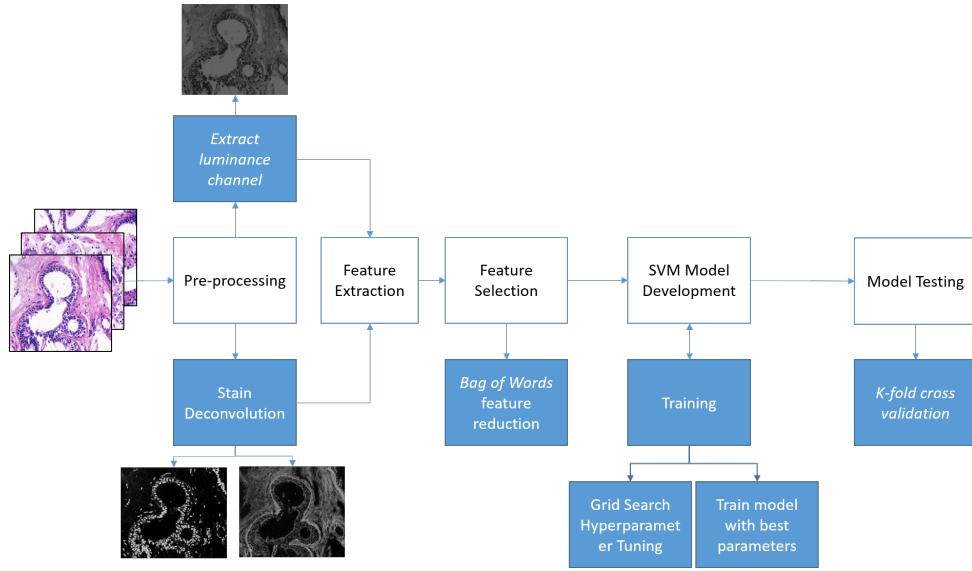


Fig. 1. Experimental design

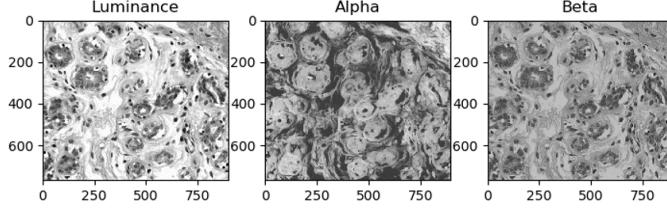


Fig. 2. $l\alpha\beta$ channel representations

$$eosin = mode_{eos} + \sigma_H(h) \quad (10)$$

With independent images of the stain components, texture features associated with nuclei and stroma can be extracted. Through feature extraction and further analysis, the usefulness of such features can be observed for patch classification. After the stain images are generated, features are extracted using Gabor filters.

D. Feature Extraction and Selection

1) *Gabor Filters*: The Gabor filter was first introduced as a one-dimensional filter which involved the multiplication of a cosine/sin wave with Gaussian windows. Later the 2D representation was proposed and is defined in the equations below:

$$g_e(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-0.5(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2})} \cos(2\pi\omega_{x_0}x + 2\pi\omega_{y_0}y) \quad (11)$$

$$g_o(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-0.5(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2})} \cos(2\pi\omega_{x_0}x + 2\pi\omega_{y_0}y) \quad (12)$$

where $(\omega_{x_0}, \omega_{y_0})$ defines the center frequency and (σ_x, σ_y) the spread of the Gaussian window [11]. In a review by

[12], Gabor filters are used for textural feature extraction in mammogram classification, and in [4] for the classification of prostate images. Gabor filters give the highest response at edges and points where intensity changes, and therefore analyzes texture by capturing edge information. In this paper, only the *real* component of the filter is used. **Figure 3** depicts the Gabor filters used in the proposed pipeline.

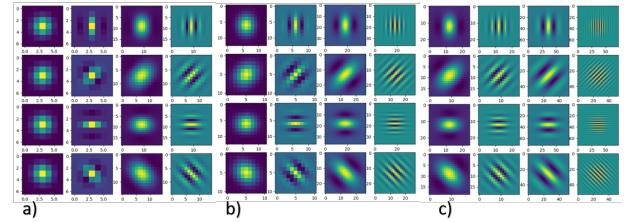


Fig. 3. Gabor Filter Banks: a) $\sigma_x, \sigma_y = 1, 3$ b) $\sigma_x, \sigma_y = 2, 6$ c) $\sigma_x, \sigma_y = 4, 12$

2) *Feature Selection*: This paper utilizes a Gabor filter bank of multiple levels in the spatial and frequency domain. This gives a total of forty-eight filters at multiple scales described by first order statistics such as mean, standard deviation, mode, skewness, and kurtosis. Therefore, in total 240 features are initially extracted. However, to avoid redundancy, similar features are grouped together using a *Bag-of-Words* methodology with 100 descriptors as the optimal dictionary size [2]. To implement the Bag-of-Words feature grouping method, a histogram of features is generated for each instance of the source data. A *dictionary* is formed by combining the histograms of each instance and grouping similar features again to reduce the dictionary to one-hundred features. New features are formed by attaining a vector that describes how many times a feature appears in an instance. The figure below depicts the format of the final feature vector along with the

class label.

TABLE I
FEATURE MATRIX AND CLASS LABEL

Image	f_1	f_2	f_n	Label
1	f_{11}	f_{12}	f_{1n}	0
2	f_{21}	f_{22}	f_{2n}	1

After features are formed, the features and class labels are split for training and testing, and classifiers are developed for each image representation.

E. Patch Classification and Cross-validation

For patch classification, the *Support Vector Machine* (SVM) is used. SVM is algorithm to find a hyperplane in the feature space that maximally separates two classes of data points. In this paper, SVM hyperparameters such as *kernel*, *gamma*, and a regularization parameter, C , are optimized using the *GridSearch* methodology.

1) *Kernel*: The *kernel* parameter in the SVM classifier is used to transform the input data to a higher feature space where it may be more optimally separable.

2) *gamma*: The *gamma* parameter of the *radial bases function* kernel is depicted below.

$$K(x_i, x_j) = e^{\gamma}(-x_i - x_j)^2 \quad (13)$$

γ defines the spread of the kernel and determines the decision region. Low values of gamma result in a broad decision boundary, whereas high gamma values results in complex decision boundaries.

3) *Regularization*: The regularization parameter, C , is a penalty parameter. When C is small, the classifier is does not penalize misclassified data greatly, while large values of C heavily penalizes misclassified data.

III. RESULTS

A. Colour Transformation and Stain Deconvolution

Figure 4 depicts the luminance, hematoxylin, and eosin images, where the first row, a), presents the original images, followed by b)-d), luminance, hematoxylin, and eosin respectively. Qualitatively, the stain deconvolution algorithm does a satisfactory job of separating stains. However, for eosin images, in regions where nuclei would be, exhibits "leftover" stain that was not captured for the hematoxylin image. Similarly, this phenomenon is observed for the nuclei images as well. Some stromal regions are included in the hematoxylin image due to error in the stain deconvolution algorithm. However, because in this paper, global features are utilized, errors such as these should not have a significant impact on feature extraction and subsequent analysis.

B. Feature Extraction

An example output from the convolution of an input image and Gabor filter is displayed in **Figure 5**. The source images are depicted in row a), b) outputs at scale (1,3), c) outputs at scale (2,6), and d) outputs at scale (4,12). The associated

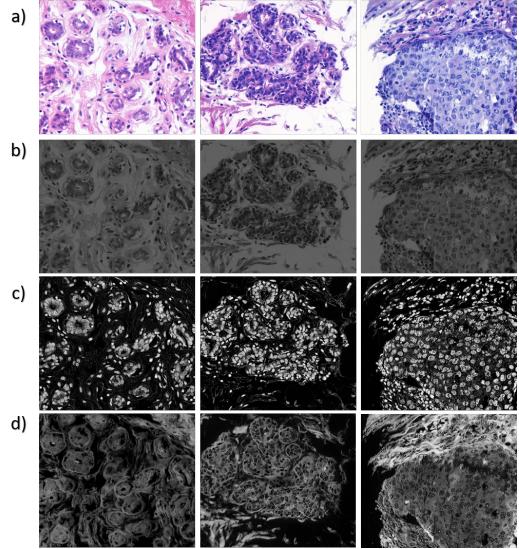


Fig. 4. Comparing the original rgb-image to luminance and stain images: a) RGB Images b) Luminance Images c) Hematoxylin Images d) Eosin Images

filters at these scales are shown at the end of each row. These spatial frequencies were implemented based on the work by [2]. Overall, at each scale the output images exhibit blurring. However, at (4,12) the filter seems to capture information related to vertical edges. Both the luminance image and the eosin image appear to have similar outputs. Perhaps this indicates the stromal features contribute to the overall response compared to the hematoxylin image. The lack of edge content in the hematoxylin output is likely a result of the lack of vertical edges in the original hematoxylin image. In contrast, the original luminance and eosin images contain vertical edges due to the stromal tissue structure. Characteristics such as these, make it possible to differentiate benign and malignant tissue.

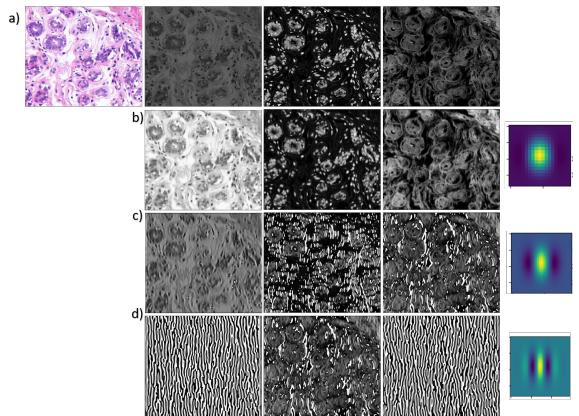


Fig. 5. Comparing image-filter outputs: a) Image Representation b) 1,3 spatial frequency responses b) 2,6 spatial frequency responses and c) 4,12 spatial frequency responses

C. Feature Selection

Once extracted, features are grouped using the Bag-of-Words methodology. **Figure 6** depicts the feature histograms

of each image representation. The luminance histogram depicts a multi-modal shape which could indicate that the chosen features differentiate the disease types well. However, for this representation a direct conclusion cannot be made because the tissue classes may be more separable in a higher dimensional space. The hematoxylin and eosin histograms depict multiple modes, but is dominated more by a single mode. The morphology of these histograms could be a result of a single tissue type. As such, much of the first order statistics could be redundant. For further analysis, the grouped features are used to train an SVM classifier.

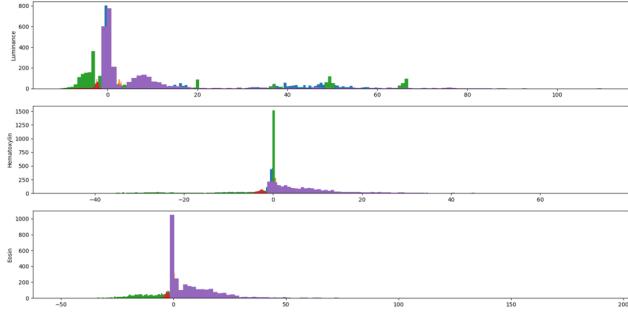


Fig. 6. Comparing the feature histograms of a) luminance, b) hematoxylin, and c) eosin image representations

D. Patch Classification and Cross-validation

1) *Model Hyperparameters and Training:* Using the Grid-Search methodology the hyperparameters were chosen for training and evaluating the SVM model. For GridSearch parameter tuning, preset values of the kernel, gamma, and C are iterated through, and the parameters that result in the highest validation metrics are selected to continue with analysis. The first evaluation was using an *rbf* kernel, γ values of $1e^{-2}, 1e^{-3}, 1e^{-4}$, and C values of 1, 10, 100, 1000. The second evaluation was using a *linear* kernel and C values of 1, 10, 100, 1000. **Table 2** below presents the optimal training parameters for the SVM for each image representation.

TABLE II
OPTIMAL TRAINING PARAMETERS

Image Type	Kernel	γ	C
luminance	rbf	0.0001	10
hematoxylin	rbf	0.0001	10
eosin	rbf	0.0001	10

2) *Testing and K-fold Cross Validation:* **Table 3, Table 4, Table 5, and Table 6** display the classification results of each image representation as well as the k-fold cross validation outcomes. Overall, each SVM model performed the same on the test data. However, based on the k-fold cross validation experiments, the hematoxylin-based SVM model demonstrated the best performance on average across the k-folds. The hematoxylin-based model achieved an average accuracy score of 0.750, compared to the luminance and eosin models which achieved 0.650 and 0.667 respectively. **Figure 7**

depicts the *receiver operating characteristics* curve. Evidently, both luminance and hematoxylin -based models achieve an AUC of 1, while the eosin-based model achieves an AUC of 0.89.

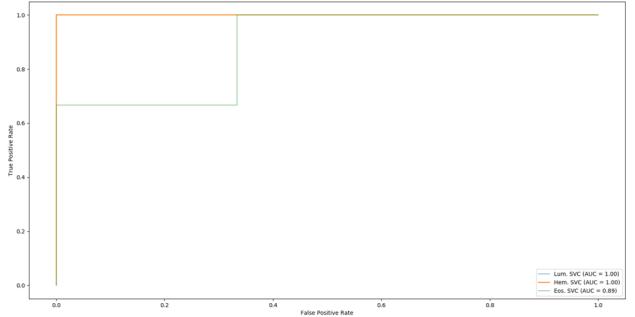


Fig. 7. Receiver Operating Characteristics for luminance, hematoxylin, and eosin -based classifiers

TABLE III
CLASSIFICATION PERFORMANCE-LUMINANCE IMAGES

Tissue	Sensitivity	Specificity	f1-score
Benign	1.000	0.670	0.800
Malignant	0.750	1.000	0.860

TABLE IV
CLASSIFICATION PERFORMANCE-HEMATOXYLIN IMAGES

Tissue	Sensitivity	Specificity	f1-score
Benign	1.000	0.670	0.800
Malignant	0.750	1.000	0.860

TABLE V
CLASSIFICATION PERFORMANCE-EOSIN IMAGES

Tissue	Sensitivity	Specificity	f1-score
Benign	1.000	0.670	0.800
Malignant	0.750	1.000	0.860

IV. DISCUSSIONS AND CONCLUSIONS

In this paper, H&E patch classification was evaluated with respect to image representation. Features from luminance, hematoxylin, and eosin images were extracted using Gabor filters, and individual SVM models were trained. The results demonstrated that while all representations resulted in the same performance, the hematoxylin model demonstrated the greatest average cross validation accuracy of 0.750. This could indicate that nuclei features are more effective when differentiating benign and malignant tissue types. However, because the methodology uses global features (evaluated over the entire patch), a more holistic representation is captured. Perhaps using object level features, extracted from individual nuclei, would be more practical for a task such as tissue classification. Based on the nature of tumors, healthy tissue and cells are mixed in with tumorous tissue and cells. For future works, nuclei based features and subsequent classification could offer more precision and localization.

TABLE VI
COMPARING K-FOLD CROSS VALIDATION PERFORMANCE

Image Representation	Average Cross Validation Accuracy
Luminance	0.650
Hematoxylin	0.750
Eosin	0.667

REFERENCES

- [1] Navid Farahani, Anil V. Parwani, and Liron Pantanowitz. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. 2015.
- [2] Mohammad Peikari, Judit Zubovits, Gina Clarke, and Anne L. Martel. Clustering analysis for semi-supervised learning improves classification performance of digital pathology. In *Proceedings of the 6th International Workshop on Machine Learning in Medical Imaging - Volume 9352*, page 263–270, Berlin, Heidelberg, 2015. Springer-Verlag.
- [3] Abdullah Nahid and Yinan Kong. Involvement of machine learning for breast cancer image classification: A survey. *Computational and Mathematical Methods in Medicine*, 2017:1–29, 12 2017.
- [4] Po-Whei Huang and Cheng-Hsiung Lee. Automatic classification for pathological prostate images based on fractal analysis. *IEEE transactions on medical imaging*, 28:1037–50, 02 2009.
- [5] Cheng Lu, David Romo, Xiangxue Wang, Andrew Janowczyk, Shridar Ganeshan, Hannah Gilmore, David Rimm, and Anant Madabhushi. Nuclear shape and orientation features from he images predict survival in early-stage estrogen receptor-positive breast cancers. *Laboratory Investigation*, 98:1, 06 2018.
- [6] Patrick Leo, George Lee, Natalie Shih, Robin Elliott, Michael Feldman, and Anant Madabhushi. Evaluating stability of histomorphometric features across scanner and staining variations: Prostate cancer diagnosis from whole slide images. *Journal of Medical Imaging*, 3:047502, 10 2016.
- [7] Harsh Mohan. *Textbook of Pathology*. Jayppe Brothers MEdical Publishers, The address, 6 edition, 2010.
- [8] A. Ford. Colour space conversions. 1998.
- [9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [10] Arnout C.C. Ruifrok and Dennis A. Johnston. Quantification of histochemical staining by color deconvolution. *Analytical and quantitative cytology and histology*, 23 4:291–9, 2001.
- [11] Konstantinos Derpanis. Gabor filters. 4 2007.
- [12] Abdullah-Al Nahid. Involvement of machine learning for breast cancer image classification: A survey. *Hindawi*, 2017:29, 12 2017.