

Automated Classification for Breast Cancer Histopathology Images: Is Stain Normalization Important?

Vibha Gupta¹(✉), Apurva Singh², Kartikeya Sharma³, and Arnav Bhavsar¹

¹ Indian Institute of Technology Mandi, Mandi, India
vibha_gupta@students.iitmandi.ac.in, arnav@iitmandi.ac.in

² Manipal Institute of Technology, Manipal, India
apurva1095@gmail.com

³ National Institute of Technology Hamirpur, Hamirpur, India
09kartikeya@gmail.com

Abstract. Breast cancer is one of the most commonly diagnosed cancer in women worldwide. A popular diagnostic method involves histopathological microscopy imaging, which can be augmented by automated image analysis. In histopathology image analysis, stain normalization is an important procedure of color transfer between a source (reference) and the test image, that helps in addressing an important concern of stain color variation. In this work, we hypothesize that if color-texture information is well captured with suitable features using data containing sufficient color variation, it may obviate the need for stain normalization.

Considering that such an image analysis study is relatively less explored, some questions are yet unresolved such as (a) How can texture and color information be effectively extracted and used for classification so as to reduce the burden on the uniform staining or stain normalization. (b) Are there good feature-classifier combinations which work consistently across all magnifications? (c) Can there be an automated way to select reference image for stain normalization?

In this work, we attempt to address such questions. In the process, we compare the independent texture and color channel information with that of some more sophisticated features which consider jointly color-texture information. We have extracted above features using images with and without stain normalization to validate the above hypothesis. Moreover, we also compare different types of contemporary classification in conjunction with the above features. Based on the results of our exhaustive experimentation we provide some useful indications.

Keywords: Histopathology image analysis · Stain normalization · Color-texture features · SVM · Random forest

1 Introduction

Breast cancer is the most common cancer in women worldwide and is the second leading cause of cancer deaths in women, after lung cancer [1]. One of the

common approaches used for cancer detection is histopathology; it is defined as the microscopic examination of the histological sections of a biopsy sample by a pathologist in order to study the effects of a particular disease. Computer-aided approaches with image analysis and machine learning can be included in digital pathology to achieve quick and reproducible results. Computer aided diagnostic (CAD) models are also important as they assist pathologists in locating and identifying abnormalities in the breast tissue images.

A concern in histopathology image based assessment is the variation in color is obtained due to a number of factors like chemical reactivity from different manufacturers, differences in color responses of slide scanners or due to the light transmission being a function of slide thickness. However, this variability only partially limits the interpretation of images by pathologists. But leads to a large variation in the efficiency of automated image analysis algorithms. This problem can be partially reduced by incorporating the use of standardized staining protocols and automated staining machines. Stain normalization algorithms [2–4] have also been recently introduced to address stain variation, with the aim of matching stain colors of whole slide image with a given template.

One of the naive options to deal with color constancy is to convert color image to grayscale. In [5] author extracted features from a grayscale version of a query image. Conversion of an image into grayscale gives us the average of concentrations of the tissue components and does not tell us the relative amounts of each of them. Further, this also does not make effective use of the color information which is present. Recent research in histopathology has confirmed that color information is quite significant in quantitative analysis.

The outstanding ability of a pathologist to identify stain components is not only due to the utilization of color information but also because of incorporating the spatial dependency of tissue structures. The use of standardized staining protocols and automated staining machines may improve staining quality by yielding a more accurate and consistent staining. However, eliminating all the underlying sources of variation is infeasible.

Indeed, methods that investigate the importance of staining in conjunction with classification framework have also been developed. In [6], authors investigate the importance of stain normalization in tissue classification utilizing convolution network. In the same way, the authors in [7] perform the classification of prostate tumor regions via stain normalization and adaptive cell density estimation. On the other hand, there also exists some work which considers the utilization of color information without the use of stain normalization. The works in [8] and [9] propose the use color information in addition to texture. Milagro et al. [8] propose the combinations of traditional texture features and color spaces. Furthermore, they have also considered different classifiers such as Adaboost learning, bagging trees, random forest, Fisher discriminant and SVM. In [9], authors utilized color and differential invariants to assign class posterior probabilities to pixels and then perform a probabilistic classification.

The above approaches suggest different views about the consideration of stain normalization for classification. Inspired by this, our primary contribution in the

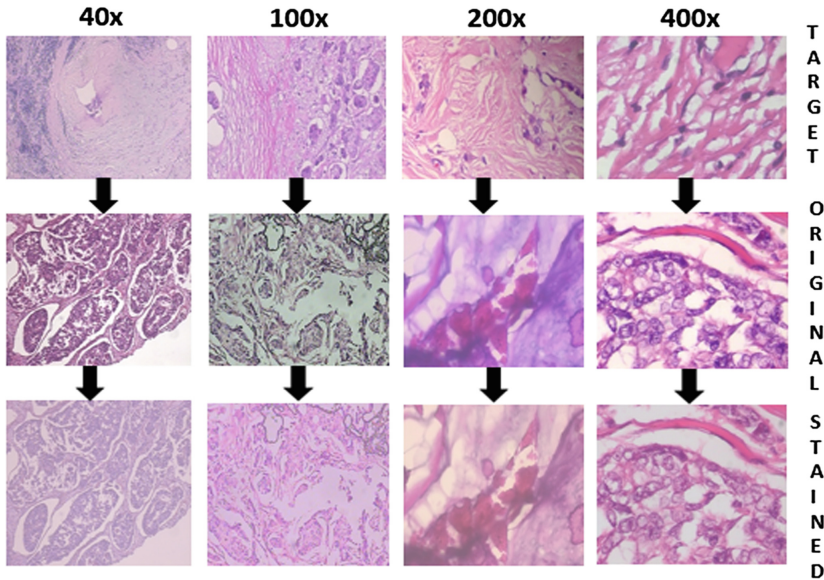


Fig. 1. Reference images chosen for normalization (top), sample of original images (middle), and obtained stain normalized images using the target images (bottom)

present work, is to explicitly provide indications towards addressing the question about how important is stain normalization for automatic classification, and whether there may exist useful features which inherently capture the color-texture variability without performing stain normalization. In other words, we attempt to justify the role of color-texture information in automated classification framework without performing stain normalization. We believe that such a study is important from a systems perspective, as it may help reduce the stain normalization overhead for automated histopathology classification systems. We note that such an indication of mitigating the need for stain normalization assumes that the training data contains images with different stain color/intensity, that helps in capturing the color-texture variability. We validate this hypothesis on a reasonably large dataset containing such images.

In this primary context, our work also has the following salient contributions about methodology and evaluation: (1) Exploring joint color-texture features and various contemporary classifiers, which makes the proposed work also serve as an extended comparative review. (2) Suggesting an automatic approach to select the reference (target) image for stain normalization where it is not available. Figure 1 shows examples of stained images of each magnification generated using target images. (3) Demonstrating an improved performance of the proposed method with joint color-texture features with respect to the state-of-the-art.

2 Proposed Approach

We now discuss the overall framework of proposed approach including: dataset description, stain normalization, feature descriptors, and contemporary classifiers. Due to lack of space, we briefly describe the features and classifiers with suitable references. Figure 3 depicts our overall framework.

2.1 Dataset Description

In this work, we use BreakHis dataset [10] that contains fairly large amount of histopathology images (7909). A detailed description is provided in Table 3.

In [10–12], authors developed the framework to classify breast cancer histopathological images utilizing the BreakHis dataset. In [10], a series of experiments utilizing six different state-of-art texture descriptors such as Local Binary Pattern (LBP), Completed Local Binary Pattern (CLBP), Threshold Adjancey Statistics (PFTAS), Grey-Level Co-occurrence Matrix (GLM), Local Phase Quantization (LPQ), Oriented FAST and rotated BRIEF (ORB), and four different classifiers were evaluated and showed the accuracy at patient level. In [11], Alexnet [13] was used for extracting features and classification. Bayramoglu et al. [12] proposed a magnification independent model utilizing deep learning to classify the benign and malignant cases. The magnification independent system is trained with images of different magnifications, and thus can handle the scale diversity in microscopic images.

2.2 Stain Normalization Procedure

Various methods [2–4] have been developed to automate the standardization process of histopathological images to reduce the effect of variation that exists in staining protocol. In [2] authors utilized chromatic and density distributions for each individual stain class in the hue-saturation-density (HSD) color model i.e. the spatial dependency of tissue structures was incorporated along with color information. In experiments, the target template was chosen based on opinion of two pathologists, who studied a large number of slides from two different laboratory. The high contrast between hematoxylin and eosin staining (H & E) and visibility of the nuclear texture were taken into consideration while choosing template image. In [3], the authors use the linear transform in a perceptual color space for matching the color distribution of an image to that of a target image. In [4], authors also utilized pathologist-preferred target image to generate structure-preserving color normalization. These stain normalization methods require prior knowledge of reference stain vectors for every dye present in the histopathological images.

Due to the unavailability of a target template in the public dataset used in this work, we automatically select the target image from the dataset. Our approach considers that the target stain chromatic information should be considered that which occurs most commonly, so that a large number of images need not be color-transformed. Thus, we suggest the following process: (1) First, all the images in dataset are converted from RGB color-space to HSV color-space.

(2) We choose H and S component for further analysis as Hue and Saturation essentially relate to the chromatic information. (3) A K-means clustering algorithm is applied to form the desired number of clusters. The number of clusters chosen is the double of the number of different Hue that are found on manual examination of different images. This is to ensure that we have a good enough separation of images of different Hue. (4) The number of stain hues in the dataset found after manual examination are five. So we chose to create 10 clusters. In the pictorial representation of Fig. 2, we show less number of clusters for better clarity. (5) We choose the cluster that has highest number of images. In this cluster we find out the mean H and S value of image which is the closest to the centroid of the cluster using Euclidean distance measure. The corresponding image is used as the target image after conversion to RGB color-space.

The proposed procedure is applied separately to images of different magnifications. At the end, we have one target image for each magnification group of images. After selecting target image, we use stain normalization method proposed by [3] to normalize the dataset. Figure 2 illustrates the overall procedure for selecting target image. Figure 2 is just a way to depict the procedure and is not a real picture of our plots.

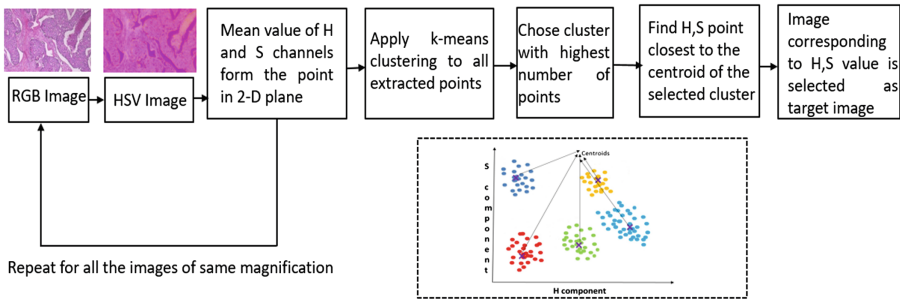


Fig. 2. Selection of target image.

2.3 Feature Descriptors and Contemporary Classifiers

Various texture features that consider the mutual dependency of color channel as well as the features that don't utilize the color information are extracted in order to support believe that we have made. Due to space constraint we are not providing the detailing of features. Gray level co-occurrence matrix (GLCM), Completed local binary pattern (CLBP) [14], and Local phase quantization (LPQ) [15] are used to extract plain texture. For capturing joint color-texture variability, features such as Opponent Colour Local Binary pattern (OCLBP) [16], Gabor features on Gaussian color model [17], Multilayer Coordinate Clusters representation (MCCR) [18], and Parameter-free Threshold Adjacency Statistics (PFTAS) [19] are utilized. Note that the choice of features and classifiers is simply based on the popularity of the traditional texture features, and considering some recently reported color-texture features.

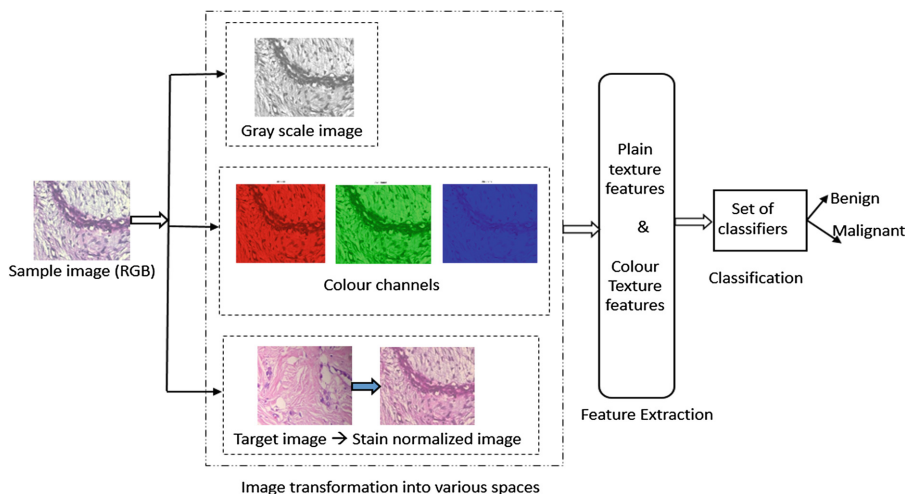


Fig. 3. Overall process of image classification.

Various contemporary classifiers [20] such as Support Vector Machine (SVM), Nearest Neighbors (NN) and Random forest (RF) are experimented with above listed texture features. For the proposed study, we utilize following variants of mentioned classifiers: (1) Linear SVM: Linear kernel, (2) Gaussian SVM: RBF kernel and kernel scale set to $2\sqrt{P}$, (3) K-NN: 100 neighbors and euclidean metric for distance measure, (4) RF: number of trees (30), maximum number of splits (20). Thus, the study also serves as a comparative review of performance for the above mentioned features and classifiers for the considered problem.

Figure 3 illustrates the overall structure utilized for the proposed study. As indicated earlier, the primary intention of this study is to consider the effect of stain normalization for automated breast cancer histopathology image classification. Thus, we test the classification performance with gray-scale image, stain normalized images, and original non-normalized images.

3 Results and Discussion

3.1 Training-Testing Protocol and Evaluation Metric

In our experiments, we have randomly chosen 58 patients (70%) for training and remaining 25 for testing (30%). This also enables fair comparison with a state-of-the-art approach [10–12]. We train the above listed contemporary classifiers using images for the chosen 58 patients, and have also used five trials of random selection of training-testing data. These trained models are tested using images of the remaining images 25 patients.

To compare the results with existing approach [10–12], we use patient recognition rate (PRR) as evaluation metric. However, some other evaluation metrics

such as recall, precision, area under the ROC curve (AUC) can also be utilized. The definition of patient recognition rate (PRR), which uses the patient score (PS), is given as follows:

$$PRR = \frac{\sum_{i=1}^N PS_i}{N} \quad PS = \frac{N_{rec}}{N_P} \tag{1}$$

where N is the total number of patients (available for testing), and N_{rec} and N_P are the correctly classified and total cancer image of patient P respectively.

3.2 Performance Evaluation and Comparison

Tables 1, 2 illustrate the performance of contemporary classifiers for 40X, 100X, 200X and 400X magnification respectively. For each magnification, seven texture features out of which three are plain texture that directly extracts the feature from gray version of image, and other four are color texture that utilizes color-channel information are extracted.

From the tables following can be observed (for most or all cases):

- (1) The performance when using color information (with or without stain normalization) is better performance than using gray level information. This highlights the importance of color in classification.

Table 1. Evaluation of color channel information along with contemporary classifiers for 40x and 100x. Best performance at each magnification is highlighted.

Image magnification			40x				100x			
Classifier and their variations			Linear SVM	Gaussian SVM	Gaussian KNN	Random forest	Linear SVM	Gaussian SVM	Gaussian KNN	Random forest
Plain texture features	GLCM	Gray scale	74.60	74.72	68.73	68.10	74.49	74.58	70.71	65.29
		Stain normalization	74.91	74.81	72.96	71.55	75.77	75.31	74.77	73.49
		Channel information	81.65	76.79	74.45	78.30	82.16	76.76	75.82	77.89
	CLBP	Gray scale	70.13	70.68	72.38	72.91	70.13	70.68	72.38	72.91
		Stain normalization	70.4	70.4	70.4	70.67	70.4	70.4	70.4	70.42
		Channel information	75.09	74.65	72.30	71.39	73.50	75.46	74.51	73.09
	LBQ	Gray scale	76.37	77.44	74.69	75.04	73.77	76.72	73.60	72.66
		Stain normalization	74.52	75.21	73.86	74.40	72.73	76.79	73.67	74.50
		Channel information	74.97	76.77	73.08	75.42	72.69	76.79	73.67	73.38
Colour texture feature	PFTAS	Gray scale	74.77	74.90	69.19	68.99	73.75	73.45	71.90	68.19
		Stain normalization	74.77	75.59	74.68	75.00	75.26	75.46	74.28	74.53
		Channel information	70.11	72.51	75.63	75.52	80.92	78.07	81.54	84.75
	Opponent LBP	Gray scale	73.86	70.22	71.62	72.36	75.42	74.58	74.62	70.55
		Stain normalization	76.09	75.25	75.54	76.05	68.66	75.88	74.27	71.92
		Channel information	86.88	86.20	85.41	85.18	86.17	86.40	85.09	87.41
	Gabor on gaussian color model	Gray scale	74.76	74.93	72.95	73.90	71.74	73.58	72.27	67.82
		Stain normalization	78.58	74.50	74.85	75.01	79.12	74.94	74.90	75.37
		Channel information	85.62	84.86	80.63	84.93	88.41	85.61	84.65	86.11
	M CCR (8)	Gray scale	73.83	70.4	67.80	68.06	69.74	70.4	68.76	65.63
		Stain normalization	72.35	74.40	72.78	71.18	71.83	73.25	70.87	75.67
		Channel information	78.93	78.50	82.63	84.42	80.40	78.96	80.30	87.15

Table 2. Evaluation of color channel information along with contemporary classifiers for 200x and 400x. Best performance at each magnification is highlighted.

Image magnification			200x				400x				
Classifier and their variations			Linear SVM	Gaussian SVM	Gaussian KNN	Random Forest	Linear SVM	Gaussian SVM	Gaussian KNN	Random Forest	
Plain texture features	GLCM	Gray scale	70.37	71.10	69.73	63.80	70.06	69.57	70.46	66.03	
		Stain normalization	71.98	72.67	70.18	68.88	69.28	70.08	72.33	70.51	
		Channel information	84.19	81.24	76.98	82.61	78.45	77.21	74.30	76.14	
	CLBP	Gray scale	62.85	68.93	66.65	63.31	65.69	69.46	68.56	63.68	
		Stain normalization	70.4	70.4	70.4	70.67	70.42	70.29	62.08	69.89	
	LBQ	Channel information	73.48	71.19	73.52	70.29	72.96	72.54	72.85	70.26	
		Gray scale	72.86	75.34	74.27	72.08	72.93	74.99	73.73	70.55	
		Stain normalization	73.76	75.50	74.48	71.94	75.97	75.48	74.28	72.52	
	Colour texture feature	PFTAS	Channel information	73.94	75.24	74.04	73.16	75.59	75.33	73.92	69.79
			Gray scale	71.71	70.56	70.91	65.13	72.22	71.92	70.10	64.00
			Stain normalization	74.17	73.08	73.89	72.73	75.54	74.79	75.19	72.30
		Opponent LBP	Channel information	73.52	75.46	74.51	72.80	77.14	74.31	77.18	80.97
Gray scale			74.04	70.20	72.19	70.04	74.60	73.30	73.38	69.05	
Stain normalization			73.55	72.42	73.11	72.49	76.59	72.93	72.86	74.01	
Gabor on gaussian color model		Channel information	88.86	87.89	85.28	87.13	87.55	87.34	84.70	86.43	
		Gray scale	65.28	67.29	67.17	61.10	71.67	69.48	68.86	67.70	
		Stain normalization	79.56	74.43	73.16	75.11	77.55	73.59	72.92	75.93	
M CCR (8)		Channel information	87.76	88.19	85.49	85.75	85.66	86.83	83.93	82.62	
		Gray scale	70.24	70.4	69.41	69.76	70.80	69.7	68.40	67.61	
		Stain normalization	72.43	73.44	74.12	75.09	74.59	75.90	76.37	76.58	
		Channel information	79.30	75.87	80.95	88.51	78.55	75.39	80.15	81.40	

(2) Comparing the cases with and without stain normalization, it is seen that the classification performance is better in case of latter. However, the difference is not too high or even quite close in many cases where traditional texture features are used. This is expected as texture information is similar, except in some cases where even independent color channels may help in capturing the color variation.

(3) However, in case of joint color-texture features, except for a very few cases, the performance without stain normalization is consistently quite high. This indicates that the joint color-texture features, which consider the mutual dependency of color channels, indeed, better capture the color-texture variation for classification.

(4) It can also be observed that opponent LBP where opponent channels are considered to extract color channel information, shows the superior performance for most (three) of magnification images. However, Gabor feature on Gaussian color model, and M CCR also yield somewhat comparable performance.

There are very few exceptions from the above observations for some feature-classifier combinations. While these need to be better scrutinized, we note that most of the results support our hypothesis of the effectiveness of color-texture information in mitigating the need for stain normalization.

Finally, Table 4 compares the results of the proposed approach obtained with joint color-texture features, with some existing state-of art methods. In Table 4,

Table 3. Detailed description.**Table 4.** Performance comparison.

	Magnifications				Total	Patient	Methods & Score		[10]	[11]	[12]	Proposed
	40x	100x	200x	400x			40x	Patient level	83.8 ± 4.1	90.6 ± 6.7	83.08 ± 2.08	86.88 ± 2.37
Benign	625	644	623	588	2480	24	100x	Patient level	82.1 ± 4.9	88.4 ± 4.8	83.17 ± 3.51	88.41 ± 2.73
Malignant	1370	1437	1390	1232	5429	58	200x	Patient level	85.1 ± 3.1	84.6 ±	84.63 ± 2.72	88.86 ± 3.76
Total	1995	2081	2013	1820	7909	82	400x	Patient level	82.3 ± 3.8	86.1 ± 6.2	82.10 ± 4.42	87.55 ± 3.01

we report the best results of the proposed method obtained across various combination of color-texture descriptors and classifiers. We can observe that, except for the case of 40x, the proposed method outperforms the existing approaches. We also note from the Table 4 that at 400x, the obtained accuracy for most methods is lower than that at 100x and 200x. The reason could be that there are relatively less number of images for 400x and perhaps more data is required to capture the finer traits at higher magnification. Furthermore, one can also observe that the proposed work yields a lesser variance in scores, in most of the cases. This comparison further shows the effectiveness of joint color-texture features for classification.

4 Conclusion

In this work, we attempt to establish the usefulness of joint color-texture information, for classification without the need for stain normalization. We have experimented with various classifiers to show the importance of independent and dependent (mutual) color-channel information and find some interesting aspects about the same. From our experiments, it is apparent that joint dependency of color-texture can better capture the color-texture variability. We have also shown the role of contemporary classifiers with these sophisticated color-texture features. We believe that this is an interesting study which points towards obviating the need of stain normalization given effective features and classifiers. We also demonstrate that employing the joint color-texture features can also outperform the state-of-the-art methods for the breast cancer histopathology image classification.

References

1. Boyle, P., Levin, B., et al.: World cancer report 2008. IARC Press, International Agency for Research on Cancer (2008)
2. Bejnordi, B.E., Litjens, G., Timofeeva, N., Otte-Höller, I., Homeyer, A., Karssemeijer, N., van der Laak, J.A.W.M.: Stain specific standardization of whole-slide histopathological images. *IEEE Trans. Med. Imaging* **35**(2), 404–415 (2016)
3. Reinhard, E., Adhikmin, M., Gooch, B., Shirley, P.: Color transfer between images. *IEEE Comput. Graphics Appl.* **21**(5), 34–41 (2001)
4. Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A.M., Esposito, I., Navab, N.: Structure-preserving color normalization and sparse stain separation for histological images. *IEEE Trans. Med. Imaging* **35**(8), 1962–1971 (2016)

5. Basavanthally, A.N., Ganesan, S., Agner, S., Monaco, J.P., Feldman, M.D., Tomaszewski, J.E., Bhanot, G., Madabhushi, A.: Computerized image-based detection and grading of lymphocytic infiltration in her2+ breast cancer histopathology. *IEEE Trans. Biomed. Eng.* **57**(3), 642–653 (2010)
6. Ciompi, F., Geessink, O., Bejnordi, B.E., de Souza, G.S., Baidoshvili, A., Litjens, G., van Ginneken, B., Nagtegaal, I., van der Laak, J.: The importance of stain normalization in colorectal tissue classification with convolutional networks. *arXiv preprint* (2017). [arXiv:1702.05931](https://arxiv.org/abs/1702.05931)
7. Weingant, M., Reynolds, H.M., Haworth, A., Mitchell, C., Williams, S., DiFranco, M.D.: Ensemble prostate tumor classification in h&e whole slide imaging via stain normalization and cell density estimation. In: Zhou, L., Wang, L., Wang, Q., Shi, Y. (eds.) *International Workshop on Machine Learning in Medical Imaging. LNCS*, pp. 280–287. Springer, Cham (2015)
8. Fernández-Carrobles, M.M., Bueno, G., Déniz, O., Salido, J., García-Rojo, M., González-López, L.: Influence of texture and colour in breast TMA classification. *PloS one* **10**(10), e0141556 (2015)
9. Amaral, T., McKenna, S., Robertson, K., Thompson, A.: Classification of breast-tissue microarray spots using colour and local invariants. In: *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008*, pp. 999–1002. IEEE (2008)
10. Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L.: A dataset for breast cancer histopathological image classification. *IEEE Trans. Biomed. Eng.* **63**(7), 1455–1462 (2016)
11. Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L.: Breast cancer histopathological image classification using convolutional neural networks. In: *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 2560–2567. IEEE (2016)
12. Bayramoglu, N., Kannala, J., Heikkilä, J.: Deep learning for magnification independent breast cancer histopathology image classification. In: *23rd International Conference on Pattern Recognition, ICPR 2016* (2016)
13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp. 1097–1105 (2012)
14. Haralick, R.M., Shanmugam, K., et al.: Textural features for image classification. *IEEE Trans. Syst. Man Cybernet.* **3**(6), 610–621 (1973)
15. Ojansivu, V., Heikkilä, J.: Blur insensitive texture classification using local phase quantization. In: Elmoataz, A., Lezoray, O., Nouboud, F., Mammass, D. (eds.) *ICISP 2008. LNCS*, vol. 5099, pp. 236–243. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-69905-7_27](https://doi.org/10.1007/978-3-540-69905-7_27)
16. Mäenpää, T., Pietikäinen, M.: Texture analysis with local binary patterns. *Handbook Pattern Recog. Comput. Visi.* **3**, 197–216 (2005)
17. Hoang, M.A., Geusebroek, J.-M., Smeulders, A.W.M.: Color texture measurement and segmentation. *Sig. Process.* **85**(2), 265–275 (2005)
18. Bianconi, F., Fernández, A., González, E., Caride, D., Calviño, A.: Rotation-invariant colour texture classification through multilayer CCR. *Pattern Recog. Lett.* **30**(8), 765–773 (2009)
19. Hamilton, N.A., Pantelic, R.S., Hanson, K., Teasdale, R.D.: Fast automated cell phenotype image classification. *BMC Bioinf.* **8**(1), 110 (2007)
20. Classification-learner-app. <https://in.mathworks.com/help/stats/classification-learner-app.html>