



City Research Online

City, University of London Institutional Repository

Citation: Bianconi, F., Kather, J. N. & Reyes-Aldasoro, C. C. (2020). Experimental Assessment of Color Deconvolution and Color Normalization for Automated Classification of Histology Images Stained with Hematoxylin and Eosin. *Cancers*, 12(11), 3337. doi: 10.3390/cancers12113337

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/25289/>

Link to published version: <https://doi.org/10.3390/cancers12113337>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk

Article

Experimental Assessment of Color Deconvolution and Color Normalization for Automated Classification of Histology Images Stained with Hematoxylin and Eosin

Francesco Bianconi ^{1,2,*} , Jakob N. Kather ³  and Constantino Carlos Reyes-Aldasoro ² 

¹ Department of Engineering, Università degli Studi di Perugia, Via Goffredo Duranti 93, 06125 Perugia, Italy

² giCentre, School of Mathematics, Computer Science & Engineering, City, University of London, Northampton Square, London EC1V 0HB, UK; reyes@city.ac.uk

³ Department of Medical Oncology and Internal Medicine VI, National Center for Tumor Diseases (NCT), University Hospital Heidelberg, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany; jakob.kather@nct-heidelberg.de

* Correspondence: bianco@ieee.org; Tel.: +39-075-585-3706

† Performed part of this work as an Academic Visitor in the School of Mathematics, Computer Science & Engineering at City, University of London.

Received: 7 October 2020; Accepted: 4 November 2020; Published: 11 November 2020



Simple Summary: The appearance of histology images stained with H&E can vary a lot as a consequence of changes in the reagents, staining conditions, preparation procedure and acquisition system. In this work we investigated whether color preprocessing – specifically color deconvolution and color normalization – could be used to correct such variability and improve the performance of automated classification procedures. Experimenting on 11 datasets, 13 image descriptors and eight color pre-processing methods we found that doing no color preprocessing was the best option in most cases.

Abstract: Histological evaluation plays a major role in cancer diagnosis and treatment. The appearance of H&E-stained images can vary significantly as a consequence of differences in several factors, such as reagents, staining conditions, preparation procedure and image acquisition system. Such potential sources of noise can all have negative effects on computer-assisted classification. To minimize such artefacts and their potentially negative effects several color pre-processing methods have been proposed in the literature—for instance, color augmentation, color constancy, color deconvolution and color transfer. Still, little work has been done to investigate the efficacy of these methods on a quantitative basis. In this paper, we evaluated the effects of color constancy, deconvolution and transfer on automated classification of H&E-stained images representing different types of cancers—specifically breast, prostate, colorectal cancer and malignant lymphoma. Our results indicate that in most cases color pre-processing does not improve the classification accuracy, especially when coupled with color-based image descriptors. Some pre-processing methods, however, can be beneficial when used with some texture-based methods like Gabor filters and Local Binary Patterns.

Keywords: histology images; H&E staining; color; texture

1. Introduction

Digital pathology plays a fundamental role in cancer diagnosis, treatment and follow-up [1–9]. This consists of a range of activities such as the acquisition, storage, sharing, analysis and interpretation

of histological images [10]. In this domain, computer-assisted classification of tissue samples has attracted considerable research interest in recent years as a means for assisting pathologists in several tasks, for instance, the classification of specimens into normal or abnormal [11–14], the grading of neoplastic tissue [15–18], the estimation of tumor proliferation [19] and the identification of tissue substructures such as epithelium, stroma, lymphocytes, necrosis, etc. [20,21]. With the growing popularity of whole-slide scanners, and consequently, the increasing availability of digital images, digital pathology has the potential not only to reduce the workload by automating several repetitive tasks, but also to increase the reproducibility of human-based evaluations.

Among the problems that so far have limited the adoption of digital pathology on a wide scale are differences in the protocols, materials and procedures for image acquisition and the little availability of large datasets of annotated images [22]. Such variations in protocols, materials and procedures can result in unlike visual appearance of the pathology slides, which can have the undesired effect of reducing the accuracy, sensibility and specificity of automated, machine-based approaches [23,24]. The problems related to stain normalization have generated considerable research interest in the last few years and several methods have been proposed in the literature [22,23,25–31]. However, few studies investigated the subject on a quantitative basis, and some reported divergent results. Furthermore, many such studies were based on a limited number of data sets—as few as one in some cases—which makes it difficult to draw general conclusions. Consequently, the effects of pre-processing methods on automated classification of H&E-stained images are not entirely clear yet. In [32,33] the authors reported improved accuracy for patch-based classification based on Convolutional Neural Networks (CNN), whereas [34] showed that color features lost distinctiveness when color normalization was applied. More recently, Hameed et al. [35] also reported that their classification performance deteriorated upon using color-normalized images. Furthermore, the combined effects color pre-processing/image descriptors (e.g., color descriptor, texture descriptors and/or convolutional network) have been addressed only in [34,36,37].

This work presents a quantitative evaluation of color deconvolution and color normalization on automated (patch-based) classification of histology images stained with hematoxylin and eosin from breast, prostate, colorectal cancer and malignant lymphoma. The present study extends the preliminary results presented in [38] and the main contribution is to provide a set of guidelines to select the appropriate combinations color pre-processing/image descriptor for histopathological image analysis. We found that in most cases color pre-processing did not improve classification accuracy, especially when coupled with color-based image descriptors convolutional networks. Some pre-processing methods, however, provided some slight gain when used with texture-based methods like Gabor filters and Local Binary Patterns. On the whole the best combinations involved the use of pre-trained networks (ResNet50/101) or color histograms as image descriptors and no color pre-processing at all.

2. Materials

We considered nine datasets of H&E-stained histological images representing different types of neoplastic diseases as detailed below. Samples images of each dataset are illustrated in Figure 1.

2.1. *Agios Pavlos (AP)*

Histological images from breast carcinoma collected within the ‘Agios Pavlos’ Department of Pathology at the General Hospital of Thessaloniki (Thessaloniki, Greece). The dataset includes 300 images (magnification 40×, dimension 1280px × 960px) of invasive ductal carcinoma (grades I, II and III) from 21 patients.

2.2. *BreakHis (BH)*

Histological samples of breast carcinoma collected at the Pathological Anatomy and Cytopathology Laboratory (P&D Lab, Paraná, Brazil) [39]. This collection features 7909 microscopy images of breast tumor tissue from eight different histological sub-types. The tissue samples were

collected from 82 patients under four magnifying factors: 40 \times , 100 \times , 200 \times and 400 \times , of which the first was the one used in this study. The dimension of the images is 700px \times 460px.

2.3. Cedars-Sinai (CS)

Histological images from patients with prostate cancer collected at the Cedars-Sinai Medical Center (Los Angeles, CA, USA) [40]. The data set features 625 images of dimension 1201px \times 1201px each containing manually annotated regions of either benign tissue, stroma and/or malignant tissue (Gleason grade from III to V). The spatial resolution is $\approx 0.5\mu\text{m}/\text{px}$. From this set we randomly extracted 256px \times 256px tiles representing clearly identifiable areas of each grade (100 tiles for each class).

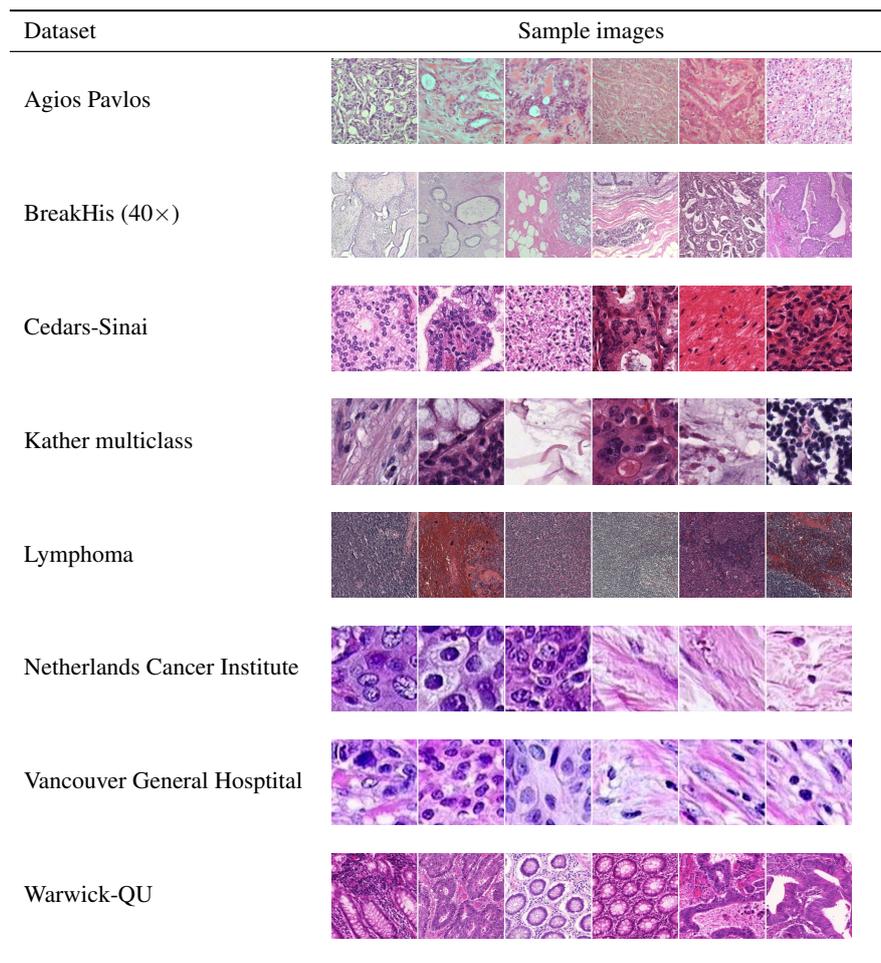


Figure 1. Six representative sample images from the datasets used in the experiments. It should be noticed the diverse gamut of colors as well as the different magnifications, density and cell density of the datasets. Cedars-Sinai by courtesy of Cedars-Sinai Medical Center (©2020 Cedars-Sinai Medical Center. All rights reserved).

2.4. HICL

Histological samples from 109 subjects with breast ductal carcinomas who received a biopsy at the Department of Pathology, University Hospital of Patras, Rio, Greece, between 2000 and 2007 [41]. The dataset comes with a manually defined, ground truth subdivision into grade I ($n = 63$), II ($n = 83$) and III ($n = 80$). The images were acquired with 40 \times magnification factor and the final dimension is 1728px \times 1296px.

2.5. *Kather Multiclass (KM)*

A dataset of histological images of colorectal cancer collected at the University Medical Center Mannheim, Heidelberg University (Heidelberg, Germany) [21,42]. The data set is composed of 5000 tissue samples (tiles) from 10 patients representing eight different tissue sub-types (see Figure 1 for details). Each tile has a dimension of 150px × 150px and spatial resolution of $\approx 0.5 \mu\text{m}/\text{px}$. The images were acquired under 20× magnification using an Aperio ScanScope (Aperio/Leica biosystems).

2.6. *Lymphoma*

Histological images of malignant lymphoma from different institutions [43,44]. This data set is part of the Benchmark Suite for Biological Image Analysis (IICBU 2008). It includes a total of 374 images organized in three classes: chronic lymphocytic leukemia ($n = 113$), follicular lymphoma ($n = 139$) and mantle cell lymphoma ($n = 122$). The dimension of the images is 1388px × 1040px. Since the samples come from different centers there is a large amount of staining variation.

2.7. *Netherlands Cancer Institute (NKI)*

Breast cancer histology images from a population of 248 patients. The images were collected at the Netherlands Cancer Institute (Amsterdam, Netherlands) [45,46]. From the predefined segmentation into epithelium and stroma which comes with the dataset we respectively extracted 1106 and 189 tile images of each class (dimension 100px × 100px).

2.8. *Vancouver General Hospital (VGH)*

This dataset has the same structure as the one described in Section 2.7, but in this case the study population comprises 328 subjects enrolled at Vancouver General Hospital (Vancouver, BC, Canada) [45,46]. With the same procedure and settings described in Section 2.7 we extracted 226 image samples of epithelium and 47 of stroma.

2.9. *Warwick-QU (WR)*

This dataset includes a total of 165 images representing colorectal tissue and is organized in two classes: benign ($n = 74$) and malignant tissue ($n = 91$). The samples were collected at the University Hospitals Coventry and Warwickshire (Coventry and Rugby, United Kingdom) [47,48]. The images were acquired at 20× magnification factor and spatial resolution of $\approx 0.62 \mu\text{m}/\text{px}$; the dimension is variable. The data set was part of the Gland Segmentation Challenge Contest (GlaS) at MICCAI 2015 (Munich, Germany, 5–9 October 2015) [49].

2.10. *Combined Datasets (AP+HICL, NKI+VGH)*

One important factor that can affect the colors of histological images are the specific conditions of the acquisition laboratory. To assess the effects of inter-laboratory variability, we generated two additional datasets by merging Agios Pavlos and HICL ($AG + HICL$) and NKI and VGH ($NKI + VGH$). These datasets were selected as they consider the same disease type and grades, and have compatible magnification factor and image resolution.

It should be noted that the images considered in this work are considerably smaller than those provided by whole-slide scanners [50,51]. Images from whole-slide scanners can span tens or hundreds of thousands of pixels, and these are typically cropped into smaller tiles and thus very large number of images can be used for studies. For reproducibility, we used the nine publicly available datasets described above.

3. Methods

3.1. Color Pre-Processing

It is convenient to classify color pre-processing methods for histological images into three categories: *color augmentation*, *color deconvolution* and *color normalization* (Figure 2).

3.1.1. Color Augmentation

Color augmentation is a type of data augmentation whereby new images are generated by applying some kind of perturbation to the colour distribution of the original ones [23,36]. Color augmentation was not considered in this study as it is intrinsically different from color deconvolution and color normalization, which were considered. The main difference is the input/output relationship: in both color deconvolution and color normalization, the relationship is one-to-one, while in color augmentation it is one-to-many. The number of output images returned by color augmentation is a parameter to set and depends on the method chosen. Testing color augmentation would therefore require a rather different set-up than the one used for color deconvolution and color augmentation.

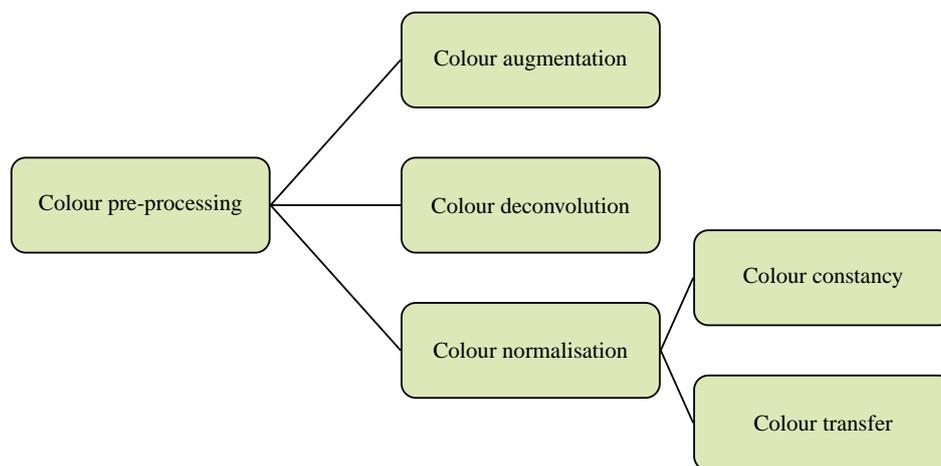


Figure 2. Color pre-processing for histological images: a taxonomy.

3.1.2. Color Deconvolution

Color deconvolution consists of decomposing the input images into separate channels, each representing the concentration of each stain used [52]. In H&E-stained images that means separating the original images into haematoxylin, eosin and background. Please note that in some cases colour deconvolution is just one step towards colour normalization [22]. In this work we considered Ruifrok and Johnston’s method [26] (‘decoRJ’ in the remainder) and Macenko’s *et al.* [25] (‘decoMC’ in the remainder)—both through the implementation provided in [53]. Figure 3 shows the effects of these methods on a set of sample images.

3.1.3. Colour Normalization

Color normalization can be further classified into *color constancy* and *color transfer*. The first derives from color constancy in vision theory, the objective of which is to assign a constant color to the same objects when acquired under different illumination conditions [54,55]. This extends seamlessly to histological images, even if, in this case, changes in color can be due both to variable illumination and, to a greater extent, to differences in tissue preparation and staining. The second, color transfer, modifies the color distribution of the input image to make it match that of a *target* image [56]. Below we describe the color constancy and color transfer methods considered in the experiments.

The colour constancy methods investigated in this work were: (1) chromaticity representation ('chroma' in the remainder), (2) grey-world normalisation ('gw') and (3) histogram equalization ('heq') [57,58]. The first simply divides the R , G and B values of each pixel of the input image by their sum $R + G + B$. The second works on the assumption that the average color in a scene is grey, and that deviations of the average color from grey are due to the light source. The input image is corrected accordingly. The third modifies the marginal distribution (histogram) of each color channel by making it approximate a uniform one. The implementation was based on the Color Constancy toolbox [59] (for chroma and gw) and Matlab's `histeq()` function histogram equalisation.

For color transfer we considered the methodologies of Khan et al. [22], Macenko et al. [25] and Reinhard et al. [56], each with four different target images denoted as T1–T4 in the remainder (see also Figure 4). Three of these images (T2–T4) are histology images, and one (T1) is not. For the latter we selected a color calibration mask (colour checker), which is an image with a large variation of colors not related to histology. The rationale was to investigate how widely the colors of the original image could vary if those of the target image were markedly different. For the implementation we used the functions available in Warwick's Stain Normalization Toolbox [53]. Figure 4 illustrates the effects of each color normalisation methods on a set of sample images.

3.2. Image Descriptors

The image descriptors that can be used for histological image analysis fall into two main categories: the traditional, 'hand-designed' methods on the one hand and the convolutional networks (CNN) on the other [60]. The first group can be further subdivided into spatial (texture), spectral (color) and hybrid methods [61] (Figure 5). For this study we considered eight 'hand-designed' descriptors and five pre-trained convolutional networks as detailed below.

3.2.1. Hand-Designed Methods (Spectral)

Three-Dimensional Color Histogram (*FullHist*)

The three-dimensional probability distribution in the color space as described in [62]. We used ten bins for each color channel giving a total of $10^3 = 1000$ features.

One-Dimensional Marginal Color Histograms (*MargHists*)

The concatenation of the three one-dimensional probability distributions of the intensity level in each color channel [63]. We used 256 bins for each color channel giving a total of $256 \times 3 = 768$ features.

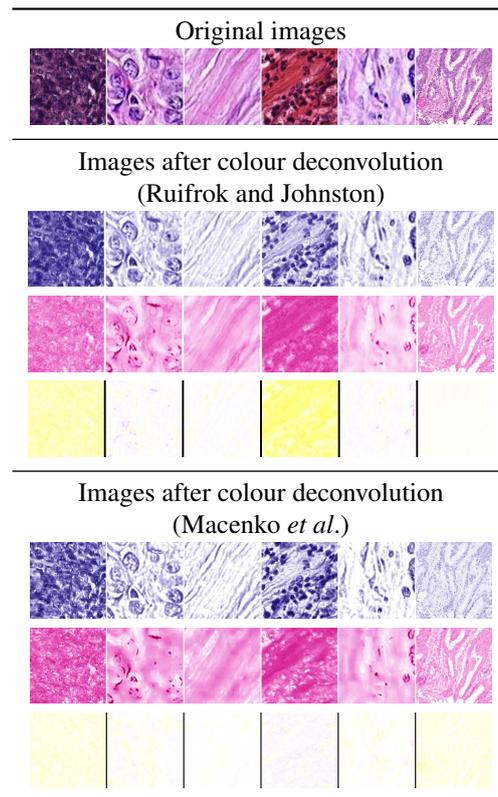


Figure 3. Effects of colour deconvolution through Ruifrok and Johnston's [26], and Macenko *et al.*'s method [25]. The top row shows the original images, then each box below reports the deconvolved haematoxylin channel (first row), the deconvolved eosin channel (second row) and the background channel (third row). The haematoxylin, eosin and background channels are rendered in pseudo-colors.

Original images					
Images after colour normalisation		Method	Target img.		
		Chromaticity repr.	None		
		Grey-world norm.	None		
		Histogram equalisation	None		
		Colour transfer (Khan <i>et al.</i>)			
		Colour transfer (Macenko <i>et al.</i>)			
		Colour transfer (Reinhard <i>et al.</i>)			

Figure 4. Illustration of the effects of color constancy and color transfer on a series of representative images with four different target images. Three targets are histological images, and one is a color checker mask used to investigate the impact caused by an image with a large and distant color variation.

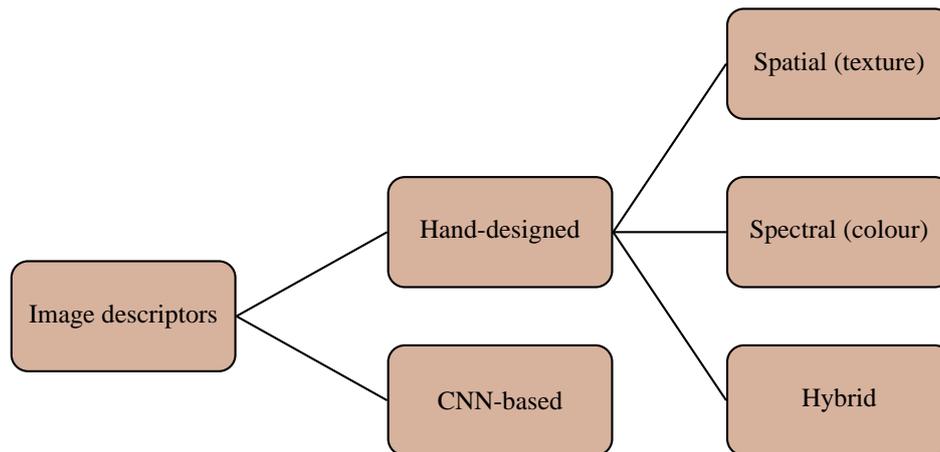


Figure 5. Taxonomy of the image descriptors used in this study.

3.2.2. Hand-Designed Methods (Spatial)

Grey-Level Co-Occurrence Matrices (GLCM)

Texture features from 12 co-occurrence matrices computed using three distances (1px, 2px and 3px) and four orientations (0° , 45° , 90° and 135°). From each matrix we extracted five statistical parameters: contrast, correlation, energy, entropy and homogeneity [64] for a total of $12 \times 5 = 60$ features. We finally applied Discrete Fourier Transform (DFT) normalization to obtain rotationally invariant features [65].

Gabor Filters (Gabor)

Texture features from a bank of 24 Gabor filters with four frequencies and six orientations. From the absolute value of each Gabor-transformed image we computed the mean and standard deviation giving a total $2 \times 4 \times 6 = 48$ features. Again, rotationally invariant features were finally obtained via DFT normalization [66].

Local Binary Patterns (LBP)

Histograms of rotation-invariant ('ri') Local Binary Patterns [67] computed using non-interpolated circular neighborhoods of eight-pixels each and resolution 1px, 2px and 3px (see also [68] for details). This configuration produces 36 features for each resolution, therefore a total of $36 \times 3 = 108$ features.

3.2.3. Hand-Designed Methods (Hybrid)

From the grey-scale texture descriptors described in Section 3.2.2 we derived marginal color versions by applying the grey-scale methods to each color channel separately and concatenating the resulting feature vectors. Consequently, the marginal color versions of Gabor, GLCM and LBP (which we indicate as 'MargGabor', 'MargGLCM' and 'MargLBP' henceforth) have feature vectors that are three times longer than those of the grey-scale counterparts.

3.2.4. Pre-Trained Convolutional Networks

We used five pre-trained convolutional networks 'off-the-shelf'—i.e., without any further re-training or fine-tuning (see also [60,69] for details on this approach). For all the models the imaging features were the L_2 -normalized output of the layer indicated in Table 1. The number of features generated by each configuration is also reported in the table.

Table 1. Round-up table of the pre-trained convolutional models considered in the study.

Model	Ref.	Layer (Name/No.)	No. of Features
InceptionV3	[70]	313	2048
ResNet50	[71]	'pool5'	2048
ResNet101	[71]	'pool5'	2048
Vgg16	[72]	'FC-4096'	4096
Vgg19	[72]	'FC-4096'	4096

3.3. Further Pre-Processing Steps

Convolutional networks have input fields of fixed shape and size, which requires the input images to be resized accordingly. To this end we cropped non-square images to a maximal centered square, then linearly resized the resulting crop to the networks' input field. Since all the networks considered here feature a square input field, the first step was required to avoid distortion. For fair play the crop was applied in any case, even though the hand-designed descriptors can cope with input images of any shape and size. Linear resize after crop was used with the networks only.

4. Experiments

To test the effectiveness of each combination of color pre-processing/image descriptor (Section 3) we conducted a series of supervised image classification experiments, each of them using the data sets previously described in Section 2. We estimated the accuracy through split-sample validation with stratified sampling; that is, for each data set analyzed, we considered a fraction (f) of the samples of each class (i.e., the train set) to construct the classifier, and then, the remaining samples (i.e., the test set) were used to calculate the accuracy. Thus, the accuracy was the percentage of samples of the test set classified as correct. To obtain a stable estimation, we repeated the random subdivision of the train and test sets hundred times and the results were averaged. We repeated the experiments using $f = 1/4$ and $f = 1/8$. The classification was based on the rule of nearest-neighbor with the L_1 ('cityblock') distance.

The experiments were implemented using Matlab[®] (The Mathworks[™], Natick, USA) and carried out on a laptop PC equipped with Intel[®] core[™] i5-3230M CPU@ 2.60GHz, 8 GB RAM and Windows 7 Professional 64-bit. Feature extraction was based on the freely available Color And Texture Analysis Toolbox for Matlab (CATAcOMB) [73] for the hand-designed descriptors, on MatConvNet [74] for the ResNet and VGG models and on Matlab's dedicated plug-in for InceptionV3.

5. Results and Discussion

5.1. Accuracy

The results for the best and second-best combinations of image descriptor and color pre-processing method for each data set are presented in Table 2. It can be observed that out of the 11 best combinations, 7 cases corresponded to the pre-trained ResNet50 and ResNet101, three cases to the joint and marginal color histograms and one to co-occurrence matrices. When considering the best and second-best cases, these corresponded to the pre-trained ResNet50 and ResNet101 in 12 cases out of 22. Regarding color pre-processing, doing nothing provided the best or second-best option in ten cases out of 22, followed by deconvolution (five) and chromaticity representation (three).

Table 2. Best (rank = 1) and second-best (rank = 2) combinations color pre-processing/image descriptor by dataset. Figures indicate accuracy, also reflected in the ball size and color (blue = low, brown = high). Values are filtered on $f = 1/4$.

Dataset	Rank	Accuracy (%)	Descriptor	Pre-Processing
AP	1	81.79	MargGLCM	decoMC
	2	81.70	FullHist	heq
AP+HICL	1	68.97	ResNet50	decoRJ
	2	67.61	FullHist	Reinhard (T1)
BH	1	90.67	ResNet101	none
	2	90.07	ResNet50	none
CS	1	87.59	FullHist	none
	2	86.39	ResNet50	none
HICL	1	51.58	ResNet101	decoMC
	2	51.51	InceptionV3	decoRJ
KM	1	92.18	FullHist	none
	2	89.03	MargHists	chroma
Lymphoma	1	85.98	MargHists	chroma
	2	84.53	FullHist	none
NKI	1	98.87	ResNet50	none
	2	98.86	ResNet50	chroma
NKI+VGH	1	98.39	ResNet50	none
	2	98.33	ResNet101	gw
VGH	1	96.10	ResNet101	none
	2	96.00	MargHists	decoRJ
WR	1	94.37	ResNet50	none
	2	94.11	ResNet50	Khan (CC140)

Figure 6 shows the accuracy for each descriptor and data set, while color indicates the pre-processing methodology. As can be observed, the performance of the color-based descriptors (i.e.: color histograms and pre-trained networks) varied significantly depending on the pre-processing method used. By contrast, the texture-based descriptors were markedly more resilient, as one would reasonably expect. Also, it should be noted that the marginal versions of the texture descriptors (MargGabor, MargGLCM and MargLBP) outperformed their grey-scale counterparts (Gabor, GLCM and LBP).

Table 3 reports the difference to the baseline (i.e., no color pre-processing) divided by descriptor and color pre-processing methodology. These values are averaged over all the data sets. The box plots of Figures 7 and 8 break down the difference by color pre-processing method, while color and shape of the markers respectively show details about the descriptor and data set. On the whole, color pre-processing caused a loss of accuracy in most cases. This was particularly true when pure color descriptors and convolutional networks were involved (Figure 8); moreover, we can see that in some cases the decrease in accuracy was very sharp. Those methodologies which rely heavily on color responded negatively to color pre-processing, which is in line with the results reported in [34]. The results also show that the outcome of color transfer methodologies (Khan's, Macenko's and Reinhard's) was pretty much independent on the target image used, regardless this being a histology image (T2–T4) or not (T1). In fact, it is quite surprising that on average T1 performed slightly better than the others (Table 3). We believe this is an important finding, because it suggests that despite the color-transformed images obtained using T2–T4 as target images 'look better' than those obtained using T1, this does not translate into a better performance of the automatic classification. A comparison among the three

methods show that Khan’s and Macenko’s had a similar performance, whereas that of Reinhard’s was markedly worse. Regarding color deconvolution, we observe (Table 3) that on average this was generally beneficial only when coupled with texture descriptors, but not in the other cases (i.e., color descriptors and pre-trained CNN).

The methods based on texture proved fairly resilient to color pre-processing (Figure 6), as it would reasonably expected. In these cases, there was even a noticeable improvement of the accuracy in some combinations of the descriptor and the pre-processing methodology. Specifically, the marginal color texture descriptors (i.e., MargGabor, MargGLCM and MargLBP) seemed to provide a positive response both to ‘chroma’ normalization and color deconvolution. The latter results looked particularly interesting, i.e., it suggests that the texture features can provide complementary information when applied to each of the channels separately, i.e., haematoxylin, eosin and background.

To reduce potential sources of bias related to the samples distribution in the training and test sets, we repeated the classification experiments using a lower train ratio ($f = 1/8$). The complete results show that no significant difference was observed with the trend with $f = 1/4$.

5.2. Computational Demand

Table 4 illustrates the average feature extraction time by descriptor and color pre-processing methods. On the whole the results indicate that there was some additional overhead, as one would reasonably expect. This was more noticeable for the color transfer methods—particularly Khan’s—than for the color constancy ones, which is consistent with the higher complexity of the first group compared with the second. Surprisingly, there was a gain in speed in some cases, as for instance with the combinations chroma normalization/GLCM and MargGLCM. A possible explanation is that by definition, chroma normalization projects the color distribution onto a plane, therefore effectively reducing the dimensionality of the color space from three to two. As for the image descriptors, it can be seen that MargHists was the quickest method, followed by FullHist, LBP and the ResNet and VGG pre-trained models. The other texture descriptors were significantly slower, as was InceptionV3.

Table 3. Difference to the baseline by image descriptor and pre-processing method. The values are averaged over the eight data set and filtered on $f = 4$. Baseline is the condition where no color pre-processing is applied.

Descriptor	Colour pre-processing method																		AVG
	chroma	ew	heq	Khan (T1)	Khan (T2)	Khan (T3)	Khan (T4)	Macenko (T1)	Macenko (T2)	Macenko (T3)	Macenko (T4)	Reinhard (T1)	Reinhard (T2)	Reinhard (T3)	Reinhard (T4)	decoMC	decoR		
FullHist	-8.01	-1.24	-1.77	-5.45	-7.35	-7.70	-7.84	-5.01	-4.88	-5.54	-5.02	-20.94	-20.94	-20.94	-20.94	-5.48	-20.94	-9.44	
MargHists	0.77	-2.26	-21.12	-5.77	-5.82	-7.53	-7.39	-4.33	-4.71	-4.21	-5.05	-16.57	-16.18	-19.65	-21.33	-3.80	-2.92	-8.22	
Gabor	-0.42	1.00	-2.47	-3.38	-2.87	-2.63	-2.71	-2.05	-1.51	-1.63	-1.79	-2.26	-2.45	-2.35	-2.13	0.06	-0.29	-1.66	
GLCM	0.66	0.37	-3.54	-2.02	-2.62	-4.26	-2.20	-1.28	-2.33	-0.82	-2.67	-1.79	-3.86	-1.85	-10.11	-0.74	-2.57	-2.31	
LBP	3.14	-0.17	-0.40	0.13	-0.50	-1.51	-1.85	-0.11	-0.31	-0.29	-0.09	0.94	-1.12	-0.56	-1.09	0.96	1.11	-0.10	
MargGabor	1.66	1.34	-0.52	-2.58	-2.16	-2.79	-3.65	1.18	-1.70	-1.87	-2.11	-2.37	-1.69	-2.31	-2.12	2.36	3.42	-0.88	
MargGLCM	1.10	0.62	-1.88	-1.86	-2.47	-4.65	-4.57	-0.86	-3.04	-2.90	-5.31	-0.90	-3.66	-1.96	-6.33	0.09	-0.74	-2.18	
MargLBP	3.43	0.27	-0.06	-0.04	-1.84	-2.55	-3.00	2.09	-0.73	-0.87	-1.09	1.77	0.09	-0.33	-1.53	1.16	1.96	-0.07	
InceptionV3	-3.30	-0.60	-0.55	-5.22	-1.81	-3.30	-3.44	-1.49	-0.68	-1.70	-0.77	-8.90	-10.79	-10.18	-9.02	-3.24	-6.58	-3.98	
ResNet50	-2.19	-1.33	-1.24	-4.98	-4.46	-5.51	-6.29	-1.37	-1.88	-1.46	-2.31	-14.70	-13.76	-14.16	-12.03	-2.39	-10.96	-5.61	
ResNet101	-3.83	-0.63	-1.63	-6.11	-5.93	-6.97	-8.23	-2.51	-1.88	-1.67	-2.88	-14.40	-14.39	-14.21	-13.95	-3.42	-11.40	-6.34	
Vgg16	-2.93	-1.13	-0.94	-4.23	-3.55	-5.38	-7.24	-2.94	-0.99	-1.21	-1.60	-17.52	-17.26	-17.47	-16.69	-5.69	-9.22	-6.44	
Vgg19	-1.16	-0.98	-1.66	-4.78	-3.33	-3.54	-5.10	-1.22	-3.13	-1.50	-4.55	-16.16	-13.77	-14.43	-13.14	-2.51	-11.37	-5.69	
AVG	-0.85	-0.36	-2.91	-3.56	-3.44	-4.49	-4.89	-1.53	-2.14	-1.97	-2.71	-8.75	-9.21	-9.26	-10.03	-1.74	-5.42	-4.07	

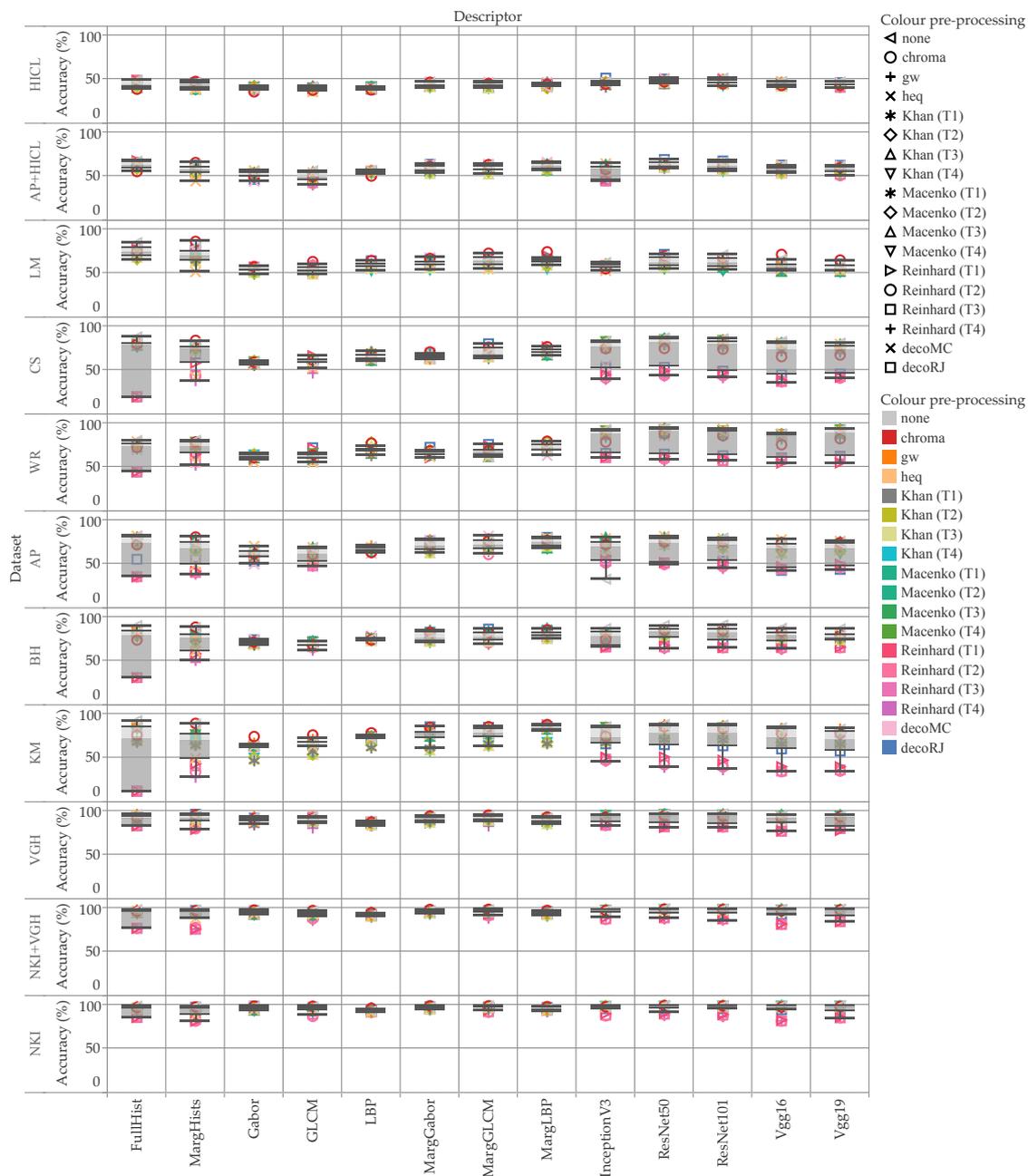


Figure 6. Accuracy by data set and descriptor; color indicates the pre-processing method. The values reported are for $f = 4$. This chart shows interesting things. First, texture-based image descriptors (e.g., LBP) are much more insensitive to color pre-processing than the other methods (e.g., FullHist). This is important when analyzing the reproducibility of the methodologies. Second, the cases where there was large variation seemed to have results on the extremes (i.e., KM / FullHist) and not a uniform distribution. Third, and perhaps the most important, the accuracy obtained in different datasets is considerably different. Compare for instance NKI which is very close to 100% with LM where most cases are around 50%. This highlights the importance of testing on more than one dataset, as the choice of dataset can result in higher or lower values of accuracy.

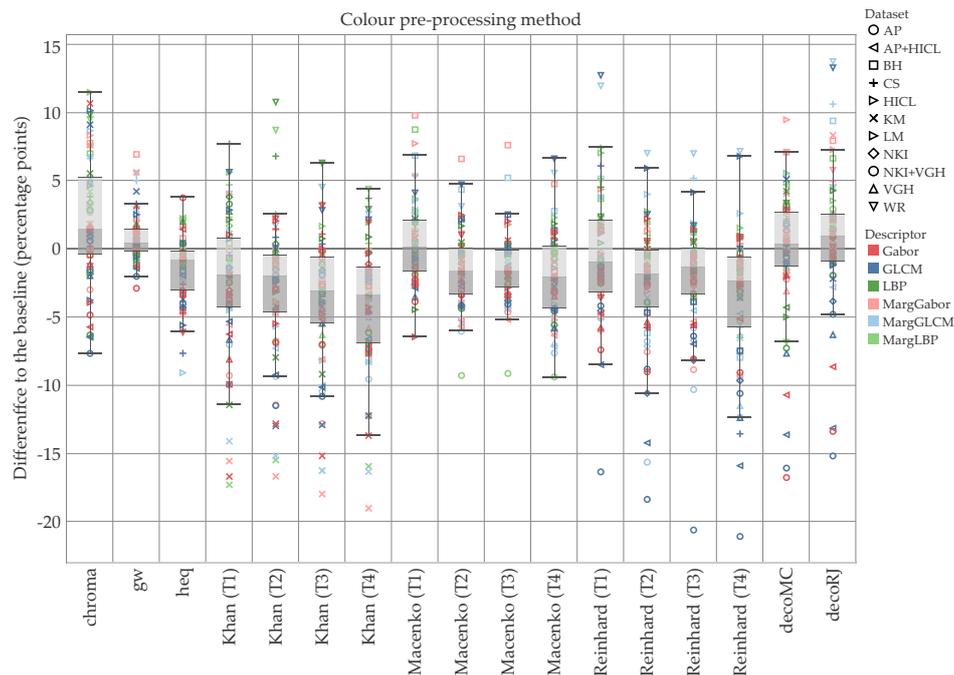


Figure 7. Difference to the baseline for each of the color pre-processing methodologies (texture and hybrid hand-designed descriptors). Color shows details about descriptor, shape about data set. The data are filtered on $f = 4$. The zero line represents the condition where no color pre-processing was applied. The use of color pre-processing caused loss of accuracy in the majority of the cases: the median of 12 of the 17 methodologies was below the baseline, and the upper quartile of 11 of the 17 was close or below the baseline. It should be noted that for both Macenko and Reinhard, the best results were recorded when T1—i.e., the non-histological target—was used. Ruifrok and Johnston’s (decoRJ) was among the highest results, together with the relatively simpler chroma and gw.

Table 4. Feature extraction time (sec/image). The values were recorded on the HICL dataset. Please note that for efficiency reasons, the color-preprocessed images were cached after the first calculation, which was carried out during the extraction of ‘Gabor’ features. Therefore, the figures in the ‘Gabor’ row include both the color pre-processing time and the feature extraction time.

Descriptor	Colour pre-processing method																	AVG	
	none	chroma	gw	heq	Khan (T1)	Khan (T2)	Khan (T3)	Khan (T4)	Macenko (T1)	Macenko (T2)	Macenko (T3)	Macenko (T4)	Reinhard (T1)	Reinhard (T2)	Reinhard (T3)	Reinhard (T4)	decoMC		decoRJ
FullHist	1.09	1.12	1.20	1.24	1.14	1.13	1.00	1.00	1.13	1.10	1.09	0.97	1.01	0.98	0.97	0.93	0.99	0.95	1.06
MargHists	0.29	0.32	0.39	0.39	0.42	0.41	0.42	0.41	0.44	0.41	0.41	0.41	0.41	0.46	0.47	0.40	0.46	0.44	0.41
Gabor	6.49	6.59	6.82	6.55	54.45	55.17	57.44	57.30	7.64	7.60	7.68	7.65	8.19	8.15	8.25	8.13	7.08	7.09	18.24
GLCM	2.23	1.44	2.46	3.00	2.71	2.68	2.72	2.56	2.72	2.66	2.67	2.61	2.82	2.65	2.76	2.43	2.57	2.40	2.56
LBP	1.17	1.34	1.35	1.32	1.32	1.35	1.29	1.29	1.37	1.35	1.34	1.36	1.32	1.32	1.35	1.44	1.58	1.66	1.36
MargGabor	23.86	22.94	20.44	20.23	20.54	20.64	20.56	19.87	19.77	19.79	19.76	19.84	20.15	19.99	19.76	19.81	22.41	21.61	20.66
MargGLCM	7.26	5.12	7.02	7.14	7.90	7.52	7.69	7.21	7.73	7.58	7.63	7.27	7.93	7.43	7.71	6.86	7.27	7.01	7.29
MargLBP	3.54	3.86	3.93	4.08	4.24	4.22	4.39	4.38	4.26	4.34	4.21	4.04	4.16	4.15	3.95	3.93	4.30	4.31	4.13
InceptionV3	7.04	7.25	7.20	6.44	6.00	5.39	5.53	5.36	5.51	5.36	5.36	5.32	5.57	5.37	5.84	5.51	5.70	6.06	5.88
ResNet50	0.61	0.65	0.63	0.62	0.62	0.64	0.67	0.66	0.66	0.68	0.66	0.66	0.66	0.68	0.65	0.64	0.66	0.65	0.65
ResNet101	1.01	1.10	1.09	1.06	1.06	1.08	1.11	1.21	1.32	1.18	1.18	1.08	1.07	1.06	1.05	1.05	1.07	1.06	1.10
Vgg16	0.81	0.87	0.87	0.85	0.86	0.88	0.81	0.85	0.82	0.82	0.82	0.85	0.85	0.87	0.85	0.83	0.81	0.83	0.84
Vgg19	0.98	1.00	0.96	0.96	0.97	0.98	0.97	1.05	1.12	1.16	1.17	1.31	1.42	1.47	1.47	1.38	1.19	1.24	1.16
AVG	4.34	4.12	4.18	4.14	7.86	7.85	8.05	7.93	4.19	4.15	4.15	4.10	4.28	4.20	4.24	4.10	4.31	4.25	5.03

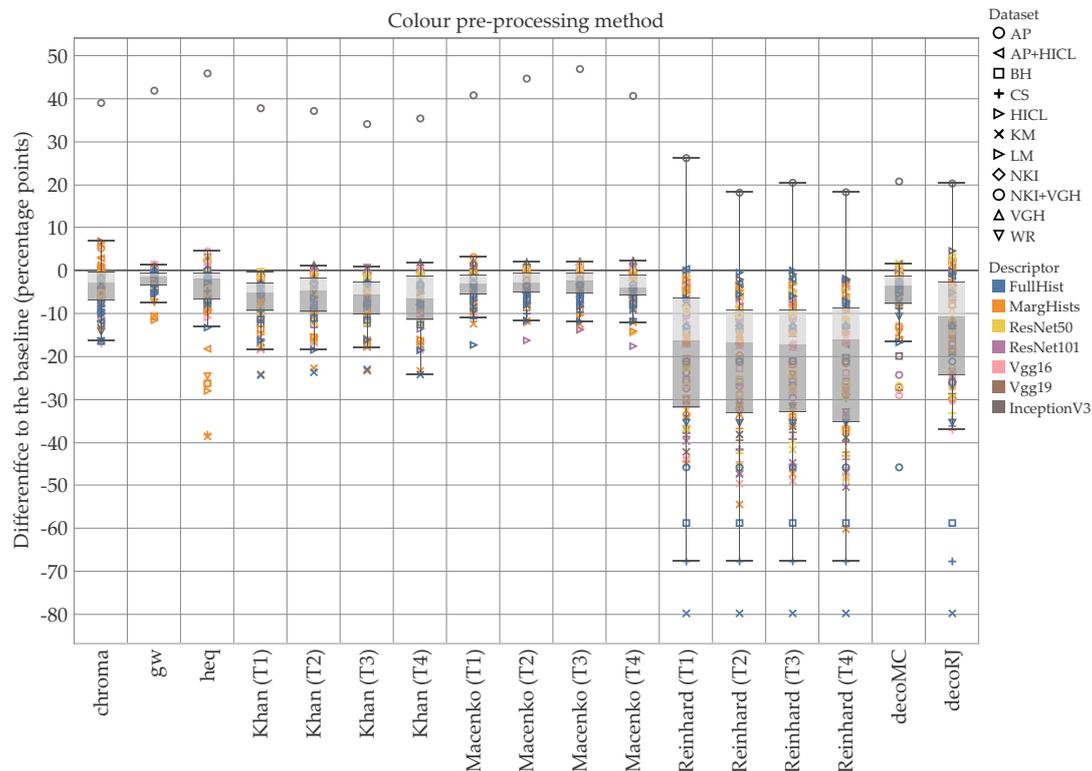


Figure 8. Difference to the baseline for each color pre-processing method (color hand-designed descriptors and convolutional networks). Color shows details about descriptor, shape about data set. The data are filtered on $f = 4$. The zero line represents the condition where no color pre-processing was applied. It should be noticed the considerable decrease of accuracy of Reinhard’s methodology, irrespective of the target image. This is due to the reliance of the methodology on color. On the other hand, results for T1 were slightly higher than T2–T4 for both Reinhard and Macenko. This is surprising as T1 is not histological and the colors are considerably different from the images to normalize. Note the presence of outliers both above and below the zero line, respectively InceptionV3/AP and FullHist/KM.

6. Conclusions

Digital pathology is a rapidly developing discipline with important implications, for instance, the management of those patients who present neoplastic disorders. Potential applications include disease classification, identification of blood vessels, mitosis detection and tissue segmentation. Crucial to all them is the classification of tissue areas into homogeneous and clinically significant regions. As a result of immuno-staining, color plays a significant role in this process, for it enables the differential visualization of tissue micro-structures such as nuclei, ribosome and cytoplasm. However, variations in tissue preparation, reagents, image acquisition settings and other factors can easily lead to significant differences between whole-slide images. To circumvent these problems several pre-processing methodologies have been investigated. Although such procedures can produce appealing results on a qualitative basis, their effects on automatic patch-based classification of histological slides are not clear.

In this work we found that color pre-processing resulted in a noticeable reduction of the accuracy in most cases, especially when coupled with image descriptors that rely heavily on the color of the image. This agrees with the results presented in [34], but differ from those appeared in [33]. In [35,36] the authors achieved the top performance without the use of color normalization, which is again consistent with the results found here. Our findings also conform with those reported by Cusano et

al. [55] for the recognition of color textures under variable lighting conditions—a problem conceptually equivalent to the one investigated here. Interestingly, some pre-processing methods (i.e., chroma and decoRJ) provided positive effects when joined with certain texture descriptors, i.e., MargGabor, MargGLCM and MargLBP. We consider that this is a novel finding that could pave the way to new investigations in future studies.

We speculate that the most interesting new investigations would be those that follow the impact of color pre-processing and pass the classification stage towards the correlation with clinical outcome. Currently, there are several reports that correlate clinical outcome with bio-markers derived from histological images [50,51,75–78], and while these studies provide promising results, it would be interesting to test if these could be affected by color pre-processing.

In conclusion, the results suggest that the application of color pre-processing methodologies for patch-based classification of H&E-stained images should be considered with care. Although our results show some dependence on the dataset used, on the whole our findings indicate that in the absence of enough data for domain-specific tuning, (1) doing nothing (no color pre-processing) is likely to be a good option in most cases (*primum non nocere*) and (2) pre-trained CNN from the ResNet family are the descriptor of choice. Otherwise, if there are enough data enough to carry out some domain-specific tuning, we recommend the color pre-processing method(s) be always evaluated along with the image descriptor(s) used.

Author Contributions: Conceptualization, F.B., C.C.R.-A. and J.N.K.; Data curation, F.B. and J.N.K.; Software, F.B.; Supervision, C.C.R.-A.; Validation, F.B., C.C.R.-A. and J.N.K.; Writing—original draft, F.B.; Writing—review & editing, F.B. and C.C.R.-A. All authors have read and agreed to the published version of the manuscript.

Funding: F.B. was partially supported by the Department of Engineering, Università degli Studi di Perugia, Italy, through the project ‘Shape, color and texture features for the analysis of two- and three-dimensional images: methods and applications’ (Fundamental Research Grants Scheme 2019).

Acknowledgments: Most charts and tables presented in the paper were generated using [Tableau Desktop](#), Professional Edition. The authors wish to thank Tableau Software LLC, CA, USA, for providing a (free) license of the tool for research purposes.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network(s)
GLCM	Grey-level Co-occurrence Matrices
H&E	Hematoxylin & Eosin
LBP	Local Binary Patterns
TCGA	The Cancer Genome Atlas
TMA	Tissue Micro-array(s)

References

1. Madabhushi, A. Digital pathology image analysis: opportunities and challenges. *Imaging Med.* **2009**, *1*, 7–10. doi:10.2217/IIM.09.9.
2. Słodkowska, J.; García-Rojo, M. Digital pathology in personalized cancer therapy. *Stud. Health Technol. Inform.* **2012**, *179*, 143–154.
3. Siregar, P.; Julen, N.; Hufnagl, P.; Mutter, G.L. Computational morphogenesis - Embryogenesis, cancer research and digital pathology. *Bio Syst.* **2018**, *169–170*, 40–54. doi:10.1016/j.biosystems.2018.05.006.
4. Williams, B.J.; Lee, J.; Oien, K.A.; Treanor, D. Digital pathology access and usage in the UK: results from a national survey on behalf of the National Cancer Research Institute’s CM-Path initiative. *J. Clin. Pathol.* **2018**, *71*, 463–466. doi:10.1136/jclinpath-2017-204808.
5. Parwani, A.V. Digital pathology enhances cancer diagnostics. *MLO Med Lab. Obs.* **2017**, *49*, 25.

6. Kwak, J.T.; Hewitt, S.M. Multiview boosting digital pathology analysis of prostate cancer. *Comput. Methods Programs Biomed.* **2017**, *142*, 91–99. doi:10.1016/j.cmpb.2017.02.023.
7. Heindl, A.; Nawaz, S.; Yuan, Y. Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology. *Lab. Investig.* **2015**, *95*, 377–384. doi:10.1038/labinvest.2014.155.
8. Pell, R.; Oien, K.; Robinson, M.; Pitman, H.; Rajpoot, N.; Rittscher, J.; Snead, D.; Verrill, C.; UK National Cancer Research Institute (NCRI) Cellular-Molecular Pathology (CM-Path) Quality Assurance Working Group's. The use of digital pathology and image analysis in clinical trials. *J. Pathol. Clin. Res.* **2019**, *5*, 81–90. doi:10.1002/cjp2.127.
9. Bera, K.; Schalper, K.A.; Rimm, D.L.; Velcheti, V.; Madabhushi, A. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.* **2019**, *16*, 703–715. doi:10.1038/s41571-019-0252-y.
10. Griffin, J.; Treanor, D. Digital pathology in clinical use: Where are we now and what is holding us back? *Histopathology* **2017**, *70*, 134–145.
11. Lutsyk, M.; Ben-Izhak, O.; Sabo, E. novel computerized method of pattern recognition of microscopic images in pathology for differentiating between malignant and benign lesions of the colon. *Anal. Quant. Cytopathol. Histopathol.* **2016**, *38*, 270–276.
12. Sudharshan, P.; Petitjean, C.; Spanhol, F.; Oliveira, L.; Heutte, L.; Honeine, P. Multiple instance learning for histopathological breast cancer image classification. *Expert Syst. Appl.* **2019**, *117*, 103–111.
13. Yao, H.; Zhang, X.; Zhou, X.; Liu, S. Parallel structure deep neural network using CNN and RNN with an attention mechanism for breast cancer histology image classification. *Cancers* **2019**, *11*, 1901.
14. Yari, Y.; Nguyen, T.V.; Nguyen, H.T. Deep learning applied for histological diagnosis of breast cancer. *IEEE Access* **2020**, *8*, 162432–162448.
15. Sparks, R.; Madabhushi, A. Statistical shape model for manifold regularization: Gleason grading of prostate histology. *Comput. Vis. Image Underst.* **2013**, *117*, 1138–1146.
16. Dimitropoulos, K.; Barmpoutis, P.; Zioga, C.; Kamas, A.; Patsiaoura, K.; Grammalidis, N. Grading of invasive breast carcinoma through Grassmannian VLAD encoding. *PLoS ONE* **2017**, *12*, e0185110.
17. Jørgensen, A.; Emborg, J.; Røge, R.; Østergaard, L. Exploiting Multiple Color Representations to Improve Colon Cancer Detection in Whole Slide H&E Stains. In Proceedings of the 1st International Workshop on Computational Pathology (COMPAY), Granada, Spain, 16–20 September 2018; Volume 11039, pp. 61–68.
18. Saxena, S.; Gyanchandani, M. Machine Learning Methods for Computer-Aided Breast Cancer Diagnosis Using Histopathology: A Narrative Review. *J. Med. Imaging Radiat. Sci.* **2020**, *51*, 182–193.
19. Martino, F.; Varricchio, S.; Russo, D.; Merolla, F.; Ilardi, G.; Mascolo, M.; dell'Aversana, G.; Califano, L.; Toscano, G.; De Pietro, G.; et al. A Machine-learning Approach for the Assessment of the Proliferative Compartment of Solid Tumors on Hematoxylin-Eosin-Stained Sections. *Cancers* **2020**, *12*, 1344.
20. Linder, N.; Konsti, J.; Turkki, R.; Rahtu, E.; Lundin, M.; Nordling, S.; Haglund, C.; Ahonen, T.; Pietikäinen, M.; Lundin, J. Identification of tumor epithelium and stroma in tissue microarrays using texture analysis. *Diagn. Pathol.* **2012**, *7*, 1–11.
21. Kather, J.; Weis, C.A.; Bianconi, F.; Melchers, S.; Schad, L.; Gaiser, T.; Marx, A.; Zöllner, F. Multi-class texture analysis in colorectal cancer histology. *Sci. Rep.* **2016**, *6*, 27988.
22. Khan, A.; Rajpoot, N.; Treanor, D.; Magee, D. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 1729–1738.
23. Komura, D.; Ishikawa, S. Machine Learning Methods for Histopathological Image Analysis. *Comput. Struct. Biotechnol. J.* **2018**, *16*, 34–42.
24. Clarke, E.L.; Treanor, D. Colour in digital pathology: a review. *Histopathology* **2017**, *70*, 153–163, doi:10.1111/his.13079.
25. Macenko, M.; Niethammer, M.; marron, J.; Borland, D.; Woosley, J.; Guan, X.; Schmitt, C.; Thomas, N. A method for normalizing histology slides for quantitative analysis. In Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI), Boston, MA, USA, 28 June–1 July 2009; pp. 1107–1110.
26. Ruifrok, A.; Johnston, D. Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.* **2001**, *23*, 291–299.

27. Zanjani, F.G.; Zinger, S.; Bejnordi, B.E.; van der Laak, J.A.W.M.; de With, P.H.N. Stain normalization of histopathology images using generative adversarial networks. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 573–577. doi:10.1109/ISBI.2018.8363641.
28. Shaban, M.T.; Baur, C.; Navab, N.; Albarqouni, S. Staingan: Stain Style Transfer for Digital Histological Images. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 953–956. doi:10.1109/ISBI.2019.8759152.
29. Roy, S.; Kumar Jain, A.; Lal, S.; Kini, J. A study about color normalization methods for histopathology images. *Micron* **2018**, *114*, 42–61. doi:10.1016/j.micron.2018.07.005.
30. Li, X.; Plataniotis, K.N. Circular Mixture Modeling of Color Distribution for Blind Stain Separation in Pathology Images. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 150–161. doi:10.1109/JBHI.2015.2503720.
31. Janowczyk, A.; Basavanthally, A.; Madabhushi, A. Stain Normalization using Sparse AutoEncoders (StaNoSA): Application to digital pathology. *Comput. Med Imaging Graph. Off. J. Comput. Med Imaging Soc.* **2017**, *57*, 50–61. doi:10.1016/j.compmedimag.2016.05.003.
32. Sethi, A.; Sha, L.; Vahadane, A.; Deaton, R.; Kumar, N.; MacIas, V.; Gann, P. Empirical comparison of color normalization methods for epithelial-stromal classification in H and E images. *J. Pathol. Inform.* **2016**, *7*, 17.
33. 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. In Proceedings of the IEEE International Symposium on Biomedical Imaging, Melbourne, Australia, 18–21 April 2017.
34. Gadermayr, M.; Cooper, S.; Klinkhammer, B.; Boor, P.; Merhof, D. A quantitative assessment of image normalization for classifying histopathological tissue of the kidney. In Proceedings of the 39th German Conference on Pattern Recognition (GCPR), Basel, Switzerland, 12–15 September 2017; Volume 10496, pp. 3–13.
35. Hameed, Z.; Zahia, S.; Garcia-Zapirain, B.; Aguirre, J.; Vanegas, A. Breast cancer histopathology image classification using an ensemble of deep learning models. *Sensors* **2020**, *20*, 4373.
36. Tellez, D.; Litjens, G.; Bándi, P.; Bulten, W.; Bokhorst, J.M.; Ciompi, F.; van der Laak, J. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Med Image Anal.* **2019**, *58*, 101544.
37. 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* **2016**, *35*, 1962–1971. doi:10.1109/TMI.2016.2529665.
38. Bianconi, F.; Kather, J.; Reyes-Aldasoro, C. Evaluation of colour pre-processing on patch-based classification of H&E-stained images. In Proceedings of the 15th European Congress on Digital Pathology, ECDP 2019, Warwick, UK, 10–13 April 2019; Volume 11435, pp. 56–64.
39. Spanhol, F.; Oliveira, L.; Petitjean, C.; Heutte, L. A Dataset for Breast Cancer Histopathological Image Classification. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 1455–1462.
40. Gertych, A.; Ing, N.; Ma, Z.; Fuchs, T.; Salman, S.; Mohanty, S.; Bhele, S.; Velásquez-Vacca, A.; Amin, M.; Knudsen, B. Machine learning approaches to analyze histological images of tissues from radical prostatectomies. *Comput. Med Imaging Graph.* **2015**, *46*, 197–208.
41. Kostopoulos, S.; Glotsos, D.; Cavouras, D.; Daskalakis, A.; Kalatzis, I.; Georgiadis, P.; Bougioukos, P.; Ravazoula, P.; Nikiforidis, G. Computer-based association of the texture of expressed estrogen receptor nuclei with histologic grade using immunohistochemically-stained breast carcinomas. *Anal. Quant. Cytol. Histol.* **2009**, *31*, 187–196.
42. Kather, J.N.; Zöllner, F.G.; Bianconi, F.; Melchers, S.M.; Schad, L.R.; Gaiser, T.; Marx, A.; Weis, C.A. Collection of Textures in Colorectal Cancer Histology. 2016. Version 1.0. Available online: <https://zenodo.org/record/53169> (accessed on 6 November 2018).
43. Shamir, L.; Orlov, N.; Mark Eckley, D.; Macura, T.; Goldberg, I. IICBU 2008: A proposed benchmark suite for biological image analysis. *Med. Biol. Eng. Comput.* **2008**, *46*, 943–947.
44. National Institute on Aging. Lymphoma. 2008. Available online: <https://ome.grc.nia.nih.gov/iicbu2008/lymphoma/index.html> (accessed on 6 November 2018).

45. Beck, A.; Sangoi, A.; Leung, S.; marinelli, R.; Nielsen, T.; Van De Vijver, M.; West, R.; Van De Rijn, M.; Koller, D. Imaging: Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci. Transl. Med.* **2011**, *3*, 108ra113.
46. Beck, A.; Sangoi, A.; Leung, S.; marinelli, R.; Nielsen, T.; Van De Vijver, M.; West, R.; Van De Rijn, M.; Koller, D. Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival: Supplementary Documents. 2011. Available online: https://tma.im/tma_portal/C-Path/supp.html (accessed on 7 November 2018).
47. Warwick-QU. 2015. Available online: <https://warwick.ac.uk/fac/sci/dcs/research/tia/glascontest/download/> (accessed on 7 November 2018).
48. Sirinukunwattana, K.; Snead, D.; Rajpoot, N. A Stochastic Polygons Model for Glandular Structures in Colon Histology Images. *IEEE Trans. Med. Imaging* **2015**, *34*, 2366–2378.
49. Sirinukunwattana, K.; Pluim, J.; Chen, H.; Qi, X.; Heng, P.A.; Guo, Y.; Wang, L.; Matuszewski, B.; Bruni, E.; Sanchez, U.; et al. Gland segmentation in colon histology images: The GlaS challenge contest. *Med. Image Anal.* **2017**, *35*, 489–502.
50. Kather, J.N.; Krisam, J.; Charoentong, P.; Luedde, T.; Herpel, E.; Weis, C.A.; Gaiser, T.; Marx, A.; Valous, N.A.; Ferber, D.; et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS Med.* **2019**, *16*, e1002730. doi:10.1371/journal.pmed.1002730.
51. Skrede, O.J.; Raedt, S.D.; Kleppe, A.; Hveem, T.S.; Liestøl, K.; Maddison, J.; Askautrud, H.A.; Pradhan, M.; Nesheim, J.A.; Albrechtsen, F.; et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet* **2020**, *395*, 350–360. doi:10.1016/S0140-6736(19)32998-8.
52. Haub, P.; Meckel, T. A model based survey of colour deconvolution in diagnostic brightfield microscopy: Error estimation and spectral consideration. *Sci. Rep.* **2015**, *5*, 12096.
53. Stain Normalisation Toolbox. Available online: <https://warwick.ac.uk/fac/sci/dcs/research/tia/software/sntoolbox/> (accessed on 8 November 2018).
54. Foster, D. Color constancy. *Vis. Res.* **2011**, *51*, 674–700.
55. Cusano, C.; Napoletano, P.; Schettini, R. Evaluating color texture descriptors under large variations of controlled lighting conditions. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **2016**, *33*, 17–30.
56. Reinhard, E.; Ashikhmin, M.; Gooch, B.; Shirley, P. Color transfer between images. *IEEE Comput. Graph. Appl.* **2001**, *21*, 34–41.
57. Cernadas, E.; Fernández-Delgado, M.; González-Rufino, E.; Carrión, P. Influence of normalization and color space to color texture classification. *Pattern Recognit.* **2017**, *61*, 120–138.
58. Finlayson, G.; Schaefer, G. Colour indexing across devices and viewing conditions. In Proceedings of the 2nd International Workshop on Content-Based Multimedia Indexing, Florence, Italy, 19–21 June 2017; pp. 215–221.
59. van de Weijer, J. Color in Computer Vision. Available online: <http://lear.inrialpes.fr/people/vandeweijer/research.html> (accessed on 1 October 2019).
60. Napoletano, P. Hand-Crafted vs Learned Descriptors for Color Texture Classification. In Proceedings of the 6th Computational Color Imaging Workshop (CCIW'17), Milan, Italy, 29–31 March 2017; Volume 10213, pp. 259–271.
61. González, E.; Bianconi, F.; Álvarez, M.; Saetta, S. Automatic characterization of the visual appearance of industrial materials through colour and texture analysis: An overview of methods and applications. *Adv. Opt. Technol.* **2013**, *2013*, 503541.
62. Swain, M.; Ballard, D. Color indexing. *Int. J. Comput. Vis.* **1991**, *7*, 11–32.
63. Pietikainen, M.; Nieminen, S.; marszalec, E.; Ojala, T. Accurate color discrimination with classification based on feature distributions. In Proceedings of the International Conference on Pattern Recognition (ICPR), Vienna, Austria, 25–29 August 1996; Volume 3, pp. 833–838.
64. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *3*, 610–621.
65. Bianconi, F.; Fernández, A. Rotation invariant co-occurrence features based on digital circles and discrete Fourier transform. *Pattern Recognit. Lett.* **2014**, *48*, 34–41.
66. Lahajnar, F.; Kovačič, S. Rotation-invariant texture classification. *Pattern Recognit. Lett.* **2003**, *24*, 706–719.
67. Ojala, T.; Pietikäinen, M.; Mäenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987.

68. Bianconi, F.; Bello-Cerezo, R.; Napoletano, P. Improved opponent color local binary patterns: an effective local image descriptor for color texture classification. *J. Electron. Imaging* **2018**, *27*, 011002.
69. Bello-Cerezo, R.; Bianconi, F.; Di maria, F.; Napoletano, P.; Smeraldi, F. Comparative Evaluation of Hand-Crafted Image Descriptors vs. Off-the-Shelf CNN-Based Features for Colour Texture Classification under Ideal and Realistic Conditions. *Appl. Sci.* **2019**, *9*, 738.
70. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
71. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
72. Chatfield, K.; Simonyan, K.; Vedaldi, A.; Zisserman, A. Return of the devil in the details: Delving deep into convolutional nets. In Proceedings of the British Machine Vision Conference, Nottingham, UK, 1–5 September 2014.
73. Bianconi, F. CATAcOMB: Colour And Texture Analysis Toolbox for MatlaB. Available online: <https://bitbucket.org/biancovic/catacomb/src/master/> (accessed on 21 September 2020).
74. Vedaldi, A.; Lenc, K. MatConvNet: Convolutional neural networks for MATLAB. In Proceedings of the 23rd ACM International Conference on Multimedia (MM 2015), Brisbane, Australia, 26–30 October 2015; pp. 689–692.
75. Shaban, M.; Khurram, S.A.; Fraz, M.M.; Alsubaie, N.; Masood, I.; Mushtaq, S.; Hassan, M.; Loya, A.; Rajpoot, N.M. A Novel Digital Score for Abundance of Tumour Infiltrating Lymphocytes Predicts Disease Free Survival in Oral Squamous Cell Carcinoma. *Sci. Rep.* **2019**, *9*, 13341. doi:10.1038/s41598-019-49710-z.
76. Mobadersany, P.; Yousefi, S.; Amgad, M.; Gutman, D.A.; Barnholtz-Sloan, J.S.; Vega, J.E.V.; Brat, D.J.; Cooper, L.A.D. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E2970–E2979. doi:10.1073/pnas.1717139115.
77. Bychkov, D.; Linder, N.; Turkki, R.; Nordling, S.; Kovanen, P.E.; Verrill, C.; Walliander, M.; Lundin, M.; Haglund, C.; Lundin, J. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci. Rep.* **2018**, *8*, 3395. doi:10.1038/s41598-018-21758-3.
78. Jiang, D.; Liao, J.; Duan, H.; Wu, Q.; Owen, G.; Shu, C.; Chen, L.; He, Y.; Wu, Z.; He, D.; et al. A machine learning-based prognostic predictor for stage III colon cancer. *Sci. Rep.* **2020**, *10*, 10333. doi:10.1038/s41598-020-67178-0.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).