Final Report

Comparative analysis of different color models for color Normalization

Prof. Pradipta Maji

Indian Statistical Institute, Baranagar, Kolkata, West Bengal 700108.

Satyam Bijendra Dubey

Vivekanand Education Society's Institute of Technology, Chembur, Maharashtra 400074.

 (Github link)

ABSTRACT

Different types of cancer can be diagnosed with the analysis of histological samples stained with hematoxylin-eosin (H&E). Through this stain, it is possible to identify the architecture of tissue components and analyze cellular morphological aspects that are essential for cancer diagnosis. One of the foremost and challenging tasks in hematoxylin and eosin-stained histological image analysis is to reduce color variation present among images, which may significantly affect the performance of computer-aided histological image analysis. For automated image analysis, these H&E stained images need to be normalized. This is because of the significant variation in image colors arising from sample preparation and imaging conditions.

In this regard, I have worked on implementing Mark Mecenko's color normalization technique. And also worked out to find which color model among RGB, HSI, and Lαβ is most effective for color normalization in the standard UCSB Dataset using a Normalized Median Intensity index for quantitative estimation of color forms.

Keywords: Histological image analysis, Color Normalization, Quantitative Analysis, Color Models.

ABBREVIATIONS

RGB	Red, Green, Blue
HSI	Hue, Saturation, Intensity
CMY	Cyan, Magenta, Yellow
O.D	Optical Density
SVD	Singular Value Decomposition
NMI	Normalized Median Intensity
ROI	Region of Interest

1. Introduction

1.1 Background

Breast Cancer is a very common type of cancer among people which causes thousands of lives. In a survey in 2018, 266,120 new cases of invasive breast cancer were diagnosed in women in the U.S, along with 63,960 new cases of non-invasive breast cancer. About 2550 new cases of invasive breast cancer were diagnosed in men. With advancements in technology, there is an increase in diagnostic tools that help detect cancer. Some standard diagnostic tools include laboratory tests, Biopsy, and Imaging tests such as X-ray, PET/CT, MRI, etc. A biopsy is a procedure in which the doctor removes a sample of tissue and then examines it to see if cancer is present. Imaging procedures create pictures of areas inside the human body that helps the doctor see whether a tumor is present whereas in most cases, doctors need to do a biopsy to make a diagnosis of cancer. Cancer is defined as a tissue mass composed of genetically modified cells that do not respond to regulatory mechanisms of cellular growth. Currently, it is considered one of the main causes of death in the world due to late diagnoses. To confirm the diagnosis of cancer, a microscopic analysis of cells and tissue components of a stained cancer sample has to be performed by a pathologist. Hematoxylin and Eosin (H&E) stains are one of the primary tissue stains for histology. This is because H stain makes nuclei easily visible in blue against a pink background of cytoplasm (and other issue regions). This enables a pathologist to easily identify and evaluate the tissue.

One of the foremost and challenging tasks in H&E stained histological image analysis is to reduce the color variation present among images, which may remarkably affect the performance of the digital histological image analysis. For automated image analysis, these H&E stained images need to be normalized. This is because of the significant variation in image colors arising from both sample preparation and imaging conditions.

Stain normalization is an important processing task for computer-aided diagnosis (CAD) systems in modern digital pathology. This task reduces the color and

intensity variations present in stained images from different laboratories. Consequently, stain normalization typically increases the prediction accuracy of CAD systems. The most important task in the normalization techniques is altering the image for better visualization where the color of one image is transferred to another i.e. the color of the target image is transferred onto the source image. If the image is processed without preprocessing, this might result in an incorrect diagnosis. There are various normalization algorithms such as Histogram Specification, Reinhard Method, Macenko Method, Stain Color Descriptor (SCD), Complete Color Normalization, Structure Preserving Color Normalization (SPCN), and many other methods.

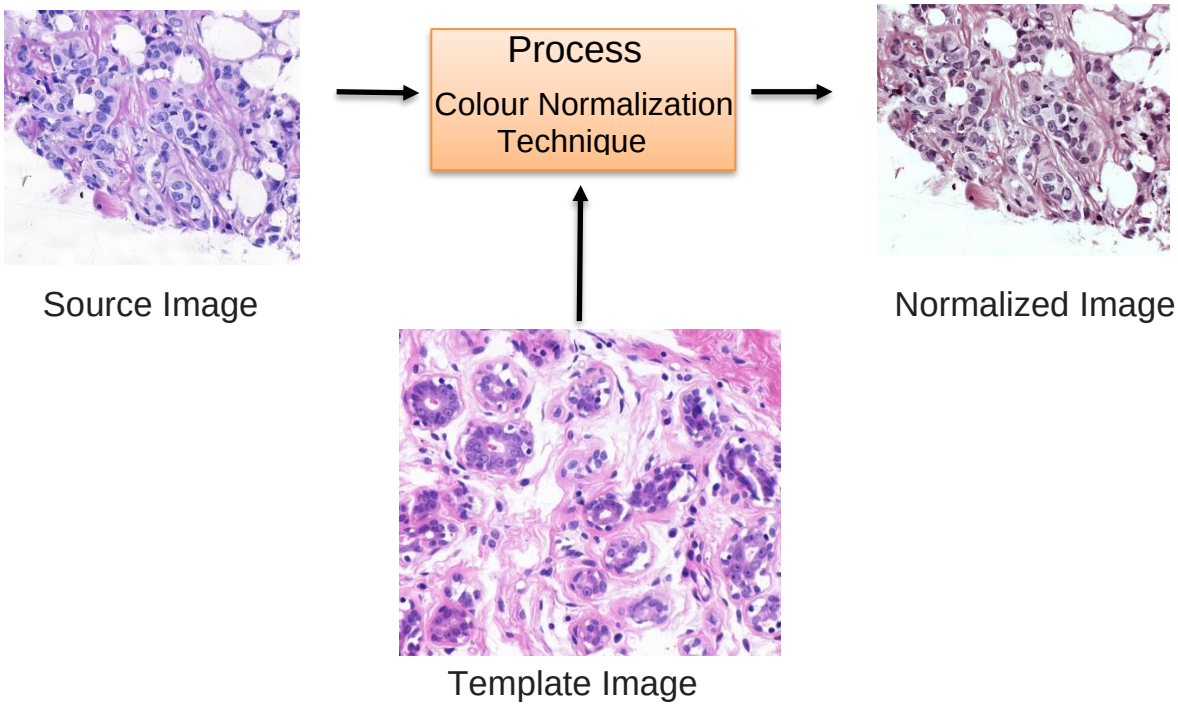


Fig. 1. Glimpse of Normalization Process

1.2 Topic of work in Brief

The Standard UCSB Data-set of histological images comprises 32 non-cancerous benign cells and 26 cancerous malignant cell images, acquired from ten hematoxylins and eosin (H&E) stained breast cancer biopsy sets. Each image has a resolution of 768×896, and associated ground-truth with nuclei considered as ROIs. So my task was to implement color normalization on this image data set concerning a target image in the ROI, in 3 different color spaces, RGB, HSI, L$\propto \beta$, and then using a quantitative index to compare these color normal forms to conclude which among the following color normal form is fruitful for color normalization.

2. Literature Review

Gonzales[1] very well explained the fundamentals of image processing and different color models like the RGB, HSI, CMY, CMYK, and Lαß and their interconversion techniques through which it was possible to easily write down my own functional code in python to convert one image format to another. It has also very well explained different image processing methods such as histogram processing, smoothing and sharpening, and color image segmentation, which has enhanced my understanding and interest in image analysis.

Color variation in histopathology images is due to the use of different scanners; during slide preparation, different equipment is used, different stain coloring, and activity from different manufacturers and batches of stains. One of the ways to reduce color variation in histopathology images is that convert RGB images into grayscale, but in grayscale, some of the information is lost. So, we have to go with color normalization techniques. Color normalization is the process where we do the mean color transformation from one image to another image. There are various algorithms for the color normalization of histopathology images like histogram specification, Reinhard method, Macenko method, stain color descriptor (SCD), complete color normalization, and structure-preserving color normalization (SPCN), and other recent color normalization methods are explored in the literature.

Reinhard et al.,[3] proposed a technique that is based on color transfer between a standard image and color varied image using the mean and variance of both the images.

In this method, the source image is altered to the target image. Here, the color distribution of the source image to that of the target image by linear transform in perceptual color space. The contrast of the source image is approximately the same as that of the target image. In this technique, the image is transformed into a lαβ space in which stains cannot be properly separated.

Macenko et al.,[4] presented an algorithm to find particular stain vectors for each image based on the colors that are present in the image. In this method, there is a specific stain vector corresponding to each of the two stains in the image and this is a fully automatic method that is suitable for analyzing multiple slides rapidly because of having very few parameters and no optimizations required. This method is performed on twelve different slides with considerable variation. For each stain, calculate the intensity histograms for all pixels that have a majority of that particular stain and estimate the 99th percentile of these intensity values for robust approximation. It gives better results for less stained images.

Vahadane et al., [8] presented a method used for color normalization is the image is decomposed into a stain density map that is sparse and non-negative. Then, the stain density maps are combined based on the stain color of a pathologist's preferred target image. Thus, altering only its color and preserving the structure. In SNMF, solution space is reduced. The structure of the source image is preserved. The computation complexity for reducing the solution space is higher. The solution provided for the optimization problem may lead to the estimation of local minima rather than global minima.

The NMI index by Basavanhally et al., [5] is the segmentation-based assessment of color consistency of the segmented region of the data-set. Using this we can find the best color normal form in the data set which eventually will improve the segmentation, and feature selection in the computer-aided diagnostics. Though, normalized median intensity (NMI) evaluates the color consistency of a specific ROI within an image, it does not capture the color consistency of the ROI among images within the same biopsy set. In order to address the above problem, a new quantitative index is introduced next to evaluate the performance of different color normalization methods by Maji et al., [6], between-image color constancy (BiCC) index. If the between-image color constancy is maintained after color normalization, it is expected that the median and maximum average intensity values, corresponding to ROIs of images I and J, would exhibit a close proximity. So, a good color normalization algorithm should make the value of BiCC as high as possible.

3. Methodology

3.1 Color Normalization

In Macenko's method [4], based on the colors that are present an algorithm to find stain vectors for each image is presented. A pixel with an Optical Density (OD) value of 0 represents that there was no light absorbed. Such pixels were removed for stability reasons. An adaptive mechanism was employed and the threshold value of $\beta = 0.15$ was found to provide the best results. To find the endpoints that correspond to the stain vectors, the geodesic direction (the shortest path between two unit-norm color vectors) is found and then the OD transformed pixels are projected onto it. The first step is to calculate the plane that these vectors form. This is done by forming a plane corresponding to the two largest singular values of decomposition. Then the OD transformed pixels are projected onto this plane. Then the angle with respect to the first SVD direction is calculated. With respect to the stain separation process, the stain vector corresponding to the minimum vector represents Haematoxylin stain and the stain vector corresponding to that of the maximum vector represents Eosin stain.

Algorithm: Macenko Method(Input Image, β, Reference Image)

1. **Optical Density of RGB image (I) is calculated OD = –log(I).**
2. **Data with Optical Density intensity less than β is removed. By default β = 0.15 experimentally).**
3. **Singular Value Decomposition on the Optical Density tuples is calculated.**

$$\text{CoVar}(x, y) = \sum (X_i - \mu(X_i))(Y_i - \mu(Y_i))/ n - 1; \quad i \text{ Where } i = 1 \text{ to } n = 6$$

$$A * v = \lambda * v$$

Any value of λ for which this equation has a solution is known as an The eigenvalue of the matrix A, where A is the OD matrix and v is the vector.

$$V = \text{eigen}(\text{Co var}(A))$$

4. **Project on the plane spanned by the eigen vectors corresponding to the two largest eigen values. A*V (sliced with largest eigen values)**

5. **Calculate the angle of each point with respect to the Singular Value Decomposition direction.**

6. **Find robust extremes (α th and (100$-\alpha$) th percentiles) of angle. By default alpha=1.**

   i = (p /100) * n, Where i = position; n = number of input elements.

   minimum angle = percentile (angle, alpha)

   maximum angle = percentile (angle, 100-alpha)

7. **The minimum and maximum vectors are found and projected back to OD space.**

   vectorMin = V*[cosine(minimumangle), sine(minimumangle)]

   vectorMax = V*[cosine(maximumangle), sine(maximumangle)]

8. **Make the vector corresponding to haematoxylin first and the one corresponding to eosin second.**

   if vectorMin(1) > vectorMax
   HE = [vectorMin vectorMax](Haematoxylin)
   else
   HE = [vectorMax vectorMin](Eosin)

9. **Determine concentrations of the individual stains.**

   C = HE \ Y.

   Y- is the matrix columns represent RGB channel, Rows represent OD intensity

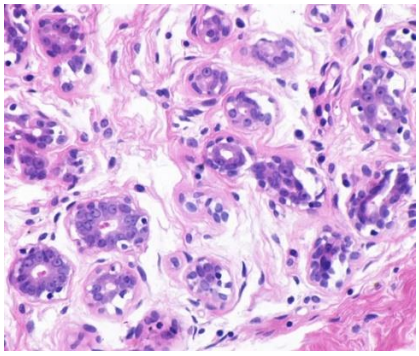
10. **Recreate the image using reference mixing matrix**

    $I_{norm} = I_0 * \exp(-HRef*C)$

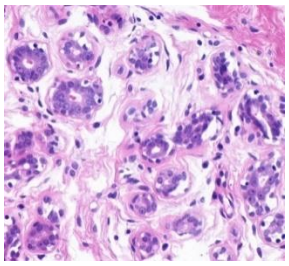
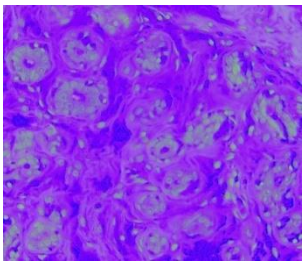
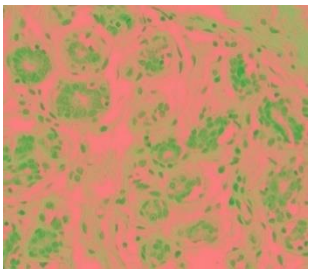
    HERef = reference H&E OD matrix

    I_0 = Transmitted Light Intensity.

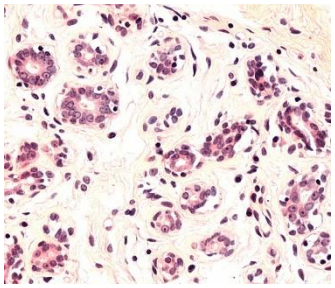
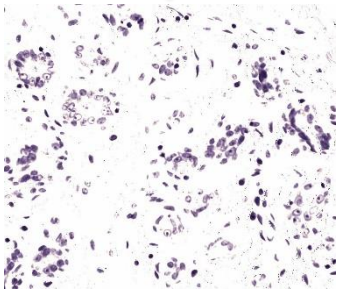
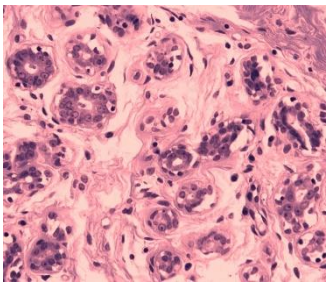
    I_{norm} = Normalized image.



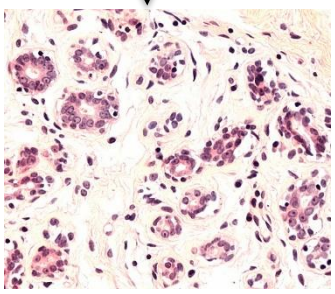
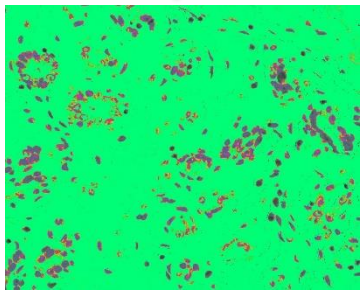
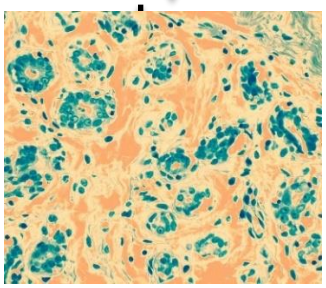
RGB **HSI** **L∝ β**

After Color Normalization

(Converted back to RGB space for NMI Computation)

RGB` HSI` L∝ β`

3.2 Quantitative Evaluation of Color Normal Form

Several metrics are used to determine the utility of color normalization. A common metric used to demonstrate color consistency, or lack thereof, is the normalized median intensity (NMI). Based on Each Normalized RGB image NMI Values are Calculated as:

$$\mathrm{NMI}(I) = \frac{\underset{i \in \mathrm{ROI}(I)}{\mathrm{median}} \{W(i)\}}{\underset{i \in \mathrm{ROI}(I)}{\max} \{W(i)\}}.$$

where W(i) denotes the average of (R, G, B) intensities for the i-th pixel corresponding to the ROI of the image I.

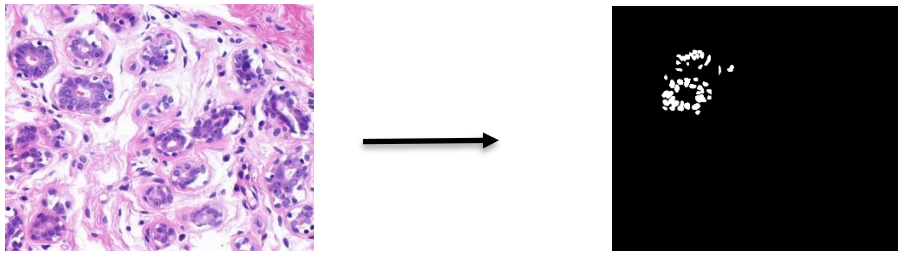


Fig. 2. Ground-Truth (Region of Interest)

So, if the color constancy is maintained after normalization, then it is expected that the median and maximum of the average intensity values, corresponding to a particular ROI of the image I, would be close enough; which indicates a high value of NMI. Also, the value of NMI ranges from 0 to 1. So, a good color normalization algorithm should make the NMI higher of a particular image, belonging to the biopsy set.

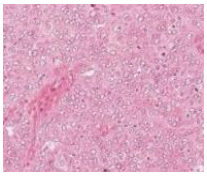
Results

Table I

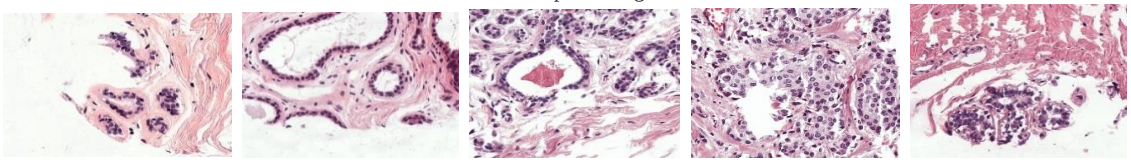
NMI Analysis of Different Color Models After Normalization

Image Name	NMI_RGB	NMI_HSI	NMI_ L$\propto$ β
ytma10_010704_benign1_ccd.tif	0.390196078	0.807843137	0.57254902
ytma10_010704_benign2_ccd.tif	0.607843137	0.745098039	0.443137255
ytma10_010704_benign3_ccd.tif	0.650980392	0.784313725	0.474509804
ytma10_010704_malignant1_ccd.tif	0.556862745	0.737254902	0.529411765
ytma10_010704_malignant2_ccd.tif	0.337254902	0.996078431	0.545098039
ytma10_010704_malignant3_ccd.tif	0.384313725	0.964705882	0.568627451
ytma12_010804_benign1_ccd.tif	0.552941176	0.764705882	0.458823529
ytma12_010804_benign2_ccd.tif	0.603921569	0.776470588	0.48627451
ytma12_010804_benign3_ccd.tif	0.537254902	0.224409449	0.454901961
ytma12_010804_malignant1_ccd.tif	0.584313725	0.976470588	0.552941176
ytma12_010804_malignant2_ccd.tif	0.501960784	0.812865497	0.490196078
ytma12_010804_malignant3_ccd.tif	0.501960784	0.756862745	0.458823529
ytma23_022103_benign1_ccd.tif	0.674509804	0.77254902	0.48627451
ytma23_022103_benign2_ccd.tif	0.631372549	0.768627451	0.498039216
ytma23_022103_benign3_ccd.tif	0.580392157	0.764705882	0.474509804
ytma23_022103_malignant1_ccd.tif	0.701960784	0.776470588	0.525490196
ytma23_022103_malignant2_ccd.tif	0.654901961	0.756862745	0.509803922
ytma23_022103_malignant3_ccd.tif	0.639215686	0.764705882	0.51372549
ytma49_042003_benign1_ccd.tif	0.635294118	0.164705882	0.517647059
ytma49_042003_benign2_ccd.tif	0.615686275	0.545098039	0.501960784
ytma49_042003_benign3_ccd.tif	0.549019608	0.792156863	0.517647059
ytma49_042003_malignant1_ccd.tif	0.466666667	0.97254902	0.549019608
ytma49_042003_malignant2_ccd.tif	0.588235294	0.964705882	0.517647059
ytma49_042003_malignant3_ccd.tif	0.533333333	0.996078431	0.541176471
ytma49_042203_benign1_ccd.tif	0.580392157	0.764705882	0.474509804
ytma49_042203_benign2_ccd.tif	0.607843137	0.062745098	0.443137255
ytma49_042203_benign3_ccd.tif	0.556862745	0.77254902	0.440944882
ytma49_042203_malignant1_ccd.tif	0.643137255	0.764705882	0.521568627
ytma49_042203_malignant2_ccd.tif	0.600000000	0.77254902	0.450592885
ytma49_042203_malignant3_ccd.tif	0.643137255	0.450980392	0.482352941
ytma49_042403_benign1_ccd.tif	0.635294118	0.023529412	0.458823529
ytma49_042403_benign2_ccd.tif	0.650980392	0.756862745	0.517647059
ytma49_042403_benign3_ccd.tif	0.682352941	0.745098039	0.51372549
ytma49_042403_malignant1_ccd.tif	0.592156863	0.741176471	0.498039216
ytma49_042403_malignant2_ccd.tif	0.533333333	0.960784314	0.568627451
ytma49_042403_malignant3_ccd.tif	0.619607843	0.764705882	0.51372549
ytma49_072303_benign1_ccd.tif	0.588235294	0.77254902	0.462745098

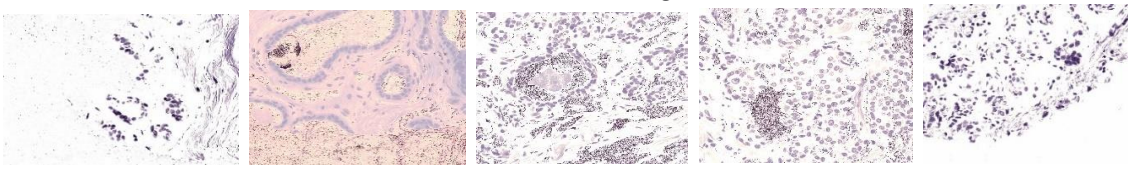
ytma49_072303_benign2_ccd.tif	0.631372549	0.764705882	0.470588235
ytma49_072303_malignant1_ccd.tif	0.709803922	0.749019608	0.541176471
ytma49_072303_malignant2_ccd.tif	0.62745098	0.768627451	0.51372549
ytma49_111003_benign1_ccd.tif	0.564705882	0.431372549	0.447058824
ytma49_111003_benign2_ccd.tif	0.62745098	0.462745098	0.51372549
ytma49_111003_benign3_ccd.tif	0.654901961	0.764705882	0.529411765
ytma49_111003_malignant1_ccd.tif	0.560784314	0.984313725	0.556862745
ytma49_111003_malignant2_ccd.tif	0.57254902	0.745098039	0.505882353
ytma49_111003_malignant3_ccd.tif	0.537254902	0.59437751	0.482352941
ytma49_111303_benign1_ccd.tif	0.654901961	0.239215686	0.517647059
ytma49_111303_benign2_ccd.tif	0.57254902	0.341269841	0.470588235
ytma49_111303_benign3_ccd.tif	0.57254902	0.428000000	0.42745098
ytma49_111303_malignant1_ccd.tif	0.607843137	0.996078431	0.541176471
ytma49_111303_malignant2_ccd.tif	0.623529412	0.577075099	0.517647059
ytma49_111303_malignant3_ccd.tif	0.607843137	0.239215686	0.490196078
ytma55_030603_benign1_ccd.tif	0.639215686	0.823529412	0.552941176
ytma55_030603_benign2_ccd.tif	0.635294118	0.996078431	0.521568627
ytma55_030603_benign3_ccd.tif	0.647058824	0.837837838	0.529411765
ytma55_030603_benign4_ccd.tif	0.674509804	0.996078431	0.537254902
ytma55_030603_benign5_ccd.tif	0.615686275	0.77254902	0.470588235
ytma55_030603_benign6_ccd.tif	0.662745098	0.556862745	0.501960784



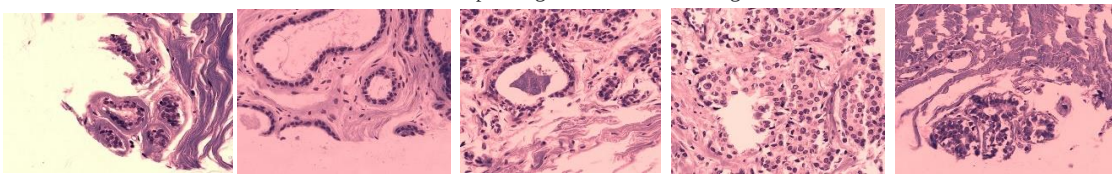
Template Image



RGB Normalized Images



Corresponding HSI Normalized Images



Corresponding L$\propto$ β Normalized Images

Fig. 3. Sample Normalized images in all three color spaces

Table II

	NMI-RGB	NMI-HSI	NMI- L∝ β
Mean	0.593340095	0.703586071	0.502998116
Median	0.607843137	0.764705882	0.511764706
S.D	0.072804652	0.241292907	0.036138956

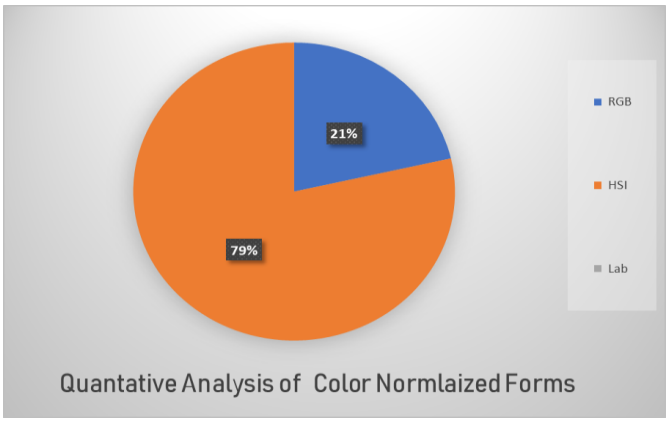


Fig. 4. Pie Chart of Max NMI values in all 3 color forms

Based on the above NMI index values calculated, 44 times the NMI value of HSI space and 14 out of 58 times the value of RGB Normalized space came out to give higher results.

Conclusion

The color normalization in histological images is a fundamental task as the presence of color inconsistency among the images may degrade the performance of automated histological image analysis. In this regard, through the mean and the standard deviation of the NMI values of the 3 color spaces, we can conclude that the HSI normalized images are more likely to give better results for segmentation and feature selection eventually increase the Classification accuracy of the model.

In the future, deep learning-based color normalization techniques can be developed to learn the amount of staining present in the huge voluminous data and to perform the color normalization process. A deep learning-based model does not require the target image for the color normalization process. Effective choice of color normalization technique will helpful for detecting cancerous and noncancerous cells.

ACKNOWLEDGEMENTS

I am grateful to the Indian Academy of Sciences for providing me an opportunity to pursue the Summer Research Fellowship (SRF-2022) under Prof. Pradipta Maji, at the Indian Statistical Institute, Kolkata. It is a pleasure for me to express my heartfelt gratitude to my guide, Prof. Pradipta Maji, at the very outset for his endless support, guidance, training, and teaching throughout my training tenure. I thank him for allowing me to carry out my research work in his Biomedical Imaging and Bioinformatics Laboratory and for introducing me to the world of front-line research in Image Processing. This internship gave me an insight into color image processing and taught me how research work is approached and implemented which has opened many new opportunities for me to pursue research in the near future. I would also like to thank Mr. Suman Mahapatra who assisted and gave timely guidance in my project along with all the members of BIBL for their guidance and support. I also express my sincere thanks to the authorities of the Indian Statistical Institute for the facilities extended to carry out my research. I also thank the office staff for their efforts to help me out in my time of need.

My special thanks to my college Vice Principal, Dr. M. Vijayalakshmi who encouraged me to pursue this research internship. And my parents who kept me motivated through these eight weeks of research Internship.

REFERENCES

[1] R. Gonzales and E. Woods, **Digital Image Processing**.

[2] B. Lakshmanan, S. Anand and T. Jenitha, "Stain Removal Through Color Normalization of Haematoxylin and Eosin Images: A Review" Journal of Physics: Conference Series, Volume 1362, International Conference on Physics and Photonics Processes in Nano Sciences 20–22 June 2019, Eluru, India

[3] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color Transfer Between Images," IEEE Computer Graphics and Applications, vol. 21, no. 5, pp. 34–41, 2001.

[4] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A Method for Normalizing Histology Slides for Quant--itative Analysis," in Proceedings of IEEE International Symposium on Biomedical From Imaging: Nano to Macro, 2009, pp. 1107-1110.

[5] A. Basavanhally and A. Madabhushi, "EM-Based Segmentation-Driven Color Standard--ization of Digitized Histopathology," in Proceedings of SPIE Medical Imaging, 2013, vol. 8676, pp. 86 760G–1–86 760G–12.

[6] P. Maji and S. Mahapatra, "Rough-Fuzzy Circular Clustering for Color Normalization of Histological Images," Fundamenta Informaticae, vol. 164, no. 1, pp. 103–117, 2019.

[7] S. Roy, A. Jain, S. Lal, J. Kini "A study about color normalization methods of Histo-pathology images," Micron,2018.

[8] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images," IEEE Transactions on Medical Imaging, vol. 35, no. 8, pp. 1962–1971, 2016.

[9] P. Maji and S. Mahapatra, Circular Clustering in Fuzzy Approximation Spaces for Color Normalization of Histological Images, IEEE Transactions on Medical Imaging, 39(5), pp. 1735--1745, May 2020.

Sources:

Breast Cancer Cell UCSB Bio-Segmentation Benchmark dataset.