

Robust Image Population Based Stain Color Normalization: How Many Reference Slides Are Enough?

Jose L. Agraz ¹, Member, IEEE, Caleb M. Grenko, Andrew A. Chen ², Angela N. Viaene ³, MacLean D. Nasrallah, Sarthak Pati ⁴, Tahsin Kurc ⁵, Joel Saltz ⁶, Michael D. Feldman ⁷, Hamed Akbari, Parth Sharma ⁸, Russell T. Shinohara ⁹, and Spyridon Bakas ¹⁰

Manuscript received 25 July 2022; revised 8 November 2022, 23 December 2022, and 1 January 2023; accepted 1 January 2023. Date of publication 5 January 2023; date of current version 7 February 2023. This work was supported in part by the National Institutes of Health (NIH) under Grants NIH/NCI:U01CA242871, NIH/NINDS:R01NS042645, NIH/NCI:U24CA189523, and NIH/NCATS:UL1TR001878 and in part by the University of Pennsylvania Institute for Translational Medicine and Therapeutics (ITMAT). The review of this article was arranged by Editor P. Bonato. (Corresponding author: Spyridon Bakas.)

Jose L. Agraz was with the Center for Biomedical Image Computing and Analytics (CBICA), Philadelphia, PA 19139 USA. He is now with the Department of Pathology and Laboratory Medicine, Perelman School of Medicine, Philadelphia, PA 19139 USA, and also with the Department of Radiology at Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19139 USA (e-mail: joseagraz1@gmail.com).

Caleb M. Grenko is with the Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania and the Center for Interdisciplinary Studies, Davidson College, NC 28035 USA (e-mail: cagrenko@davidson.edu).

Andrew A. Chen is with the Penn Statistical Imaging and Visualization Endeavor (PennSIVE), University of Pennsylvania, Philadelphia, PA 19139 USA (e-mail: andrewac@pennmedicine.upenn.edu).

Angela N. Viaene is with the Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA 19139 USA (e-mail: viaenea@email.chop.edu).

MacLean D. Nasrallah and Michael D. Feldman are with the Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19139 USA (e-mail: macleannp@gmail.com; Michael.Feldman2@pennmedicine.upenn.edu).

Sarthak Pati is with the Symphony Health, Philadelphia, PA 19139 USA and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19139 USA, and also with the Department of Radiology at Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19139 USA (e-mail: patis@upenn.edu).

Tahsin Kurc and Joel Saltz are with the Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY 11794-0751 USA (e-mail: tahsin.kurc@stonybrook.edu; joel.saltz@stonybrookmedicine.edu).

Hamed Akbari is with the CBICA and the Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19139 USA (e-mail: hamed.akbari@uphs.upenn.edu).

Parth Sharma is with the Symphony Health, Philadelphia, PA 19139 USA (e-mail: parth.sharma.usa1111@gmail.com).

Russell T. Shinohara is with the CBICA and the Penn Statistical Imaging and Visualization Endeavor (PennSIVE), University of Pennsylvania, Philadelphia, PA 19139 USA (e-mail: rshi@upenn.edu).

Spyridon Bakas is with the CBICA, and the Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19139 USA, and also with the Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19139 USA (e-mail: sbakas@upenn.edu).

This article has supplementary downloadable material available at <https://doi.org/10.1109/OJEMB.2023.3234443>, and a Code Ocean Python capsule at <https://codeocean.com/capsule/3803185>, provided by the author.

Digital Object Identifier 10.1109/OJEMB.2023.3234443

Abstract—Histopathologic evaluation of Hematoxylin & Eosin (H&E) stained slides is essential for disease diagnosis, revealing tissue morphology, structure, and cellular composition. Variations in staining protocols and equipment result in images with color nonconformity. Although pathologists compensate for color variations, these disparities introduce inaccuracies in computational whole slide image (WSI) analysis, accentuating data domain shift and degrading generalization. Current state-of-the-art normalization methods employ a single WSI as reference, but selecting a single WSI representative of a complete WSI-cohort is infeasible, inadvertently introducing normalization bias. We seek the optimal number of slides to construct a more representative reference based on composite/aggregate of multiple H&E density histograms and stain-vectors, obtained from a randomly selected WSI population (WSI-Cohort-Subset). We utilized 1,864 IvyGAP WSIs as a WSI-cohort, and built 200 WSI-Cohort-Subsets varying in size (from 1 to 200 WSI-pairs) using randomly selected WSIs. The WSI-pairs' mean Wasserstein Distances and WSI-Cohort-Subsets' standard deviations were calculated. The Pareto Principle defined the optimal WSI-Cohort-Subset size. The WSI-cohort underwent structure-preserving color normalization using the optimal WSI-Cohort-Subset histogram and stain-vector aggregates. Numerous normalization permutations support WSI-Cohort-Subset aggregates as representative of a WSI-cohort through WSI-cohort CIELAB color space swift convergence, as a result of the law of large numbers and shown as a power law distribution. We show normalization at the optimal (Pareto Principle) WSI-Cohort-Subset size and corresponding CIELAB convergence: a) Quantitatively, using 500 WSI-cohorts; b) Quantitatively, using 8,100 WSI-regions; c) Qualitatively, using 30 cellular tumor normalization permutations. Aggregate-based stain normalization may contribute in increasing computational pathology robustness, reproducibility, and integrity.

Index Terms—Power law, Pareto principle rule, Stain normalization bias problem, Ivy GAP, Whole slide image.

Impact Statement—Glioblastoma is an aggressive cancer with a tragic 5-year survival rate, difficult to understand and treat. By eliminating the normalization bias problem, we further advance automated disease characterization, understanding, and outcomes.

I. INTRODUCTION

THERE is a perennial interest in digital pathology whole slide image (WSI) stain color normalization, with regards to personalized medicine in automated disease diagnosis

and patient management [1], [2], [3], [4], [5], [6], [7], [8]. State-of-the-art stain separation normalization methods use a single WSI as the reference standard for normalization of a WSI-cohort [2], [9], [10], [11], [12]. However, using a single WSI as a reference resembling an entire WSI-cohort is infeasible, as WSI regions of interest (ROI) are morphologically unique, and hence introducing an unintended color bias, resulting in inferior algorithm generalizability [13], [14], [15], [16], [17]. Often, multiple WSI are selected, the mean stain segmentation accuracy recorded, and the highest scoring images used as reference for WSI-cohort color normalization [18]. Alternatively, a random WSI is selected as reference for WSI-cohort color normalization potentiating resulting in worse color bias. Both WSI selection methods are impractical, error prone, and do not completely address the color normalization bias problem. Most importantly, these WSI selection approaches deteriorate neural network model generalization, while magnifying data domain shift, making unseen tumor data classification significantly challenging [19], [20], [21], [22], [23], [24].

We discuss a population (WSI-cohort subset) based hematoxylin and eosin (H&E) WSI-Cohort stain color normalization, not in terms of a single WSI, as a stain color normalization reference standard, but by considering the WSI-cohort subset H&E density histograms and Stain Vectors aggregates. The WSI-cohort subset size where stain CIELAB color space (also referred to as Lab) intensity channel (LabIC) [25] converges and beyond are representative of the entire WSI-cohort and the WSI-cohort subset aggregate suitable as a color normalization reference standard. We build upon an initial study of WSI color convergence tasked for a smaller Ivy GAP [26], [27], [28] WSI-cohort ($n = 509$) [29]. The present work enhances the initial approach using a larger Ivy GAP WSI-cohort ($n = 1864$), demonstrating LabIC convergence and developing an optimization method based on the Pareto principle 80/20 rule [30], [31], [32], [33]. The WSI-cohort LabIC convergence curve follows the law of large numbers [34], [35], [36], shown as the Wasserstein distance standard deviation between random WSI pairs, and is described by a right long-tailed power law distribution [32].

There is an extraordinary amount of research in stain color normalization [37], [38], [39] falling into two categories; color deconvolution and statistical pixel clustering. The color deconvolution method separates stains, normalizing stains individually, and rejoining stains into a single image. However, the method assumes accurate stain separation limited by a tissue structure spatial dependency. From creation, color deconvolution has evolved from a high-performance stain normalization requiring prior reference WSI stain vectors [40], to more accurate methods based on deconvolution in the optical density [11] and in CIELAB color spaces with a lesser tissue structure spatial dependency [12].

In addition to the law of large numbers, power law, and Pareto principle theory, we use high performance computing (HPC) for the analysis of large amounts of WSIs (i.e., Big Data) through stain deconvolution, histogram matching, and probability distributions distance measurements. Finally, we present color intensity convergence through WSI, ROIs, and

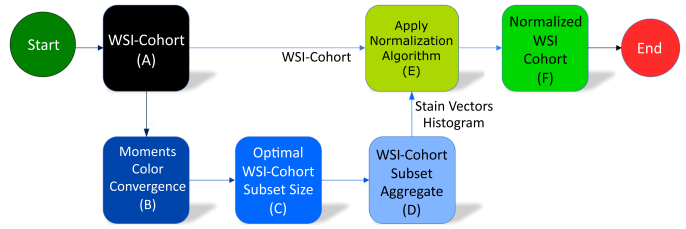


Fig. 1. Ivy GAP Cohort Color Normalization Work Flow Analysis. (A) Input: WSI-cohort, (B) Moments LabIC convergence approximation, (C) Optimization using the Pareto principle 80/20 rule, WSI-cohort subset size of randomly selected WSIs, (D) WSI-cohort subset histogram and stain vectors aggregates, (E) WSI-cohort color normalization by reference histogram and stain vectors transfer, (F) Output: Normalized Ivy GAP WSI-cohort.

large patch normalization permutations, as well as stain color normalization based on an optimal WSI-cohort subset size at the Pareto principle 80/20 rule. The above method may theoretically be used for stains, diseases, and normalization methods other than H&E, glioblastoma, and stain deconvolution method. In these new conditions the key question is how quickly stains color intensity convergence reaches an asymptote.

II. MATERIALS AND METHODS

The proposed method aims to find the optimal WSI-cohort subset size representative of a complete WSI-cohort using the Pareto principle 80/20 rule. The resulting optimal WSI-cohort subset aggregate (histogram and stain vectors) used as a normalization reference standard, when normalizing a WSI-cohort using the structure-preserving color normalization algorithm [12] (Fig. 1).

A. WSI-Cohort Subset Selection

A WSI-cohort of 1864 Ivy GAP [26], [27], [28] WSIs was selected (Fig. 2(A)), the background removed (Fig. 2(B)), and stain deconvolution separated in H&E WSI-cohorts [9], [10], [11], [12], [16] (Fig. 2(C)). Then, 200 new WSI-cohort subsets created using randomly selected WSI-pairs with WSI replacement. These subsets varied in size from 1 to 200 WSI-pairs and the mean Wasserstein Distance (W_p) calculated (Fig. 2(D)). The process was repeated for the 200 WSI-cohort subsets and stains, for a total of 1,000 permutations per WSI-cohort subset and respective standard deviations (S_{pi}) calculated (Fig. 2(E)). Finally, the stains standard deviations plotted, verifying S_{pi} power law behavior. The optimal cohort WSI subset size was calculated at the Pareto principle 80/20 rule [30] (Fig. 2(F)) and the Pearson correlation Test applied to H&E stain power law curves for curve similarity quantification.

B. Wasserstein Distance for Normalization Evaluation

We compute the Wasserstein distance for each WSI-pair across two different cohorts of equal size. In computing the average Wasserstein Distance across increasingly large cohorts, we show that results follow the law of large numbers, as results converge to a stable value. In general, the law of large numbers

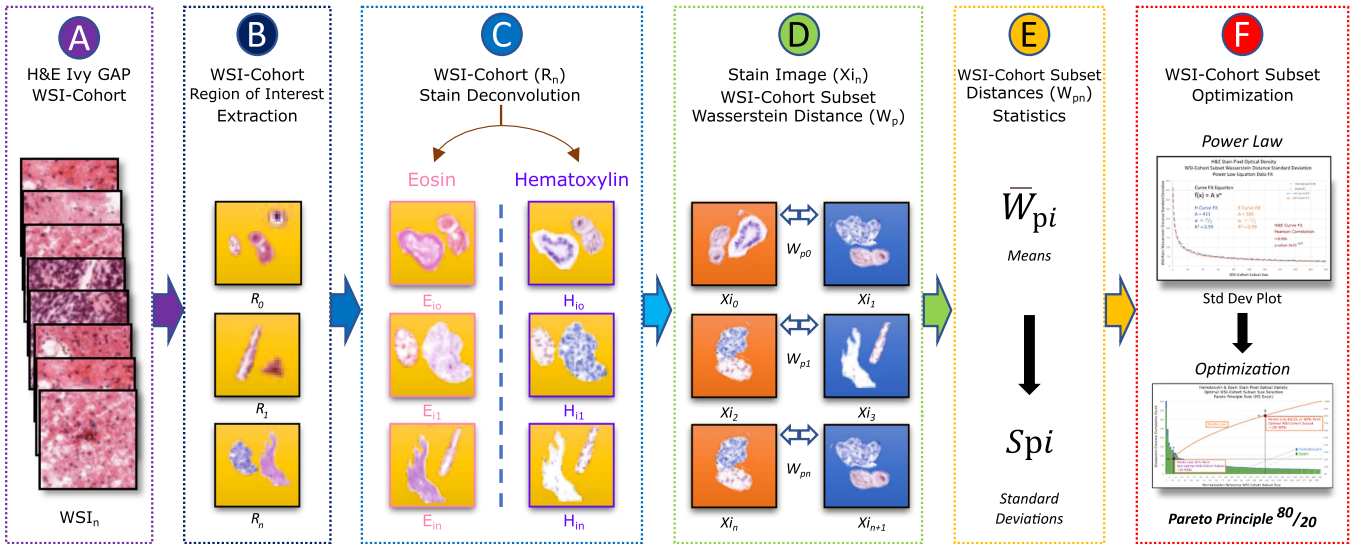


Fig. 2. Optimal WSI-cohort subset size process. (A) Ivy GAP 1864 WSI-cohort. The WSI-cohort was validated by discarding WSIs with markings and abnormal discoloration. (B) Background isolated ROIs. (C) Deconvolution separated H&E stains. (D) 200 WSI-cohort subsets creation. (E) WSI-pairs Wasserstein Distance means and WSI-cohort subset standard deviations calculation. Repeated for 1000 mean data points per WSI-cohort subset. (F) Pareto Principle approximation.

states that the sample mean converges to the true average value, as the number of samples increases. Here, samples consist of pairwise Wasserstein distances between individual WSIs. The mean of pairwise distances is often called the Gini’s coefficient and is used as a standard measure of spread [46], [47], [48]. Let D be a random variable representing the Wasserstein distance between a pair of WSI from separate cohorts with true expected value $E(D) = \mu$. For independent and identically distributed samples from D , the strong law of large numbers states that the sample average of Wasserstein distances will converge almost surely to the expected value [49]. That is, as the sample size n goes to infinity, the sample average of Wasserstein distances \overline{D}_n follows:

$$\overline{D}_n \xrightarrow{a.s.} \mu \tag{1}$$

C. Optimal WSI-Cohort Subset Size and the Pareto Principle

At larger WSI-cohort sizes, the effectiveness of normalization is reflected in the decreasing spread of Wasserstein distances measured across WSI-pairs. We hypothesize that the WSI-cohort Wasserstein distance standard deviation may be described by a power law

$$p(x) = Ax^{-\alpha} \tag{2}$$

where $p(x)$ is the distribution, A is the asymptote, and α is the data estimated exponent. This equation has been used as a model for many phenomena in nature and society [41].

We select an optimal cohort size as one that provides sufficient reduction of Wasserstein distance standard deviation. We adapt the widely used in finance Pareto analysis technique [33] available as a function in Microsoft Excel. Following Pareto analysis, we choose an optimal cohort size that reduces the standard deviation by 80% of the maximum reduction, measured at the

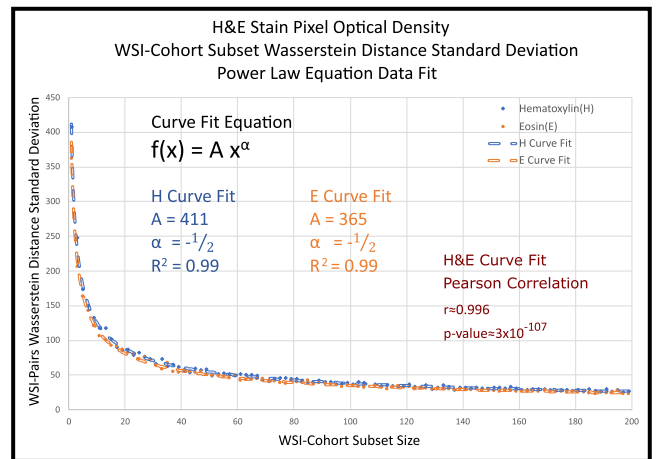


Fig. 3. H&E pixel optical density Wasserstein distance standard deviation curves. Curves show the convergence (decay) in Wasserstein distance standard deviation between WSI-pairs. H&E power law curves (2) are of similar shape with parameters: $A = 411$ and $A = 365$ for Hematoxylin and Eosin respectively, with $\alpha = -\frac{1}{2}$ and $R^2 = 0.99$, and Pearson correlation $r = 0.996$ with a p -value < 0.05 .

largest cohort sizes of 200 WSIs for H&E stains. This optimal cohort size provides sufficient stability in the pairwise distances between WSI, reflecting controlled differences between normalized WSIs [32].

D. Optimal WSI-Cohort Subset Normalization Parameters

After WSI-cohort stain separation and the optimal WSI-cohort subset size (S_n) are found, a S_n of random WSI were selected and labeled as the optimal WSI-cohort subset. Then, the optimal WSI-cohort subset \hat{W} (dictionary learning) and \hat{H} (sparse coding) matrices were calculated [12]. Finally, the

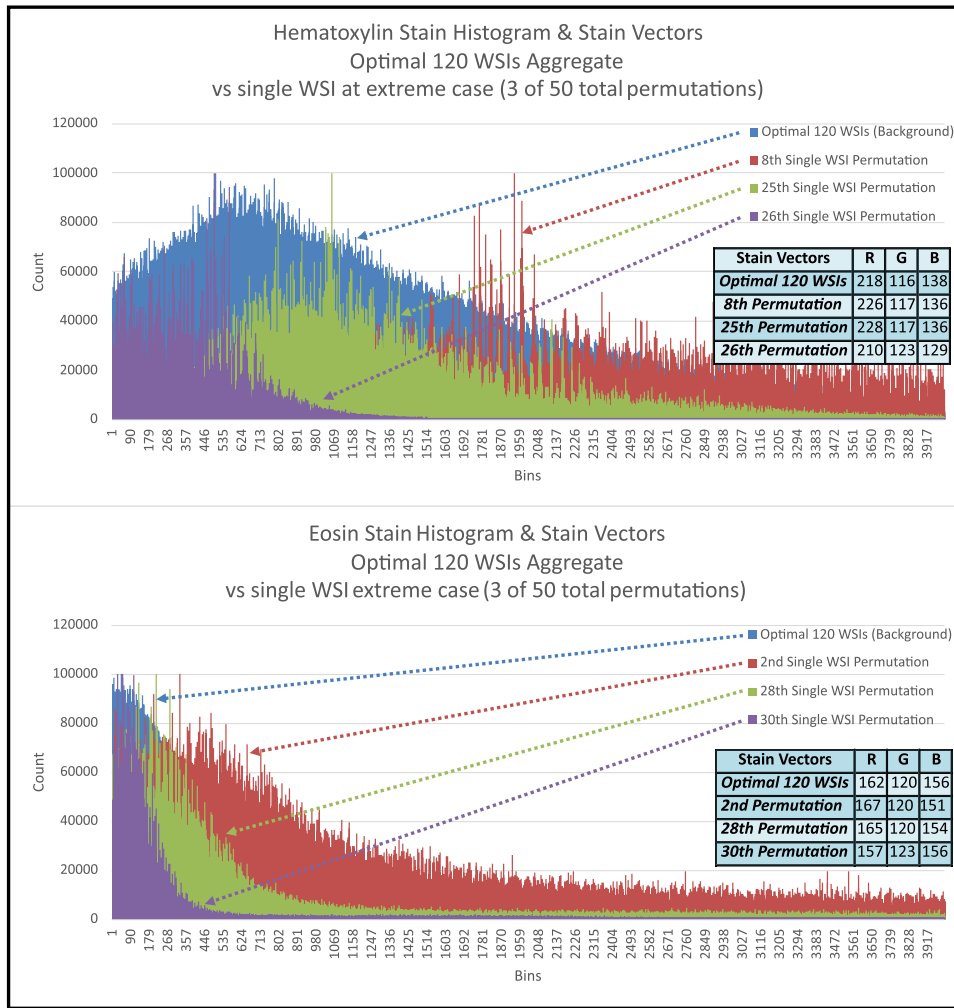


Fig. 4. H&E histograms and corresponding stain vectors, at 120 WSI-Cohort subset optimal size and three permutations of single WSI at extreme cases for comparison. Top:Hematoxylin and Bottom: Eosin.

optimal WSI-cohort subset W_{op} and H_{op} calculated by

$$W_{op} = \frac{1}{S_n} \sum_{k=1}^{S_n} W_k \quad (3)$$

$$H_{op} = \frac{1}{S_n} \sum_{k=1}^{S_n} H_k \quad (4)$$

$$H_k = \frac{H_i - H_{\min}}{H_{\max} - H_{\min}} * C \quad (5)$$

Where the histogram H bin size is taken from the image with the largest Knuth optimal bin size [51] and the resulting bin size increased to next power of 2 for hardware efficiency. Then, the histogram normalized to a constant C . Lastly, image normalization is performed through H_{op} and W_{op} matching

E. Quantitative WSI-Cohort LabIC Convergence

First, the WSI-cohort subset aggregated stain vectors and histogram were calculated, followed by complete 1864 WSI-cohort normalization, WSIs transformation to CIELAB color space,

and respective intensity channel extraction. Then, the intensity channel means per WSI and standard deviation per WSI-cohort were calculated. The process was repeated for 50 permutations and five WSI-cohort subset sizes (1, 10, 120, 1000, and 1864 WSIs), resulting in an analysis of 500,000 WSIs. Finally, the Levene's test (Levene, 1960) was applied between adjacent WSI-cohort subset-based normalizations and range measured for variance quantification across WSI-cohort subsets.

F. Quantitative ROI LabIC Convergence

First, the WSI-cohort subset aggregated stain vectors and histogram were calculated, followed by Ivy GAP WSI file names 300933007 and 101713101 normalization, WSIs transformation to CIELAB color space, and respective LabIC extraction. Subsequently, the LabIC means per WSI and standard deviation per cohort calculated. The process was repeated for 225 permutations and five WSI-cohort subset sizes (1, 10, 100, 1000, and 1864 WSIs), and nine WSI annotations or ROIs (Leading Edge, Infiltrating Tumor, Cellular Tumor, Tumor Perinecrotic Zone, Pseudopalisading Cells around Necrosis, Tumor

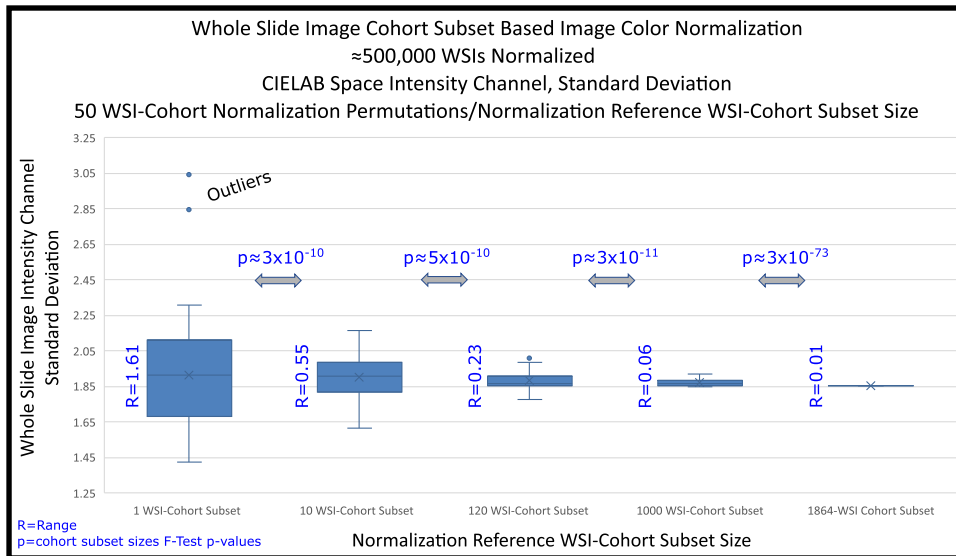


Fig. 5. LabIC convergence analysis for various WSI-cohort subsets. Nearly 500,000 WSIs processed in a HPC environment for two months. Applied Levene's test to adjacent WSI-cohort subsets. p-values < 0.05 show a statistical difference while decreasing ranges show convergence (exponential decay).

Microvascular Proliferation, Hyperplastic Blood, Necrosis, and Pseudopalisading Cells with no visible Necrosis), resulting in an analysis of 10,000 ROIs. Finally, the Levene's test was applied between adjacent WSI-cohort subset-based normalizations and range measured for variance quantification across WSI-cohort subsets.

G. Qualitative Large Patch LabIC Convergence

In addition to quantitative convergence representation, we show a qualitative representation of convergence using different WSI-cohort subset sizes. First, the WSI-cohort subset aggregated stain vectors and histogram were calculated, followed by Ivy GAP WSI file name 300933007 normalization and 5,000 by 5,000-pixel Cellular Tumor ROI patch extraction ($x = 4300$, $y = 6750$), corresponding LabIC extracted, and shown as a heatmap. The process was repeated for six permutations and five cohort subset sizes used as normalization reference (1, 10, 100, 1000, and 1800 WSIs).

III. RESULTS

A. WSI-Cohort Color Moments Convergence

The H&E stain color WSI-pairs Wasserstein distance standard deviation are shown as a power law curves (2) and curve fitting simulations found to achieve $A = 411$ and $A = 365$, for the Hematoxylin and Eosin stains respectively, with an $\alpha = -\frac{1}{2}$ and $R^2 = 0.99$, where R^2 is the multiple correlation coefficient. Furthermore, H&E stain curves Pearson correlation Test, two-tailed distribution, shows $r = 0.996$ with a p-values = 3×10^{-107} (Fig. 3). Lastly, the optimal 120 WSI-Cohort subset aggregated histogram and three permutations of single WSI histograms at extreme cases for Hematoxylin (8th, 25th, and 26th permutation)

and Eosin stains (2th, 28th, 30th permutation) are shown along with corresponding stain vectors are shown (Fig. 4) [1].

B. Quantitative WSI-Cohort Normalization Convergence

The color normalization for the complete WSI-cohort using five WSI-cohort subset sizes (1, 10, 100, 1000, and 1864 WSIs), as normalization references, yielded Levene's test p-values < 0.5 and decreasing range values results (Fig. 5), and values shown in Supplementary Information (SI), Tables I and II respectively.

C. Quantitative WSI-Cohort ROI Normalization Convergence

Quantitative normalization LabIC convergence for nine ROIs using four WSI-cohort subsets sizes (1,10,100,1000 WSIs), as normalization references, yielded Levene's test p-values < 0.5 and decreasing ranges results (Fig. 6), values shown in SI, Tables III and IV respectively.

D. Qualitative Large Patch Convergence

Qualitative representation of Cellular Tumor ROI large patch normalization for RGB and LabIC convergence using five WSI-cohort subset sizes (1, 10, 100, 1000, 1800 WSIs), as normalization references, are shown (Fig. 7). RGB and LabIC patches become more homogeneous as the size of the normalization reference WSI-cohort subset increases

E. Pareto Principle 80/20 Rule Normalization

The Pareto Principle 80/20 rule was applied to the H&E stains power law curves (Fig. 3), yielding optimal WSI-cohort subset size at the 80% point, or 120 WSIs (Fig. 8). For comparison,

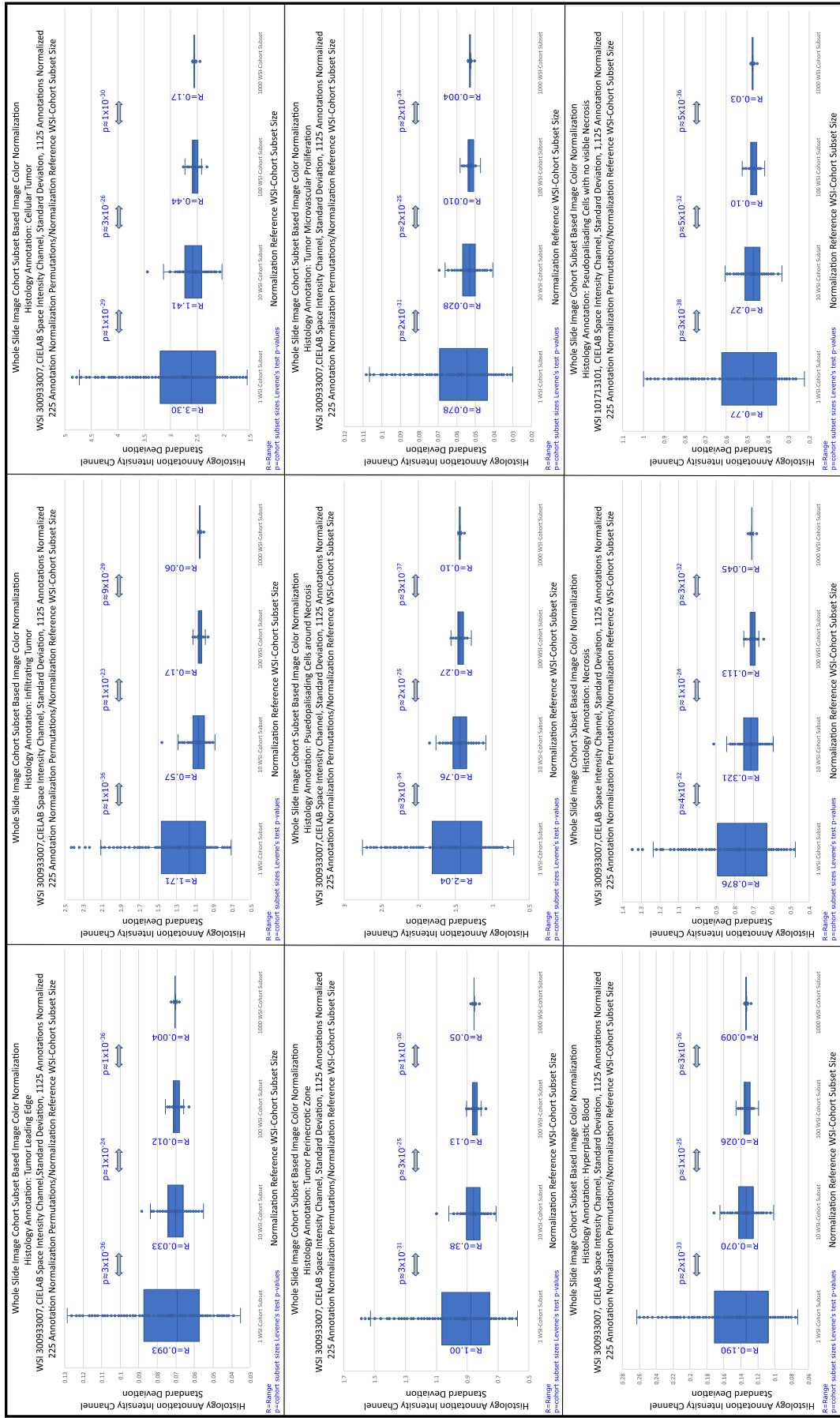


Fig. 6. Nine glioblastoma ROIs LabIC convergence. ROIs were normalized using four WSI-cohort subset sizes (1, 10, 100, and 1000 WSIs), as normalization reference. Applied Levene's Test to adjacent WSI-cohort subsets, p-values and ranges shown.

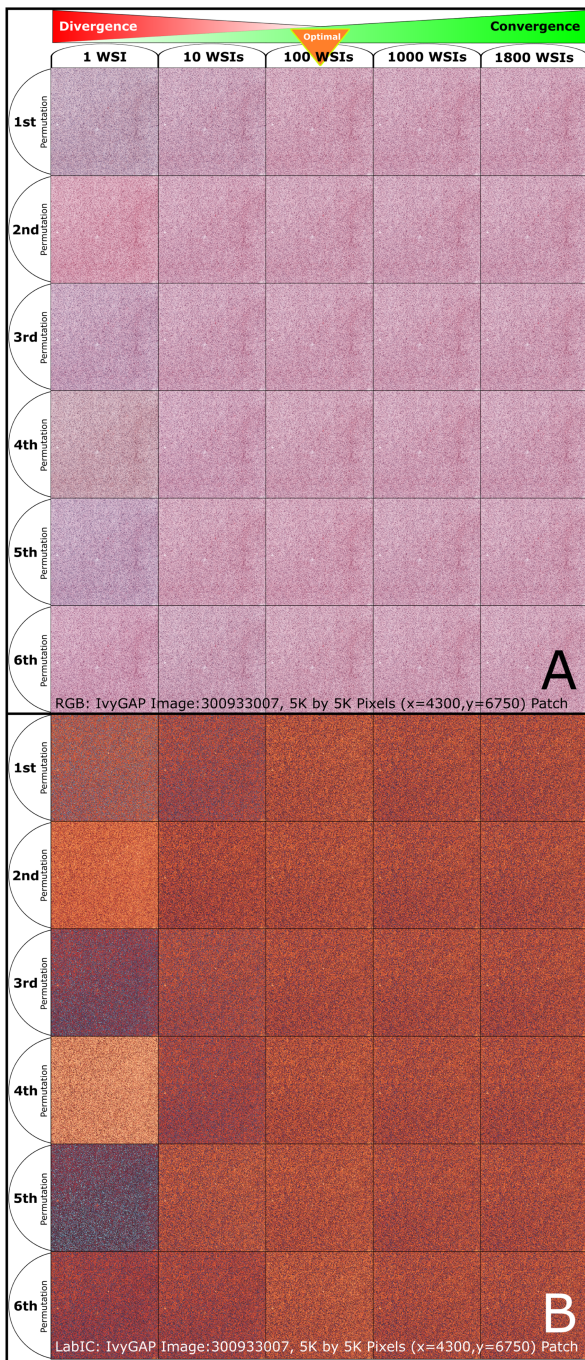


Fig. 7. Qualitative Cellular Tumor ROI large patch normalization convergence. (A) RGB, large patch normalization convergence increases as the number of WSIs increase in a WSI-cohort subset used as a normalization reference. (B) LabIC, clearer representation of large patch normalization convergence. Large patch normalized using five WSI-cohort subsets sizes (1, 10, 100, 1000, and 1864 WSIs), as a normalization reference shown for six permutations using WSI(s) patches with the most extreme results.

antithetical points were selected at a) single WSI, b) sub-optimal WSI-cohort subset size at 20%, or 10 WSIs and c) Full WSI-Cohort or 1864 WSIs. Converging Wasserstein distance standard deviation values shown in SI, Table V.

IV. DISCUSSION

Histological WSI often show color nonconformity caused by materials, equipment, and staining protocols differences. Although, pathologists compensate for these irregularities, these inaccuracies hinder automated computational analysis by accentuating data domain shift and algorithm generalizability that aid the diagnoses and treatment of disease.

We have shown the H&E density histogram and stain vectors composite or aggregates of a number of randomly selected WSI population (WSI-cohort subset) within a WSI-cohort, can be used as the normalization reference standard, since the WSI-cohort LabIC convergence follows the law of large numbers theorem and power law distribution. The WSI-cohort subset's stain vectors and histogram aggregates are representative of a given WSI-cohort with greater fidelity than standard approaches using a single WSI normalization reference. This new approach yields effective Ivy GAP 1864 WSI-cohort normalization results without color nonconformity and the unintended color bias from single WSI color normalization.

Using the Wasserstein distances means, we have calculated the cohort subsets' standard deviation curves. These curves are shown as a power law (Pareto Distribution), for both H&E stains (2). Worth noticing, a single curve is needed to show power law trend, as both H&E power curves are statistically identical, sharing the same decay (α), and a strong Person correlation (Figs. 3 & 8). Furthermore, standard deviations distance curves beyond the shown image-pairs yielded negligible distances contributions (not shown).

Quantitative, WSI-cohort normalization LabIC convergence is shown two-fold through Wasserstein Distance standard deviation values in the LabIC: a) at the complete WSI and b) WSI ROI levels. More specifically, the standard deviations quantification was performed at four WSI-cohort subsets (1, 10, 100, and 1000 WSIs) with 50 permutations/cohort-subset and five WSI-cohort subsets (1, 10, 120, 1000, 1864 WSIs) with 225 permutations/cohort-subset, for the complete WSI and nine WSI ROI respectively. All standard deviations show LabIC convergent behavior with Levene's test p-values < 0.05 , and decaying ranges, as WSI-cohort subsets size increase (Figs. 5 & 6). In addition, we have shown qualitative normalization convergence validation at the patch level by observing a normalized cancer tumor ROI patch (5,000 by 5,000-pixels) in RGB space and LabIC heatmaps at four WSI-cohort subsets (1, 10, 100, 1000, 1800 WSIs) for six permutations (Fig. 7)

Moreover, we have determined the optimal cohort subset size, as the normalization reference standard, by utilizing the Pareto principle. The Pareto principle 80/20 rule states 80% of the effects are the result of 20% of the causes. For normalization convergence comparison, we normalized the Ivy GAP 1864 WSI-cohort utilizing the followings as a normalization reference standard: a) a single WSIs, b) sub-optimal WSI-cohort subset (10 WSIs), c) optimal WSI-cohort subset (120 WSIs), and d) full WSI-cohort (1864 WSIs), and show Levene's test p-values and ranges. The sub-optimal and optimal cohort subsets were found by applying the Pareto Principle at 20% and 80/20 rule respectively, where variances converge at optimal cohort subset

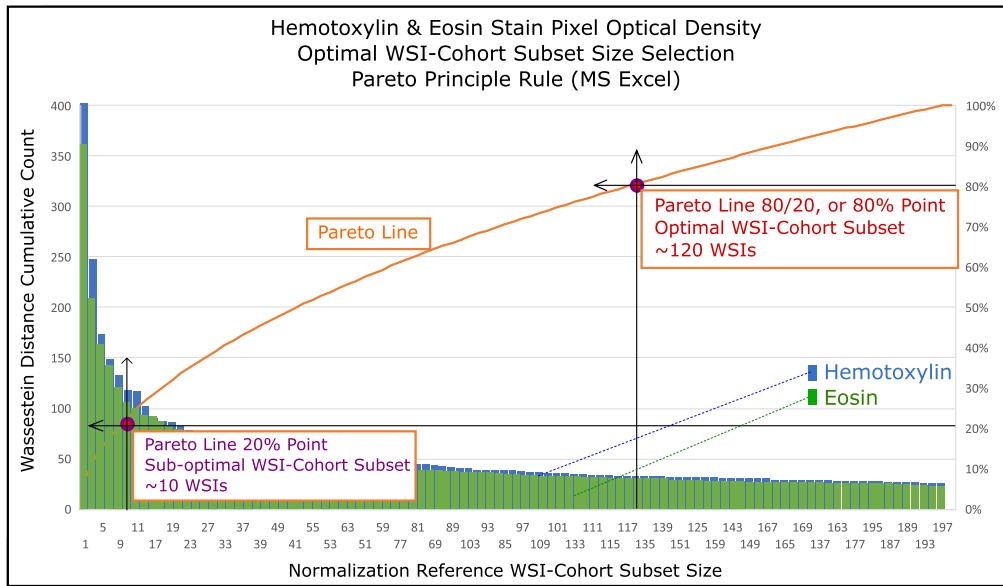


Fig. 8. Pareto Principle determination of sub-optimal and optimal H&E WSI-cohort sizes using Wasserstein distances standard deviations.

size through full cohort. These results demonstrate 120 random WSIs as optimal WSI-cohort subset size as a normalization reference (Fig. 8).

The determination of 120 random WSIs, as the optimal cohort subset size normalization reference, instead of single WSI, suggests a new robust approach in image pre-processing for network model development, as well as a more accurate clinical evaluation. Furthermore, our analysis shows no statistical difference in the Wasserstein distance standard deviations between H&E stains (Fig. 3). Thus, using a single stain (H or E) analysis could yield faster results than utilizing both H&E datasets.

The method is limited by the large amount of HPC resources required for computations and aggravated by the large size of the images in the Ivy-GAP WSI-cohort (1864 images). The following calculations show the longest computation effort by order of difficulty: a) the asymptote for the Wasserstein Distance [44] standard deviations between WSI-pairs (Fig. 3), b) the optimal histogram bin size [51] for the WSI-cohort, and c) demonstrating quantitative normalization convergence in WSIs (Fig. 5) and RIOs (Fig. 6) in multiple WSI-cohort subsets. Processing the 500,000 WSIs used in calculations required 75% of CBICA's HPC for over a two-month period. Future explorations in python-dask parallel-computing and sophisticated color convergence optimization algorithms will be particularly rewarding.

V. CONCLUSION

This paper presents a technique for employing a 120 WSIs as the color normalization reference standard for structure-preserving color normalization of a given WSI-cohort. The technique demonstrates that a WSI-cohort subset stain vectors and histogram aggregates are representative of a given WSI-cohort

with greater fidelity than a single WSI. Furthermore, the use of either H or E Wasserstein distance standard deviations is all that is required to deduce the optimal WSI-cohort subset of 120 WSIs for WSI-cohort normalization.

The analysis shows that the common practice of using a single WSI, as reference standard, results in a significant skewed WSI-cohort normalization. However, utilizing an optimal WSI population (WSI-cohort subset) size of 120 WSIs (Pareto principle 80/20 rule), as the reference standard, results in an accurate WSI-cohort color normalization without the inherent color normalization bias.

Theoretically, this approach could minimize discrepancies in glioblastoma histological evaluations. Future implementations utilizing a fully parallel-based-computation color convergence approach and improved stain separation methods could also result in a more efficient analysis.

SUPPLEMENTARY MATERIALS

We provide quantitative data in table form showing LabIC convergence at two levels: WSI-cohort and individual WSI-cohort ROIs. In addition, included are: a) video, still, and Power Point presentation Graphical abstracts and b) Python Code Ocean capsule to evaluate results with sample images, test cases, and a Power Point presentation describing Python code functions. The Code Ocean python capsule is available at <https://codeocean.com/capsule/3803185>. Note: Code Ocean is a cloud-based computational reproducibility platform fully integrated with IEEE Xplore.

ACKNOWLEDGMENT

The authors would like to thank Jesse Yonkovich and Robert Pozos Ph.D. for taking time away from their busy schedule to contribute to this document.

REFERENCES

- [1] K. Jafari-Khouzani and H. Soltanian-Zadeh "Multiwavelet grading of pathological images of prostate," *IEEE Trans. Biomed. Eng.*, vol. 1, no. 6, pp. 697–704, Jun. 2003.
- [2] A. Janowczyk, A. Basavanthally, and A. Madabhushi, "Stain normalization using sparse autoencoders (StNoSA): Application to digital pathology," *Computerized Med. Imag. Graph.*, vol. 57, pp. 50–61, 2017.
- [3] A. J. Méndez et al., "Computer-aided diagnosis: Automatic detection of malignant masses in digitized mammograms," *Med. Phys.*, vol. 25, no. 6, pp. 957–964, 1998.
- [4] T. Jinshan, R. M. Rangayyan, J. Xu, I. E. Naqa, and Y. Yang "Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances," *IEEE Trans. Inf. Technol. Biomed.* vol. 13, no. 2, pp. 236–251, Mar. 2009.
- [5] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.
- [6] B. D'Alessandro, A. P. Dhawan, and N. Mullani, "Computer aided analysis of epi-illumination and transillumination images of skin lesions for diagnosis of skin cancers," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2011, pp. 3434–3438.
- [7] M. Veta, J. P. W. Pluim, P. J. van Diest, and M. A. Viergever, "Breast cancer histopathology image analysis: A review," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1400–1411, May 2014.
- [8] N. Stathonikos, M. Veta, A. Huisman, and P. J. van Diest, "Going fully digital: Perspective of a Dutch academic pathology lab," *J. Pathol. Inform.*, vol. 4, 2013, Art. no. 15.
- [9] J. Niethammer et al., "Appearance normalization of histology slides," *Computerized Med. Imag. Graph.*, vol. 43, pp. 89–98, 2015.
- [10] A. M. Khan, N. Rajpoot, D. Treanor and D. Magee, "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1729–1738, Jun. 2014
- [11] M. Macenko et al., "A method for normalizing histology slides for quantitative analysis," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano to Macro*, 2009, pp. 1107–1110.
- [12] A. Vahadane et al., "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Trans. Med. Imag.*, vol. 35, no. 8, pp. 1962–1971, Aug. 2016.
- [13] I. D. Peter et al. "Leveraging random forests for interactive exploration of large histological images," *Med. Image Comput. Comput. Assist. Interv.*, vol. 17, pp. 1–8, 2014.
- [14] R. S. Crowley, G. J. Naus, S. J. and C. P. Friedman, "Development of visual diagnostic expertise in pathology an information processing study," *J. Amer. Med. Inform. Assoc.*, vol. 10, pp. 39–51, 2003.
- [15] D. Grube, "Constants and variables in immunohistochemistry," *Arch. Histol. Cytol.*, vol. 67, no. 2, 2004.
- [16] P. J. Tadrous, "Digital stain separation for histological images," *J. Microsc.*, vol. 240, pp. 164–172, 2010.
- [17] G. Mcnamara, "Color balancing histology images for presentations and publication," *J. Histotechnol.*, vol. 28, 2005, pp. 81–88.
- [18] R. A. Hoffman, S. Kothari, and M. D. Wang, "Comparison of normalization algorithms for cross-batch color segmentation of histopathological images," in *Proc. IEEE 36th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2014, pp. 194–197.
- [19] K. Stacke, G. Eilertsen, J. Unger and C. F. Lundstrom, "A closer look at domain shift for deep learning in histopathology," in *Proc. MICCAI Comput. Pathol. Workshop COMPAY*, 2019.
- [20] W. M. Kouw and M. Loog, "A review of domain adaptation without target labels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 766–785, Mar. 2021.
- [21] A. Kurakin, I. J. Goodfellow and S. Bengio, "Adversarial examples in the physical world," *Artif. Intell. Saf. Secur.*, Chapman and Hall/CRC, 2018, pp. 99–112.
- [22] S. Ben-David, B. John, K. Crammer, A. Kulesza and F. Pereira, "A theory of learning from different domains," *Mach. Learn.*, vol. 79, pp. 151–175, 2010.
- [23] D. Tellez, M. Balkenhol, N. Karssemeijer, G. Litjens, J. van der Laak, and F. Ciompi, "H and E stain augmentation improves generalization of convolutional networks for histopathological mitosis detection," *Proc. SPIE*, vol. 10581, pp. 264–270, 2018.
- [24] K. Stacke, G. Eilertsen, J. Unger and C. Lundström, "Measuring domain shift for deep learning in histopathology," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 2, pp. 325–336, Feb. 2021.
- [25] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Comput. Graph. Appl.*, vol. 21, no. 5, pp. 34–41, Jul./Aug. 2001.
- [26] R. B. Puchalski, "An anatomic transcriptional atlas of human glioblastoma," *Science*, vol. 360, pp. 660–663, 2018.
- [27] R. B. Puchalski, "IvyGAP website," 2014 [Online]. Available: <https://glioblastoma.alleninstitute.org>
- [28] R. B. Puchalski, "Supplementary materials for: An anatomic transcriptional atlas of human glioblastoma," *Science*, vol. 360, pp. 660–663, 2018.
- [29] C. M. Grenko, A. N. Viaene, M. P. Nasrallah, M. D. Feldman, H. Akbari, and S. Bakas, "Towards population-based histologic stain normalization of glioblastoma," in *Proc. Int. MICCAI Brainlesion Workshop*, 2020, pp. 44–56.
- [30] D. Mineva, "Application of the pareto principle in the contractual process for oncological disease," *Knowl.-Int. J.*, vol. 38, no. 4, pp. 985–989, 2020.
- [31] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 1521–1528.
- [32] M. E. J. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary Phys.*, vol. 46, pp. 323–351, 2005.
- [33] G. Karuppusami and R. Gandhinathan, "Pareto analysis of critical success factors of total quality management," *JOOR*, vol. 18, pp. 372–385, 2006.
- [34] K. Yao and J. Gao, "Law of large numbers for uncertain random variables," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 3, pp. 615–621, Jun. 2016.
- [35] F. Dekking, C. Kraaikamp, H. Lopuhaa and L. Meester, *A Modern Introduction to Probability and Statistics: understanding Why and How*. Berlin, Germany:Springer, 2005.
- [36] I. Fazekas and O. Klesov, "A general approach to the strong law of large numbers," *Theory Probability Appl.*, vol. 45, pp. 436–449, 2014.
- [37] D. P. Onder, S. M. Zengin and S. M. Sarioglu, "A review on color normalization and color deconvolution methods in histopathology," *Appl. Immunohistochemistry Mol. Morphol.*, vol. 22, no. 10, pp. 713–719, 2014.
- [38] S. Kothari, J. Phan, R. Moffitt, T. Stokes, S. Hassberger and Q. Chaudry, "Automatic batch-invariant color segmentation of histological cancer images," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano to Macro*, 2011, pp. 657–660.
- [39] D. Magee, D. Treanor, D. Crellin, M. Shires, K. Mohee, and P. Quirke, "Colour normalisation in digital histopathology images," in *Proc. Opt. Tissue Image Anal. Microsc., Histopathol. Endoscopy*, 2009, pp. 100–111.
- [40] J. D. A. Ruifrok AC, "Quantification of histochemical staining by color deconvolution," *Anal. Quantitative. Cytol. Histol.*, vol. 23, pp. 291–299, 2001.
- [41] E. H. Moore, "On the reciprocal of the general algebraic matrix," *Bull. Amer. Math. Soc.*, vol. 26, pp. 394–395, 1920.
- [42] A. Bjerhammar, "Application of calculus of matrices to method of least squares: with special reference to geodetic calculations," *Elander*, 1951.
- [43] R. Penrose, "A generalized inverse for matrices," *Math. Proceed. Cambridge Philos. Soc.*, vol. 51, pp. 406–413, 1955.
- [44] L. Ruschendorf, "The Wasserstein distance and approximation theorems," *Wahrscheinlichkeitstheorie Verw Gebiete*, vol. 70, pp. 117–129, 1985.
- [45] S. T. Rachev and L. Ruschendorf, *Mass Transportation Problems*, vol. 1. Berlin, Germany:Springer, 1998, pp. 325–336.
- [46] T. F. Liao, "Measuring and analyzing class inequality with the gini index informed by model-based clustering. Sociological Methodology," *Sociol. Methodol.*, vol. 36, no. 1, 2006, pp. 201–224.
- [47] T. Sitthiyot and K. Holasut, "A simple method for measuring inequality," *Palgrave Commun.*, vol. 6, 2020, Art. no. 112.
- [48] F. De Maio, "Income inequality measures," *J. Epidemiol. Community Health*, vol. 61, no. 10, pp. 849–852, 2007.
- [49] N. Etemadi, "An elementary proof of the strong law of large numbers," *Zeitschrift Für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 55, pp. 119–122, 1981.
- [50] H. Levene, *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Redwood City, CA, USA: Stanford Univ. Press, 1960.
- [51] Kevin H. Knuth, "Optimal data-based binning for histograms," *Digit. Signal Process.*, vol. 95, 2019, Art. no. 102581.