

RESEARCH ARTICLE

Nuclear IHC enumeration: A digital phantom to evaluate the performance of automated algorithms in digital pathology

Muhammad Khalid Khan Niazi^{1*}, Fazly Salleh Abas², Caglar Senaras¹, Michael Pennell³, Berkman Sahiner⁴, Weijie Chen⁴, John Opfer⁵, Robert Hasserjian⁶, Abner Louissaint, Jr.⁶, Arwa Shana'ah⁷, Gerard Lozanski^{7‡}, Metin N. Gurcan^{1‡}

1 Center for Biomedical Informatics, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, **2** Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia, **3** Department of Biostatistics, College of Public Health, The Ohio State University, Columbus, Ohio, United States of America, **4** US Food and Drug Administration, Silver Spring, Maryland, United States of America, **5** Department of Psychology, The Ohio State University, Columbus, Ohio, United States of America, **6** Department of Pathology, Harvard Medical School, Boston, Massachusetts, United States of America, **7** Department of Pathology, The Ohio State University, Columbus, Ohio, United States of America

‡ These authors are joint senior authors on this work.

* Mniazi@wakehealth.edu



OPEN ACCESS

Citation: Niazi MKK, Abas FS, Senaras C, Pennell M, Sahiner B, Chen W, et al. (2018) Nuclear IHC enumeration: A digital phantom to evaluate the performance of automated algorithms in digital pathology. *PLoS ONE* 13(5): e0196547. <https://doi.org/10.1371/journal.pone.0196547>

Editor: Kristen C. Maitland, Texas A&M University, UNITED STATES

Received: August 14, 2017

Accepted: April 15, 2018

Published: May 10, 2018

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: The data can be found in the following Github repository: <https://github.com/cialab/Synthetic-Image-Model/tree/master/1000x1000>.

Funding: This work was supported by R01CA134451 (PIs: MG, GL), U24CA199374 (PIs: MG, AMadabhushi, AMartel), U01 CA198945 (PI: AB) from the National Cancer Institute. The work was also supported by OSU CCC Pelotonia Award (PI: Gurcan). The funders had no role in study

Abstract

Automatic and accurate detection of positive and negative nuclei from images of immunostained tissue biopsies is critical to the success of digital pathology. The evaluation of most nuclei detection algorithms relies on manually generated ground truth prepared by pathologists, which is unfortunately time-consuming and suffers from inter-pathologist variability. In this work, we developed a digital immunohistochemistry (IHC) phantom that can be used for evaluating computer algorithms for enumeration of IHC positive cells. Our phantom development consists of two main steps, 1) extraction of the individual as well as nuclei clumps of both positive and negative nuclei from real WSI images, and 2) systematic placement of the extracted nuclei clumps on an image canvas. The resulting images are visually similar to the original tissue images. We created a set of 42 images with different concentrations of positive and negative nuclei. These images were evaluated by four board certified pathologists in the task of estimating the ratio of positive to total number of nuclei. The resulting concordance correlation coefficients (CCC) between the pathologist and the true ratio range from 0.86 to 0.95 (point estimates). The same ratio was also computed by an automated computer algorithm, which yielded a CCC value of 0.99. Reading the phantom data with known ground truth, the human readers show substantial variability and lower average performance than the computer algorithm in terms of CCC. This shows the limitation of using a human reader panel to establish a reference standard for the evaluation of computer algorithms, thereby highlighting the usefulness of the phantom developed in this work. Using our phantom images, we further developed a function that can approximate the true ratio from the area of the positive and negative nuclei, hence avoiding the need to detect individual nuclei. The predicted ratios of 10 held-out images using the function (trained on 32 images) are within $\pm 2.68\%$ of the true ratio. Moreover, we

design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

also report the evaluation of a computerized image analysis method on the synthetic tissue dataset.

I. Introduction

Immunohistochemical (IHC) staining of tissue sections is routinely used in pathology to aid in diagnosis and to characterize malignant tumors in humans and in animals [1–3]. It also plays a vital role in selecting an appropriate systemic therapy for cancer patients [2]. Certain stains are accepted in medical practice as important predictive markers of tumor aggressiveness (Ki67) or tumor type (ER, PR and HER2Neu) and are used to determine the type and intensity of therapy for a given patient [4]. All these markers are used according to specific guidelines where the intensity of stains and the number of positive cells are expressed as percentage of all malignant cells.

Considering the importance of these markers, technical aspects of stains such as tissue fixation, tissue processing and quality of stains are becoming strictly regulated and monitored to assure reproducible results [5]. In the majority of cases, IHC stain interpretation is rendered by a trained pathologist using a light microscope and a manual method, which consists of counting each positively stained cell [5]. As expected, the manual method of enumeration of positive cells suffers from poor reproducibility even in the hands of an expert pathologist [6, 7]. Moreover, due to human reader limitations, true counting of IHC positive and negative cells can only be performed for a limited number of cells [8–11]. Thus, by necessity IHC stains are counted only in small fragments of tumors selected by the pathologist. Such random sampling may be representative of tumors with a homogenous pattern of staining; however, the majority of malignant tumors show a heterogeneous pattern of staining with some areas enriched in positive cells and others with only a few positive cells. Consequently, the counting of IHC positive cells using random sampling of small areas of IHC stained heterogeneous tumors is often biased and does not represent the true nature of the tumor [12]. Oversampling of predominantly positive areas results in overestimation, and oversampling of IHC poor areas results in underestimation of IHC positive cells.

An alternative approach to quantify percentage of positive cells in a high number of cells is estimation based on visual perception of areas without actually counting the positive cells [6]. Due to the limited ability of the human reader to efficiently and reproducibly count hundreds or thousands of cells, the pathologist replaces counting with a more crude method such as estimation (gestalt) [9]. Not surprisingly, such a method introduces substantial bias and variability, and results in poor reproducibility. Considering these difficulties, it is logical that a marked effort is underway to develop computer algorithms to accurately, precisely and reproducibly enumerate IHC stained cells using high resolution images of histological sections. Using different approaches, several competing algorithms have been developed and are currently being tested for clinical use in pathology. Comparison of accuracy and reproducibility between these platforms is based on a reference standard established by human reader interpretation of IHC stains. Paradoxically, all computer method generated results are still compared to results that are generated by a human reader [13]. Using reference standards based on human reader interpretation of images hampers standardization of computer IHC methods mainly due to high inter- and intra-reader variability of humans.

We have developed an *in silico* phantom comprised of both IHC positive and IHC negative (counterstained) cells with a known ground truth, thereby avoiding the need for a reference

standard built by human readers. This phantom is generated by our specialized software that uses large collections of high resolution images of different negative and positive cells and generates virtual tissue sections with a known percentage of positive and negative cells. The proportion can be varied from less than 1% to 100% of positive cells. The distribution of positive and negative cells can be adjusted to mimic true histological sections where positive cells can be distributed evenly throughout the tissue or may show focal clustering depending on the intention of the operator. A cluster of cells can be comprised of positive cells with different intensities of staining, different patterns of staining, different cell size with cells showing good separation and cells overlapping (to provide a more challenging task of cell segmentation for computer algorithms).

II. Methods

We extracted a group of Ki-67 positive and negative nuclei from images of Ki-67 stained lymphoma biopsies [6, 14]. Each group consists of single nuclei as well as nuclei clumps having two to five nuclei. For simplicity we call these nuclei (single or clumped) “objects.” In total, we have 50 Ki-67 positive objects and 40 Ki-67 negative objects. Fig 1 shows examples of some of these positive and negative objects. The IRB was approved (Study Number: 2007C0069) by the Office of Responsible Research Practices and The Institutional Review Board at the Ohio State University. It is worth mentioning that all images used in this study were fully anonymized.

Let L_+ and L_- denote the positive and negative objects, respectively. Here $i = 1, 2, 3, \dots, 50$ and $j = 1, 2, 3, \dots, 40$. To create a synthetic image I_S , we perform the following steps.

1. Create canvas of 1000x1000 pixels.
2. For each pixel in the canvas, generate a three element vector (with elements corresponding to RGB colors) using a uniformly distributed pseudorandom integer generator between 190 and 250. These values serve as background for the canvas.
3. Select 900 equally spaced points/locations on the canvas. Let (x_a, y_b) represents the set of 900 locations, where $a = 1, 2, 3, \dots, 30$ and $b = 1, 2, 3, \dots, 30$.
4. Using a uniformly distributed pseudorandom integer generator, create a set of 900 pairs of random integers between -11 and +11. Let (Rx_a, Ry_b) represent the set of 900 value pairs.
5. Disturb the 900 equally spaced points by adding them with (Rx_a, Ry_b) , i.e. $(Px_a, Py_b) = (Rx_a, Ry_b) + (x_a, y_b)$. This will perturb the points slightly and will remove the systematic look of computer generated points.
6. To insert G% positive nuclei to the canvas, randomly draw positive objects with replacement from L_+ until the total number of nuclei reaches $G \times 900 / 100$. For instance, to create an image with 40% positive nuclei, we will keep on drawing positive objects until we have drawn a total of 360 nuclei. Then the selected objects are randomly scaled to resemble all different cell sizes in real Ki67 stained tissues.
7. Let us assume that we require N positive objects to create an image with G% positive nuclei. We place these N positive objects in randomly selected N locations from (Px_a, Py_b) .
8. We repeat step 6 and 7 to place negative objects on the canvas. The only difference is that we exclude the N positions already filled by positive objects during extraction of random integers. As a final step, we perform Gaussian filtering on the edges of all positive and negative objects in the canvas.

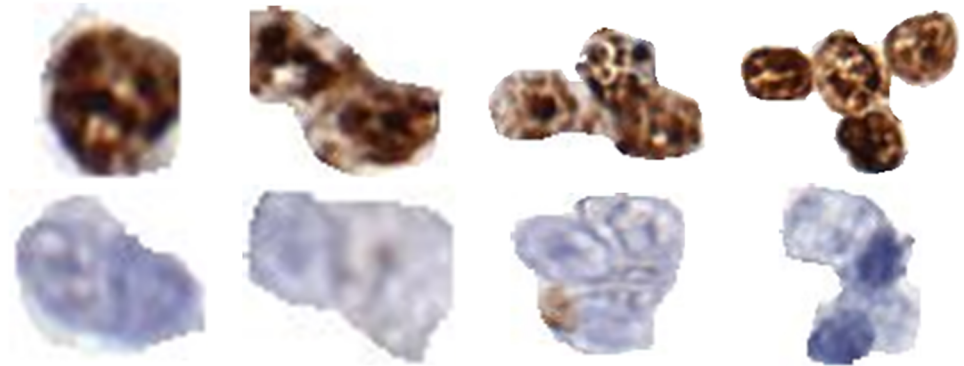


Fig 1. Examples of positive and negative objects used in creation of synthetic images. All objects were cropped from Follicular Lymphoma whole slide images stained for Ki67.

<https://doi.org/10.1371/journal.pone.0196547.g001>

Using this method we created a set of 42 images with different concentrations of positive and negative nuclei. Fig 2 shows an example of two synthetic images created using this method. The 42 synthetic images created using the proposed method were used to investigate the following three questions on Estimation, Mapping, and Validation which are outlined next.

II.a Estimation

Is a visual estimation of the ratio of positive to total number of nuclei by a group of board certified pathologists consistent with the true ratio? Let r_n denote true ratio (i.e., the ratio of positive to total number of nuclei during synthetic image generation)

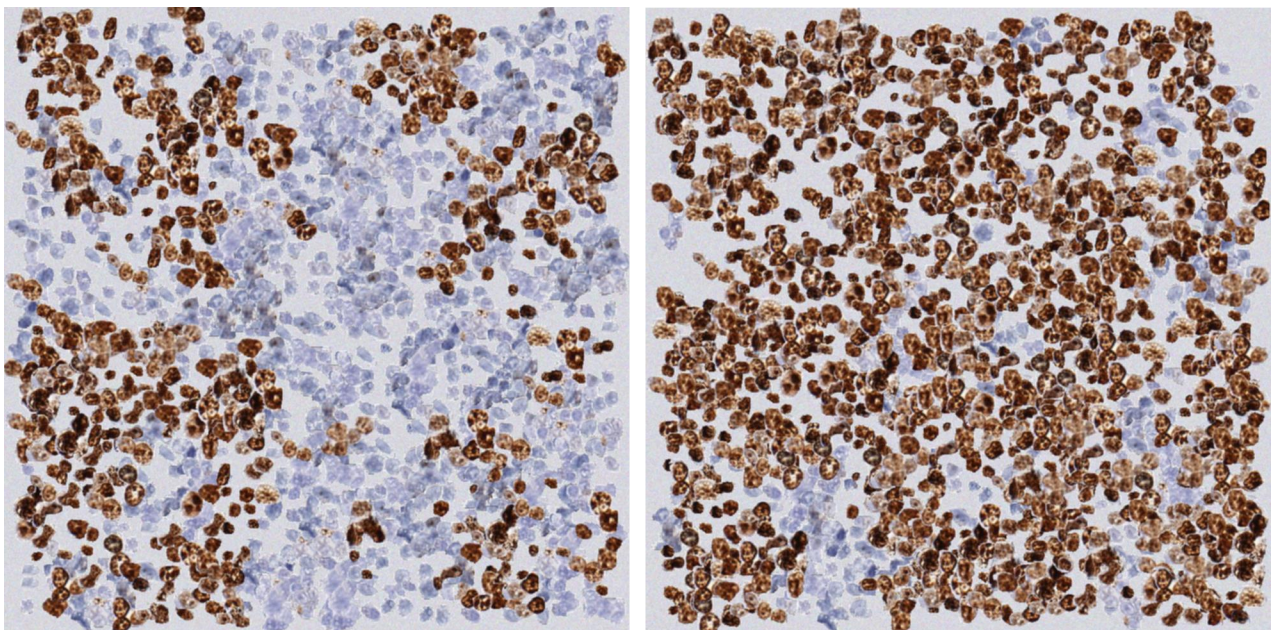


Fig 2. Examples of synthetic images generated by the proposed method. Left) An example image with only a few Ki67 positive nuclei. Right) An example image with higher concentration of Ki67 positive nuclei.

<https://doi.org/10.1371/journal.pone.0196547.g002>

Table 1. Division of synthetic images into 10 subsets. Here SS_i correspond to the i^{th} subset. The second row contains the ratio of positive to all nuclei within each SS_i . The third row contains the number of images in each subset.

SS_1	SS_2	SS_3	SS_4	SS_5	SS_6	SS_7	SS_8	SS_9	SS_{10}
$r_n < 0.1$	$0.1 \leq r_n < 0.2$	$0.2 \leq r_n < 0.3$	$0.3 \leq r_n < 0.4$	$0.4 \leq r_n < 0.5$	$0.5 \leq r_n < 0.6$	$0.6 \leq r_n < 0.7$	$0.7 \leq r_n < 0.8$	$0.8 \leq r_n < 0.9$	$r_n \geq 0.9$
4	4	7	4	5	6	3	3	3	3

<https://doi.org/10.1371/journal.pone.0196547.t001>

To answer the first question, a total of four board certified pathologists took part in the study. We shared the images through a customized web application, which allowed the pathologists to remotely view the images and provide their visual approximation of r_n , which is denoted as P_n^j ($j = 1, \dots, 4$). The pathologists also provided their visual approximation of the following ratio:

$$r_t = \frac{\text{Area of positive nuclei}}{\text{Area of the image}} \tag{1}$$

II.b Validation

Can one of our recent nuclei detection algorithms [15] approximate r_n ?

In order to evaluate the robustness of the computer based solutions on the artificial dataset, we tested it against one of our recently published nuclei detection algorithms [15]. The algorithm starts by calculating two score maps for blue and brown objects, which correspond to negative and positive nuclei respectively. After calculating the score maps, the algorithm analyzes the connected components for different threshold values and finds the optimal threshold value, which maximizes the number of separated nuclei. Finally, the algorithm uses this coarse segmentation result and splits the merged nuclei by detecting elliptical arcs and applying a super pixel method [16].

II.c Mapping

Can we find a function $\Psi(r_a)$ which maps $r_a \rightarrow r_n$, where r_a corresponds to the ratio of the area of the positive nuclei to the total area of nuclei? The aim is to determine if we can replace the computationally expensive nuclei detection step with the less expensive area estimation.

In order to answer the second question, we divided the 42 synthetic images into 10 subsets ($SS_1, SS_2, SS_3, \dots, SS_{10}$) as shown in Table 1. The third row in Table 1 contains the number of images in each subset. To approximate a function that maps $r_a \rightarrow r_n$, we randomly held out one image from each subset (a total of 10 images which we will use for testing) and computed the following difference for the remaining images (a total of 32 images which we used for training):

$$D = r_a - r_n \tag{2}$$

III. Results

For ease of comparison and to avoid redundancy, we combined the results of Subsections II.a (Estimation) and II.b (Validation) under Subsection III.a. The results for Subsection II.c are outlined in Subsection III.b.

III.a Estimation & validation

The results of our experiment are shown in Tables 2, 3 and 4. Table 2 shows the summary statistics of the pathologists and the computer algorithm in estimating r_n (the ratio of true

Table 2. Summary statistics of pathologists' and computer's estimation of r_n (the ratio of true number of positive to total number of nuclei). Here $P^1_{r_n}, P^2_{r_n}, P^3_{r_n}, P^4_{r_n}, P^c_{r_n}$ denote the estimation of r_n by pathologist 1, pathologist 2, pathologist 3, pathologist 4 and the computer algorithm, respectively.

		$P^1_{r_n}$	$P^2_{r_n}$	$P^3_{r_n}$	$P^4_{r_n}$	$P^c_{r_n}$
r_n	Mean ± SD	51.05 ± 31.4	54.2 ± 30.2	53.6 ± 30.6	49.1 ± 32.5	43.1 ± 27.4
	Bias ± SD	5.5 ± 8.4	8.6 ± 12.8	8.1 ± 7.5	3.5 ± 9.3	-2.5 ± 2.2
	CCC	0.94	0.86	0.93	0.95	0.993
	(95% C.I.)	(0.90, 0.97)	(0.77, 0.92)	(0.88, 0.96)	(0.91, 0.97)	(0.987, 0.996)
	Rank Corr.	0.98	0.90	0.99	0.97	0.998
	(95% C.I.)	(0.96, 0.99)	(0.83, 0.95)	(0.98, 0.99)	(0.94, 0.98)	(0.996, 0.999)
	Pearson Corr.	0.97	0.91	0.97	0.97	0.997
	(95% C.I.)	(0.94, 0.98)	(0.83, 0.95)	(0.95, 0.99)	(0.94, 0.98)	(0.994, 0.998)

<https://doi.org/10.1371/journal.pone.0196547.t002>

Table 3. Summary statistics of pathologists' and computer's estimation of r_t (the ratio of the true area of positive nuclei to total image area). Here $P^1_{r_t}, P^2_{r_t}, P^3_{r_t}, P^4_{r_t}, P^c_{r_t}$ represent the estimation of r_t by pathologist 1, pathologist 2, pathologist 3, pathologist 4 and the computer algorithm, respectively. The CCC is lowest for the computer because of the near perfect negative correlation between bias and the number of nuclei (see Fig 5).

		$P^1_{r_t}$	$P^2_{r_t}$	$P^3_{r_t}$	$P^4_{r_t}$	$P^c_{r_t}$
r_t	Mean ± SD	45.8 ± 30.6	43.8 ± 26.2	46.5 ± 28.5	38.7 ± 29.9	18.0 ± 8.4
	Bias ± SD	14.5 ± 15.6	12.5 ± 13.8	15.2 ± 13.1	7.4 ± 15.5	-13.3 ± 7.7
	CCC	0.68	0.68	0.69	0.76	0.53
	(95% C.I.)	0.57, 0.76)	(0.56, 0.78)	(0.59, 0.77)	(0.68, 0.82)	(0.41, 0.63)
	Rank Corr.	0.98	0.90	0.99	0.97	0.992
	(95% C.I.)	(0.97, 0.99)	(0.82, 0.94)	(0.97, 0.99)	(0.94, 0.98)	(0.985, 996)
	Pearson Corr.	0.97	0.90	0.98	0.95	0.995
	(95% C.I.)	(0.94, 0.98)	(0.82, 0.94)	(0.97, 0.99)	(0.91, 0.97)	(0.991, 0.997)

<https://doi.org/10.1371/journal.pone.0196547.t003>

number of positive to total number of nuclei in an image). The four pathologists' estimation of r_n is represented by $P^i_{r_n}$, where $i \in \{1,2,3,4\}$, while the estimation by the computer algorithm is represented by $P^c_{r_n}$. Likewise, Table 3 shows the summary statistics of the four pathologists and the computer algorithm in estimating r_t (ratio of the area of the positive nuclei to the total area of the image). Here the total area of the image corresponds to the image size. Similarly, Table 4

Table 4. Summary statistics of pathologists' and computer's estimation of r_a (the ratio of the true area of positive nuclei to area covered by all nuclei). Here $P^1_{r_a}, P^2_{r_a}, P^3_{r_a}, P^4_{r_a}, P^c_{r_a}$ represent the estimation of r_a by pathologist 1, pathologist 2, pathologist 3, pathologist 4 and the computer algorithm, respectively. The aim of computing this statistics was to check if pathologists' are unconsciously computing r_a when asked to approximate r_n .

		$P^1_{r_a}$	$P^2_{r_a}$	$P^3_{r_a}$	$P^4_{r_a}$	$P^c_{r_a}$
r_a	Mean ± SD	51.05 ± 31.4	54.2 ± 30.2	53.6 ± 30.6	49.1 ± 32.5	51.8 ± 27.8
	Bias ± SD	-6.0 ± 8.2	-2.83 ± 12.9	-3.4 ± 6.1	-7.9 ± 10.2	-5.2 ± 4.1
	CCC	0.94	0.90	0.97	0.91	0.97
	(95% C.I.)	(0.90, 0.97)	(0.82, 0.94)	(0.95, 0.98)	(0.85, 0.94)	(0.95, 0.98)
	Rank Corr.	0.98	0.91	0.99	0.97	0.999
	(95% C.I.)	(0.96, 0.99)	(0.83, 0.95)	(0.98, 1.00)	(0.95, 0.99)	(0.997, 0.999)
	Pearson Corr.	0.97	0.90	0.98	0.96	0.99
	(95% C.I.)	(0.95, 0.98)	(0.83, 0.95)	(0.97, 0.99)	(0.92, 0.98)	(0.98, 0.99)

<https://doi.org/10.1371/journal.pone.0196547.t004>

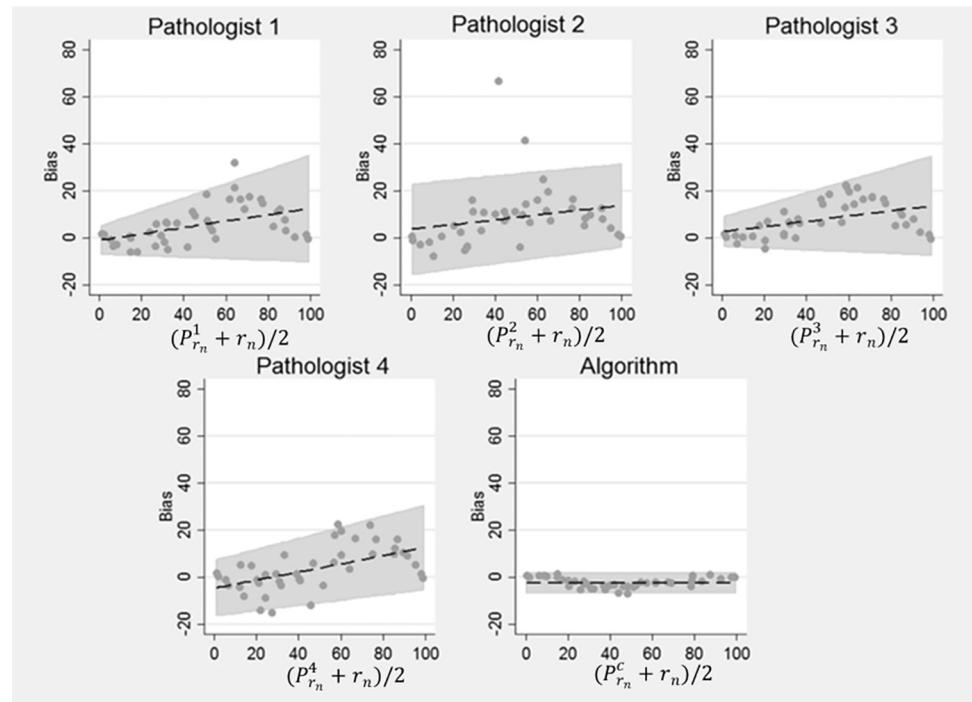


Fig 3. Bland Altman Plots for r_n . Bias and variability of the pathologists' estimates of r_n increased with the percentage of positive nuclei. However, the bias and variability of the computer's estimate was relatively smaller than the pathologists' estimate of r_n .

<https://doi.org/10.1371/journal.pone.0196547.g003>

shows the summary statistics of the four pathologists and the computer algorithm in estimating r_a (the ratio of the true area of positive nuclei to area covered by all nuclei). It is worth mentioning that we never explicitly asked the pathologists to estimate r_a . We used pathologists' estimates of r_n as their surrogate estimate of r_a in computing the performance metrics in Table 4 and our aim here was to check if the pathologists were unconsciously estimating r_a when requested to estimate r_n . CCC and C.I. in all three (Tables 2, 3 and 4) corresponds to Concordance Correlation Coefficient and Confidence Interval, respectively. Similarly Figs 3, 4 and 5 show the results in terms of Bland-Altman plots. These plots enable us to better understand bias and variability among pathologists and the computer algorithms relationship to ground truth.

We also performed a separate analysis for each pathologist to allow the relationships to differ by reader. As expected, a pathologist's estimate of the percentage of positive nuclei can (almost) be perfectly explained by either the true percentage of positive cells r_n and r_a (see Table 5 for details).

III.b Mapping

Table 6 shows the approximation of r_n for the 10 test images. Here \hat{r}_n represents the approximation of r_n through Eq 5. Table 6 was generated using leave-one-out cross validation methodology. Fig 6 shows a plot of D (as dots) for 32 training images along with a function that approximates these values. The horizontal axis in Fig 6 corresponds to r_a while the vertical axis represents D . The objective is to find a function that maps r_a to r_n . The dotted lines in Fig 6 correspond to the 95% confidence interval. The continuous solid line represents a function

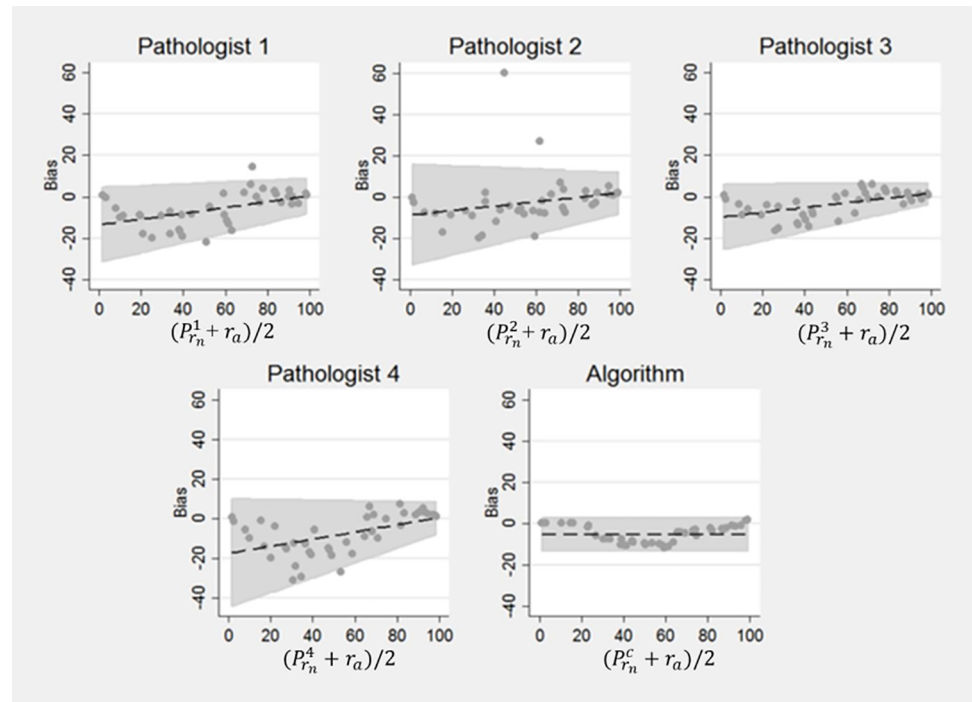


Fig 4. Bland Altman Plots for r_a . The absolute bias and the variability of the pathologists' estimates of r_a decreased with the increasing percentage of positive nuclei. This shows that for images with higher concentrations of positive nuclei, pathologists' estimates deviate from accuracy in ways that are not present in the algorithm's estimates.

<https://doi.org/10.1371/journal.pone.0196547.g004>

that approximates D using the following 2nd degree polynomial:

$$D \approx \Psi(r_a) = a \times (r_a)^2 + b \times r_a + c, \tag{3}$$

where $a = -0.007$, $b = 0.64$, $c = 1.26$

We determined the coefficients a, b, c by solving the following minimizing problem:

$$\min_x \|D(x) - r_a\|_2 \tag{4}$$

Here x corresponds to the vector of coefficients a, b, c , i.e., $x = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$.

This means that r_n can be approximated by substituting D with $\Psi(r_a)$ in Eq 2, i.e.

$$\Psi(r_a) \approx D = r_a - r_n \tag{5}$$

$$r_n \approx r_a - \Psi(r_a) \tag{6}$$

IV. Discussion and conclusions

The bias and agreement of the pathologists relative to the ground truth and those of our computer algorithm [15] are described in Tables 2–4 and the Bland-Altman Plots in Figs 3, 4 and 5. The level of agreement is captured in Tables 2–4 by the value of the CCC. Generally, the

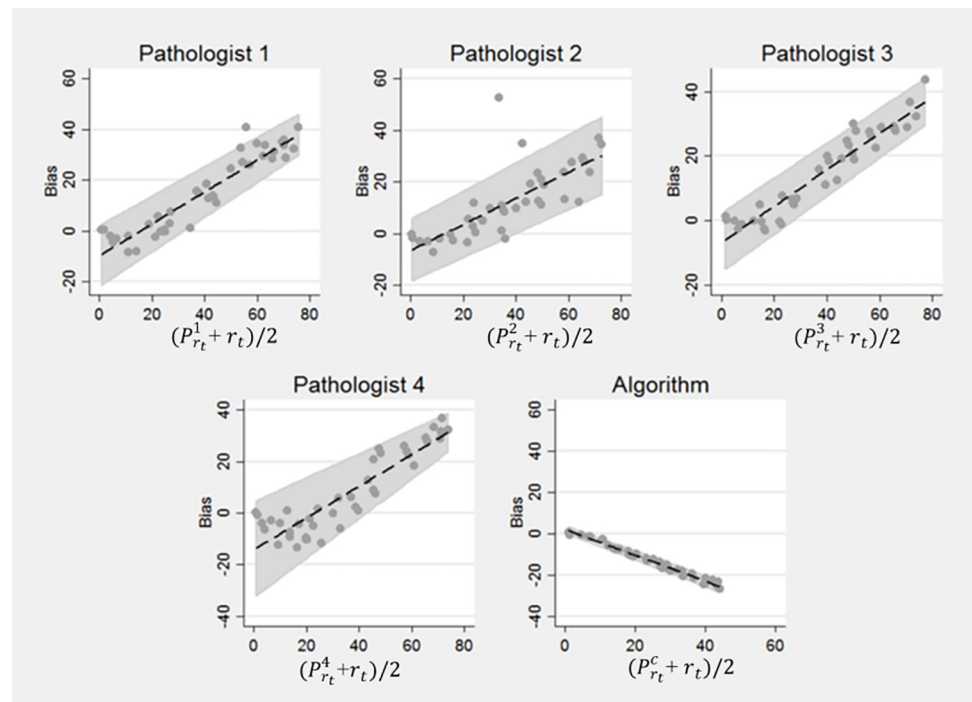


Fig 5. Bland Altman Plots for r_t . There was a strong increasing linear relationship between pathologist bias in estimating r_t and the percentage of positive nuclei. In contrast, there was an almost perfect inverse linear relationship between the bias of the computer's estimates and percentage of positive nuclei.

<https://doi.org/10.1371/journal.pone.0196547.g005>

larger values of CCC indicate better reproducibility between two variables. Pathologist bias and agreement with r_n and r_a were less biased and in better agreement than their estimates of r_t . However, variability and absolute bias of the pathologists' estimates of r_n increased as the number of positive cells increased (Fig 3) while the opposite was true when pathologist estimates were compared to r_a (Fig 4). There was a strong increasing linear relationship between pathologist bias in estimating r_t and the percentage of positive cells (Fig 5). The computer algorithm exhibited low bias and almost perfect agreement with r_n and r_a , but there was an almost perfect decreasing linear relationship between the computer algorithm bias and the percentage of positive nuclei. In contrast to the pathologist estimates, the variability of the computer's estimates was smaller and was independent of the number of positive nuclei. From the summarized statistics in Table 2, it is clear that visual estimations by pathologists are not a reliable means to approximate r_n .

Table 6 suggests that $\Psi(r_a)$ provides a reasonable approximation of r_n . The standard deviation of the differences between approximated values of r_n and true r_n is small (2.68%). State of

Table 5. Multiple regression analysis of pathologist r_n . Total R^2 is the proportion of the total variability explained by the true r_n and r_a . R^2 is the proportion of variability explained by either r_n or r_a alone. Semi-Partial R^2 is the additional proportion of total variability explained by the r_n or r_a beyond the amount explained by the other factor.

Pathologist	Total R^2	r_n		r_a	
		R^2	Semi-Partial R^2	R^2	Semi-Partial R^2
1	0.9449	0.9421	0.0005	0.9444	0.0027
2	0.8190	0.8164	0.0003	0.8187	0.0026
3	0.9699	0.9692	0.0023	0.9677	0.0007
4	0.9333	0.9146	0.0040	0.9293	0.0187

<https://doi.org/10.1371/journal.pone.0196547.t005>

Table 6. Approximation of r_n for the 10 test images. Here \hat{r}_n represents the approximation of r_n through Eq 5.

Test image	$r_a(\%)$	$r_n(\%)$	$\hat{r}_n(\%)$
1	61.99	44.11	46.98
2	49.27	35.18	32.86
3	97.47	97.82	97.89
4	29.92	18.21	15.56
5	42.92	28.68	26.63
6	75.11	58.87	63.83
7	43.82	29.13	27.48
8	55.72	38.95	39.75
9	14.92	8.70	5.62
10	63.95	47.81	49.35
Standard deviation of the difference between r_n and \hat{r}_n			$\pm 2.68\%$

<https://doi.org/10.1371/journal.pone.0196547.t006>

the art nuclei detection algorithms result in around 5–10% false detection rate [13, 17–19]. So, from that perspective, this is an extremely useful result as it 1) has the potential to replace computationally expensive nuclei detection algorithm with inexpensive area estimation, and 2) promises to produce more accurate or at least as good results as any state of the art nuclei detection method. But it still requires extensive validation and testing on a real dataset. At the moment, the synthetic nature of our images allowed us to compute r_a by simply counting the number of positive and negative nuclei. However, the generalization of the mapping function to real data requires a selection of an automatic segmentation algorithm to approximate r_a .

Accurate IHC quantification is essential; improper quantification of IHC stains can deny a patient an important therapy option or result in unnecessary treatment. Unfortunately, IHC quantification can vary significantly between readers, as well as within readers [20, 21], for reasons that are neither fully understood nor easily addressed. Both quality assurance and pathologist training programs designed to standardize IHC stain interpretation suffer from lack of absolute standards with known ground-truth. Our Phantom generator allows creation of unlimited quantities and varieties of digital images that can be used for quality assurance programs, training of pathology residents and development of computer algorithms that will help to solve problem of poor reproducibility of IHC stains interpretation and quantification in clinical practice of pathology.

IHC stains are used to characterize the repertoire of antigens (mostly proteins) that are expressed by malignant cells. Certain antigens are very specific and some are less specific but sensitive in determination of cell types and properties. For example most of carcinoma cells will express various keratin molecules and appear negative for common leukocyte antigen CD45. On the other hand, most leukemia/lymphoma cells will be negative for keratin and will be positive for CD45 antigen. Similarly different antigens are expressed by sarcoma cells. These antigens are negative in carcinoma and lymphoma cells while antigens expressed by carcinoma and lymphoma cells are negative in sarcoma cells. IHC stains are essential in detecting these differences and thus very useful in differentiating between carcinoma, lymphoma and sarcoma cells.

Computationally, this finding suggests savings on computation time, considering that most nuclei detection algorithms [19, 22] take hours (if not days) to approximate the number of nuclei to facilitate the computation of r_n . This mapping function if used in conjunction with the method in [6, 14] has the potential to reduce this time to minutes. However, this function (*i.e.g* $\Psi(A)$) is specific to images of Ki-67 stained follicular lymphoma biopsies. Its generalization to other tumors may require re-estimation of $\Psi(A)$. It is important to perform a detailed analysis

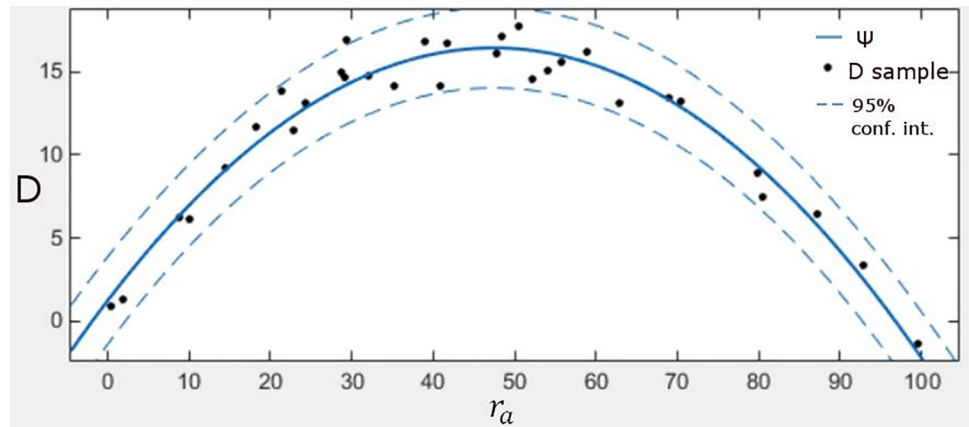


Fig 6. The plot shows a function which facilitates the mapping of $r_a \rightarrow r_n$. The horizontal axis corresponds to different values of r_a while the vertical axis represents D, i.e. error. Each individual dot represents the error between r_a and r_n for the training images. The solid line corresponds to the mapping function Ψ which facilitates the mapping of r_a to r_n while the dotted lines represent the confidence interval.

<https://doi.org/10.1371/journal.pone.0196547.g006>

of Fig 6 to understand the quadratic behavior of the mapping function, Ψ . When there are only a few positive nuclei, it results in the smallest error. This is evident from the y-axis of the plot in Fig 6; the errors are small when there are few positive nuclei and increase as the percentage of positive nuclei approaches 50%. The function starts to decrease in a quadratic fashion after this point until it reaches a point where the image only contains positive nuclei. So, it is essentially a function of homogeneity; it results in least amount of error when the image contains one type of nuclei. The function increases with the increase in heterogeneity.

By generating a series of virtual tissues (phantoms) with different proportions and distribution of positive and negative cells, the exact number of positive and negative cells is known, therefore, it can be used as a gold standard for the following purposes:

- An IHC standard for the industry to test computer algorithms for enumeration of IHC stained cells.
- IHC standards for pathology quality assurance programs such as those administered by 9AP and similar organizations for programs used to standardize breast cancer pathology, lung cancer pathology, lymphoma pathology, etc.
- Using these Phantom standards with different proportions of positive and negative cells, one can print hundreds of high resolution slides on glass that can be used for testing with light microscopy and for testing of high resolution slide scanners.
- By further expanding this model, one can generate 3D phantoms of tissue with different proportions of negative and positive cells. Using 3D printers, one can print artificial tissue using a cartridge of collagen, or other matrix, with a positively and negatively stained suspension of cells. These 3D printed tissues could be used to standardize histology processing of tissue fixation and tissue cutting. Using 3D printers with cartridges of collagen or other Matrix and unstained cells with a known immunophenotype, one can generate (print) 3D artificial tissues that can be used as a standard for tissue fixation, tissue processing, IHC staining using different IHC platforms, image acquisition, and image IHC analysis.

The main objective of this study was to provide a reliable approach to generate ground truth in the study and analysis of IHC stained tissues. In most studies of IHC stained images,

obtaining the ground truth is a major challenge due to various factors such as limited availability of experts to perform annotations, inter- and intra reader-variability and limited datasets. Often, the ground truth is either a generalization of reader consensus or the average of readings among multiple readers which may be biased, hence disputed. This method allows generating virtual images mimicking IHC stained tissue with exact proportions of positive and negative nuclei with controlled distribution of stained nuclei. These virtual tissue images can be used as the gold standard for IHC quantification for both computer based image analysis methods and quality assurance programs for manual evaluation of IHC stained tissue by pathologists. It creates virtual IHC stained sections with a known percentage of positive and negative cells and eliminates bias that is associated with IHC standards where the ground truth is based on manual counting of positive nuclei by pathologists. The study also showed that it is possible to approximate the ratio of positive to total number of nuclei from areas of positive and negative nuclei. The study also evaluated the performance of one of our recently developed nuclei detection algorithms [15].

In the future, we are planning on generating whole slide images which contain structures like blood vessels and other types of nuclei. We are also planning on extending it to other disease types including lymphoma, different types of breast cancer, lung cancer, prostate cancer, etc. For each tumor type phantom tissue can be further complicated by introduction of other objects such as lymphocytes, histiocytes, blood vessels, nerve bundles, fibrotic fibers, and artifacts such as hemorrhage of necrosis.

The current approach lacks the ability to generate complex structures like ducts and lobules in breast. Moreover, it is relatively hard to generate a particular distribution of nuclei with the current design. Similarly, it requires a bit of tuning if the objective is to orient nuclei in a certain direction.

Supporting information

S1 File. Interactive Image Compression for Big Data Image Analysis: Application to Hot-spot Detection in breast cancer.

(PDF)

Acknowledgments

The project described was supported in part by Awards Number R01CA134451 (PIs: Gurcan, Lozanski), U24CA199374 (PIs: Gurcan, Madabushi, Martel), and U01 CA198945 (PI: Bilgin) from the National Cancer Institute, and OSU CCC Pelotonia Award. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute, or the National Institutes of Health.

Author Contributions

Conceptualization: Metin N. Gurcan.

Data curation: Muhammad Khalid Khan Niazi, Robert Hasserjian, Abner Louissaint, Jr., Arwa Shana'ah, Gerard Lozanski.

Formal analysis: Muhammad Khalid Khan Niazi, Michael Pennell, Berkman Sahiner, Weijie Chen, Metin N. Gurcan.

Funding acquisition: Metin N. Gurcan.

Methodology: Metin N. Gurcan.

Project administration: Metin N. Gurcan.

Software: Muhammad Khalid Khan Niazi, Fazly Salleh Abas, Caglar Senaras.

Validation: Muhammad Khalid Khan Niazi, Caglar Senaras, Michael Pennell, Berkman Sahiner, Weijie Chen, Robert Hasserjian, Abner Louissaint, Jr., Arwa Shana'ah, Metin N. Gurcan.

Visualization: Muhammad Khalid Khan Niazi.

Writing – original draft: Muhammad Khalid Khan Niazi, Michael Pennell, Metin N. Gurcan.

Writing – review & editing: Muhammad Khalid Khan Niazi, Fazly Salleh Abas, Caglar Senaras, Michael Pennell, Berkman Sahiner, Weijie Chen, John Opfer, Robert Hasserjian, Abner Louissaint, Jr., Arwa Shana'ah, Gerard Lozanski, Metin N. Gurcan.

References

1. Zaha DC. Significance of immunohistochemistry in breast cancer. *World journal of clinical oncology*. 2014; 5(3):382. <https://doi.org/10.5306/wjco.v5.i3.382> PMID: 25114853
2. Niazi MKK, Downs-Kelly E, Gurcan MN, editors. Hot spot detection for breast cancer in Ki-67 stained slides: image dependent filtering approach. *SPIE Medical Imaging*; 2014 2014: International Society for Optics and Photonics.
3. Di Cataldo S, Ficarra E, Acquaviva A, Macii E. Automated segmentation of tissue images for computerized IHC analysis. *Computer methods and programs in biomedicine*. 2010; 100(1):1–15. <https://doi.org/10.1016/j.cmpb.2010.02.002> PMID: 20359767
4. de Matos LL, Trufelli DC, de Matos GL, da Silva Pinhal A. Immunohistochemistry as an important tool in biomarkers detection and clinical practice. *Biomarker insights*. 2010; 5:9. PMID: 20212918
5. Taylor C, Levenson RM. Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment II. *Histopathology*. 2006; 49(4):411–24. <https://doi.org/10.1111/j.1365-2559.2006.02513.x> PMID: 16978205
6. Niazi MKK, Pennell M, Elkins C, Hemminger J, Jin M, Kirby S, et al., editors. Entropy based quantification of Ki-67 positive cell images and its evaluation by a reader study. *SPIE Medical Imaging*; 2013: International Society for Optics and Photonics.
7. Niazi MKK, Downs-Kelly E, Gurcan MN, editors. Hot spot detection for breast cancer in Ki-67 stained slides: image dependent filtering approach. *SPIE Medical Imaging*; 2014: International Society for Optics and Photonics.
8. Niazi MKK, Satoskar AA, Gurcan MN, editors. An automated method for counting cytotoxic T-cells from CD8 stained images of renal biopsies. *SPIE Medical Imaging*; 2013: International Society for Optics and Photonics.
9. Revkin SK, Piazza M, Izard V, Cohen L, Dehaene S. Does subitizing reflect numerical estimation? *Psychological Science*. 2008; 19(6):607–14. <https://doi.org/10.1111/j.1467-9280.2008.02130.x> PMID: 18578852
10. Mandler G, Shebo BJ. Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, 111, 1–22. *Journal of Experimental Psychology: General*. 1982;111(1).
11. Frith CD, Frit U. The solitaire illusion: An illusion of numerosity. *Attention, Perception, & Psychophysics*. 1972; 11(6):409–10.
12. Belhomme P, Toralba S, Plancoulaine B, Oger M, Gurcan MN, Bor-Angelier C. Heterogeneity assessment of histological tissue sections in whole slide images. *Computerized Medical Imaging and Graphics*. 2015; 42:51–5. <https://doi.org/10.1016/j.compmedimag.2014.11.006> PMID: 25475487
13. Xu H, Lu C, Berendt R, Jha N, Mandal M. Automatic nuclei detection based on generalized laplacian of gaussian filters. *IEEE Journal of Biomedical and Health Informatics*. 2016; 21(3):826–37. <https://doi.org/10.1109/JBHI.2016.2544245> PMID: 28113876
14. Niazi MKK, Lin Y, Liu F, Ashoka A, Marcellin M, Tozbikian G, et al. Interactive Image Compression for Big Data Image Analysis: Application to Hotspot Detection in Breast Cancer. Submitted for Journal publication. 2017.
15. senaras; C, Pennell; M, Chen; W, Sahiner; B, Shana'ah; A, Louissaint; A, et al., editors. FOXP3-stained image analysis for follicular lymphoma: Optimal adaptive thresholding with maximal nucleus coverage. *SPIE Medical Imaging* 2017; Orlando, USA.

16. Pătrăucean V, Gurdjos P, Von Gioi RG. A parameterless line segment and elliptical arc detector with enhanced ellipse fitting. *Computer Vision—ECCV*. Springer; 2012. p. 572–85.
17. Chen T, Ched'hotel C, editors. Deep learning based automatic immune cell detection for immunohistochemistry images. *International Workshop on Machine Learning in Medical Imaging*; 2014: Springer.
18. Mao Y, Yin Z, Schober JM, editors. Iteratively training classifiers for circulating tumor cell detection. *IEEE 12th International Symposium on Biomedical Imaging (ISBI)* 2015: IEEE.
19. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*. 2009; 2:147–71. <https://doi.org/10.1109/RBME.2009.2034865> PMID: 20671804
20. Hermans J, Krol A, Van Groningen K, Kluin PM, Kluin-Nelemans J, Kramer M, et al. International Prognostic Index for aggressive non-Hodgkin's lymphoma is valid for all malignancy grades. *Blood*. 1995; 86(4):1460–3. PMID: 7632953
21. Metter GE, Nathwani BN, Burke JS, Winberg CD, Mann RB, Barcos M, et al. Morphological subclassification of follicular lymphoma: variability of diagnoses among hematopathologists, a collaborative study between the Repository Center and Pathology Panel for Lymphoma Clinical Studies. *Journal of Clinical Oncology*. 1985; 3(1):25–38. <https://doi.org/10.1200/JCO.1985.3.1.25> PMID: 3965631
22. Sirinukunwattana K, Raza SEA, Tsang Y-W, Snead DR, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE transactions on medical imaging*. 2016; 35(5):1196–206. <https://doi.org/10.1109/TMI.2016.2525803> PMID: 26863654