HUMAN BIOLOGY
2024, VOL. 94, ISSUE 4
ORIGINAL ARTICLE

OPEN ACCESS

Consecutive Cytokeratin Immunochemistry-Supervised Algorithm for Predicting Tumor Areas in Ki67 Breast Cancer Images

Chien-Hui Wu[1], Min-Hsiang Chang[1], Hsin-Hsiu Tsai[2], Mei-Lin Yang[2], and Yi-Ting Peng[2]*

ABSTRACT

Automatic Ki67 index (KI) assessment has become popular in breast cancer research; however, the results are easily influenced by non-tumor cells. This can be addressed by using neural networks to predict tumor areas. Compared to human annotation, cytokeratin immunostaining can more accurately highlight epithelial regions and provide reliable ground truth. We built an immunohistochemistry (IHC)-supervised neural network using the ground truth extracted from consecutive cytokeratin-stained slides, which could predict the tumor area in Ki67 images of breast cancer. The effect of masks on KI quantification was evaluated in 20 patients with breast carcinoma. Set A (three cases) was used to measure the similarity of adjacent whole-slide images (WSIs). A UNet++ (with an EfficientNet-b7 backbone) model was built using Set B (67 cases) for tumor area prediction. The KI in Set C (20 cases) was quantified with and without the application of tumor-area masks, and the KI difference was computed. The mean intersection over union of the epithelial masks extracted from adjacent cytokeratin sections was 0.72 (0.68–0.76). After training and validating in 49 cases, the intersection over union in the test set was 0.44–0.73. At the tile image-level, KI difference was −42.5–41.7%. Images with the highest difference usually contained numerous lymphocytes or vessels, and the masks prevented disguised cells from being counted. At the WSI-level, the hotspot location changed in 18/20 cases, but hotspot KI changed insignificantly (−1.0% on average). The global KI changed less (0.9% on average). Thus, consecutive IHC staining provides substantial, precise, and reliable ground truths that trained the algorithm efficiently. This IHC-supervised training workflow can be applied to other targets by replacing IHC antibodies. Furthermore, the virtual tumor areas improved Ki67 counting by excluding the non-tumor areas at the tile image-level, although the effect on the WSI-level was insignificant.

INTRODUCTION

Ki67 is a protein expressed in the nucleus during all phases of the cell cycle, except G0. The Ki67 index (KI), i.e., the percentage of immunoreactive tumor cells among tumor cells, is a measure of tumor proliferation Gerdes et al. (1984). KI is widely used in the subclassification of hormone-positive breast carcinomas. With a cutoff point of 14%, tumors are classified as luminal type A if the KI is < 14%. Otherwise, the lesions are classified as luminal type B Goldhirsch et al. (2011). This helps clinicians to decide whether chemoradiotherapy is appropriate for the patient Thomssen et al. (2021), Nielsen et al. (2021). Thus, it is critical that KI is counted precisely. However, in most cases in daily routine, the KI is visually estimated by pathologists. Even if it is carefully counted according to the standard international protocol, intraobserver and interobserver variations remain considerable Dowsett et al. (2011), Polley et al. (2015). In recent decades, digital image analysis has increasingly been adopted by pathologists and researchers for automatic KI assessment, using either commercially available or public-domain image analysis software Tuominen et al. (2010), Acs et al. (2019), Abubakar et al. (2019). In breast carcinoma, using an automated scoring system, the intraclass correlation coefficient for average KI scores across 14 laboratories can reach 0.83 (highly correlated), and even higher (0.89) for laboratories using scanners from a single vendor Acs et al. (2019), Rimm et al. (2019). This indicates that assessing KI by means of algorithms is more reliable than assessing it based on human experience alone. The software often relies on color deconvolution algorithms to detect DAB chromogen-stained (brown) and hematoxylin-stained (blue) cell nuclei Ruifrok et al. (2001), Saha et al. (2017), Lakshmi et al. (2020). However, the staining results of the nuclei of tumor cells and non-tumor cells, such as lymphocytes and stromal cells, can cause KI to be underestimated or overestimated. Therefore, when manipulating these applications, pathologists are required to pre-annotate invasive carcinoma areas to avoid interference from non-tumor cells and optimize KI quantification Volynskaya et al. (2019). To address this problem, some algorithms have attempted to exclude non-tumor cells based on

[1]Department of Anatomical Pathology, Far Eastern Memorial Hospital, No. 21, Section 2, Nanya S. Road, Banqiao District, New Taipei City, 220, Taiwan.
[2]AI Lab, Quanta Computer Inc., No. 211, Wenhua 2nd Rd., Guishan Dist., Taoyuan City 333, Taiwan.

Correspondence to: Min-Hsiang Chang, Department of Anatomical Pathology, Far Eastern Memorial Hospital, No. 21, Section 2, Nanya S. Road, Banqiao District, New Taipei City, 220, Taiwan.
E-mail: magieva_tw@protonmail.com, Tel: +886-921546377.

their shape and size; however, this approach has not achieved satisfactory results Valkonen et al. (2020). Another solution is the virtual double-staining method: Ki67-stained images are analyzed using a cytokeratin (CK) mask extracted from an adjacent CK-stained slide Koopman et al. (2018), Røge et al. (2016), Wessel Lindberg et al. (2017). Although technically feasible, this method requires consecutive CK-stained slides and is, therefore, unsuitable for daily clinical practice. A recently reported solution involves training a neural network using fluorochromogenic CK-Ki67 double-staining to predict CK masks from Ki67-stained images, which provides a convenient alternative to the ground truth annotated manually by an expert Valkonen et al. (2020). This neural network could predict the epithelial regions in Ki67-, estrogen receptor-, and progesterone receptor-stained slides and refine the quantification results. However, fluorochromogenic CK–Ki67 double-staining costs much more than single staining and is not efficient for deep learning, which consumes a lot of data. Inspired by these innovative studies, we observed the feasibility of immunohistochemistry (IHC-supervised) learning Valkonen et al. (2020), Brázdil et al. (2022), Hong et al. (2021). IHC-supervised learning trains a neural network to predict the target-stained mask, given a source-stained image. In this case, the ground truth was obtained from IHC rather than from human annotations, which has several advantages, such as reliability, precision, and efficiency. First, IHC highlights and distinguishes cells according to their unique proteins, which indicates the nature of the cells, rather than according to morphology, on which pathologists rely heavily. Secondly, the ground truth extracted from an IHC image is intricate and detailed. Finally, and most critically, for data-consuming deep learning, a large amount of IHC ground truth can be provided quickly and economically.

In this study, to explore the potential of IHC-supervised learning, we combined two virtual double-staining methods. A neural network was developed with consecutive CK immunostaining and was used to predict the tumor area in breast cancer Ki67 images. Finally, we evaluated the effect of tumor-area masks on KI quantification.

MATERIAL AND METHODS

This study was approved by the Research Ethics Review Committee of the Far Eastern Memorial Hospital (No. 111205-F/DPAI-005). All slides and blocks in the study were used in a clinical diagnosis setting, were retrieved from the repository, and did not contain any personal information. All methods were performed in accordance with the relevant guidelines and regulations. Therefore, the requirement for obtaining informed patient consent was waived by Research Ethics Review Committee of the Far Eastern Memorial Hospital.

Case collection

We collected data on 94 breast carcinoma cases diagnosed at the Department of Anatomical Pathology of Far Eastern Memorial Hospital between January 2020 and September 2022. Patients with lobular carcinoma, mucinous carcinoma, and tumor areas with < 10 high-power fields were excluded. Thus, 90 patients were eventually included in the study. Among these cases, three (Set A) were used to measure the similarity of adjacent whole-slide images (WSIs), 67 (Set B) were used for IHC-supervised learning and testing, and 20 (Set C) were used to evaluate the effect of tumor-area masks on KI counting. Furthermore, considering a widely adopted usage scenario in which users have already selected the tumor area for model prediction, it is more practical to focus on images containing tumors. To analyze mask performance in the tumor area, Set B-1 was built from the test set of Set B. First, the WSIs were divided into 1,936 × 1,216 image tiles. The tile size was equivalent to a resolution of 1080 p (full high-definition monitors and digital camera for optical microscope), which is commonly used in hospitals, and our KI quantification algorithm was designed for this image size. Second, if a tile contained over 10% of the tumor area, then it was added to Set B-1. Table 1 shows the details of all cases.

Specimen preparation, staining, and scanning

The specimens used in the routine diagnostic setting were retrieved from the repository of the Anatomical Pathology Department of the Far Eastern Memorial Hospital. Tumor tissue was fixed in 10% neutral-buffered formalin and was embedded in paraffin. Tissues were cut into 5-μm-thick sections and mounted on hydrophilic slides. Ki67 and CK immunostaining was performed using the Benchmark Ultra automated staining system (Ventana Medical Systems, Oro Valley, AZ, USA). The slides were heated to 96℃ for 34 min for antigen retrieval, followed by incubation with anti-Ki67 (30-9, Ventana Medical Systems) or anti-CK (AE1/AE3, Ventana Medical Systems) antibody and the ultra-View Universal DAB Detection Kit (Ventana Medical Systems). Hematoxylin was used as a counterstain. The slides were scanned using a Hamamatsu S210 microscope (Hamamatsu Photonics, Hamamatsu, Japan) at 40× magnification (0.23 μm per pixel resolution). WSIs were saved in NDPI format.

Data preparation for IHC-supervised learning

To facilitate training of the neural network by IHC-supervised learning, Ki67-stained WSIs and their consecutive CK-stained WSIs in Set B were processed, including image alignment, tumor area annotation, and ground truth mask extraction (Figure 1). As tissue shapes and locations in consecutive slides were similar but not perfectly identical. We selected CK images as references and then rotated and translated the

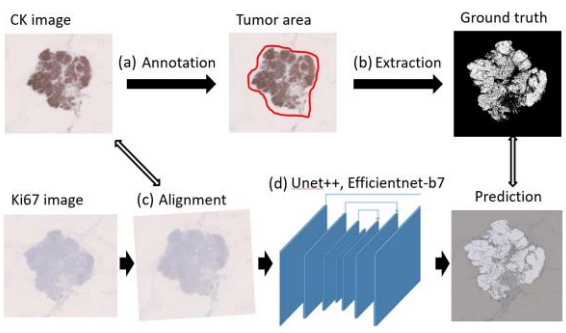
Table 1: Data details.

Case set	Case count	Specimen type	Histological grade			Slides
		Excision/Biopsy	1	2	3	
Total	90		13	52	25	
Set A	3	3/0	1	1	1	Pairs of consecutive CK slides
Set B	67	41/26	9	40	18	
Training	36	20/16	5	23	8	Pairs of Ki67 and corresponding CK slides
Validation	13	09-Apr	2	8	3	
Testing / Set B-1	18	12-Jun	2	9	7	
Set C	20	20/0	3	11	6	Ki67 slides

CK, cytokeratin

corresponding Ki67 images to align to these references. Ki67 and CK images were rescaled with scale factors equal to 1/500, 1/100, 1/50, and 1/7 for varying resolutions, and binarized for simplification. Then, the optimal angle and offsets of the x- and y-axes were found through gradient descent to optimize the loss between Ki67- and CK-stained binary images. Second, the rough tumor area was annotated to exclude unwanted regions, such as the skin, benign breast glands, or inks on the margins. Annotation was performed by a licensed pathologist (CMH) using the Smart Pathology System (version 1.2.0; Quanta Computer, Taoyuan, Taiwan). Finally, masks were extracted from the CK images and intersected with human-annotated tumor areas to construct the ground truth for supervised learning. Color deconvolution was introduced to extract the CK masks, followed by a Gaussian filter, which is often used for denoising. Subsequently, the Otsu thresholding algorithm was used to transform the masks into binary masks.

Figure 1: Workflow of data annotation and algorithm training. (a) Tumor area was annotated by a pathologist (red circle) in cytokeratin (CK) whole slide images (WSIs). (b) CK mask was extracted and intersected to human annotation. (c) Input Ki67 image was aligned to the CK image. (d) UNet++ network, using EfficientNet-b7 as the backbone, was trained to predict the tumor area in Ki67 WSIs.



Owing to the large size of WSIs, GPU memory is a critical issue. To avoid this issue, the WSIs were compressed 1/7 times on each side and divided into image tiles of 276 × 172 (1/7 of 1,936 × 1,216 approximately) pixels for training. This is the same shape commonly used in monitors and digital cameras for optical microscopes. During inference, a WSI was divided into image tiles, and the predictions of these tiles were stitched back in the same order to reconstruct the whole-slide tumor area mask.

Measurement of similarity of adjacent WSIs

In IHC-supervised learning, the ground truth was obtained from adjacent slides of the input images. Therefore, before model development, we measured the similarity between adjacent WSIs. To this end, we extracted CK masks from pairs of adjacent CK-stained WSIs in Set A and aligned these masks using a previously described method (Figure 1b and c). The similarity of these masks was measured using intersection over union (IoU). The IoU was calculated as the intersection area of the two masks divided by their union area Valkonen et al. (2020). The IoU approaches 1 as the overlapping area approximates 100%.

IHC-supervised learning

The algorithm architecture was UNet++ based on an EfficientNet-b7 backbone (Figure 1d). The core framework was based on the traditional UNet model, and UNet++ advanced this model by incorporating a nested architecture inspired by successful features of other networks, notably the feature pyramid network. Zhou et al. (2018). This enhanced structure can utilize skip connections more effectively, thus optimizing the model for complex tasks in medical image segmentation. Exponential moving average and augmentation methods, including flipping, translation, rotation, scaling, and color jittering, were applied during training to improve the robustness of

of the model. IoU was chosen to evaluate the accuracy of comparing the predicted tumor-area masks with the corresponding ground truth masks.

KI counting and visual validity assessment

We used another 20-breast carcinoma Ki67 slides (Set C) to evaluate the influence of tumor-area masks on Ki67 quantification. The KI was analyzed from two different perspectives. At the image tile-level, the WSIs were divided into image tiles of 1,936 × 1,216 pixels (Figure 2). Using the Smart Pathology System, KIs before (version 1.2.0) and after (version 1.3.0) the application of tumor-area masks was computed. According to the global method recommended by the International Ki67 in Breast Cancer Working Group, representative fields should contain at least 100 tumor cells for scoring. Therefore, image tiles containing fewer than 100 tumor cells were excluded. In addition, we applied the concept of "Pareto hotspot" Plancoulaine et al. (2015), in which the 90th percentile of the entire KI range rather than the highest one is chosen as the hotspot KI. Another technical reason is that the previous algorithm (ver 1.2.1) easily misrecognized red blood cells as positive tumor cells, which caused extraordinarily high KIs in non-tumorous vascular regions. As a result, KI values higher than the average KI plus two standard deviations were considered outlier data and were excluded from the analysis. In each case, 10 tiles with the highest KI difference (200 tiles in total) were evaluated visually by four pathologists. The pathologists gave mask quality scores ranging from 1 to 5 (1 being totally inaccurate and 5 being a good tumor mask and improve KI quantification).

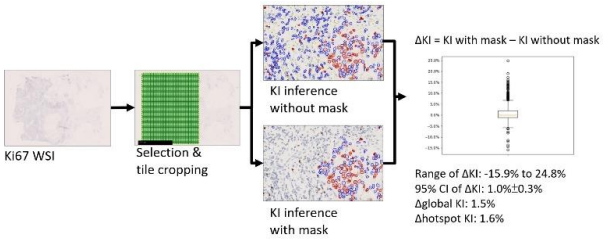
At the WSI-level, since comprehensive KI quantification of whole-tumor sections is possible, we could directly set the global KI as the KI of the entire tumor in the WSI, unlike in the global method of the International Ki67 in Breast Cancer Working Group (either unweighted or weighted). Hotspot KI was also calculated. Here, a hotspot refers to the tile with the highest KI in the selected tiles of WSI. The related analyses are presented in Table 2 and Table S1. Python software (ver. 3.6.10; https://www.python.org/) was used to calculate the KI data. Excel (Microsoft Corp., Redmond, WA, USA) was used for the statistical analysis of the difference in KI (ΔKI = KI with masks – KI without masks, Figure 2).

RESULT

Analysis of adjacent WSIs

The mean IoU for masks of adjacent slides in Set A was 0.72 (range: 0.68–0.76). This indicated that the upper bound of the IoU between two adjacent slides was approximately 0.72. In set B, by color deconvolution, the IoU between hematoxylin-stained area in Ki67 WSI and DAB-stained CK WSI was 0.60 in average (range: 0.09-0.9). Combine with the study by Valkonen et al. (2020), their IoU could reach 0.7, we considered an IoU of 0.58, which is 80% of 0.72, as the endpoint of the IHC-supervised learning.

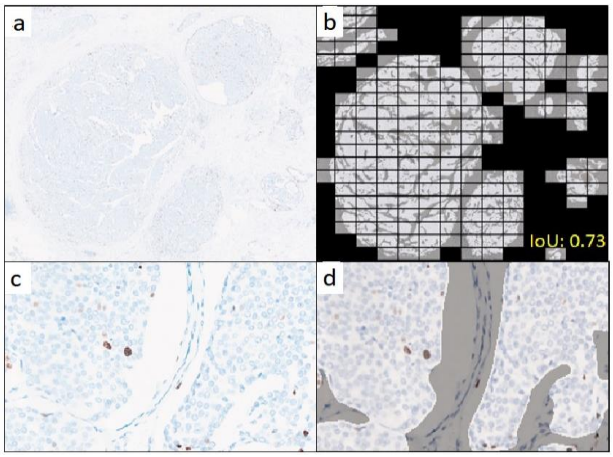
Figure 2: Ki67 index (KI) difference (ΔKI) analysis. The tumor region in Ki67-stained whole slide images (Case 9 as an example) was selected and divided into tiles. Each tile was inferenced with and without masks. Then, the ΔKIs were computed and summarized as a boxplot.



Evaluation of tumor area-predicting algorithm

After training, validation, and testing of Set B, the IoUs in validation set ranged from 0.02 to 0.79 (95% confidence interval: 0.56 ± 0.13), and in the test set ranged from 0.36 to 0.69 (95% confidence interval: 0.57 ± 0.05). For testing set B-1, the IoUs ranged from 0.44 to 0.73 (95% confidence interval: 0.62 ± 0.04, Figure 3), which reached the endpoint of neural network development.

Figure 3: Tumor area predicted by the neural network. (a) Overview of a breast carcinoma test case from Set B. (b) Predicted tumor regions of the corresponding test case from Set B-1. The intersection of the union (IoU) is 0.73. (c) Zoomed-in image. (d) Predicted tumor regions of the zoomed-in image.



Effect of predicted tumor-area masks in Ki67 counting

At the tile image level, ΔKI ranged from −42.5% to 41.7% (Table 2 and Figure 4a). In the visual validation by pathologists, the mean scores were 4.14 (σ = 1.22), 3.75 (σ = 1.43), 4.56 (σ = 0.77), and 4.01 (σ = 0.81) (Figure 4c). The images with large differences

Table 2: Ki67 index assessment with and without tumor-area masks. After applying the predicted tumor-area masks, the number of selected tiles decreased because the masks eliminated more non-tumor areas.

		Without mask			With mask			KI Difference			
Case	Tile Numbers	Selected tiles	Global KI	Hotspot KI	Selected tiles	Global KI	Hotspot KI	95%CI of Δ KI in tiles	Range of Δ KI in tiles	Δ global KI	Δ hotspot KI
1	305	184	8.50%	20.40%	183	8.00%	20.3%*	-0.5%±0.1%	-4.4%~2.1%	-0.50%	-0.10%
2	1295	683	14.2%	25.30%	653	13.9%	25.50%	-0.3%±0.1%	-7.5%~10.4%	-0.30%	0.10%
3	3290	2390	9.90%	35.70%	2332	9.80%	30.50%	-0.5%±0.1%	-42.5%~15.3%	-0.10%	-5.20%
4	2420	1513	25.7%	46.50%	1503	28.0%	49.50%	-2.1%±0.2%	-35.8%~28.3%	2.30%	3.10%
5	660	423	3.40%	10.30%	420	3.40%	10.70%	0.2%±0.1%	-4.1%~5.6%	0.10%	0.40%
6	2376	1512	8.10%	20.10%	1486	7.90%	18.80%	-0.2%±0.1%	-12.0%~8.2%	-0.20%	-1.30%
7	1242	621	23.1%	46.70%	607	25.6%	52.60%	2.0%±0.3%	-7.3%~28.4%	2.50%	5.90%
8	3066	733	12.2%	29.40%	712	12.5%	28.50%	0.2%±0.1%	-9.3%~16.7%	0.20%	-0.90%
9	3003	724	21.6%	49.50%	704	23.1%	51.10%	1.0%±0.3%	-7.3%~24.8%	1.50%	1.60%
10	1518	865	7.70%	20.80%	830	7.50%	15.70%	0.2%±0.1%	-4.8%~9.5%	-0.20%	-5.10%
11	1296	512	9.70%	21.10%	502	9.20%	20.40%	-0.6%±0.1%	-6.6%~3.4%	-0.60%	-0.70%
12	1650	616	7.60%	43.80%	607	7.80%	29.50%	0.0%±0.2%	-29.1%~5.0%	0.20%	-14.30%
13	2600	1356	26.4%	53.10%	1355	35.0%	66.70%	9.2%±0.5%	-32.1%~41.7%	8.60%	13.60%
14	1363	676	11.4%	22.20%	660	11.1%	18.90%	-0.3%±0.2%	-35.6%~7.7%	-0.30%	-3.30%
15	2829	1464	25.1%	49.20%	1447	26.0%	51.40%	0.6%±0.1%	-8.7%~16.8%	0.90%	2.20%
16	551	342	20.1%	37.20%	340	21.9%	39.10%	1.3%±0.3%	-3.4%~22.1%	1.70%	1.90%
17	3080	1790	30.6%	52.80%	1761	32.4%	53.90%	1.3%±0.1%	-24.2%~24.4%	1.80%	1.00%
18	1733	866	7.70%	37.10%	850	7.40%	21.20%	-0.5%±0.1%	-33.2%~7.3%	-0.30%	-16.00%
19	1070	393	2.70%	14.40%	344	3.00%	14.6%*	-0.1%±0.1%	-3.1%~3.4%	0.30%	0.20%
20	1649	968	17.4%	39.40%	618	16.9%	35.10%	-0.1%±0.3%	-30.5%~28.0%	-0.40%	-4.20%
95%CI			14.6% ± 3.9%	33.7% ± 6.4%		15.5% ± 4.6%	32.7% ± 7.6%			0.9% ± 1.0%	-1.0% ± 3.0%

* Hotspot location did not change after masking in these two cases.

(Figure 5) contained numerous lymphocytes or red blood cells (RBCs), which were easily misrecognized as negative or positive tumor cells before masking. The predicted tumor areas excluded lymphocyte-rich stroma and blood vessels, preventing inclusion of these disguised cells in the Ki67 count.

At the WSI-level, the masking effect was less significant. Δglobal KI values ranged from −0.6% to 8.6% (95% confidence interval: 0.9% ± 1.0%). Δhotspot KI values ranged from −16.0% to 13.6% (95% confidence interval: −1.0% ± 3.0%). Further details are shown in Figure 4b. The location of the hotspots (Figure 6) changed in 18 of the 20 cases, with the exclusion of Cases 1 and 19.

For clinical classification, we used 14% as the threshold for luminal type A and luminal type B. Based on global KI before masking, 9 of 20 patients were in the luminal type B group, and Case 2 was reclassified as luminal type A (from 14.2% to 13.9%) after applying the masks. In contrast, based on hotspot KI, 19 of 20 cases were in the luminal type B group, and none were downgraded due to mask application.

Figure 4: Ki67 Index (KI) difference (ΔKI) after applying virtual tumor masks at tile image level and at whole slide image (WSI) level and the visual validity assessment. (a) Boxplot showing ΔKI in the tiles of each case. Case 13 showed the largest change. (b) Graph exhibiting the distribution of ΔGlobal KI and ΔHotspot KI for all 20 cases. (c) Four pathologists gave mask quality scores ranging from 1 to 5 in 200 image tiles with highest ΔKI in Set C.

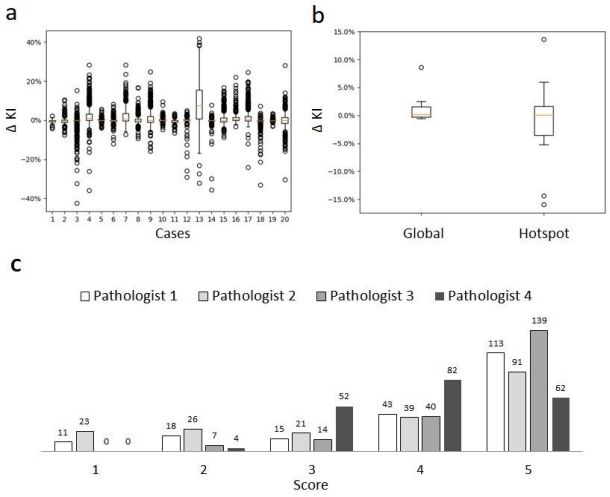


Figure 5: Ki67 assessment before and after masking. With tumor-area masks, most of the non-tumor cells were filtered and the Ki67 Index (KI) could be counted more precisely. Here, cells are categorized into positive tumor cells (red), negative tumor cells (blue), and stromal cells (yellow). Stromal cells are not visualized after masking.

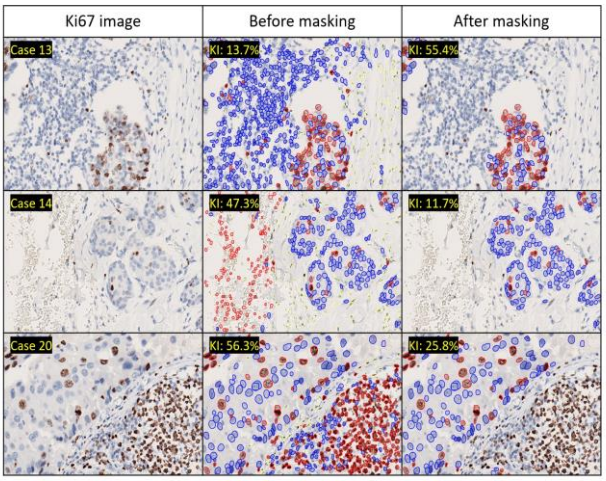
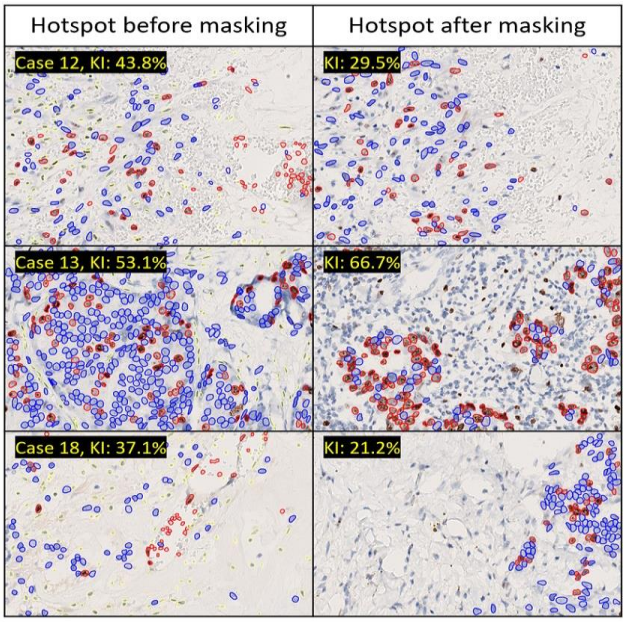


Figure 6: Hotspot changes before and after masking.



DISCUSSION

In this study, we built an IHC-supervised neural network for predicting tumor areas in Ki67 breast cancer images, which enhanced automatic KI assessment. To develop image analysis algorithms in digital pathology, we usually rely on pathologist annotation as the ground truth, which involves labor-intensive preparation of sufficient data for neural network construction. In addition, human-labeled ground truth cannot be precise at the pixel level, particularly for indicating the borders of complex areas, such as tumor edges or nuclear contours. Limited and inaccurate ground truth datasets hinder deep learning. To improve the quantity and quality of the ground truth, we chose to use consecutive CK immunostaining, which can highlight the epithelial area more reliably, accurately, and efficiently than can be achieved by humans Valkonen et al. (2020), Brázdil et al. (2022), Hong et al. (2021). In our approach, pathologists only depicted the tumor area vaguely and roughly to

exclude non-tumor epithelial areas, such as the skin and most benign tissues. By combining this with the masks extracted from CK images, we were able to harvest a large number of fine-quality datasets in a short period.

High-quality datasets accelerated the development of our neural network. In digital pathology image analysis, a maturing neural network usually involves hundreds or thousands of cases. However, in this study, we only fed data of 49 cases into the network and were able to achieve our training goal (the IoU reached 0.62 on average, which was 86% of the upper bound). These results showed that IHC-supervised learning is not only efficient in data collection but also in algorithm building. Despite the promising results of our algorithm, another algorithm created by Valkonen et al. (2020), which also predicted epithelial area in Ki67 stains of breast cancers, achieved a mean IoU of up to 0.69, which was higher than ours. This could be because they used a fluoro-chromogenic dual stain (Ki67 and CK) in the same slide rather than consecutive slides. Thus, the CK fluorescent stain can provide more accurate ground truth for epithelial masks. However, this method required sophisticated staining protocols and a scanner capable of capturing fluorescent images. Consecutive CK stains require simpler protocols and equipment, meaning that it is more accessible and economical for pathology laboratory with limited budget. Another reason for low performance is morphology. We observed the 3 cases with the lowest IoU in the test set (Table S1), it is noticed that these tumors were mainly composed of small cells nests (less than 10 cells). This increased deviations in tumor area between the Ki67 stains and paired CK slides. Therefore, the IoU decreased because the ground truth extracted from CK slides cannot perfectly reflect the epithelial area in Ki67 stains. Even the predicted mask worked effectively. Lastly, IoU was also lower in core biopsy specimen than excised tumors because the tissue shape and location were more different in the nearby slides. IHC-supervised learning appears to be efficient, so it is possible to apply this pipeline to other types of immunostaining and study topics. For example, if we replaced Ki67 with hematoxylin and eosin images and CK with CD45 to highlight lymphocytes, it may be possible to build a neural network to predict and measure tumor-infiltrating lymphocyte (TIL) areas in hematoxylin and eosin slides. Furthermore, by combining these algorithms, neural networks can segment different areas in breast cancer images, revealing the spatial information of intratumoral immune responses Hong et al. (2021). Another trial used lymph node hematoxylin and eosin images as the input and CK as the ground truth. The algorithm creates a virtual CK stain to predict the epithelial components, which are usually metastatic carcinomas, in the lymph nodes in the visual assessment of the validity of the prediction mask at the image tile level, the average score from the pathologists was from 3.75 to 4.56. This indicated that the virtual tumor-area masks predicted by our algorithm were adequate, and they optimized automated KI quantification in most tiles. First, by excluding non-tumor areas at the WSI level, fewer image tiles were required for KI assessment, which saved computing resources and time. Second, although some recent KI-counting algorithms can differentiate stromal cells from tumor cells at the cellular level, they still struggle to eliminate these disguised cells, particularly lymphocytes and endothelial cells. However, our overlying virtual tumor masks worked as a filter at the structural level to restrict the counting area. Similar to the approach used by pathologists, the process involved identifying the tumor area prior to counting the cells. Third, in our study, masks were the most affected in TIL-rich breast cancer (Case 13), increasing the KI by 9.2% on average. This is because RBCs and TILs are easily mistaken for tumor cells, which interferes with the KI quantification in TIL-rich areas (Figure 4). The masks eliminated these areas and corrected the KI.

The effect of the tumor-area masks became less significant at the WSI level. Although the hotspot locations changed in most cases after mask application, the difference in KI between the previous and new hotspots (ΔHotspot KI) was small, except in the TIL-rich case (Case 13). The change in global KI was even smaller. This might be because the RBC- and TIL-rich areas, where the algorithm worked best, were small in most cases. The tumor-area masks did not significantly influence the clinical classification: only one case (Case 7) was downgraded after correction of the global KI.

This study had some limitations. First, we only investigated carcinomas of no special type. These masks may not work in some common subtypes of breast cancer, such as lobular carcinoma or mucinous carcinoma. The second limitation is that consecutive CK images were easily obtained and were acceptable for training, but they were not perfect ground truths, and deviations existed (IoU = 0.72 on average). Although Ki67-CK double-staining could address this problem, the cost and difficulty of ground-truth extraction also increased. Increasing the case count for training might improve performance. Finally, we did not correlate KI with genetic profiles or clinical course. Changes in KI by applying the mask seem limited, although more cases and long-term surveys are needed to determine the clinical utility of corrected KI.

In conclusion, consecutive CK staining, with little human annotation input, can provide a substantial, precise, and reliable ground truth, which was able to train a neural network algorithm efficiently. Furthermore, the predicted tumor areas improved the KI counting by excluding non-tumor areas at the tile image level, although the effect on the WSI level was not significant. Future studies are warranted to verify our findings and determine their clinical significance.

DECLARATIONS

Acknowledgements

The authors thank Dr. Chu Hsiu-Yi for her support and assistance in writing, and Dr. Chou Yueh-Hung and Dr. Li Yueh for their effort in visual validity assessment.

Ethical Approval

This study was approved by the Research Ethics Review Committee of the Far Eastern Memorial Hospital (No. 111205-F/DPAI-005). All slides and blocks in the study were used in the clinical diagnosis setting and were retrieved from the repository. The slides and images were kept anonymous and did not contain any personal information. Therefore, the requirement for obtaining informed patient consent was waived.

Authors' contributions

Study Design: CMH and WCH. Data acquisition: CMH and WCH. Creation of software used in this study: HHT and CMH. Data analysis: HHT and CMH. Manuscript writing: WCH, CMH, HHT, and PYT. All authors have read and approved the final manuscript.

Funding

The project was funded by Quanta Medical Technology Foundation (QMTF).

Declaration of Competing Interest

The authors declare no competing interests.

Data Availability Statement

The datasets used and analyzed in the current study are available from the corresponding author upon reasonable request.

REFERENCES

1. Gerdes J, Lemke H, Baisch H, et al. 1984 Oct. Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. J Immunol. 133(4):1710-5.

2. Goldhirsch A, Wood WC, Coates AS, et al. 2011 Aug. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann Oncol. 22(8):1736-47.

3. Thomssen C, Balic M, Harbeck N, et al. 2021 Apr. St. Gallen/Vienna 2021: A brief summary of the consensus discussion on customizing therapies for women with early breast cancer. Breast Care (Basel). 16(2):135-143.

4. Nielsen TO, Leung SCY, Rimm DL, et al. 2021 Jul 1. Assessment of Ki67 in breast cancer: Updated recommendations from the International Ki67 in Breast Cancer Working Group. J Natl Cancer Inst. 113(7):808-819.

5. Dowsett M, Nielsen TO, A'Hern R, et al. 2011 Nov 16. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. J Natl Cancer Inst. 103(22):1656-64.

6. Polley MYC, Leung SCY, Gao D, et al. 2015 Jun. An international study to increase concordance in Ki67 scoring. Mod Pathol. 28(6):778-86.

7. Tuominen VJ, Ruotoistenmäki S, Viitanen A, et al. 2010. ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. Breast Cancer Res. 12(4):R56.

8. Acs B, Pelekanou V, Bai Y, et al. 2019 Jan. Ki67 reproducibility using digital image analysis: an inter-platform and inter-operator study. Lab Invest. 99(1):107-117.

9. Abubakar M, Figueroa J, Ali HR, et al. 2019 Sep. Combined quantitative measures of ER, PR, HER2, and KI67 provide more prognostic information than categorical combinations in luminal breast cancer. Mod Pathol. 32(9):1244-1256.

10. Rimm DL, Leung SCY, McShane LM, et al. 2019 Jan. An international multicenter study to evaluate reproducibility of automated scoring for assessment of Ki67 in breast cancer. Mod Pathol. 32(1):59-69.

11. Ruifrok AC, Johnston DA. 2001 Aug. Quantification of histochemical staining by color deconvolution. Anal Quant Cytol Histol. 23(4):291-9.

12. Saha M, Chakraborty C, Arun I, et al. 2017 Jun 12. An advanced deep learning approach for Ki-67-stained hotspot detection and proliferation rate scoring for prognostic evaluation of breast cancer. Sci Rep. 7(1):3213.

13. Lakshmi S, Ritwik KVS, Vijayasenan D, et al. 2020 Jul. Deep learning model-based Ki-67 index estimation with automatically labelled data. Annu Int Conf IEEE Eng Med Biol Soc. 2020:1412-1415.

14. Volynskaya Z, Mete O, Pakbaz S, et al. 2019 Mar 8. Ki67 quantitative interpretation: Insights using image analysis. J Pathol Inform. 10:8.

15.Valkonen M, Isola J, Ylinen O, et al. 2020 Feb. Cytokeratin-supervised deep learning for automatic recognition of epithelial cells in breast cancers stained for ER, PR, and Ki-67. IEEE Trans Med Imaging. 39(2):534-542.

16. Koopman T, Buikema HJ, Hollema H, et al. 2018 May. Digital image analysis of Ki67 proliferation index in breast cancer using virtual dual staining on whole tissue sections: clinical validation and inter-platform agreement. Breast Cancer Res Treat. 169(1):33-42.

17. Røge R, Riber-Hansen R, Nielsen S, et al. 2016 Jul. Proliferation assessment in breast carcinomas using digital image analysis based on virtual Ki67/cytokeratin double staining. Breast Cancer Res Treat. 158(1):11-19.

18. Wessel Lindberg AS, Conradsen K, Larsen R, et al. 2017 Jun. Quantitative tumor heterogeneity assessment on a nuclear population basis. Cytometry A. 91(6):574-584.

19. Brázdil T, Gallo M, Nenutil R, Kubanda A, et al. 2022 Mar. Automated annotations of epithelial cells and stroma in hematoxylin-eosin-stained whole-slide images using cytokeratin re-staining. J Pathol Clin Res. 8(2):129-142.

20. Hong Y, Heo YJ, Kim B, et al. 2021 Sep 28. Deep learning-based virtual cytokeratin staining of gastric carcinomas to measure tumor–stroma ratio. Sci Rep. 11(1):19255.

21. Zhou Z, Siddiquee MMR, Tajbakhsh N, et al. 2018 Sep. UNet++: A nested U-Net architecture for medical image segmentation. Deep Learn Med Image Anal Multimodal Learn Clin Decis Support (2018). 11045:3-11.

22. Plancoulaine B, Laurinaviciene A, Herlin P, et al. 2015 Oct 19. A methodology for comprehensive breast cancer Ki67 labeling index with intra-tumor heterogeneity appraisal based on hexagonal tiling of digital image analysis data. Virchows Arch. 467:711-722.

