


Recognizing basal cell carcinoma on smartphone-captured digital histopathology images with a deep neural network*

Y.Q. Jiang ¹, J.H. Xiong,² H.Y. Li,³ X.H. Yang,¹ W.T. Yu,¹ M. Gao,¹ X. Zhao,² Y.P. Ma,² W. Zhang,¹ Y.F. Guan,³ H. Gu⁴ and J.F. Sun¹

¹Department of Dermatopathology and ⁴Department of Physiotherapy, Institute of Dermatology, Peking Union Medical College & Chinese Academy of Medical Sciences, Nanjing, 210042, China

²Beijing Tulip Partners Technology Co., Ltd, Beijing, China

³Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, 48109, U.S.A.

Linked Comment: Lee and Soyer. *Br J Dermatol* 2020; **182**:540–541.

Summary

Correspondence

Jianfang Sun; Heng Gu.

E-mail: sunjf@163.com; guheng@aliyun.com

Accepted for publication

22 April 2019

Funding sources

This work was supported by the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (CIFMS-2017-I2M-1-017).

Conflicts of interest

None to declare.

Y.Q.J., J.H.X. and H.Y.L. contributed equally to this work.

*Plain language summary available online

DOI 10.1111/bjd.18026

Background Pioneering effort has been made to facilitate the recognition of pathology in malignancies based on whole-slide images (WSIs) through deep learning approaches. It remains unclear whether we can accurately detect and locate basal cell carcinoma (BCC) using smartphone-captured images.

Objectives To develop deep neural network frameworks for accurate BCC recognition and segmentation based on smartphone-captured microscopic ocular images (MOIs).

Methods We collected a total of 8046 MOIs, 6610 of which had binary classification labels and the other 1436 had pixelwise annotations. Meanwhile, 128 WSIs were collected for comparison. Two deep learning frameworks were created. The 'cascade' framework had a classification model for identifying hard cases (images with low prediction confidence) and a segmentation model for further in-depth analysis of the hard cases. The 'segmentation' framework directly segmented and classified all images. Sensitivity, specificity and area under the curve (AUC) were used to evaluate the overall performance of BCC recognition.

Results The MOI- and WSI-based models achieved comparable AUCs around 0.95. The 'cascade' framework achieved 0.93 sensitivity and 0.91 specificity. The 'segmentation' framework was more accurate but required more computational resources, achieving 0.97 sensitivity, 0.94 specificity and 0.987 AUC. The runtime of the 'segmentation' framework was 15.3 ± 3.9 s per image, whereas the 'cascade' framework took 4.1 ± 1.4 s. Additionally, the 'segmentation' framework achieved 0.863 mean intersection over union.

Conclusions Based on the accessible MOIs via smartphone photography, we developed two deep learning frameworks for recognizing BCC pathology with high sensitivity and specificity. This work opens a new avenue for automatic BCC diagnosis in different clinical scenarios.

What's already known about this topic?

- The diagnosis of basal cell carcinoma (BCC) is labour intensive due to the large number of images to be examined, especially when consecutive slide reading is needed in Mohs surgery.
- Deep learning approaches have demonstrated promising results on pathological image-related diagnostic tasks.
- Previous studies have focused on whole-slide images (WSIs) and leveraged classification on image patches for detecting and localizing breast cancer metastases.

What does this study add?

- Instead of WSIs, microscopic ocular images (MOIs) photographed from microscope eyepieces using smartphone cameras were used to develop neural network models for recognizing BCC automatically.
- The MOI- and WSI-based models achieved comparable areas under the curve around 0.95.
- Two deep learning frameworks for recognizing BCC pathology were developed with high sensitivity and specificity.
- Recognizing BCC through a smartphone could be considered a future clinical choice.

Basal cell carcinoma (BCC) is the most common skin cancer, with a rapidly rising incidence.¹ The diagnosis of BCC is straightforward for experienced pathologists but labour intensive due to the large number of cases, especially when a consecutive slide reading is needed in Mohs surgery.² Furthermore, the diagnosis is challenging when BCC areas vary in architecture and size, or become obscured due to inflammation. Therefore, a fast and rigorous computational method for automatic BCC diagnosis is needed.

Computer-aided diagnosis of diseases has been developed for decades to assist the analysis of microscopic images in pathology.^{3,4} Most previous work has focused on 'feature engineering', which required hand-crafted features specified by domain experts.^{5,6} Recent progress in deep convolutional neural networks (CNNs) has shown significantly improved performance on a wide range of computer vision tasks, including image recognition, face recognition, object detection and semantic segmentation.^{7–11} Deep learning-based solutions have also demonstrated promising results on pathological image-related tasks. The CAMELYON challenge is an academic competition in recognizing and localizing breast cancer metastases in whole-slide images (WSIs).¹² Wang *et al.* first demonstrated the power of deep learning in breast cancer pathological diagnoses.¹³ The group of Norouzi, who ranked first in the challenge leaderboard in 2018, pointed out that cancer metastases could be detected from whole-slide pathology images through some carefully designed networks.¹⁴ These previous studies focused on WSIs and leveraged classification on image patches for detecting and localizing metastases.

In this study, microscopic ocular images (MOIs) photographed from microscope eyepieces using smartphone cameras were used to develop neural network models for recognizing BCC automatically. We first investigated a CNN classification model, which did not perform well in detecting BCC in hard cases (see details in the Materials and methods). We further found that the hard cases could be well recognized by pixelwise segmentation methods. Hence, a cascade framework was created, with an initial CNN classification model and a subsequent semantic segmentation model. Alternatively,

we created another framework of the semantic segmentation model only. This segmentation framework achieved better performance on BCC recognition, yet required more computational resources. Further analysis demonstrated that the performances in recognizing BCC were comparable based on WSIs or MOIs. This indicates that recognizing BCC through a smartphone could be considered a future clinical choice, especially in some screening cases.

Materials and methods

General datasets introduction

Three histopathology image datasets with different types of annotations were collected. In total 8046 MOIs were gathered, within which 6610 images were labelled with a BCC/non-BCC binary classification tag (positive or negative), regarded as the microscopic ocular image classification (MOIC) dataset. Images of BCC areas or normal tissues without BCC were considered as positive or negative samples, respectively. Other skin diseases were not considered or involved in our dataset. The other 1436 images with pixelwise annotations were regarded as the microscopic ocular image segmentation (MOIS) dataset.

For comparison, we also collected 128 WSIs with pixelwise annotations, which would be cut into patches to train and evaluate the models. In these segmentation datasets, BCC tissue was pixelwise annotated. Pathologists outlined the BCC tissue with closed-contour segmentation, which contained further detailed information, including the shape, area and location of the BCC tissues. The segmentation model trained on these datasets was accurately able to dissect BCC tissue from the normal tissue by predicting a pixel-level binary classification in a given test image. The annotation area and prediction area were further compared and used to calculate the confidence score for BCC classification at the image level.

MOIC and MOIS images were photographed from microscope eyepieces using smartphone cameras at 10 × resolution as previously described,¹⁵ resulting in round images of 3200 × 2500 pixels in size. WSI data were obtained for 128

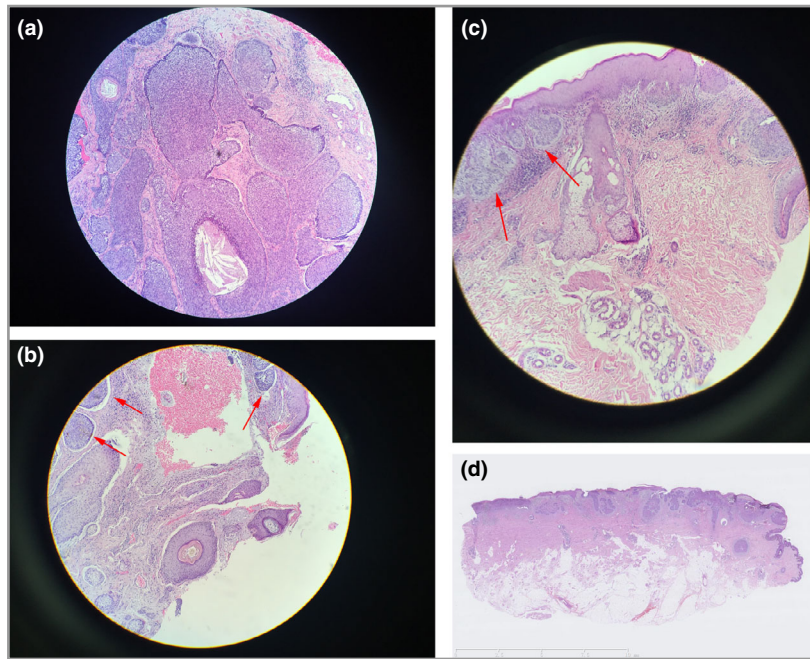


Fig 1. Examples of microscopic ocular image classification and microscopic ocular image segmentation images and a whole-slide image. (a, b, c) Examples of pathology images photographed from the microscope eyepiece using smartphones ($\times 10$ magnification). In hard examples (b, c), red arrows point to the small foci of basal cell carcinoma. (d) Whole-slide image.

patients at $40 \times$ resolution. As the obtained tissue size varies among patients, the size of the WSI varies from $63\,488 \times 53\,248$ to $22\,784 \times 19\,840$ pixels. Examples of MOIC/MOIS and WSI are shown in Figure 1. Some test images had an extremely small area of BCC, were not well focused or were obscured due to inflammation. These images were regarded as 'hard' cases, and others were considered as 'standard' cases. The reason to identify the hard cases is that these cases degraded the performance, and an example is given in Table S1 (see Supporting Information) to illustrate this phenomenon. In this work, the cases with low prediction confidence were considered as hard cases.

All cases were diagnosed at the Institute of Dermatology, Peking Union Medical College & Chinese Academy of Medical Sciences, from January 2013 to June 2018. All images were manually labelled by pathologists at the Institute of Dermatology. All methods and procedures were performed in accordance with relevant guidelines and regulations. Written informed consent was obtained from all participants, and the study was carried out with the approval of the Institutional Review Board of the Institute of Dermatology, Peking Union Medical College & Chinese Academy of Medical Sciences (no. 2013-LC/KY-033).

Basal cell carcinoma classification and segmentation framework

Figure 2 shows the flowchart of CNN classification and semantic segmentation systems. The three systems accept WSIs or MOIs as the inputs and return recognition results.

Cascade framework

The cascade framework consists of a general screening stage and an in-depth analysis stage, as shown in Figure 2(c). In the first stage of general screening, an image classification network was trained and a confidence filter strategy was used to determine the existence of BCC with high confidence. The images with lower confidence (classification-predicted confidence between 0.1 and 0.9) were regarded as hard cases and sent to the in-depth analysis stage. At the in-depth analysis stage, a semantic segmentation model was built, which segmented the BCC regions in hard cases. The segmentation results were further used to decide the existence of BCC. Only when the area of BCC was higher than the threshold was the image classified as positive.

General screening

GoogleNet inception v3 architecture with input size of 961 and global average pooling was used to build the classification network.^{16,17} The network weights were pretrained on the ImageNet dataset and fine-tuned on the MOIC and WSI data.¹⁸

The 6610 MOIC images were partitioned into three subsets to train (796, 12.0%), validate (88, 1.3%) and test (5726, 86.6%) our model. As the effective area of the MOIC images is circular, and to balance the numbers of positive and negative samples, we augmented the 553 positive samples by rotating an image with 60 degrees internal, as well as the 243 negative samples with 30 degrees internal, resulting in a total

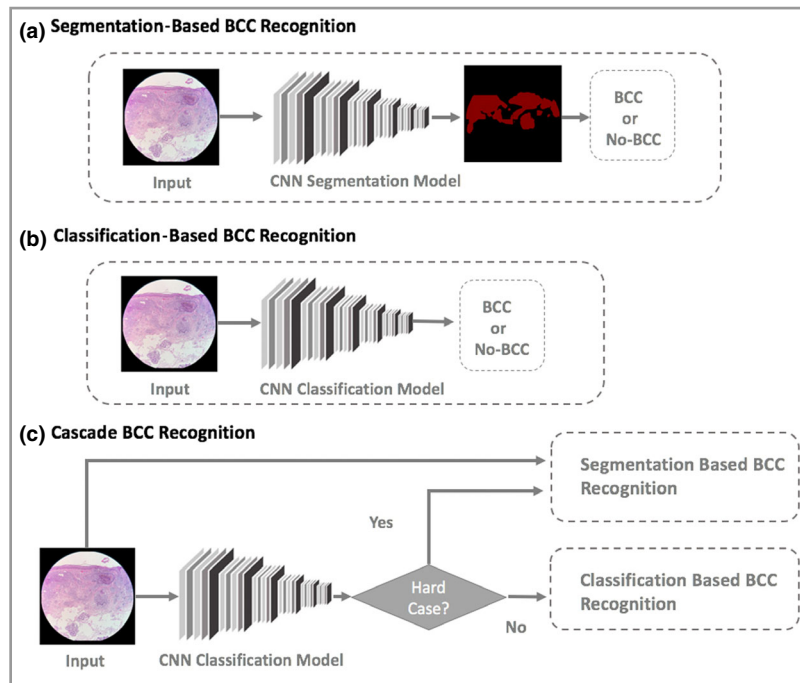


Fig 2. Schematic diagrams of basal cell carcinoma (BCC) recognition systems. (a) Segmentation-based BCC recognition system. (b) Classification-based BCC recognition system. (c) Cascade BCC recognition system. Both the classification and segmentation models can classify BCC images, and the segmentation model could further locate the BCC region. The cascade system returns the recognition result for a standard case (the high confidence image) or transfers the hard case (the low-confidence image) to a segmentation model for further analysis.

of 3318 positive samples and 2916 negative samples as the training set. The model was trained on the Caffe platform and input mirror was also involved during training.¹⁹

To compare the BCC recognition performance of models trained on smartphone-captured images or clinical WSIs, we further developed a classification model using the WSI dataset. As WSIs were huge and contained large invalid areas (e.g. nontissue regions), we employed a field of valid picking algorithm to extract valid tissue areas. Then we cut the valid area into small patches of size 1024×1024 pixels. In total, 7084 patches were obtained (2789 BCC patches and 4295 normal tissue patches), 70% of which were used for model training (4604 patches) and validation (355 patches), and the other 30% (2125 patches) for model testing. Similarly to the data augmentation in MOIs, the WSI patches were also randomly mirrored and rotated, and the positive and negative samples were balanced via random sampling. The final training set consisted of 16 860 positive and 16 241 negative samples.

In-depth analysis

To detect and locate BCC accurately in hard cases, we employed a semantic segmentation model at the in-depth analysis stage. The model was built on the DeeplabV3 network structure²⁰ with Resnet 101⁷ as the backbone. The model backbone was pretrained on ImageNet and fine-tuned with the MOIS dataset.

To build the semantic segmentation model, 1436 MOIS images with 1024×1024 resolution were pixelwise labelled by histopathologists. Each MOSI image was split into four 512×512 patches and 5744 patches were obtained. We used 4030 patches as the training set and augmented them by rotating the patch with 30 degree internal and mirroring the patch, resulting in a total of 28 140 training and 2008 validation patches. The other 1714 patches were used for testing. We used a threshold to determine the existence of BCC, and only images with BCC segmentation areas larger than the threshold were considered as positive.

The runtimes of the models mentioned above were tested. For the semantic segmentation model we used Tensorflow 1.10.1.²¹ For the classification model we used the Caffe framework. Both classification and segmentation models were trained and tested on an NVIDIA Titan X GPU $\times 2$ with CUDA version 9.0 cuDNN 7.0 (NVIDIA, Cambridge, U.K.). The employed deep learning networks of DeepLab Resnet V3 and GoogleNet Inception V3 are publicly available at <https://github.com/rishizek/tensorflow-deeplab-v3> and https://github.com/smichalowski/google_inception_v3_for_caffe, respectively.

Performance evaluation

Sensitivity, specificity and area under the curve (AUC) were used to evaluate the overall performance of BCC recognition. The receiver operating characteristic (ROC) curve represents

a dynamic trade-off between sensitivity and specificity, which can be controlled by adjusting the threshold of binary classification. An algorithm with an AUC of 1 represents a perfect prediction performance. The baseline AUC of random prediction is 0.5. The mean intersection over union was used to evaluate the performance of segmentation and localization.

Statistical comparison test

We performed statistical analysis to compare the WSI, MOIC and MOIS models. For each model we randomly sampled 80% of the test images and used this subset to calculate the AUC. We repeated this a total of 20 times and obtained 20 AUC values for each model. Then we performed the pairwise Wilcoxon signed-rank test based on the 20 AUC values of different models, which does not require stringent assumptions.

Results

Performance comparison of models trained on the microscopic ocular image or whole-slide image datasets

To our knowledge, previous studies of recognizing specific tissues in pathological images focused mainly on WSIs. In order to demonstrate the efficiency of a CNN-based method in recognizing BCC on MOIs, we trained models on WSI and MOI images separately, using the same neural network architecture.

As shown in Table 1 and Figure S1 (see Supporting Information), the performances of the models trained on MOIs and WSIs are comparable. The AUCs of the WSI (40 ×), WSI (10 ×) and MOI models are all above 0.95. The pairwise Wilcoxon signed-rank test revealed a significant difference for both the MOI segmentation model (AUC = 0.987) and the MOI classification model (AUC = 0.976) when compared with the WSI (10 ×) model (AUC = 0.955) (both P-values < 0.001). The MOI models outperformed the WSI (10 ×) model. Therefore, the deep learning model trained on MOIs could be an alternative and efficient way of diagnosing BCC in pathological images.

Semantic segmentation improves basal cell carcinoma recognition in hard cases

As mentioned in Materials and methods, the classification model filtered out hard cases with low confidence at the first stage. Subsequently, the semantic segmentation model further analysed these hard cases at the in-depth analysis stage. The generalization ability of a model was evaluated by cross-dataset AUCs, as shown in Table 1. Of note, the segmentation model not only had the best generalization ability (0.945 across dataset), but it also accurately classified hard cases within the dataset, with an AUC of 0.933. In contrast, the classification model at the first stage did not detect BCC very well in hard cases, with an AUC of only 0.701. As the sizes of BCCs in the hard cases were relatively small, the classification model missed some small BCC regions. The segmentation model captured the information of annotations at the pixel level, which allowed the neural network to learn about the context among pixels, resulting in higher performance in hard cases.

The ROC curves of these models are shown in Figure 3. As mentioned above, the ROC curve demonstrates a trade-off between sensitivity and specificity. We selected three operation points on the ROC curve to demonstrate the model performance. The first operation point, on the right, emphasizes sensitivity and sacrifices specificity. The second operation point, in the middle, makes the best trade-off between the overall sensitivity and specificity. The third operation point, on the left, emphasizes specificity. The high sensitivity and specificity indicate that the semantic segmentation model is able to recognize BCC accurately based on smartphone-captured pathological MOIs.

The runtimes of the classification, cascade and segmentation frameworks

To evaluate the runtimes, the classification and segmentation models were tested on an NVIDIA Titan X GPU. In total 100 MOIs were used to calculate the prediction runtimes. The average runtime of the MOI segmentation model is 500.2 ± 163.2 ms per image ($n = 100$), whereas the MOI classification

Table 1 Area under the curve (95% confidence interval) of deep learning models and test datasets

Model	Test dataset		
	Across dataset	Within dataset	Within-dataset hard cases
WSI (40 ×)	0.531 (0.516–0.542)	0.982 (0.969–0.993)	0.871 (0.736–0.980)
WSI (10 ×)	0.783 (0.771–0.795)	0.955 (0.945–0.964)	0.552 (0.405–0.648)
MOI classification	0.885 (0.871–0.899)	0.976 (0.968–0.986)	0.701 (0.608–0.784)
MOI segmentation	0.945 (0.935–0.954)	0.987 (0.984–0.990)	0.933 (0.906–0.959)

The whole-slide image (WSI; 10 × magnification images) classification model, WSI (40 × images) classification model and microscopic ocular image (MOI) classification model are presented. Each model is tested within dataset (the same dataset on which the model is trained), across dataset and as its own hard case. The hard cases processed by the MOI segmentation model are the same as those of the MOI classification model. Further information is given in Figure S1 (see Supporting Information). 95% confidence intervals were calculated using 2000 bootstrap samples.

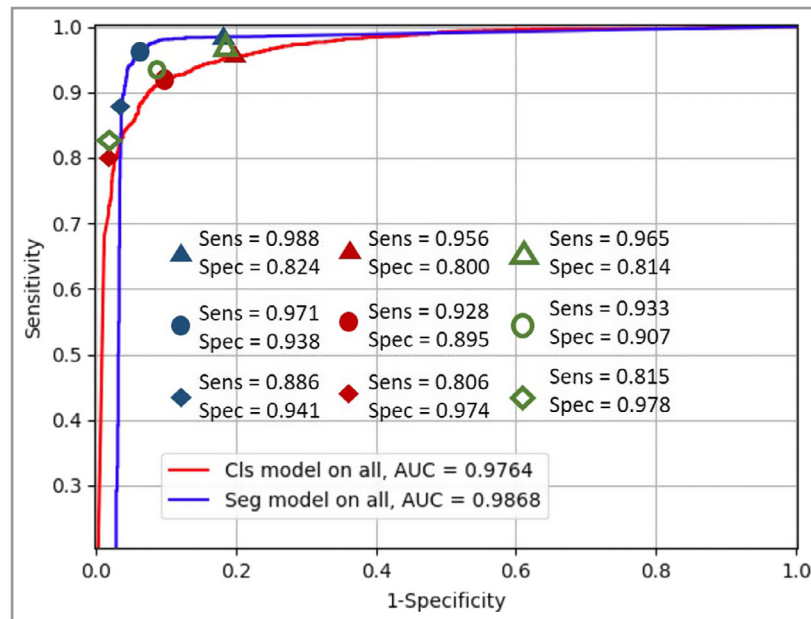


Fig 3. Receiver operating characteristic curves of the classification (Cls) and segmentation (Seg) models. The area under the curve (AUC) of the classification system (red) on both standard and hard cases is 0.976. The segmentation system (blue) on all cases achieves the highest AUC of 0.987. The green symbols show the performance of the cascade framework. The blue diamond, square and triangle highlight the high-specificity operating point, the high-accuracy operating point and the high-sensitivity operating point, respectively. Sens, sensitivity; Spec, specificity.

model requires only 89.3 ± 23.9 ms ($n = 100$) and the cascade framework needs 127.5 ± 43 ms per image on average ($n = 100$). When tested on a CPU (Intel Xeon E5 2620), the runtime of the segmentation model is 15.3 ± 3.9 s ($n = 100$) per image, but the classification model and cascade model are much faster, with runtimes of 2.7 ± 0.7 s and 4.1 ± 1.4 s, respectively ($n = 100$). These results show that the semantic segmentation model requires more time to make predictions.

Locating basal cell carcinoma regions with semantic segmentation

We also demonstrate the performance of locating a BCC area through semantic segmentation. The training data were labelled with BCC as the foreground and all other tissues as the background. The mean intersection over union reached 0.863 on the 431 testing images. Examples of gold-standard images and predictions are shown in Figure 4.

Discussion

Deep learning algorithms have shown promising results in digital pathology.²² The deep neural network extracts semantic features from the supervision of large-scale datasets. It recognizes and segments mitotic figures or small foci of metastatic cancers, which can be missed even by experienced pathologists. Reading pathology images also requires considerable time and effort.^{14,23,24} In contrast, automatic recognition of cancerous regions of digital pathology frees histopathologists from laborious labelling and allows them to concentrate on

difficult cases and work as supervisors for straightforward cases with higher efficiency.²²

The digital WSI is a fundamental data resource for novel automatic recognition techniques. Previous deep learning approaches for pathological image analysis have focused on metastatic foci in breast, sentinel lymph node and prostate cancer.^{14,25,26} Researchers integrated image patch classification from WSIs and postprocess to recognize specific metastasis in high-resolution images, such as at $40 \times$ resolution. Liu *et al.*¹⁴ reported a sensitivity of 92.4% and image-level AUC scores above 97% on both the CAMELYON16 test set and an independent set of 110 slides. They found that a combination of multiscale image patches at $10 \times$, $20 \times$ and $40 \times$ did not bring significant improvement in classification performance. The network, trained on image patches either at $40 \times$ or in multiscale resolution, concentrates on the local details or the cellular level to distinguish the cancer metastatic area.

During the practice of cutaneous pathologists diagnosing BCC, it is the architecture rather than the cellular structure that is more important in making diagnostic decisions. Based on this observation, we designed our deep neural network to identify BCC at $10 \times$ resolution instead of $20 \times$ or $40 \times$ magnification because higher magnification cannot distinct BCC from normal follicular structures, where the pleomorphism or atypia may be subtle in BCC. To obtain multiple $10 \times$ digital pathology images we used a smartphone camera to take pictures through microscope eyepieces, which was fast and convenient, instead of a slide scanning machine. To evaluate the effectiveness of smartphone-captured digital pathology

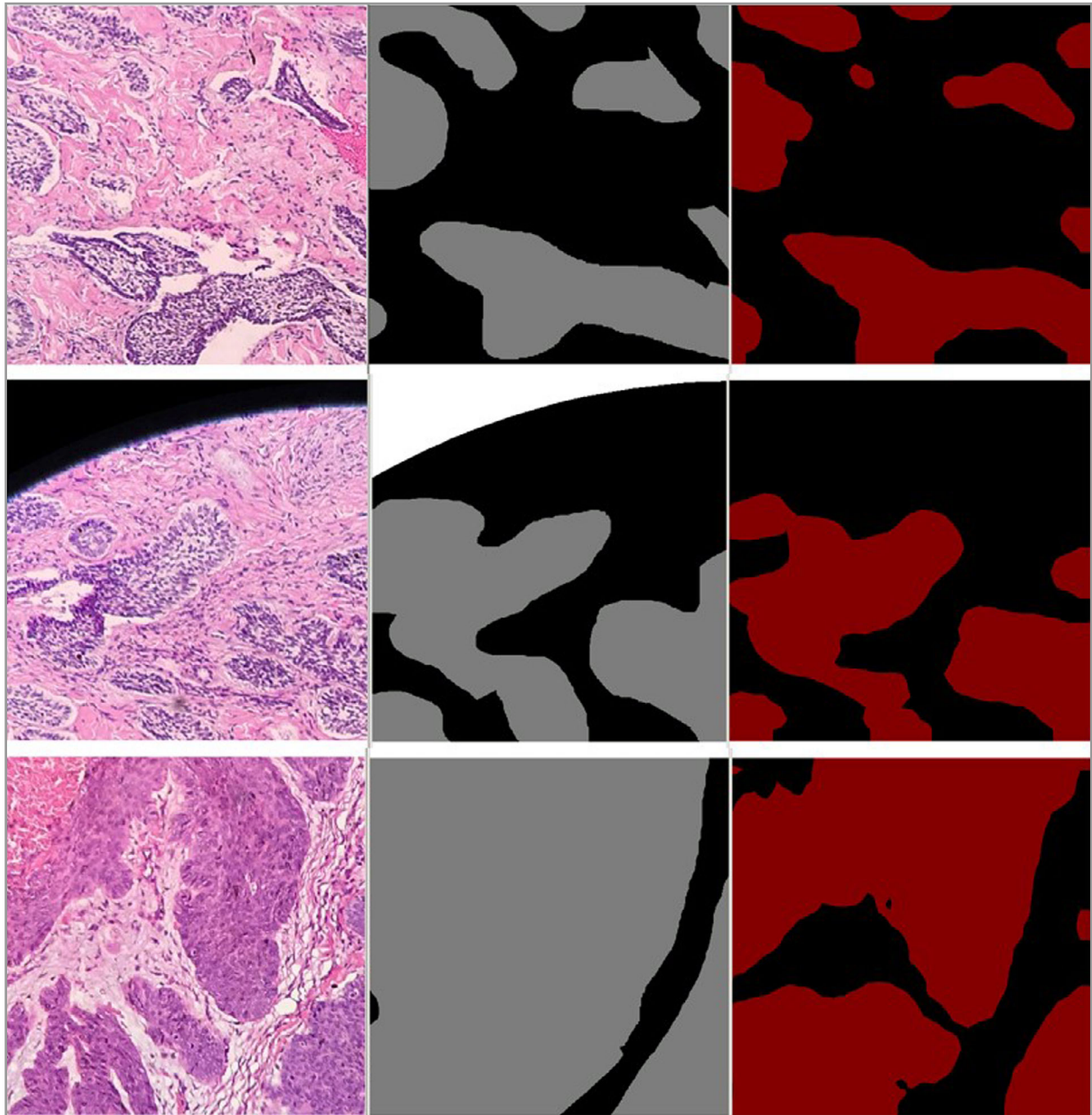


Fig 4. Segmentation of basal cell carcinoma (BCC) with microscopic ocular image segmentation images. The left column is the original microscope ocular image. The middle column is the manual labels (BCC tissue in grey and normal tissue in black). The right column is the prediction of the microscope ocular segmentation model (BCC tissue in red).

images in detecting BCC, we built models using either smartphone-captured MOIs or regular WSIs. Notably, the results reveal that the smartphone image-based model achieves comparably good performance in detecting BCC. The AUC of the MOIC model reaches 0.976, while the AUCs of the WSI (10 \times) and WSI (40 \times) models are 0.955 and 0.982, respectively. In addition, the smartphone image-based model can be generalized to WSIs, indicating the robustness of our method.

The classification model trained on MOIC data is able to detect 'standard' cases of BCC, yet it misclassifies the 'hard' cases, which are small BCC areas or BCCs obscured by

inflammation. We introduced the cascade framework to distinguish the standard cases from the hard ones based on the prediction confidence, then applied the segmentation model to the hard cases to improve the final prediction performance. It is speculated that the improved performance after using segmentation is attributed to the pixelwise annotation, as the information related to the annotation confidence, such as the area and shape of BCC, is given in segmentation. The hard cases should be annotated with low annotation confidence. Segmentation presents the area and the shape of the tissue, which indicate how confident the labelling can be.

We further developed an end-to-end segmentation model to predict BCC directly. This model achieves the best prediction performance, with a mean intersection over union of 0.863 and classification AUC of 0.987, comparable with previous studies. The mean intersection over union of the DeepLab V3 model on the Pascal VOC 2012 validation dataset is 0.764.²⁰ A similar study that employed DeepLab V3 on skin lesion segmentation achieved mean intersections over union of 0.816, 0.885 and 0.774 on three open datasets.²⁷ Regarding classification accuracy, Cruz-Roa *et al.* proposed a deep learning architecture that has 0.927 specificity and 0.869 sensitivity.²⁸ In comparison, our segmentation model achieves 0.969 specificity and 0.939 sensitivity. Compared with the cascade framework, this end-to-end model achieves higher prediction performance, yet it requires more computational time and resources. Regarding the runtime of these different frameworks, the 'segmentation' framework requires 15.3 ± 3.9 s to predict one image, whereas the classification framework needs only 2.7 ± 0.7 s and the 'cascade' framework needs 4.1 ± 1.4 s. For devices that have limited computational resources (e.g. mobile devices), the classification model is far more time efficient. When the computational resources are sufficient, the MOI segmentation model is a better alternative to achieve higher prediction performance. These two approaches can be adopted under different circumstances.

Of note, the adoption of WSIs in disease diagnosis by pathologists has been slow worldwide,²⁹ as it is limited by the cost of the system, digital slide storage and inability to handle high-throughput routine work. So, remarkably differently from the previous WSI-based dataset, the resources of the dataset we used to train the CNN algorithm were designed creatively, and the digital pathology images were obtained from microscope eyepieces using smartphone cameras at $10 \times$ resolution. The images presented efficient characteristic pattern features rather than cytological features, which are more practical in BCC diagnosis for histopathologists. The experiment showed the effectiveness and efficiency of this architecture. It would generate a novel computer-aided diagnostic method to extend the diagnostic capability of hospitals without WSI scanning facilities. The advantages of using MOIs include low cost, easy smartphone storage and capability to handle high-throughput routine work. All of these advantages of MOIs will benefit pathologists and patients in practice. To the best of our knowledge, this is the first attempt to use a deep neural network to recognize histopathology images photographed from microscope eyepieces. The faster classification and cascade models also provide feasible alternatives for diagnosing BCC when computational resources are limited.

In conclusion, in this study we developed deep neural network models for BCC recognition, based on smartphone-captured images of digital pathology through the eyepiece of a microscope. Compared with the model trained on WSIs, we showed that the MOI-based model achieves better performance on detecting hard BCC cases, and this model is generalizable across types of MOIs and WSIs. The accurate prediction performance of the MOI-based model and the

feasibility of collecting MOIs will benefit the diagnosis of BCC in practice.

References

- 1 Goldenberg G, Karagiannis T, Palmer JB *et al.* Incidence and prevalence of basal cell carcinoma (BCC) and locally advanced BCC (LABCC) in a large commercially insured population in the United States: a retrospective cohort study. *J Am Acad Dermatol* 2016; **75**:957–66.
- 2 Shi Y, Jia R, Fan X. Ocular basal cell carcinoma: a brief literature review of clinical diagnosis and treatment. *Onco Targets Ther* 2017; **10**:2483–9.
- 3 Ghaznavi F, Evans A, Madabhushi A, Feldman M. Digital imaging in pathology: whole-slide imaging and beyond. *Annu Rev Pathol* 2013; **8**:331–59.
- 4 Gurcan MN, Boucheron LE, Can A *et al.* Histopathological image analysis: a review. *IEEE Rev Biomed Eng* 2009; **2**:147–71.
- 5 Weaver DL, Krag DN, Manna EA *et al.* Comparison of pathologist-detected and automated computer-assisted image analysis detected sentinel lymph node micrometastases in breast cancer. *Mod Pathol* 2003; **16**:1159–63.
- 6 Gandomkar Z, Brennan PC, Mello-Thoms C. Computer-based image analysis in breast pathology. *J Pathol Inform* 2016; **7**:43.
- 7 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Available at: <https://arxiv.org/abs/1512.03385> (last accessed 5 June 2019).
- 8 Sun Y, Liang D, Wang X, Tang X. Deepid3: face recognition with very deep neural networks. Available at: <https://arxiv.org/abs/1502.00873> (last accessed 5 June 2019).
- 9 Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. Available at: <https://ieeexplore.ieee.org/document/7485869> (last accessed 5 June 2019).
- 10 He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. Available at: <https://arxiv.org/abs/1703.06870> (last accessed 5 June 2019).
- 11 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Available at: <https://arxiv.org/abs/1411.4038> (last accessed 5 June 2019).
- 12 Pinchaud N, Hedlund M. CAMELYON17 Grand Challenge. Available at: <https://camelyon17.grand-challenge.org> (last accessed 5 June 2019).
- 13 Wang D, Khosla A, Gargeya R *et al.* Deep learning for identifying metastatic breast cancer. Available at: <https://arxiv.org/abs/1606.05718> (last accessed 5 June 2019).
- 14 Liu Y, Gadepalli K, Norouzi M *et al.* Detecting cancer metastases on gigapixel pathology images. Available at: <https://arxiv.org/abs/1703.02442> (last accessed 5 June 2019).
- 15 Zhou C, Yu Y, Xue R, Elston DM. High-quality digital photomicrography utilizing a smartphone without adapter. *J Cutan Pathol* 2016; **43**:82–4.
- 16 Szegedy C, Vanhoucke V, Ioffe S *et al.* Rethinking the inception architecture for computer vision. Available at: <https://arxiv.org/abs/1512.00567> (last accessed 5 June 2019).
- 17 Lin M, Chen Q, Yan S. Network in network. Available at: <https://arxiv.org/abs/1312.4400> (last accessed 5 June 2019).
- 18 Deng J, Dong W, Socher R *et al.* ImageNet: a large-scale hierarchical image database. Available at: <https://ieeexplore.ieee.org/document/5206848> (last accessed 5 June 2019).
- 19 Jia Y, Shelhamer E, Donahue J *et al.* Caffe: convolutional architecture for fast feature embedding. Available at: <https://arxiv.org/abs/1408.5093> (last accessed 5 June 2019).
- 20 Chen LC, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. Available at: <https://arxiv.org/abs/1706.05587> (last accessed 5 June 2019).

- 21 Abadi M, Barham P, Chen J *et al.* Tensorflow: a system for large-scale machine learning. Available at: <https://arxiv.org/abs/1605.08695v2> (last accessed 5 June 2019).
- 22 Kayser K, Görtler J, Bogovac M *et al.* AI (artificial intelligence) in histopathology – from image analysis to automated diagnosis. *Folia Histochem Cytobiol* 2009; **47**:355–61.
- 23 Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. *Med Image Comput Comput Assist Interv* 2013; **16**:411–18.
- 24 Veta M, van Diest PJ, Jiwa M *et al.* Mitosis counting in breast cancer: object-level interobserver agreement and comparison to an automatic method. *PLOS ONE* 2016; **11**:e0161286.
- 25 Ehteshami Bejnordi B, Veta M, Johannes van Diest P *et al.* Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017; **318**:2199–210.
- 26 Litjens G, Sánchez CI, Timofeeva N *et al.* Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep* 2016; **6**:26286.
- 27 Goyal M, Yap MH. Region of interest detection in dermoscopic images for natural data-augmentation. Available at: <https://arxiv.org/abs/1807.10711> (last accessed 5 June 2019).
- 28 Cruz-Roa AA, Arevalo Ovalle JE, Madabhushi A, González Osorio FA. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. *Med Image Comput Comput Assist Interv* 2013; **16**:403–10.
- 29 Navid F, Anil VP, Liron P. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. *Pathol Lab Med Int* 2015; **7**:23–33.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Fig S1. Receiver operating characteristic curves of the segmentation model and the classification model on hard cases.

Table S1 Number of microscopic ocular image classification hard cases, proportion of hard cases correctly predicted and incorrect predictions for the diagnosis of basal cell carcinoma.