DR YIQUN JIANG (Orcid ID : 0000-0003-4965-0242)

Article type   : Original Article

Recognizing Basal Cell Carcinoma on Smartphone-Captured Digital Histopathology Images with Deep Neural Network

Running head: Recognizing BCC on Smartphone-Captured Digital Histopathology with DNN

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. Gu[4]** and J.F. Sun[1]**

*These authors contributed equally to this work.
**Joint corresponding authors.

[1]Department of Dermatopathology, Institute of Dermatology, Peking Union Medical College & Chinese Academy of Medical Sciences, Nanjing 210042, China
[2]Beijing Tulip Partners Technology Co., Ltd., Beijing, China
[3]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA
[4]Department of physiotherapy, Institute of Dermatology, Peking Union Medical College & Chinese Academy of Medical Sciences, Nanjing 210042, China

Corresponding authors:
Dr. Jianfang Sun
Department of Dermatopathology
Institute of Dermatology
Peking Union Medical College & Chinese Academy of Medical Sciences

12 Jiangwangmiao Rd

Nanjing, 210042

China

Phone: +86 25 85478015

Email: sunjf@163.com

Dr. Heng Gu

Department of Physiotherapy

Institute of Dermatology

Peking Union Medical College & Chinese Academy of Medical Sciences

12 Jiangwangmiao Rd

Nanjing, 210042

China

Phone: +86 25 85478012

Email: guheng@aliyun.com

What's already known about this topic?

- The diagnosis of BCC is labor-intensive due to the large number of images to be examine, especially when a 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 (WSI) and leveraged classification on image patches for detecting and localizing breast cancer metastases.

What does this study add?

- Instead of WSIs, microscopic ocular images (MOI) 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 AUCs around 0.95.
- Two deep learning frameworks for recognizing BCC pathologically were developed with high sensitivity and specificity.
- Recognizing BCC through smartphone could be considered as a future clinical choice.

Abstract

Background Pioneering effort has been made to facilitate the pathology recognition in malignancies based on whole slide images(WSI) through deep learning approaches. It remains unclear whether we can accurately detect and locate 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 (MOI).

Methods We collected a total of 8046 MOIs, 6610 of which had binary classification labels and the other 1436 had pixel-wise annotations. Meanwhile, 128 WSIs were collected for comparison. Two deep learning frameworks were created. "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. "Segmentation" framework directly segmented and classified all images. Sensitivity, specificity, and 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 second (s) per image, whereas the "cascade" framework was 4.1 ± 1.4 s. Additionally the "segmentation" framework achieved 0.863 mean intersection over union (mIoU)_.

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

Key words: deep convolutional neural network, basal cell carcinoma, digital histopathology images, disease diagnosis and localization

Introduction

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 labor-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 BBC 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 focused on "feature engineering", which required hand-craft features specified by domain experts [5-6]. Recent progress in deep convolutional neural network (CNN), 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. CAMELYON challenge is an academic competition of 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]. Norouzi M, who ranked first in the challenge leaderboard in 2018, pointed out that cancer metastases could be detected from whole slide pathology images through some careful designed networks [14]. These previous studies focused on whole slide images 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 **Methods**). We further found the hard cases could be well recognized by pixel-wise 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 segementation 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 whole slide images or microscope ocular images. This indicates that recognizing BCC through smartphone could be considered as a future clinical choice, especially in some screening cases.

Methods

General Datasets Introduction

Three histopathology image datasets with different types of annotations were collected. A total of 8046 microscopic ocular images were gathered, within which 6610 images were labeled with 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 pixel-wise

annotations were regarded as microscopic ocular image segmentation (MOIS) dataset. For comparison, we also collected 128 whole slide images with pixel-wise annotations, which would be cut into patches to train and evaluate models. In these segmentation datasets, BCC tissue was pixel-wise 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 able to accurately 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 referred [15], resulting in rounded images of size 3200×2500 pixels. WSI data were obtained for 128 patients at 40× resolution. As the obtained tissue size varies among patients, the size of WSI varies from 63488×53248 to 22784×19840. Examples of MOIC/MOIS and WSI are shown in Figure 1. Some test images had an extremely small area of BCC, or were not well focused, or 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 supplementary material to illustrate this phenomenon (eTable 1). 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 labeled 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 Institute of Dermatology, PUMC & CAMS (No. 2013-LC/KY-033).

BCC Classification and Segmentation Framework

Figure 2 shows the flowchart of CNN classification and semantic segmentation systems. The three systems accept whole slide images or microscope ocular image 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. 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, the image was classified as positive.

General Screening

GoogleNet inception v3 architecture with input size of 961 and a global average pooling was used to build the classification network [16-17]. The network weights were pre-trained 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. Since the effective area of the MOIC images is a circular area, 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 of 3318 positive samples and

2916 negative samples as the training set. The model was trained on 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 whole slide images, we further developed a classification model using the WSI dataset. Since WSIs were huge and contained large invalid areas (e.g. non-tissue regions), we employed a FOV (field of valid) picking algorithm to extract valid tissue areas. Then we cut the valid area into small patches of size 1000x1000. In total, 7084 patches were obtained (2789 BCC patches and 4285 normal tissue patches), 70% of which were used for model training (4604 patches) and validation (355 patches), and the other 30% (2125 patches) were used for model testing. Similar 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 16860 positive and 16241 negative samples.

In-depth analysis

To accurately detect and locate BCC in hard cases, we employed a semantic segmentation model at the in-depth analysis stage. The model was built on DeeplabV3 network structure [20] with Resnet 101 [7] as backbone. The model backbone was pre-trained on ImageNet and finetuned with the MOIS dataset.

To build the semantic segmentation model, 1436 MOIS images with 1024×1024 resolution were pixel-wise labeled 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 28140 training and 2008 validation patches. The rest 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 NVIDIA Titan-X GPU ×2 with CUDA version 9.0 cuDNN 7.0. 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 AUC were used to evaluate the overall performance of BCC recognition. The ROC curve represents a dynamic tradeoff 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, MOI classification, and MOI segmentation models. For each model, we randomly sampled 80% of the test images and used this subset to calculate the AUC. We repeated this for 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 MOI or WSI dateset

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

As shown in Table 1, the performances of models trained on MOI and WSI images are comparable. The AUCs of the WSI (40x), WSI (10x) 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 (10x) model (AUC=0.955) (p-value = 1.451e-11 and 2.902e-11, respectively). The MOI models outperformed the WSI (10x) model. Therefore, the deep

learning model trained on microscope ocular images could be an alternative and an efficient way of diagnosing BCC in pathological images.

Semantic segmentation improves BCC recognition in hard cases

As we mentioned in **Methods**, the classification model filtered out hard cases with low confidence at the first stage. Subsequently, the semantic segmentation model further analyzed these hard cases at the in-depth analysis stage. The generalization ability of a model was evaluated by cross-dataset AUCs in Table 1. Of note, the segmentation model not only had the best generalization ability (0.945 across dataset), but also accurately classified hard cases within dataset with the AUC of 0.933. In contrast, the classification model at the first stage did not detect BCC very well in hard cases, with the AUC of only 0.701. Since the sizes of BCC 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 we mentioned above, the ROC curve demonstrates a trade-off between the sensitivity and specificity. We select three operation points on ROC curve to demonstrate the model performance. The first operation point on the right emphasizes on 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 on specificity, on the contrary. The high sensitivity and specificity indicate that the semantic segmentation model is able to accurately recognize BCC based on smartphone-captured MOI pathological images.

The runtimes of the classification, cascade and segmentation frameworks

To evaluate the runtimes, the classification and segmentation models were tested on NVIDIA Titan X GPU. A total of 100 MOI images were used to calculate the prediction runtimes. The average runtime of the MOI segmentation model is 500.2 ± 163.2 millisecond (ms) per image (n=100), whereas the MOI classification 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 CPU (Intel Xeon E5 2620), the runtime of the segmentation model is 15.3 ± 3.9 second (s) (n=100) per image, but the classification model and cascade

model are much faster, with the runtime of 2.7 ± 0.7s and 4.1 ± 1.4s respectively (n=100). These results show that the semantic segmentation model requires more time to make predictions.

Locating BCC regions with semantic segmentation

We also demonstrate the performance of locating BCC area through semantic segmentation. The training data were labeled with BCC as foreground and all other tissues as background. The mIoU reaches 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 data sets. 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 labeling and allows them to concentrate on difficult cases, and work as supervisors for straightforward cases with higher efficiency [22].

The digital whole slide image 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 cancer, sentinel lymph node, and prostate cancer [14,25-26]. Researchers integrated image patch classification from WSI and post-process to recognize specific metastasis in high resolution images, such as 40× 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 multi scale image patches in 10×, 20× and 40× did not bring significant improvement in classification performance. The network trained on image patches, either in 40× or in multi scale resolution, concentrates on the local details or the cellular level to distinguish the cancer metastatic area.

During the practice of cutaneous pathologist in diagnosis of BCC, it is the architecture rather than the cellular structure that is more important in making diagnostic decision. Based upon

this observation, we designed our deep neural network to identify BCC on 10× resolution instead of 20× or 40× 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× digital pathology images, we used smartphone camera to take pictures through microscope eyepieces instead of slide scanning machine, which was fast and convenient. To evaluate the effectiveness of smartphone-captured digital pathology images in detecting BCC, we built models using either smartphone-captured microscopic ocular images or regular whole slide images. Notably, the results manifest 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×) and WSI (40×) models are 0.955 and 0.982, respectively. In addition, the smartphone image-based model can be generalized to whole slide images, indicating the robustness of our method.

The classification model trained on MOIC data is able to detect "standard" BCC cases, yet it mis-classifies the "hard" cases which are small BCC areas or obscured due to 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 pixel-wised annotation, since the annotation confidence related information, 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 indicates how confident the labeling can be.

We further developed an end-to-end segmentation model to directly predict BCC. This model achieves the best prediction performance, with the mean IoU of 0.863 and classification AUC of 0.987, comparable to previous studies. The mean IoU of the DeepLab V3 model on Pascal VOC 2012 validation dataset is 0.764 [20]. A similar study that employs DeepLab V3 on skin lesion segmentation achieves mean IoU of 0.816, 0.885 and 0.774 on three open datasets [27]. In the case of 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 framework, the "segmentation" framework requires 15.3 ± 3.9 second (s) to predict one image, whereas the classification framework only needs 2.7 ± 0.7s, and the "cascade" framework needs 4.1 ± 1.4s. For devices which have limited computational resource (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, limited by the cost of the system, digital slide storage, inability to handle high-throughput routine work, the adoption of WSI in disease diagnosis by pathologists has been slow worldwide [29]. So remarkably differed from the previous WSI based dataset, the resources of dataset we used to train the CNN algorithm were creatively designed 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 diagnosis method to extend the diagnosis capability for hospitals without WSI scanning facilities. The advantages of using MOI images includes low cost, easily smartphone storage and capability on handling high-throughputs routine work. All these advantage of MOI images will benefit pathologists and patients in practice. To our best knowledge, this is the first attempt to use 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 the computations resources are limited.

Conclusion

In this study, we developed deep neural network models for BCC recognition, based on smartphone-captured images of digital pathology through the ocular of a microscope. Compared with the model trained on WSI, we showed that the MOI-based model achieves better performance on detecting hard BCC cases and this model is generalizable across MOI and WSI image types. The accurate prediction performance of the MOI-based model and the feasibility of collecting MOI images 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 AcadDermatol.* 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; 10:2483-2489. doi: 10.2147/OTT.S130371.

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(11): 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 24 January 2018).

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 24 January 2018).

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 24 January 2018).

10. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. Available at: https://arxiv.org/abs/1703.06870 (Last accessed 24 January 2018).

11. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Available at: https://arxiv.org/abs/1411.4038 (Last accessed 24 January 2018).

12. Pinchaud N, Hedlund M. CAMELYON17 GRAND CHALLENGE. Available at: https://camelyon17.grand-challenge.org (Last accessed 24 January 2018).

13. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH. Deep learning for identifying metastatic breast cancer. Available at: https://arxiv.org/abs/1606.05718 (Last accessed 24 January 2018).

14. Liu Y, Gadepalli K, Norouziet M, et al. Detecting Cancer Metastases on Gigapixel Pathology Images. Available at: https://arxiv.org/abs/1703.02442 (Last accessed 24 January 2018).

15. Zhou C, Yu Y, Xue R, Elston DM. High-quality digital photomicrography utilizing a smartphone without adapter. *J CutanPathol.* 2016; 43(1): 82-84.

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 24 January 2018).

17. Lin M, Chen Q, Yan S. Network in network. Available at: https://arxiv.org/abs/1312.4400 (Last accessed 24 January 2018).

18. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. Available at: https://ieeexplore.ieee.org/document/5206848 (Last accessed 24 January 2018).

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 24 January 2018).

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 24 January 2018).

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 24 January 2018).

22. Kayser K, Görtler J, Bogovac M, et al. AI (artificial intelligence) in histopathology--from image analysis to automated diagnosis. *Folia HistochemCytobiol.* 2009; 47(3): 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(Pt 2): 411-8.

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(8):e0161286.

25. Ehteshami BB, Veta M, Johannes van DP, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA.* 12,318(22):2199-2210. doi: 10.1001/jama.2017.14585.

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 24 January 2018).

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.

Figure Legends

Figure 1. Examples of MOIC and MOIS images and WSI. (a,b,c) Pathology image examples photographed from the microscope eyepiece using smartphones. In hard examples, red arrows point at the small foci of BCC. (d) Whole slide image.

Figure 2. Schematic diagrams of 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 standard case (the high confidence) or transfers the hard case (the low confidence image) to a segmentation model for further analysis.

Figure 3. Receiver operating characteristic curves of the classification and segmentation model. The 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 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.

Figure 4. Segmentation of BCC with MOIS images. The first column is the original microscope ocular image. The second column is the manual labels (BCC tissue in grey and normal tissue in black). The last column is the prediction of the microscope ocular segmentation model (BCC tissue in red).

Table

Table 1. AUC (95% Confidence Interval) of deep learning models and test datasets. WSI (10X images) classification model, WSI (40X images) classification model and MOI classification model are presented. Each model is tested within dataset (the same dataset on which the model is trained), across dataset and its own hard-case. The hard cases processed by MOI segmentation model is same with that of MOI classification model. Further information is given in supplementary materials (eFigure 1).

Test Dataset / Model	Across Dataset	Within Dataset	Within Dataset Hard Cases
WSI (40X)	0.531 (0.516, 0.542)	0.982 (0.969, 0.993)	0.871 (0.736, 0.980)
WSI (10X)	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)

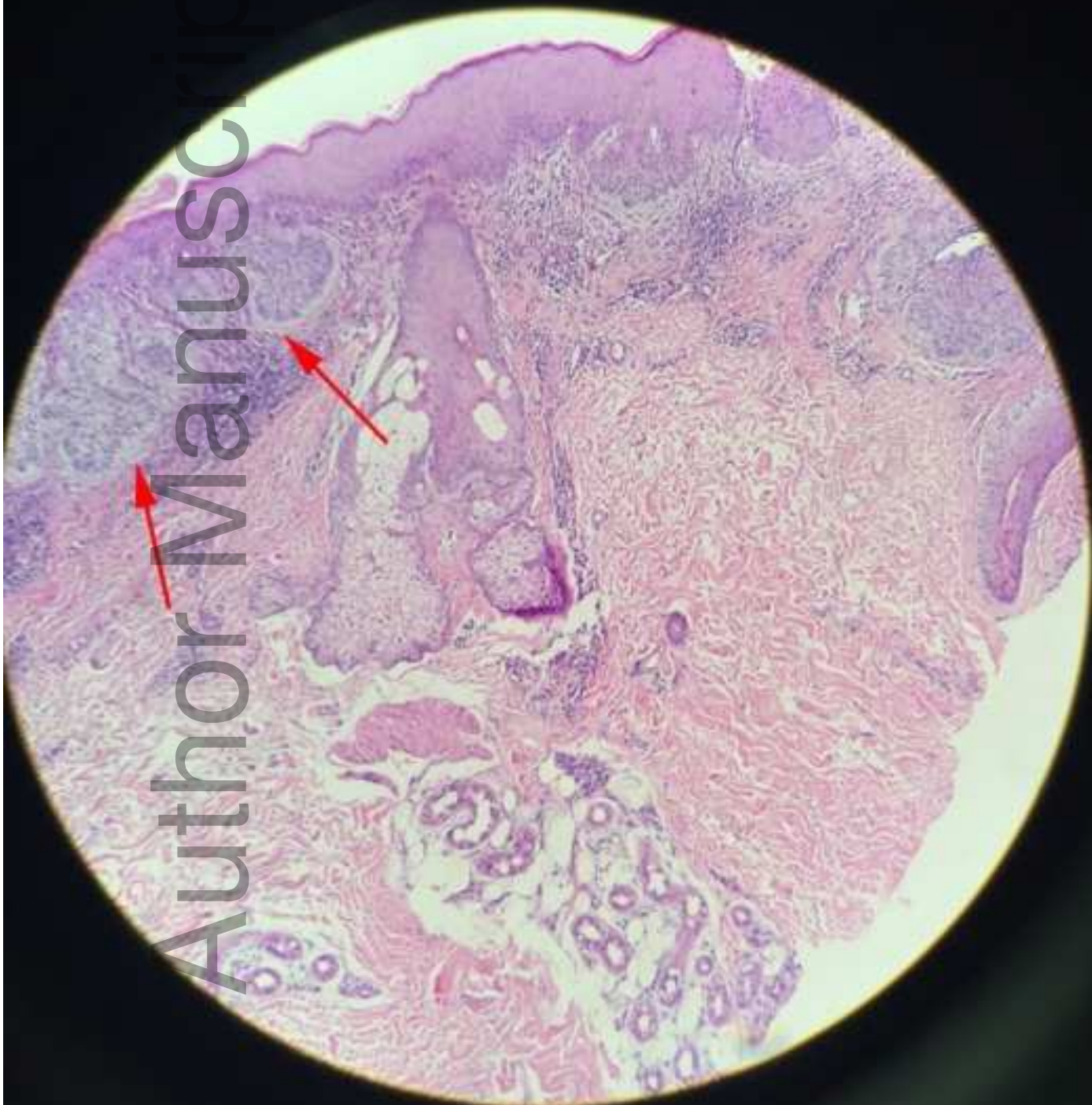*95% CIs were calculated using 2,000 bootstrap samples.
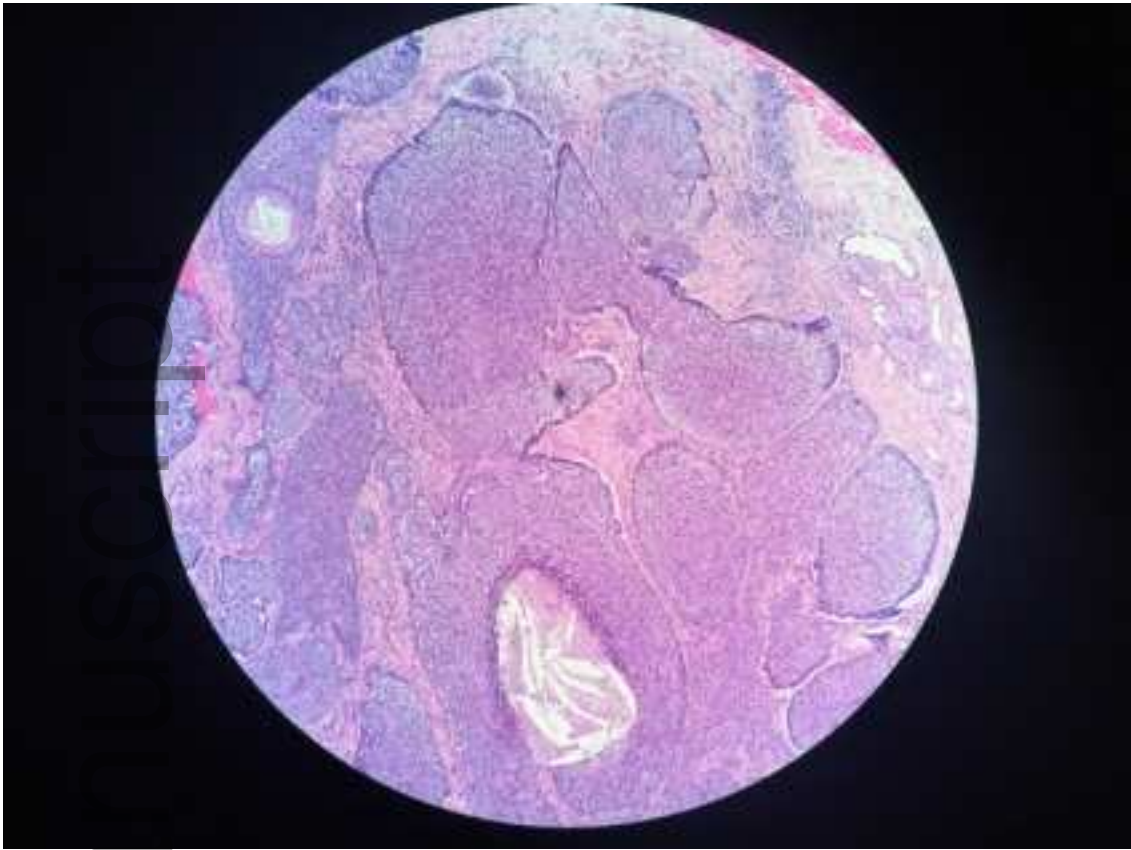
bjd_18026_f1a.jpg
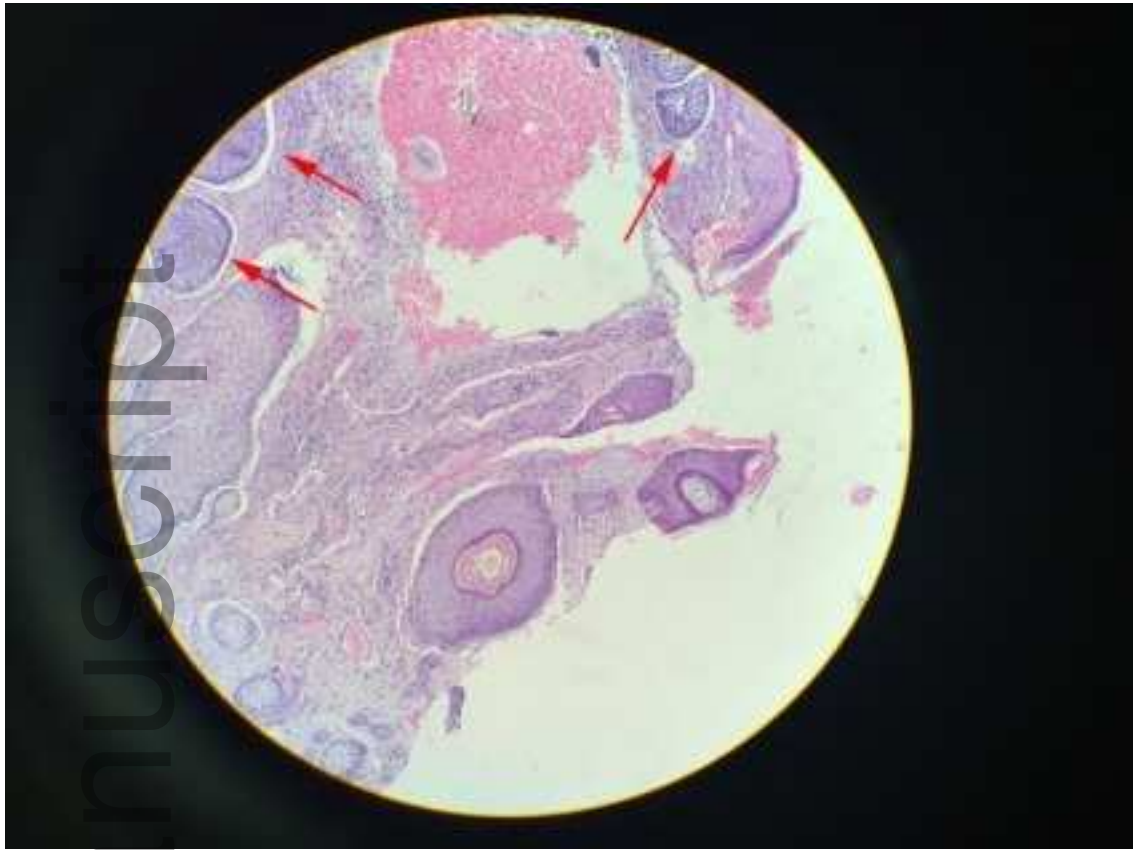
bjd_18026_f1b.jpg
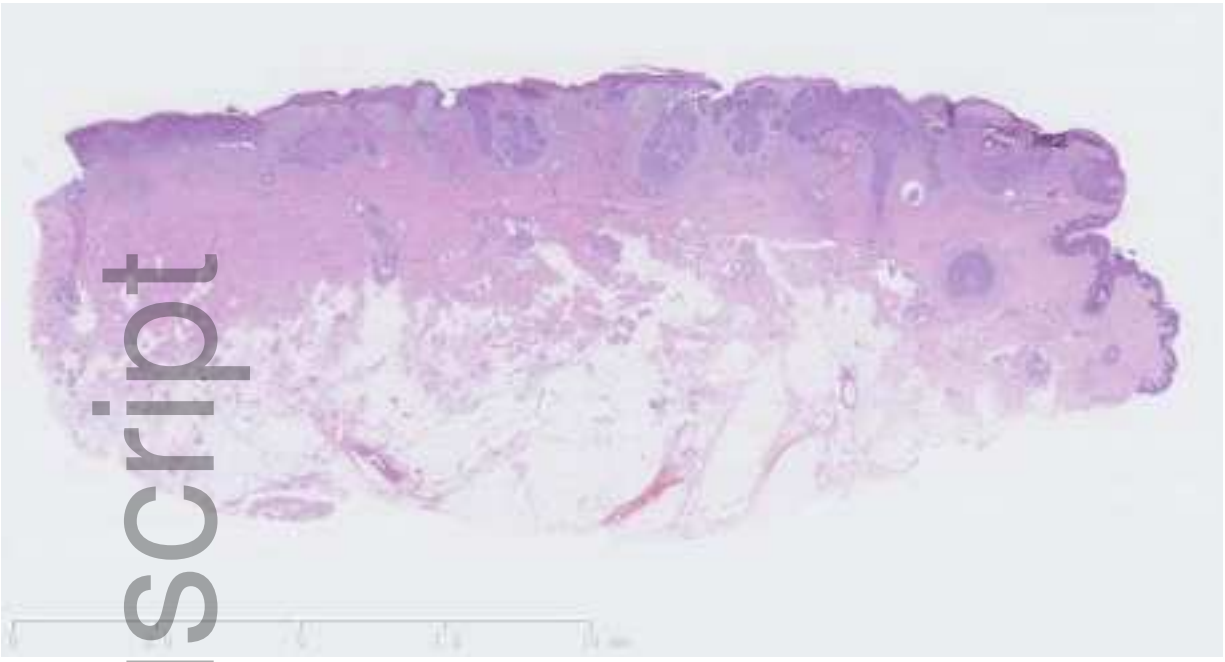
bjd_18026_f1c.jpg
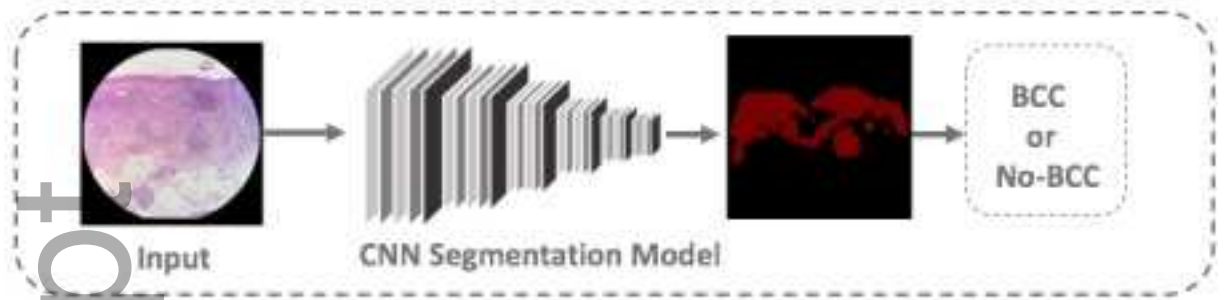
bjd_18026_f1d.jpg

(a) Segmentation Based BCC Recognition

bjd_18026_f2a.jpg

(b) Classification Based BCC Recognition

bjd_18026_f2b.jpg

(c) Cascade BCC Recognition
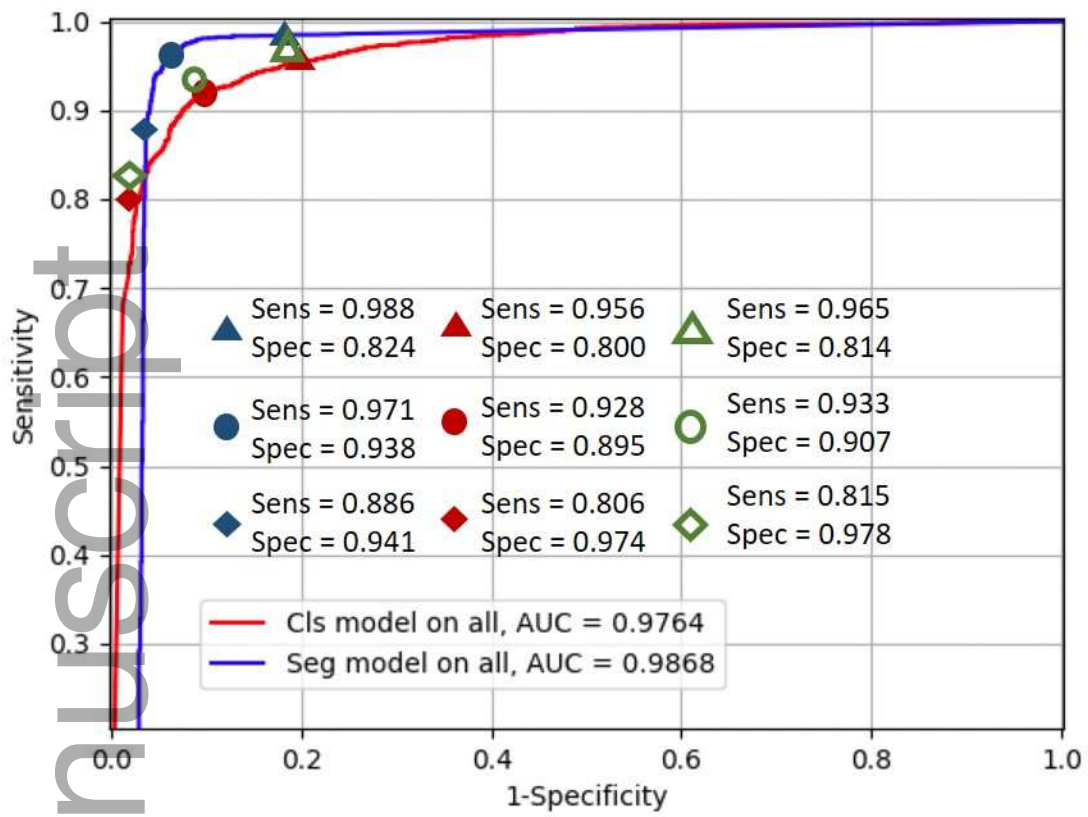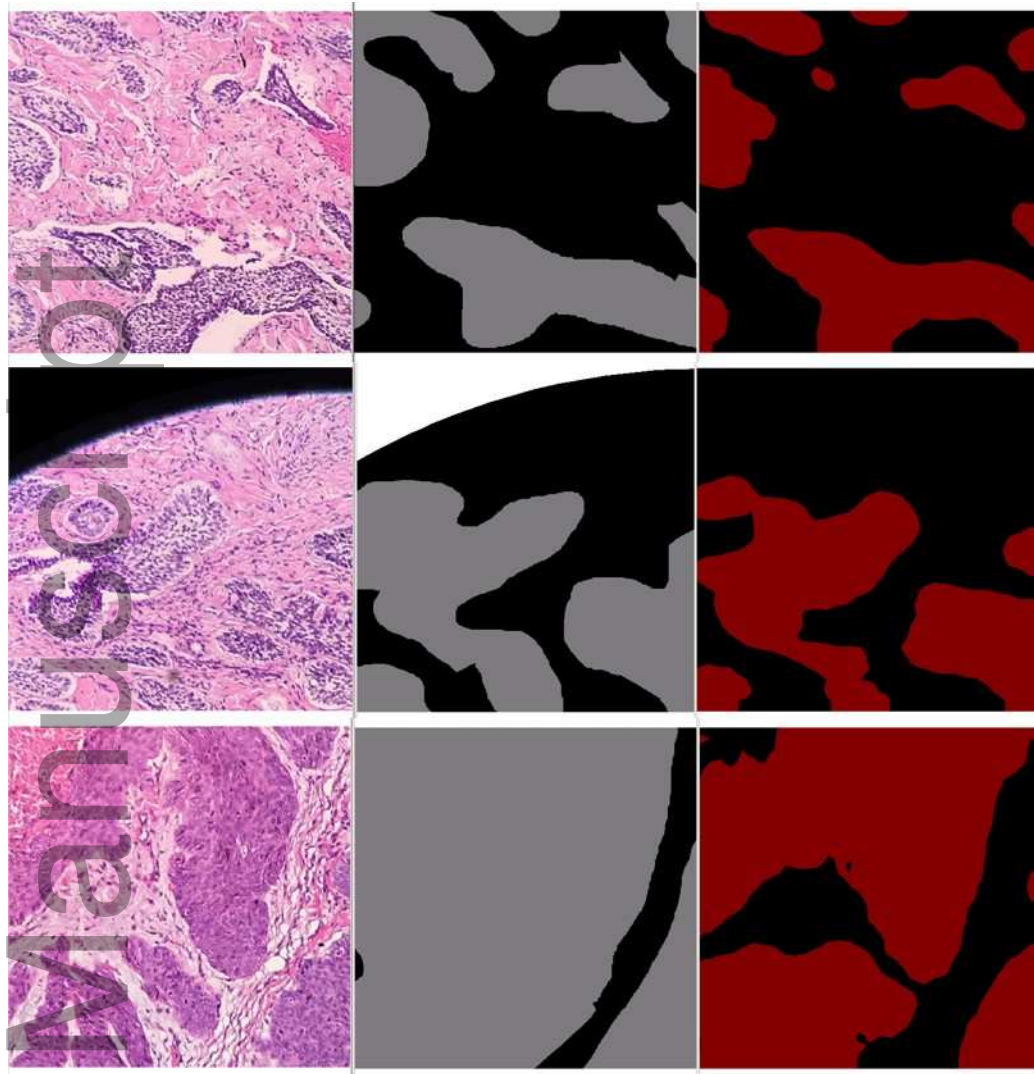
Input

CNN Classification Model

Hard Case?

Yes

No

Segmentation Based BCC Recognition

Classification Based BCC Recognition
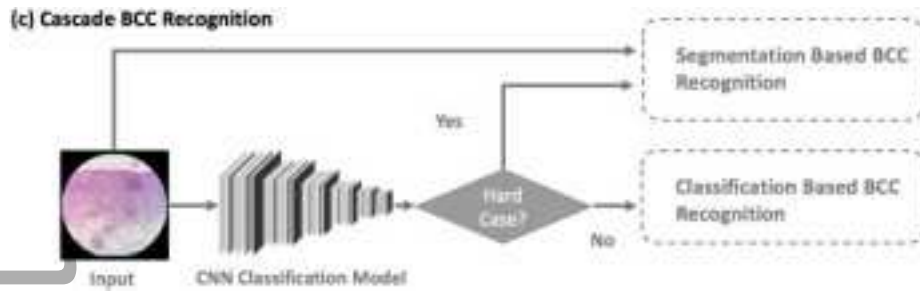
bjd_18026_f2c.jpg

bjd_18026_f3.jpg

bjd_18026_f4.jpg