

## FULL ARTICLE

# Detecting mouse squamous cell carcinoma from submicron full-field optical coherence tomography images by deep learning

Chi-Jui Ho<sup>1</sup>  | Manuel Calderon-Delgado<sup>2</sup>  | Chin-Cheng Chan<sup>1</sup>  |  
Ming-Yi Lin<sup>3</sup> | Jeng-Wei Tjiu<sup>3</sup> | Sheng-Lung Huang<sup>1,2</sup>  | Homer H. Chen<sup>1,2\*</sup> 

<sup>1</sup>Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan

<sup>2</sup>Graduate Institute of Photonics and Optoelectronics, National Taiwan University, Taipei, Taiwan

<sup>3</sup>Department of Dermatology, National Taiwan University Hospital, Taipei, Taiwan

## \*Correspondence

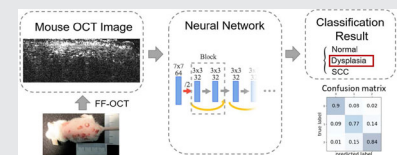
\*Homer H. Chen, Department of Electrical Engineering, National Taiwan University, Taipei 10617, Taiwan.  
Email: homer@ntu.edu.tw

## Funding information

Ministry of Science and Technology, Taiwan, Grant/Award Numbers: 103-2325-B-002-044, 107-2634-F-002-017; National Taiwan University, Grant/Award Number: 109L891707

## Abstract

The standard medical practice for cancer diagnosis requires histopathology, which is an invasive and time-consuming procedure. Optical coherence tomography (OCT) is an alternative that is relatively fast, noninvasive, and able to capture three-dimensional structures of epithelial tissue. Unlike most previous OCT systems, which cannot capture crucial cellular-level information for squamous cell carcinoma (SCC) diagnosis, the full-field OCT (FF-OCT) technology used in this paper is able to produce images at sub-micron resolution and thereby facilitates the development of a deep learning algorithm for SCC detection. Experimental results show that the SCC detection algorithm can achieve a classification accuracy of 80% for mouse skin. Using the sub-micron FF-OCT imaging system, the proposed SCC detection algorithm has the potential for in-vivo applications.



## KEYWORDS

computer-aided diagnosis, convolutional neural network, deep learning, optical coherence tomography, squamous cell carcinoma

## 1 | INTRODUCTION

Cancer accounts for one-quarter of the deaths caused by noncommunicable diseases, yearly killing millions of people worldwide, although 30% to 50% of such casualties could be prevented.<sup>1</sup> There exist hundreds of types of cancers, which can be classified according to their origin or the type of tissue affected. However, as cancer evolves, it can metastasize, spreading to other tissues and organs. Therefore, an early diagnose will determine the success of treatment and, eventually, the chance of survival.<sup>2</sup> This paper is primarily concerned with skin cancer detection by OCT and deep learning.

Skin cancer, which can be classified into melanoma and nonmelanoma skin cancer (NMSC), is among the most common types of cancers.<sup>3</sup> Melanoma skin cancers are more aggressive and pose a higher risk but are less frequent and easier to detect due to their characteristic pigmentation and irregular shapes.<sup>4</sup> Two common types of NMSC are basal cell carcinoma (BCC) and squamous cell carcinoma (SCC), the latter having a lower incidence rate.<sup>5</sup> Other types of NMSC account for less than 1% of the cases altogether. Despite the lower incidence rate, SCC is more likely to spread to other tissues and metastasize than BCC and hence more important to detect in the early stage.<sup>6</sup> However, it is hard to diagnose. A recent

study reported that the number of NMSC cases has grown drastically.<sup>7</sup> However, many review papers manifest the lack of study in cutaneous SCC.<sup>8–10</sup> In this work, we focus on SCC detection.

The gold standard in clinical assessment of skin cancer is hematoxylin and eosin (H&E) stain histology due to its high specificity.<sup>11</sup> However, H&E stain histology requires the use of chemicals that may influence the structure of the tissue to be examined. In addition, H&E techniques require the excision of tissue, which has the risk of complications due to scarring, bleeding, and infections and thus increases the diagnose time.

To avoid these problems, noninvasive diagnosis has been developed, such as dermoscopy, reflectance confocal microscopy (RCM), and optical coherence tomography (OCT). Dermoscopy is widely used in the diagnose of pigmented skin lesions, such as melanoma, and in recent years of nonpigmented lesions.<sup>12</sup> However, it can only capture two-dimensional superficial images at a small magnification (typically a 10-fold magnification), which makes small structures such as papillary vessels and cell nuclei difficult, if not impossible, to identify. On the other hand, RCM offers micrometer lateral resolution with an intrinsic tradeoff between axial resolution and imaging depth. The former is typically 3  $\mu\text{m}$  to 5  $\mu\text{m}$  and the latter 100  $\mu\text{m}$  to 200  $\mu\text{m}$ .<sup>13</sup> Thus, RCM is often used to visualize en-face planes, not cross-sectional images. In histological practice, viewing cross-sectional images is easier to read.

OCT is based on the interference of low-coherent light. It combines a sub-micrometer axial resolution with a mid-range imaging depth (0.2–2.0 mm) and is able to perform in-vivo measurements in a noninvasive way.<sup>14</sup> Full-field OCT (FF-OCT) is a variant of OCT that has the ability to capture a data volume within a few minutes or even seconds in a single A-scan using a camera sensor instead of a single photo-diode.<sup>15</sup>

The advantages of OCT make it one of the most plausible replacements for H&E histology and a perfect solution for the evaluation of different diseases of epithelial tissue. Past applications of OCT include diagnose,<sup>16, 17</sup> margin delimitation in intervention assessment,<sup>18</sup> and follow-up checks.<sup>19</sup> However, as a new technology, OCT image understanding is not a familiar task for physicians; it takes time to train. Therefore, manual instructions for classification of NMSC have been created.<sup>20, 21</sup> Automatic classification based on the support-vector machine has also been developed for BCC detection involving manual<sup>22</sup> or automatic<sup>23, 24</sup> feature extraction. However, these methods were developed with small datasets. Besides, they cannot provide interpretable information.

To address the above issues, fully automated algorithms based on convolutional neural networks (CNNs) have been developed. This type of algorithm has achieved

great success in the field of OCT imaging for retina.<sup>25, 26</sup> In contrast, the field of OCT imaging for skins is still under development and very few studies have been reported using machine learning for automatic segmentation<sup>27</sup> or classification<sup>28</sup> of this kind of images, perhaps, because of the difficulty in obtaining good quality images from an inhomogeneous, turbid media with a high and varied scattering distribution.<sup>29</sup> As a result, the relevant literature is scarce. The most relevant approach we can find is the CNN-based method for BCC classification.<sup>28</sup> However, besides the fact that it is about BCC detection instead of SCC detection, the images used are 2D slices from ex vivo samples acquired by Mohs surgery, which cannot be extended to in-vivo experiments.

In this paper, we propose an algorithm for diagnosis of SCC using an FF-OCT system and a CNN-based classifier. Although the images used in this study are taken from excised tissue, in-vivo FF-OCT images have a similar quality and can also be used for this study. The typical OCT light intensity is about 8 W/cm<sup>2</sup>, which can be easily obtained by focusing a red laser pointer. The resulting system is able to differentiate between normal, dysplasia, and cancerous samples. The success of our SCC detection system is due in part to the exploitation of sub-cellular features captured in the FF-OCT images and in part to the adaptation of multi-level receptive fields to extract features at different scales. We explain the system behavior by analyzing how various skin features contribute to SCC detection through heat map visualization.

The remainder of this paper is organized as follows. Section 2 describes the hardware configuration of the FF-OCT system used in this work, the SCC and its characteristics, and the steps to induce cancer and classify the tissue. In Section 3, we discuss how to improve the quality of raw images, how to screen the tomograms, and how the CNN-based methods perform automatic image classification. We summarize our experimental results in Section 4 and discuss the visualization of SCC detection in Section 5.

## 2 | BACKGROUND

In this section, we review SCC, FF-OCT, and deep learning methods for image classification.

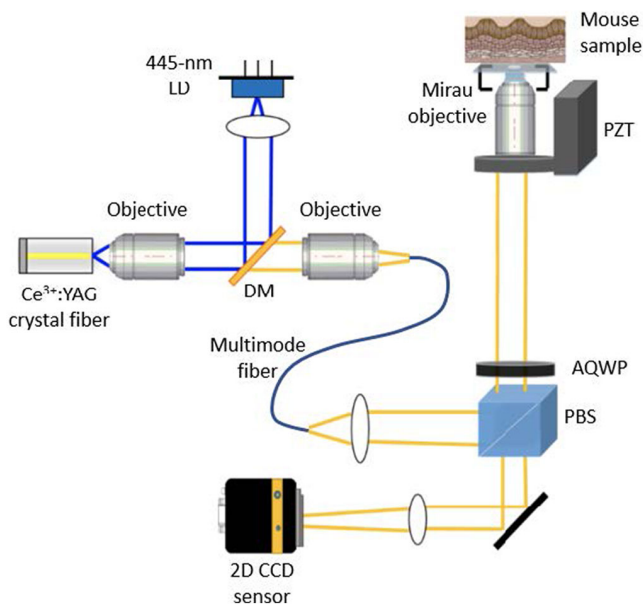
### 2.1 | Full-field OCT

FF-OCT is a noninvasive imaging technique for data collection from optical scattering media. It uses low-coherence interferometry to analyze the reflected light from objects for scan of sub-surface tissue, such as retina and

skin. Unlike the common OCT technique that scans along the axial direction to capture cross-sectional images, FF-OCT captures a sequence of en-face images and stacks them into a 3D volume. FF-OCT has two main advantages over the common OCT technique. First, FF-OCT acquires the whole field of view in a single scan, which increases the imaging speed and avoids inaccuracies in the lateral scan needed by OCT. Second, the FF-OCT has better sensitivity and higher resolution along the axial direction since it has longer exposure time per en-face image.

The device used in this work to acquire volumetric data of mouse skin is a Mirau-based FF-OCT system as shown in Figure 1. The light source is a cerium-doped yttrium aluminum garnet (Ce:YAG) single-clad crystal fiber. The fiber, which is fabricated by laser-heated pedestal growth, generates a Gaussian spectrum with central wavelength at 560 nm and a full-width half-maximum of 98 nm.<sup>30</sup> The Mirau interferometer is filled with silicon oil to mimic the refractive index of the sample tissue to reduce the optical path difference between the reference and sample arms and thus provides a more accurate depth estimation of the structures in the images. Furthermore, the use of a single objective (Olympus UMPLFLN20XW) stabilizes the system and reduces the aberrations generated by small differences in the objectives when using a Michelson interferometer.

The system has an isotropic resolution of 0.9  $\mu\text{m}$  in air, and the voxel size is  $0.45 \times 0.45 \times 0.2 \mu\text{m}^3$ , with the



**FIGURE 1** Schematic diagram of the FF-OCT system. AQWP, achromatic quarter-wave plate; CCD, charge coupled device; DM, dichroic mirror; FF-OCT, full-field optical coherence tomography; LD, laser diode; PBS, polarizing beam splitter; PZT, piezoelectric transducer

pixel depth being the smallest in size. The dimensions of en-face images are  $648 \times 488$  pixels ( $291.6 \times 219.6 \mu\text{m}^2$ ), whereas the depth varies between tomograms, ranging from 200 to 600 pixels ( $40\text{--}120 \mu\text{m}$ ).

## 2.2 | Animal model

In this work, all the images were collected from an internal study that took place from January 2015 to August 2016. At the beginning of the study, the subjects were 30 FVB/N mice aging 6 to 8 weeks, a well-known animal model for the growth of cutaneous SCC because of its easy growth.<sup>31</sup>

In the first week of the experiment, the mice's backs were shaved, and an immunosuppressor solution (100  $\mu\text{g}$  7.12-Dimethylbenz[a]anthracene dissolved into 0.2 mL of acetone), serving as a tumor initiator, was applied to their skin. During the following 20 weeks, a tumor promoter (25 mg 12-O-Tetradecanoylphorbol-13-acetate in 100 mL acetone) was applied weekly. The abdominal skin of the mice was left untreated to act as control samples.

Figure 2 shows the mouse back skin, presenting an inflamed epidermis and protruding tumor nests. For each mouse, if several tumors grew over 5 mm in diameter, we sacrificed it and then excised skin samples. These samples were covered with wax and formalin, and classified as either normal, dysplasia, or SCC, according to their location and appearance. Images of these samples were taken by means of FF-OCT and H&E stain.

The structure of mice skin consists of three main layers from shallow to deep: epidermis, dermis, and subcutaneous tissue. Since SCC only affects the epithelial tissue, we only focus on the first two layers. The epidermis can be further divided into stratum basale (SB), stratum spinosum (SS), stratum granulosum (SG), and stratum corneum (SC) from deep to shallow.<sup>32</sup> In normal skin, basal cells in the SB produce new cells called



**FIGURE 2** Photograph of the growth of SCC in the back skin of a mouse after sacrifice. The smallest division of the ruler is 1 mm. SCC, squamous cell carcinoma



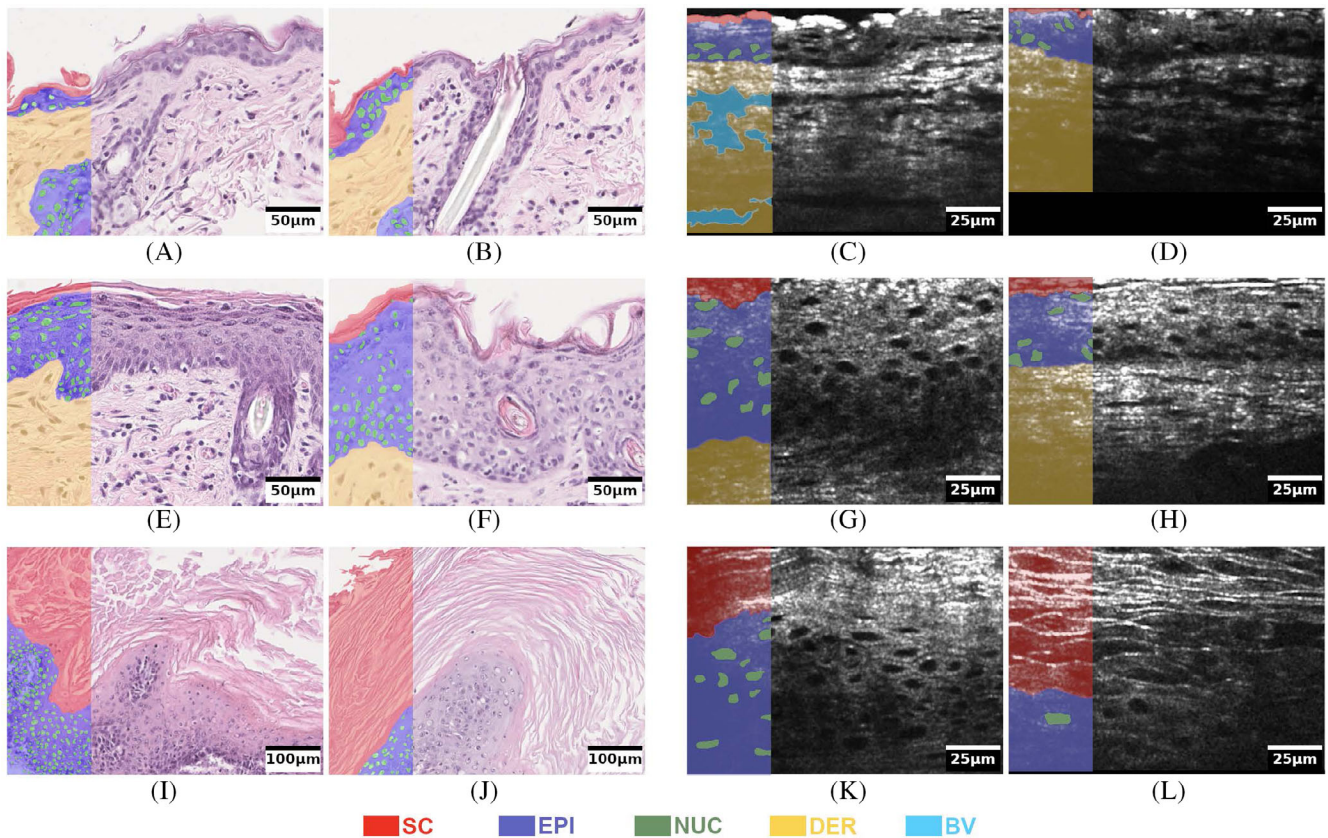
keratinocytes. As time goes, the keratinocytes produced earlier are displaced to the SS by the new ones. Then, the keratinocytes grow in the SS and migrate to the SG. At the SG, they undergo keratinization and become flattened as they move to the surface. After this process, they form a hard and thin structure, SC, consisting of squamous sheets. The SC protects the skin from the outside invasions. Cornified cells are constantly shed off at a rate equal to the production rate of basal cells. The dermis, which primarily consists of dense irregular connective tissue (mainly collagen, elastic fibers, and extracellular matrix), hosts structures such as hair follicles, blood vessels, and immune cells.<sup>32</sup>

When skin is affected by SCC, the SC becomes thicker since the production of cells in the SB is higher than the loss in the SC. Furthermore, keratinocytes grow in size and adopt different shapes. As a result, the dermal-epidermal junction (DEJ) is found at a greater depth than that in normal skin. This difference can be clearly observed in OCT images of normal and SCC skin, shown

in Figure 3A–D and Figure 3I–L, respectively. On the other hand, dysplasia is an intermediate state between normal skin and SCC. In fact, when normal cells undergo mutations by external agents, such as chemicals or radiation, there exist mechanisms that can repair the damage and prevent the development of cancer. If the damage is irreversible, the immune system will try to eliminate these affected cells. Dysplasia might then appear in different stages, depending on the local reaction to the treatment. In the images from our dataset, this intermediate stage shows a thicker epidermal layer with bigger keratinocytes and a slightly thicker or detached SC, as shown in Figure 3E–H.

### 2.3 | Deep learning on image classification

CNN-based algorithms have recently achieved great success in image classification. When training a CNN-based



**FIGURE 3** Cross-sectional images of mouse skin obtained by H&E staining (left side) and our FF-OCT system (right side). Two examples of each SCC diagnosis class are provided to show the image diversity. Note that images of the left and right sides do not correspond to each other. The first row, A–D, corresponds to normal skin, the second row, E–H, corresponds to dysplasia, and the last one, I–L, to SCC. For illustration purpose, the left portion of each image is manually colored to show the following structural components. BV, blood vessel; EPI, epidermis; FF-OCT, full-field optical coherence tomography; DER, dermis; H&E, hematoxylin and eosin; NUC, nuclei; SC, stratum corneum; SCC, squamous cell carcinoma

classifier, we iteratively update its weights by backpropagation until a predetermined condition such as low gradient of training loss is satisfied. Krizhevsky et al. pioneered a deep CNN structure by stacking multiple convolutional layers; the resulting network is called AlexNet.<sup>33</sup> AlexNet outperformed traditional methods in the ImageNet large scale visual recognition competition.<sup>34</sup> However, plain networks like AlexNet have a low convergence rate when the depth of the network increases because of gradient vanishing.<sup>35, 36</sup> One method to alleviate gradient vanishing is batch normalization,<sup>37</sup> which regularizes the distribution of every convolutional layer to ensure an effective backpropagation. To further increase the convergence rate in deep networks, He et al. proposed the residual neural network (ResNet)<sup>38</sup> that uses short-cuts to connect shallow convolutional layers with deep ones. When the depth of a network increases, ResNet converges faster than plain networks do.

Although CNN-based classifiers have achieved superior performance, their decision policy is less transparent than that of traditional methods. Therefore, model interpretation methods have been developed. One simple method is directly visualizing the trained weights, but it only works for the first layer.<sup>39</sup> To realize the features learned by deeper layers, Zhou et al. visualized class-specific feature maps by a localization method called class activation mapping (CAM).<sup>40</sup> However, this method only works for the specific network architecture that performs global average pooling right before the softmax layer. Moreover, CAM only enables the visualization of the last convolutional layer. To interpret the model more generally, Selvaraju et al. proposed a method called gradient-weighted CAM (GRAD-CAM).<sup>41</sup> GRAD-CAM works for a wide-variety of CNN-based architectures and applications, including image classification and segmentation. Moreover, it is able to interpret the network at every convolutional layer, enabling the observation of feature activation at all levels. Specifically, the heat maps generated at the shallow layers of the network show the activation of low-level features instead of high-level features, because the shallow layers focus on local features. On the other hand, the heat maps generated at deep layers show high-level features aggregated from local features; fine features are missing because the heat maps at deep layers are low resolution.

### 3 | PROPOSED METHODS

In this section, we first describe how data are preprocessed and screened. Then, we describe the design of the CNN-based classifier.

#### 3.1 | Image preprocessing

In order to reduce the size of the tomograms and obtain an isotropic voxel size, several methods are applied to the images acquired from our system using the software suite FIJI/ImageJ.<sup>42, 43</sup> At first, a 3D mean filter of diameter 1  $\mu\text{m}$  is applied to reduce the noise while preserving the features resolved by our system. Next, the three dimensions are scaled to obtain a cubic voxel of 0.5  $\mu\text{m}$  for each side, effectively reducing the size of our images by a factor of three and increasing the number of images that can fit into the RAM. Then, the images are saved in the nearly raw raster data (NRRD) format,<sup>44</sup> which saves the data in a single file, together with the necessary information for convenient usage.

Besides filtering, we note that the tomograms in the dataset have different depths since most of the tomograms were taken individually or in small groups and that, the scanning depth was independently selected for each of these runs. To unify the size of images input to our training algorithm, we pad all the tomograms to the same size. The length, width, and depth of each padded tomogram are 576, 439, and 240 pixels, respectively. Moreover, due to the memory constraint, we feed cross-sectional images instead of the whole tomogram to our training algorithm. Therefore, the size of each FF-OCT image becomes  $576 \times 240$  pixels.

#### 3.2 | Data selection

Since the dimensions of the samples were bigger than the field of view of the FF-OCT system, tomograms were taken in a sequential, automatic way by using a translational stage. While this speeds up the acquisition and increases the total number of tomograms, some of the volumes contain irrelevant information for classification or have poor quality due to the noise from vibration or darkness resulted from, for example, dirt on the sample surface. Therefore, the tomograms are manually screened, and the main criteria of tomogram selection were the visibility of cells in the epidermis and the DEJ. The final number of tomograms available to our study is 297, of which 100 are from normal skin, 97 from dysplasia, and 100 from SCC. For each class, we reserve around 15% of the tomograms for testing and use the remaining 85% for training. Since there are 439 cross-sections across the Y-axis of each tomogram, we use a total of 130 383 images.

#### 3.3 | Disease classifier

We adopt ResNet<sup>38</sup> as our base approach for disease classification. Normal ResNet consists of four stages, each of which has a different number of blocks. Each block contains two convolutional layers and a skip-

connection from the input layer to the output layer. Various versions of ResNet have been developed. The shallowest version, ResNet-18, is chosen in this work. To further reduce the computation, we halve the number of filters in every convolutional layer. The resulting network is called Pruned-ResNet-18, which contains 32, 64, 128, and 256 filters in the four stages. The network architecture is shown in Figure 4.

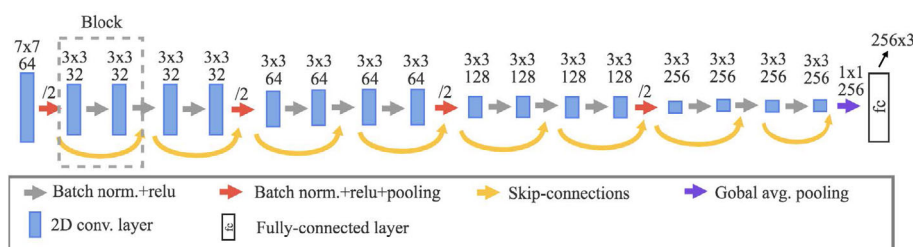
One important characteristic of this CNN-based classifier is that the receptive field of a convolutional kernel can capture features of different sizes.<sup>45</sup> We find that the features of SCC have to be captured by a large receptive field, while the features of dysplasia and normal skin can be captured by a small receptive field, as shown in Figure 5. Therefore, convolutional kernels with different sizes of receptive field are used in the network to capture features of different sizes. To investigate further, we compared Pruned-ResNet-18 with ResNet-18, Pruned-ResNet-5, and AlexNet. Pruned-ResNet-5 consists of only the first five layers of Pruned-ResNet-18 for feature extraction. In comparison, AlexNet applies five convolutional layers without skip-connection to extract features and uses pooling to increase the receptive field in each layer. The receptive fields of these models are illustrated in Figure 5. Since the receptive fields of ResNet-18 and Pruned-ResNet-18 are equal in size and the receptive field of Pruned-ResNet-5 corresponds to the first five layers of Pruned-ResNet-18, only the receptive field of Pruned-ResNet-18 is shown. The size of the receptive field matters in this work because it determines the extent of fine-to-coarse features to be captured by the network (see Section 4).

## 4 | EXPERIMENTS

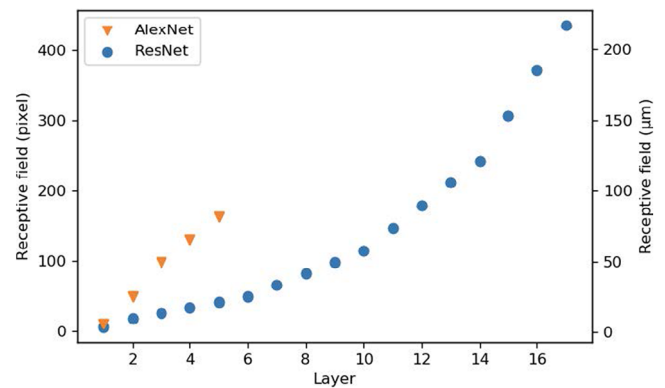
We compare the SCC detection performance of Pruned-ResNet-18 and three other networks. The details of the setup and implementation of this experiment are described in this section.

### 4.1 | Experimental setup

Although we use 2-D images for model training, the actual data for testing are 3D tomograms. Therefore, in the testing stage, we adopt the winner-take-all policy<sup>46</sup> to aggregate the



**FIGURE 4** An illustration of Pruned-ResNet-18. The kernel size and the number of filters are shown above each convolutional layer. The labels on red arrows denote the down sampling rate of pooling layers



**FIGURE 5** The receptive field of each layer of AlexNet (orange) and ResNet (blue). One pixel corresponds to  $0.5 \mu\text{m}$  in images

results of SCC detection for images in the same volume to a single value. In other words, every image in a volume is first assigned a label corresponding to an SCC diagnosis class. Then, the label that receives the most votes becomes the label of the whole volume. The accuracy of the labeling is expressed as a normalized confusion matrix.<sup>47</sup>

We apply k-fold cross-validation<sup>48</sup> to measure the performance of our method with different training and testing data. Specifically, we partition the whole dataset into 10 subsets. In each training process, one subset is reserved for validation, and the others are used for training. This procedure is repeated until all the 10 subsets have been validated.

### 4.2 | Implementation

All the classifiers are trained and tested using PyTorch,<sup>49</sup> a deep learning framework that enables fast implementation. The batch size is set to 32, the cross-entropy is used as the loss function, and Adam<sup>50</sup> is used as the optimizer. In model training, we record the validation accuracy of the following 30 epochs after the training accuracy exceeds 99%. Then, we take the average of the 10 highest values as the overall classification accuracy. Our training was performed using an NVIDIA TitanX with an approximate training time of 10 minutes/epoch.



**TABLE 1** Overall accuracy of a 10-fold cross validation

Network	Parameters	Fold index										Avg	Stdev
		1	2	3	4	5	6	7	8	9	10		
AlexNet	126 M	0.824	0.833	0.707	0.834	0.679	0.760	0.785	0.741	0.800	0.829	0.779	0.056
Pruned-ResNet-5	1.5 M	0.787	0.831	0.771	0.810	0.753	0.790	0.828	0.712	0.848	0.836	0.797	0.043
Pruned-ResNet-18	5.5 M	0.781	0.845	0.745	0.827	0.774	0.779	0.841	0.804	0.868	0.853	<b>0.812<sup>a</sup></b>	<b>0.041</b>
ResNet-18	11 M	0.792	0.852	0.746	0.821	0.777	0.781	0.854	0.754	0.861	0.859	0.810	0.045

<sup>a</sup>The highest average accuracy and lowest standard deviation are shown in boldface.

### 4.3 | Results

Table 1 shows that the overall accuracy of Pruned-ResNet-18 is over 80% and higher than the other networks. The superior performance of Pruned-ResNet-18 demonstrates the importance of the receptive field. As shown in Figure 5, the receptive fields of Pruned-ResNet-18 allow multiple-level feature extraction, which is difficult to do with AlexNet and Pruned-ResNet-5. This powerful strength makes Pruned-ResNet-18 superior to the other two networks in terms of overall accuracy. Furthermore, Pruned-ResNet-18 uses fewer parameters than ResNet-18 without sacrificing accuracy. These favorable results suggest that the classifier can reduce the computational burden of SCC detection. The multiply-add calculation and the computing time of pruned-ResNet-18 are 1.41 G and 21.44 ms/batch, respectively, as opposed to 4.88 G and 39.37 ms/batch of regular ResNet-18. The computing time is obtained on our server with Nvidia Tesla v100.

Figure 6 shows the confusion matrixes of the 10-fold cross-validation. In most cases, the accuracy of dysplasia is lower than the other two classes. Specifically, normal and SCC images are seldom misclassified with each other. However, it is easy to misclassify dysplasia images as normal or SCC. One possible reason is that there is no clean cut between the three SCC diagnosis classes because the tissue evolution from normal to dysplasia and finally to SCC is a continuous process. Therefore, dysplasia, which is an intermediate state, is subject to the highest misclassification rate. On the other hand, it is easy for the classifier to distinguish between normal and SCC since they are two disjoint classes in the tissue evolution process.

### 4.4 | Down-sampling analysis

To investigate how image resolution affects detection accuracy, we retrain the classifier with low-resolution images. These low-resolution images are generated from preprocessed images by average pooling with window sizes ranging from 2 to 32 pixels. The resulting validation accuracy is plotted against image resolution in Figure 7. Note that the image resolution is inversely proportional to the pixel size. We can see that the detection accuracy drops with image resolution. That is, higher image resolution leads to higher validation accuracy. This verifies that FF-OCT image resolution is important to SCC detection. If cellular-level information is not available, the classifier can only rely on coarse image structure information for SCC detection, resulting in performance drop.

## 5 | DISCUSSIONS

In this section, we use heat maps to explain how Pruned-ResNet-18 extracts features and how our algorithm can be further enhanced. In addition, we describe possible extensions to improve the performance of the classifier.

### 5.1 | Heat maps

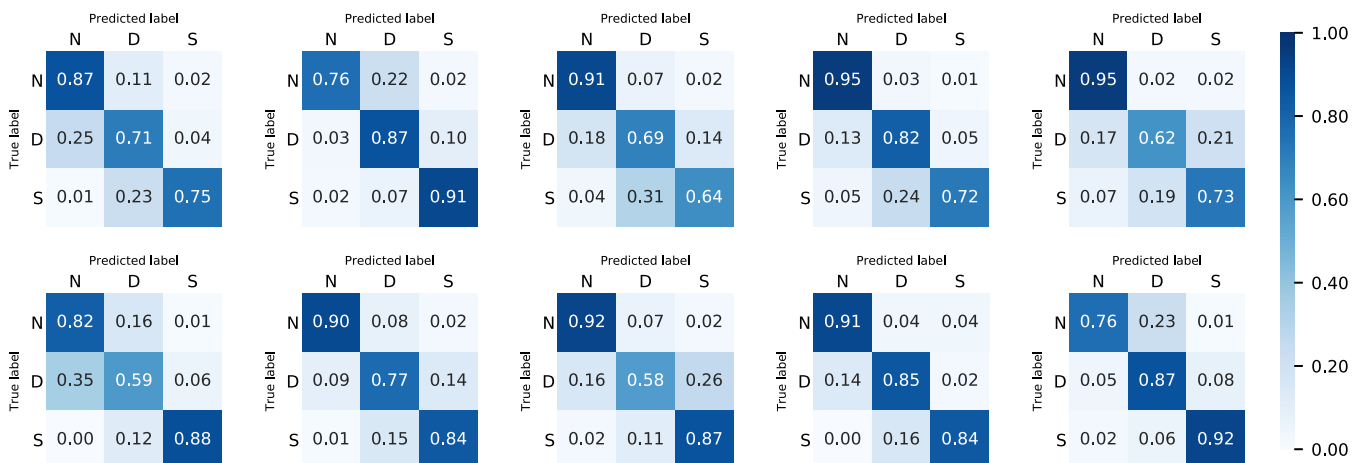
The Pruned-ResNet-18 model detects SCC by capturing features in different layers. In other words, these features are at different scales. This is done by feeding FF-OCT images of all three SCC diagnosis classes to the model and using GRAD-CAM to depict corresponding heat maps.<sup>41</sup> We analyze the heat maps generated at all four stages of our model. The physical sizes of the receptive field in the four stages are  $21 \times 21 \mu\text{m}^2$ ,  $50 \times 50 \mu\text{m}^2$ ,  $105 \times 105 \mu\text{m}^2$ , and  $217 \times 217 \mu\text{m}^2$ , from shallow to deep. The size of the largest cell nuclei in the epidermis is about  $20 \mu\text{m}$  in the present dataset, so the heat map generated at the first stage shows the extracted features at the cellular level. The other three maps show how the cellular-level features are aggregated at different scales.

Figure 8 shows the input images, one from each SCC diagnosis class, and the resulting heat maps. From left to right, the results of images of normal, dysplasia, and SCC skin are shown. HM1 denotes the heat map generated at the first stage, HM2 denotes the heat map generated at the second stage, and so on. Note that the heat maps shown in Figure 8 are upsampled to match the input image size.

The first column in Figure 8 corresponds to a normal sample. From HM1, we can see that the network extracts the SC (yellow ellipses) and the DEJ (yellow rectangles) in the first stage. Note that the features are discrete since the receptive field at this stage is smaller than the size of the whole SC and DEJ. As the receptive field increases in the second and third stages, the features are grouped into larger patches, as shown in HM2 and HM3. From HM4, we can see that the DEJ and its surrounding epidermis and dermis constitute the high intensity patches (red rectangles).

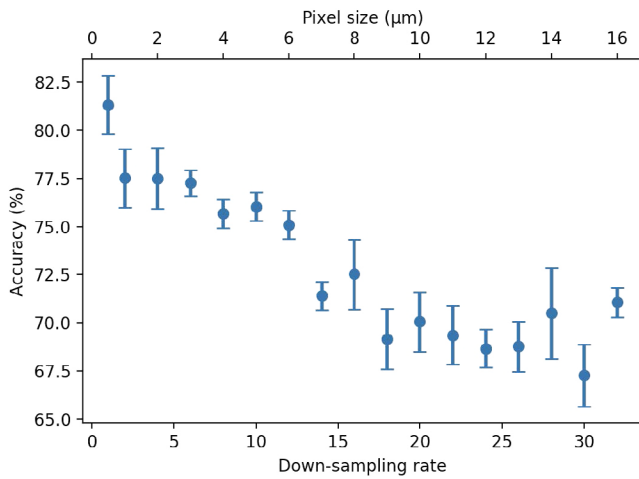
The second column in Figure 8 corresponds to a dysplasia sample. From the HM1 (and the input image with human label) in this column, we can see that the intensity of nuclei (yellow circles) in the epidermis is high, so is the intensity of capillary vessels (yellow ellipses), DEJ (yellow rectangles), and SC. The high intensity patches become the regions around the deepest nuclei in HM2 and HM3. From HM4, we can see that the high intensity patches include the DEJ and its surrounding epidermis and dermis (red rectangles). The results suggest that the deepest nuclei determine the location of the DEJ. It seems that the high intensity regions around the DEJ may encode rich information, such as morphology and distribution of the basal cells, the thickness of epithelium, and micro vasculature.

The last column in Figure 8 corresponds to an SCC sample. From its HM1, we can see that the high intensity regions include the top sheets of the SC (yellow ellipses), the nuclei (yellow circles), and the lower boundary of the SC (yellow rectangles). Similar to the normal case, the extracted features in HM1 are grouped into larger regions in HM2 and HM3. In particular, the two main high



**FIGURE 6** Normalized confusion matrices of a 10-fold cross-validation. The normal, dysplasia, and SCC classes in each matrix, are coded N, D, and S, respectively. SCC, squamous cell carcinoma



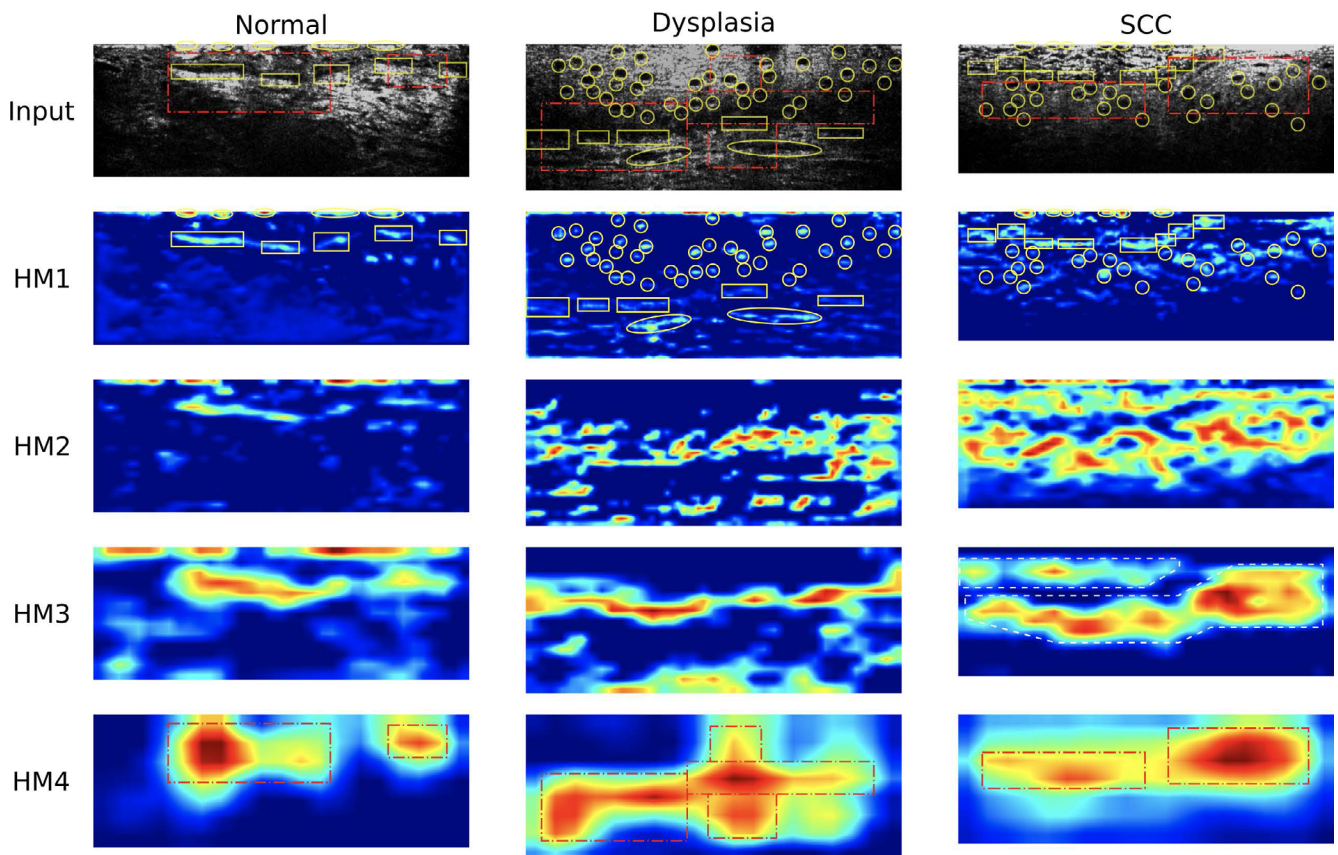


**FIGURE 7** The effect of image resolution on the SCC detection accuracy. Each vertical bar represents the standard deviation of validation accuracy. To make the chart easy to read, both down sampling rate and pixel size are shown. Note that the down-sampling rate is proportional to pixel size. The pixel size at the original image resolution is  $0.5 \mu\text{m}$

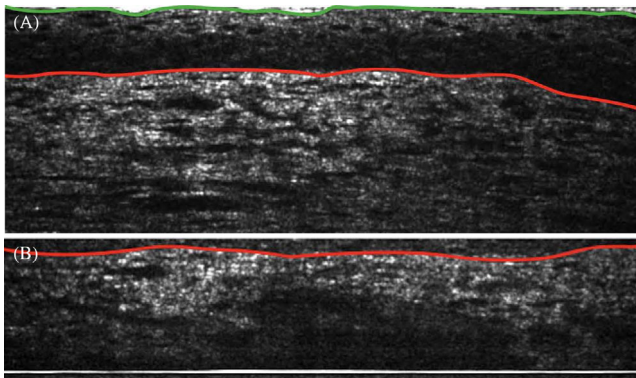
intensity regions in HM3 correspond to the SC and the epidermis (white polygons). From HM4, we can see that the main high intensity region is located at the epidermis (red rectangles), where a large amount of cells reside. The results are consistent with the fact that thick SC and large nuclei are typical features of SCC.

We also observe that the nuclei are detected in the HM1 of abnormal skin (dysplasia and SCC) but not in normal skin. This can be explained by noting that the size of the nuclei in normal skin is about four times smaller than the receptive field in the first stage, so the nuclei may firstly be captured by the first or second convolutional layer of the stage and then aggregated to localize the DEJ.

The heat maps of most images in the dataset have the following characteristic. In HM1, similar local features in images are separately extracted. In heat maps of deeper layers, these local features are progressively aggregated into high-level features. We also observe that most of the low-intensity regions, which correspond to areas with irrelevant information in the input images, do not contribute to the final SCC detection since they fail to aggregate into meaningful high-level features. This shows that



**FIGURE 8** An example input OCT image of each SCC diagnosis class and its associated heat maps at all four stages (HM1 to HM4) of Pruned-ResNet-18. The features extracted in the heat maps are marked, one single color for each stage. SCC, squamous cell carcinoma; OCT, optical coherence tomography



**FIGURE 9** Two example OCT images obtained by starting the scan above, A, and within, B, the epidermis. The green curve marks the junction between SC and epidermis, and the red curve marks the DEJ. DEJ, dermal-epidermal junction; OCT, optical coherence tomography

the network is able to select relevant information from the images for SCC detection.

## 5.2 | Possible extensions

The characteristics of images in our dataset vary. As shown in Figure 9, the depth of a tomogram in our dataset is different from one to another. For some samples, the data scanning starts from a position above the skin surface. Therefore, the superficial structures (SC, epidermis, DEJ, etc.) are visible in the image, see Figure 9A. However, for samples within the skin, the actual thickness of the epidermis, the SC or the surface roughness cannot be determined from the image data, see Figure 9B. Because a majority of the tomograms in our dataset contain the entire epidermal layer, our classifier may easily misidentify epidermis as SC and dermis as epidermis for images that only contain part of epidermis. When dermis is misidentified as epidermis, the classifier considers that the image has a very thick epidermis because the dermis normally occupies a large area of the image. Consequently, a normal image is classified as an abnormal one, suggesting that the data collection process (including the setting of scanning parameters) should be carefully done so as to prevent such misclassification. In other words, the detection accuracy can be further improved by enhancing the quality of the dataset.

We also believe that the CNN-based medical application considered in this work can be further extended if more information about the tomograms is available. For example, if the thickness of each layer of a tomogram is given in the dataset, we can train a segmentation model

by taking such information into consideration. Furthermore, the proposed system can be extended to estimate additional characteristics such as the attenuation coefficient, the mean intensity, etc., of each tomogram layer.

## 6 | CONCLUSIONS

Although it is rarely life-threatening, skin cancer accounts for at least 40% of cancer cases. As skin cancer develops, the morphology of skin layers provides an important clue for early diagnosis. In this paper, we have described a CNN-based classifier called Pruned-ResNet-18 that provides accurate and interpretable SCC detection for mouse skins. It achieves over 80% detection accuracy. This is made possible partly by the employment of an FF-OCT imaging system at sub-micron resolution and partly by the employment of a bottom-up feature extraction mechanism built into the classifier. How the feature extraction progresses, especially how the cellular-level features are captured, can be well explained through the heat maps at the four stages of the classifier, making our deep learning algorithm interpretable. The importance of an FF-OCT system at submicron resolution for SCC detection is further illustrated by a down-sampling analysis. The analysis shows that cellular-level features are critical to the success of SCC detection.

## DATA AVAILABILITY STATEMENT


The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

Chi-Jui Ho  <https://orcid.org/0000-0002-6034-8477>

Manuel Calderon-Delgado  <https://orcid.org/0000-0002-4864-1670>

Chin-Cheng Chan  <https://orcid.org/0000-0002-7220-4884>

Sheng-Lung Huang  <https://orcid.org/0000-0001-6244-1555>

Homer H. Chen  <https://orcid.org/0000-0002-8795-1911>

## REFERENCES

- [1] World Health Organization, "World Health Organization - Cancer," 2019. <https://www.who.int/cancer/en>. Accessed January 8, 2020.
- [2] World Health Organization, "Guide to early cancer diagnosis," [https://www.who.int/cancer/publications/cancer\\_early\\_diagnosis/en](https://www.who.int/cancer/publications/cancer_early_diagnosis/en). Accessed January 8, 2020.
- [3] World Cancer Research Fund, American Institute for Cancer Research, "Skin cancer report," <https://www.wcrf.org/dietandcancer/skin-cancer>. Accessed January 8, 2020.

- [4] J. F. Thompson, D. L. Morton, B. B. R. Kroon, *Textbook of Melanoma: Pathology, Diagnosis and Management*, 1st ed., London: CRC Press, **2003**.
- [5] J. G. Muzic, A. R. Schmitt, A. C. Wright, D. T. Alniemi, A. S. Zubair, J. M. Olazagasti, et al., *Mayo Clinic Proc.* **2017**, *92*(6), 890. <https://doi.org/10.1016/j.mayocp.2017.02.015>.
- [6] American Cancer Society, "Basal and Squamous Cell Skin Cancer Stages" <https://www.cancer.org/cancer/basal-and-squamous-cell-skin-cancer/detection-diagnosis-staging/staging.html> Accessed January 8, **2020**.
- [7] S. K. T. Que, F. O. Zwald, C. D. Schmults, *J. Am. Acad. Dermatol.* **2018**, *78*(2), 237. <https://doi.org/10.1016/j.jaad.2017.08.059>.
- [8] N. Reddy, B. T. Nguyen, *Br. J. Dermatol.* **2019**, *180*(3), 475. <https://doi.org/10.1111/bjd.17201>.
- [9] L. Ferrante di Ruffano, J. Dinnes, J. J. Deeks, N. Chuchu, S. E. Bayliss, C. Davenport, et al., *Cochrane Database Syst. Rev.* **2018**, *12*. <https://doi.org/10.1002/14651858.CD013189>.
- [10] L. van Manen, J. Dijkstra, C. Boccara, E. Benoit, A. L. Vahrmeijer, M. J. Gora, J. S. D. Mieog, *J. Cancer Res. Clin. Oncol.* **2018**, *144*(10), 1967. <https://doi.org/10.1007/s00432-018-2690-9>.
- [11] W. A. Wells, P. E. Barker, C. MacAulay, M. Novelli, R. M. Levenson, J. M. Crawford, *J. Biomed. Opt.* **2007**, *12*(5), 051801. <https://doi.org/10.1117/1.2795569>.
- [12] J. Kato, K. Horimoto, S. Sato, T. Minowa, H. Uhara, *Front. Med.* **2019**, *6*(180), 180. <https://doi.org/10.3389/fmed.2019.00180>.
- [13] A. Levine, O. Markowitz, *J. Am. Acad. Dermatol. Case Rep.* **2018**, *4*(10), 1014. <https://doi.org/10.1016/j.jdc.2018.09.019>.
- [14] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, et al., *Science* **1991**, *254*(5035), 1178. <https://doi.org/10.1126/science.1957169>.
- [15] A. Dubois, A. C. Boccara, Full-Field optical coherence tomography. in *Optical Coherence Tomography: Technology and Applications* (Eds: W. Drexler, J. G. Fujimoto), Berlin, Heidelberg: Springer Berlin Heidelberg, **2008**, p. 565. [https://doi.org/10.1007/978-3-540-77550-8\\_19](https://doi.org/10.1007/978-3-540-77550-8_19).
- [16] Y. Q. Xiong, Y. Mo, Y. Q. Wen, M. J. Cheng, S. T. Huo, X. J. Chen, Q. Chen, *J. Biomed. Opt.* **2018**, *23*(2), 1. <https://doi.org/10.1117/1.JBO.23.2.020902>.
- [17] J. Olsen, J. Holmes, G. B. Jemec, *J. Biomed. Opt.* **2018**, *23*(4), 1. <https://doi.org/10.1117/1.JBO.23.4.040901>.
- [18] K. X. Wang, A. Meekings, J. W. Fluhr, G. McKenzie, D. A. Lee, J. Fisher, et al., *Dermatol. Surg.* **2013**, *39*(4), 627. <https://doi.org/10.1111/dsu.12093>.
- [19] L. Themstrup, C. A. Banzhaf, M. Mogensen, G. B. Jemec, *Photodiagnosis Photodyn. Ther.* **2014**, *11*(1), 7. <https://doi.org/10.1016/j.pdpdt.2013.11.003>.
- [20] M. A. Boone, M. Suppa, A. Marneffe, M. Miyamoto, G. B. Jemec, V. Del Marmol, *J. Eur. Acad. Dermatol. Venereol.* **2016**, *30*(10), 1714. <https://doi.org/10.1111/jdv.13720>.
- [21] S. Batz, C. Wahrlich, A. Alawi, M. Ulrich, J. Lademann, *Skin Pharmacol. Physiol.* **2018**, *31*(5), 238. <https://doi.org/10.1159/000489269>.
- [22] T. M. Jorgensen, A. Tycho, M. Mogensen, P. Bjerring, G. B. Jemec, *Skin Res. Technol.* **2008**, *14*(3), 364. <https://doi.org/10.1111/j.1600-0846.2008.00304.x>.
- [23] L. Duan, T. Marvdashti, A. Lee, J. Y. Tang, A. K. Ellerbee, *Biomed. Opt. Express* **2014**, *5*(10), 3717. <https://doi.org/10.1364/BOE.5.003717>.
- [24] T. Marvdashti, L. Duan, S. Z. Aasi, J. Y. Tang, A. K. Ellerbee Bowden, *Biomed. Opt. Express* **2016**, *7*(9), 3721. <https://doi.org/10.1364/BOE.7.003721>.
- [25] L. Huang, X. He, L. Fang, H. Rabbani, X. Chen, *IEEE Signal Processing Letters* **2019**, *26*(7), 1026. <https://doi.org/10.1109/lsp.2019.2917779>.
- [26] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, et al., *Nat. Med.* **2018**, *24*(9), 1342. <https://doi.org/10.1038/s41591-018-0107-6>.
- [27] M. Calderon-Delgado, J. W. Tjiu, M. Y. Lin, S. L. Huang, *Int. Conf. Numerical Simulation Optoelectronic Devices* **2018**, 31. <https://doi.org/10.1109/NUSOD.2018.8570241>.
- [28] D. Mandache, E. Dalimier, J. R. Durkin, C. Boccara, J. C. Olivo-Marin, V. Meas-Yedid, *IEEE 15th Int. Symp. Biomed. Imaging* **2018**, 784. <https://doi.org/10.1109/ISBI.2018.8363689>.
- [29] C. Mignon, D. J. Tobin, M. Zeitouny, N. E. Uzunbajakava, *Biomed. Opt. Express* **2018**, *9*(2), 852. <https://doi.org/10.1364/BOE.9.000852>.
- [30] C. C. Tsai, C. K. Chang, K. Y. Hsu, T. S. Ho, M. Y. Lin, J. W. Tjiu, S. L. Huang, *Biomed. Opt. Express* **2014**, *5*(9), 3001. <https://doi.org/10.1364/BOE.5.003001>.
- [31] H. Hennings, A. B. Glick, D. T. Lowry, L. S. Krsmanovic, L. M. Sly, S. H. Yuspa, *Carcinogenesis* **1993**, *14*(11), 2353. <https://doi.org/10.1093/carcin/14.11.2353>.
- [32] J. A. McGrath, J. Uitto, Structure and function of the skin. in *Rook's Textbook of Dermatology*, 9th ed. (Eds: C. E. M. Griffiths, J. Barker, T. Bleiker, R. Chalmers, D. Creamer), Chichester, West Sussex: Wiley Blackwell, **2016**. <http://www.rooksdematology.com>.
- [33] A. Krizhevsky, I. Sutskever, G. E. Hinton, *Proc. 25th Int. Conf. Neural Info. Process. Sys.* Red Hook, NY:Curran Associates Inc; **2012**, p. 1097.
- [34] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, et al., *Int. J. Comput. Vision* **2015**, *115*(3), 211. <https://doi.org/10.1007/s11263-015-0816-y>.
- [35] Y. Bengio, P. Simard, P. Frasconi, *IEEE Trans. Neural Netw.* **1994**, *5*(2), 157. <https://doi.org/10.1109/72.279181>.
- [36] X. Glorot, Y. Bengio, *Proc. Thirteenth Int. Conf. Artificial Intelligence Statistic* **2010**, *9*, 249.
- [37] S. Ioffe, C. Szegedy, *Proc. 32nd Int. Conf. Int. Conf. Machine Learning* **2015**, *37*, 448. <https://doi.org/10.5555/3045118.3045167>.
- [38] K. He, X. Zhang, S. Ren, J. Sun, *IEEE Conf. Comput. Vis. Pattern Recognit.* **2016**, 770. <https://doi.org/10.1109/CVPR.2016.90>.
- [39] A. Dosovitskiy, P. Fischer, E. Ilg, P. Häusser, C. Hazirbas, V. Golkov, et al., *2015 IEEE Int. Conf. Comput. Vis* **2015**, 2758. <https://doi.org/10.1109/ICCV.2015.316>.
- [40] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, *IEEE Conf. Comput. Vis. Pattern Recognit.* **2016**, 2921. <https://doi.org/10.1109/CVPR.2016.319>.
- [41] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, *IEEE Int. Conf. Comput. Vis.* **2017**, 618. <https://doi.org/10.1109/ICCV.2017.74>.
- [42] M. D. Abramoff, P. J. Magalhães, S. J. Ram, *Biophotonics Int.* **2004**, *11*(7), 36.
- [43] J. Schindelin, I. Arganda-carreras, E. Frise, V. Kayni, M. Longair, T. Pietzsch, et al., *Nat. Methods* **2012**, *9*(7), 676.

- [44] L. Wei, E. Roberts, *Sci. Rep.* **2018**, 8(1), 7313. <https://doi.org/10.1038/s41598-018-25458-w>.
- [45] W. Luo, Y. Li, R. Urtasun, R. Zemel, *Proc. 30th Int. Conf. Neural Information Processing Sys.* Red Hook, NY: Curran Associates Inc; **2016**, 4905. <https://doi.org/10.5555/3157382.3157645>.
- [46] J. Lazzaro, S. Ryckebusch, M. A. Mahowald, C. A. Mead, Winner-take-all networks of  $O(N)$  complexity. in *Advances in Neural Information Processing Systems 1*, Morgan-Kaufmann, San Francisco, CA, USA **1989**, p. 703.
- [47] S. V. Stehman, *Remote Sens. Environ.* **1997**, 62(1), 77. [https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7).
- [48] S. Arlot, A. Celisse, *Statist. surv.* **2010**, 4, 40. <https://doi.org/10.1214/09-SS054>.
- [49] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, et al., *NIPS Autodif Workshop* **2017**, 8024-8035.
- [50] D. Kingma, J. Ba, *Proc. 3rd Int. Conf. Learning Representations* **2015**, 1.

**How to cite this article:** Ho C-J, Calderon-Delgado M, Chan C-C, et al. Detecting mouse squamous cell carcinoma from submicron full-field optical coherence tomography images by deep learning. *J. Biophotonics*. 2020;e202000271. <https://doi.org/10.1002/jbio.202000271>