

## Full length article

Deep learning enhances polarization speckle for *in vivo* skin cancer detection

Yuheng Wang <sup>a,b,c,d</sup>, Daniel C. Louie <sup>a,b,c,d</sup>, Jiayue Cai <sup>e,f,\*</sup>, Lioudmila Tchvialeva <sup>b,c</sup>, Harvey Lui <sup>b,c,d</sup>, Z. Jane Wang <sup>e</sup>, Tim K. Lee <sup>a,b,c,d</sup>

<sup>a</sup> School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada

<sup>b</sup> Department of Dermatology and Skin Science, University of British Columbia, Vancouver, BC, Canada

<sup>c</sup> Photomedicine Institute, Vancouver Coast Health Research Institute, Vancouver, BC, Canada

<sup>d</sup> Departments of Cancer Control Research and Integrative Oncology, BC Cancer, Vancouver, BC, Canada

<sup>e</sup> Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada

<sup>f</sup> School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China

## ARTICLE INFO

## Keywords:

Deep learning  
Polarization speckle  
Skin cancer detection  
Machine learning

## ABSTRACT

Polarization speckle is a growing research field, especially for the purposes of biomedical imaging and diagnostics. The statistical uniformity of speckle creates novel opportunities for AI methods to analyze speckle patterns, as they are difficult for a human to interpret. We employed deep learning and traditional machine learning methods (Support Vector Machine, K Nearest Neighbor, and Random Forests) to analyze polarization speckle images generated from the general categories of malignant and benign lesions, a classification task previously found to be very difficult. Their performance in classifying the patterns were compared both to each other and to the previous statistical methods of speckle analysis. A collection of 122 malignant and 196 benign skin lesion speckle images were augmented for deep learning using patch cropping, an advantageous method given the patterns' statistical homogeneity. The machine learning technique performed with a diagnostic accuracy of less than 65%, a result comparable to those of previous statistical methods, despite the fact that the same machine learning technique could achieve a high 90% accuracy in the simpler classification task of differentiating malignant melanoma and the benign lookalike seborrheic keratosis. However, ResNet, our chosen deep learning architecture, achieved the best performance of 82% diagnostic accuracy in the general classification task of malignant and benign. These results show that deep learning extracts previously hidden information from polarization speckle, which further develops towards a rapid, real-time, analysis technique for an automatic skin cancer detection system.

## 1. Introduction

Polarization speckle is an information-rich optical interference pattern that arises from scattered coherent or partially-coherent light [1,2]. Polarization describes the oscillation orientation of the electric field of a light wave, which can range between linear, elliptical, and circular oscillations. When coherent light is backscattered from a rough opaque surface, an interference pattern with random amplitude and phase but almost uniform polarization is formed. This is classical intensity speckle which has been applied to surface roughness quantification [3,4] and extended to medical applications [5–9]. However, when coherent light is backscattered from a turbid medium such as skin

tissue, its phase and polarization undergo a random modification. The resultant speckle field could possess random polarization properties [1]. This spatial distribution of polarization states is termed polarization speckle to differentiate from intensity speckle. Compared to imaging modalities such as reflectance microscopy, which rely on absorption-related phenomena to generate contrast, polarization speckle encodes both light wave amplitude and phase transformation. As light travels in a turbid medium such as skin, it is sensitive to scattering and absorption events, resulting in the formation of speckle patterns with statistical features informed by the scatterers within the tissue [10]. In a previous study, we demonstrated that polarization speckle generation and detection for *in vivo* skin analysis are achievable with relatively simple

\* Corresponding author.

E-mail address: [jiayuec@ece.ubc.ca](mailto:jiayuec@ece.ubc.ca) (J. Cai).

equipment and at rapid measurement speeds [2]. This strength makes polarization speckle suitable for skin cancer detection applications *in vivo*, and especially those to be integrated into busy medical clinics.

Skin cancer is the most common malignancy in the western world among fair-skinned populations. In the USA, there are over 5 million new cases of skin cancer diagnosed annually, exceeding all other cancers combined [11]. The most common skin cancers are basal cell carcinoma, squamous cell carcinoma and malignant melanoma. The latter is the most deadly form of skin cancer. In this year, it is estimated that there are over 100,000 new cases of melanoma in the USA [11]. When melanoma is diagnosed and treated at an early stage, the disease has a 5 year relative survival rate of 99%; however, the rate drops to 25% when the disease is only diagnosed at a late stage [11]. Therefore, early diagnosis is important for the prognosis of the disease. Unfortunately, melanoma is often confused with benign lesions such as seborrheic keratosis and nevus, even for experienced dermatologists. An *in vivo* detection tool for malignant and benign lesions would help physicians, especially non-specialist general practitioners.

A significant challenge of applying a polarization speckle technique to skin cancer detection is the limited ability to interpret the patterns. From an image analysis perspective, speckle patterns have unique composition. When captured, these patterns are stochastic and do not form an image with features easily recognizable to humans. Due to the stochastic phenomena behind their generation, free space speckle patterns are understood to lack local image features. The relevant information the patterns contain are global to the image, encoded in the statistical properties of speckle. Current methods of extracting information from speckle have been limited to the calculation of statistical moments. In our previous work on detecting skin cancer, we demonstrated that the statistical moment of a polarization speckle field could be used to differentiate malignant melanoma, a smooth lesion, from seborrheic keratosis, a rough lookalike benign skin lesion. However, the statistical moment approach was not sensitive enough to discriminate between larger categories of cancerous and benign skin lesions.

Recently, deep learning strategies based on convolutional neural network (CNN) architectures have been the technique of choice for image classification, especially in skin lesion color image processing, as it has demonstrated an excellent ability to classify the hidden features of complex images by learning from a set of examples without actually extracting features. Esteva et al. [12] used a pre-trained Inception-V3 model to train more than 120,000 color images including both dermoscopic and clinical images and achieved a dermatologist-level diagnostic. Gessert et al. [13] introduced EfficientNet B0-B6 incorporating meta information to the ISIC 2019 and achieved the highest accuracy of 63.6% for eight different classes' skin diseases detection. Akram et al. [14] came up with an improved deep learning structure using Entropy-Controlled Neighborhood Component Analysis (ECNCA) and layers from DenseNet 201, Inception-ResNet-V2, and Inception-V3, and achieved over 95% accuracy in differentiating melanoma with other diseases. Combining general color clinical images, dermoscopic images, and patient metadata, Kawahara et al. [15] achieved 73.7% accuracy in detecting melanoma on the EDRA dataset using matched multi-modal neural networks. Hekler et al. [16] introduced CNN methods in the modality of histopathological images for melanoma detection and the performance of 59.2% accuracy outperformed pathologists. Despite encouraging development in applying deep learning to dermoscopic and color clinical images, the technique faces a challenge of differentiating melanoma and seborrheic keratosis. These two types of skin lesions often show similar visual appearance. A recent study analyzing clinical morphology of skin lesions suggested that nearly 20% of melanomas could be misdiagnosed as seborrheic dermatosis [17]. The result was in line with Matsunaga et al.'s study, which applied ResNet to classify melanoma and seborrheic keratosis and achieved an accuracy of 83.1% by augmenting the 2017 ISIC Challenge data set with a private dataset [18].

While the above mentioned studies on clinical and dermoscopic

images have seen some success, a dimension missing from these images is the surface profile of the lesion. Dermatologists often take surface roughness into account in their diagnoses, and this information is only conveyed in photographs through uncertain qualities of dullness or glossiness. In [19] we demonstrate that speckle patterns encode surface scattering and could be used for more precise skin roughness measurement. In this paper we attempt to enhance the interpretation of speckle, a new modality that could potentially be combined with others to help improve skin cancer detection rates. While a variety of strategies exist for dermoscopic and clinical images, it is yet unknown how to use them to analyze speckle. Thus, in this paper, we focus on speckle analysis techniques.

We find that speckle lends itself well to deep learning interpretation. It is well known that deep learning methods require large amounts of images to train a network, and as such, it can be assumed that applying it to speckle would be challenging. We have only a limited number of speckle images, given that it is a new field. However, we take advantage of the statistical uniformity of speckle patterns in order to solve this challenge. In the same way that the statistics of a sufficiently large sample can represent a population, a sufficiently large cropped area of a speckle pattern can represent an entire speckle pattern. Since a cropped section of a single speckle image contains as much statistical information as the entire image, the image quality is not degraded by patch-cropping, flipping, and rotational data augmentation techniques common in deep learning methods. However, it should be noted that speckle images should not be scaled as that would distort their information. The practical advantage of this is that deep learning studies can be performed on speckle images with a much smaller collected image set than in other studies. In terms of interpretability, machine learning and deep learning methods offer the means to interpret speckle in ways that were previously undiscovered.

We hypothesize that deep learning could enhance the polarization speckle techniques for the general classification of cancerous skin lesions from benign skin lesions. In this paper, we will apply a deep learning method based on ResNet [20,21] to analyze polarization speckle patterns for skin cancer detection. In addition, we will compare the deep learning results with traditional machine learning methods, Support Vector Machine (SVM) [22], K Nearest Neighbor (KNN) [23], Random Forests (RF) [24]), and our previous statistical moment method [2].

The organization of the paper is as follows: Section 2 introduces our custom-built polarization speckle prototype along with our previous clinical study. The materials and methods are presented in Section 3, which describes pre-processing, the deep learning and machine learning classification methods for polarization speckle, along with the data set used in the study. Section 4 reports the performance of our deep learning- and machine learning-based skin cancer detection systems. Section 5 further analyzes the reported methods in order to demonstrate that the deep learning approach is a suitable method for classifying speckle images into cancer vs. benign. A discussion and the conclusion are in Section 6 and 7, respectively.

## 2. Polarization speckle prototype and previous clinical study

Our group designed and constructed a custom-built polarization speckle prototype for the detection of skin cancer. (See Fig. 1 for the schematic of the prototype.) The details of the prototype have been published elsewhere [2–4] and only a brief description is presented here for completeness. The device consists of two independent laser light sources, blue and red fiber-coupled diode lasers ( $\lambda = 405$  nm and 663 nm, 22 mW), which are operated independently and sequentially. In this scheme, a laser is pulsed, passing through a collimator and linear polarizer as it illuminates a skin lesion. The backscattered light, in the form of a polarization speckle field with random polarizations, amplitudes and phases, is split with a non-polarizing beam splitter and directed to two CCD cameras (Matrix Vision GmbH, mvBlueFOX-

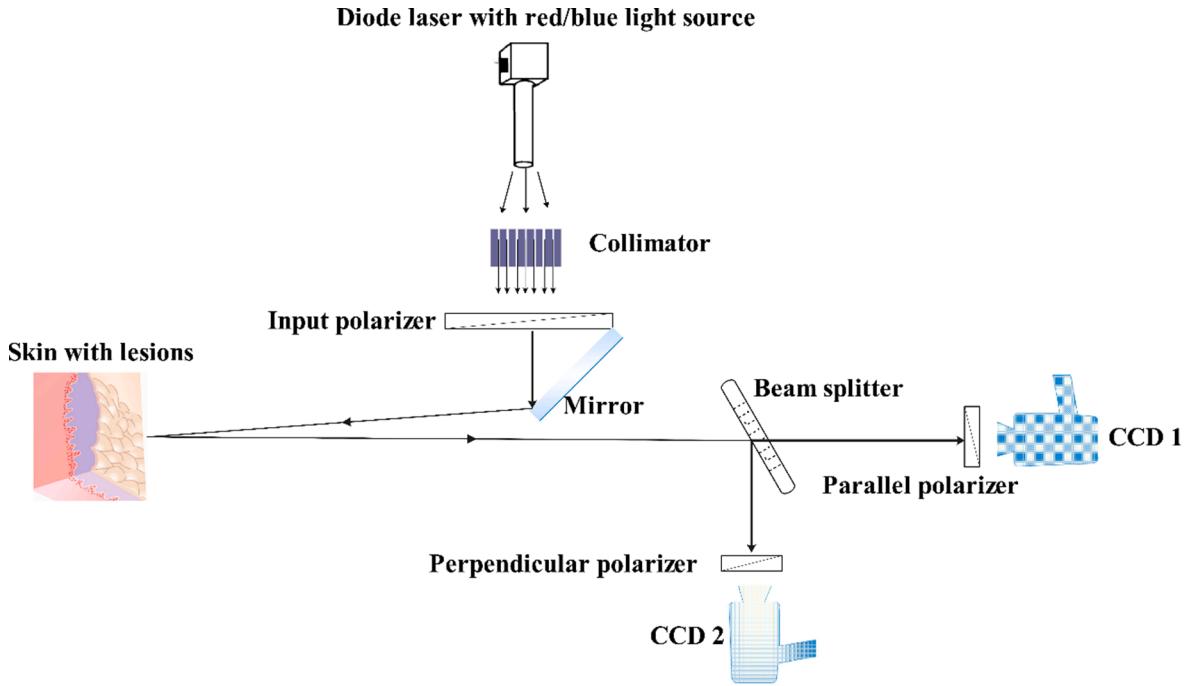


Fig. 1. An overview of the polarization speckle imaging system.

M124G) set at an equal distance. In front of one of the cameras is a polarizer orientated parallel to that of the incident light. This camera was used to measure the reduced linearly polarized light after scattering, capturing a co-polarized speckle image. The second camera was equipped with a polarizer oriented perpendicular to the incident light,

capturing a cross-polarized speckle image. The integration time for each camera/polarizer system is less than 5 ms. In addition to speckle acquisition, images with no illumination were taken to account for black current and ambient light. The prototype was designed such that the speckles are uniformly distributed over the camera's detection field,

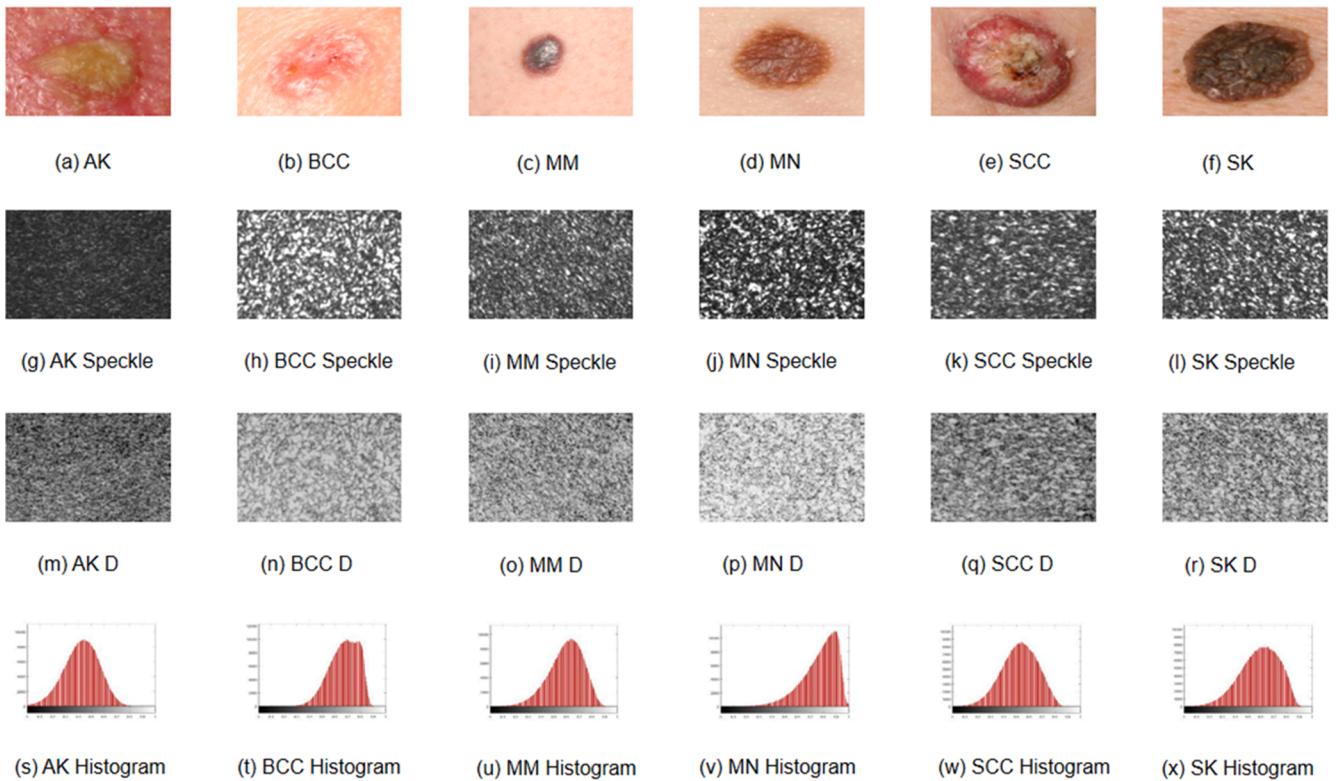


Fig. 2. Six types of skin diseases with their matched raw speckle images, depolarization ratio  $D$ , and distribution of  $D$ . (a)-(f) Images of cancerous, pre-cancerous, and benign skin diseases included in this study. AK: actinic keratosis, BCC: basal cell carcinoma, MM: malignant melanoma, MN: melanocytic nevus, SCC: squamous cell carcinoma, SK: seborrheic keratosis (g)-(l) Representative speckle images for corresponding diseases. (m)-(r) Depolarization ratio  $D$  calculated according to Eq. (1). (s)-(x) Corresponding distribution of  $D$ . The histograms demonstrate that the distribution shape of depolarization ratio is different for the lesion types.

collected as  $1024 \times 768$  pixel images. Corresponding pixels from the two patterns can be used to determine the depolarization ratio  $D$  of the backscattered light as:

$$D_{(x,y)} = \frac{I_{\parallel(x,y)} - I_{\perp(x,y)}}{I_{\parallel(x,y)} + I_{\perp(x,y)}} \quad (1)$$

where  $I_{\parallel}$  and  $I_{\perp}$  are the corresponding intensity values at the pixel  $(x,y)$  of the co-polarized and cross-polarized speckle images.

Our clinical study included malignant melanoma (MM), basal cell carcinoma (BCC), squamous cell carcinoma (SCC), melanocytic nevus (MN), and seborrheic keratosis (SK). Melanoma, BCC, and SCC are common types of skin cancer, while MN and SK are benign lesions that are often misdiagnosed as melanoma. Speckle patterns of these lesions were acquired by our prototype, from which the depolarization ratio  $D(x,y)$  was derived using Eq. (1) [2]. Fig. 2 shows representative examples of studied skin diseases and their corresponding speckle patterns, along with their depolarization ratio patterns and histograms. As illustrated by the histograms, each speckle pattern exhibits different distribution shapes in their depolarization ratio patterns, and suggest that statistical moments of these patterns could be used to differentiate the lesion types. Our previous study results show that the fourth-order statistical moment is an excellent feature to classify melanoma and seborrheic keratosis with a ROC accuracy (AUC) of 86.5% using the red light source. However, when the same technique is used in a more challenging problem of discriminating skin cancer (MM, BCC, and SCC) against benign lookalike skin lesions (MN and SK), it only achieved a disappointing ROC accuracy of 66.4%. In this study, we will use recent analytical techniques to analyze polarization speckle patterns to improve the accuracy of the cancer vs. benign classification task.

### 3. Materials and methodology

#### 3.1. Clinical data acquisition and preprocessing

In this study, we collected speckle patterns from six types of skin lesions that comprised all five lesion types mentioned in Section 2, along with actinic keratosis (AK), a pre-cancerous lesion. A total of 318 lesions, 122 malignant and 196 benign skin lesions, were acquired from volunteer patients of the Skin Care Centre of Vancouver General Hospital. After the lesions were imaged, the patients were directed back to their physicians and continued with the standard care of their skin conditions. Cancerous and pre-cancerous skin lesions in this data set included 37 malignant melanoma (MM), 30 squamous cell carcinoma (SCC), 11 actinic keratosis (AK) and 44 basal cell carcinoma (BCC.). Ninety-one melanocytic nevus (MN) and 105 seborrheic keratosis (SK) were categorized as benign lesions. All cancerous cases were biopsied, but benign cases might not have been biopsied after being diagnosed clinically by expert dermatologists. If a biopsy was taken of a lesion and it proved to be benign, it was categorized under its correct histologically-determined label. During image collection, lesions were measured in three different locations within the lesion, or five locations in the case of the less common melanoma and nevus. Following collection, 15 poor quality images due to factors such as over exposure or blur were excluded. A total of 1,087 valid speckle images were finally

included in the analysis, and their breakdown is listed in Table 1. The clinical study was approved by the UBC Clinical Research Ethical Board (study number H06-70281).

The black current was subtracted pixel-by-pixel from their respective parallel and perpendicular speckle images, in order to eliminate the effects of black current noise and/or ambient light according to Eq. (2).

$$\begin{cases} I_{\parallel(x,y)} = I_{\parallel Ori(x,y)} - I_{\parallel black(x,y)} \\ I_{\perp(x,y)} = I_{\perp Ori(x,y)} - I_{\perp black(x,y)} \end{cases} \quad (2)$$

where  $I_{\parallel Ori}$  and  $I_{\perp Ori}$  are the corresponding intensity values at the pixel  $(x,y)$  of the original co-polarized and cross-polarized speckle images.  $I_{\parallel black}$  and  $I_{\perp black}$  are their respective black current noise value.

After the black image removal, the depolarization ratio  $D$ , is calculated based on Eq. (1). Depolarization ratio  $D$  indicates the fraction of depolarization registered at each pixel location.

Next, to enhance the contrast of the speckle images and emphasize their polarization features, gamma adjustment was applied [25]. As one of the histogram modification (HM) techniques, gamma correction has a varying adaptive parameter  $\gamma$ . In our study, we set  $\gamma = 1$  so that HM saturates lowest and highest 1% of intensities in the image and remaps all the internal values linearly to maximize contrast. This step could be formulated as shown in Eq. (3).

$$D_{out(x,y)} = \left( \frac{D_{in(x,y)} - min(D_{in})}{max(D_{in}) - min(D_{in})} \right)^{\gamma} \quad (3)$$

where  $D_{in}$  and  $D_{out}$  represent the input and the output of the gamma correction function. Min () and Max () are the highest and lowest intensities in the image. Fig. 3 shows a representative example of depolarization ratio  $D$  before and after the gamma correction.

#### 3.2. Patch strategy and augmentation

Each original  $D$  image was cropped into nine  $341 \times 256$  pixel non-overlapping patches. We then applied conventional data augmentation techniques including right-left flipping, top-bottom flipping and random rotation [26]. All  $D$  patches were further cropped to  $224 \times 224$  which was the standard input size of many deep neural architectures [27]. After these data augmentation steps, 100,000 patches were created. Here, we kept the same number of augmented patches for malignant and benign skin lesions, to take care of the skin cancer diseases that have fewer images to prevent a data unbalancing problem.

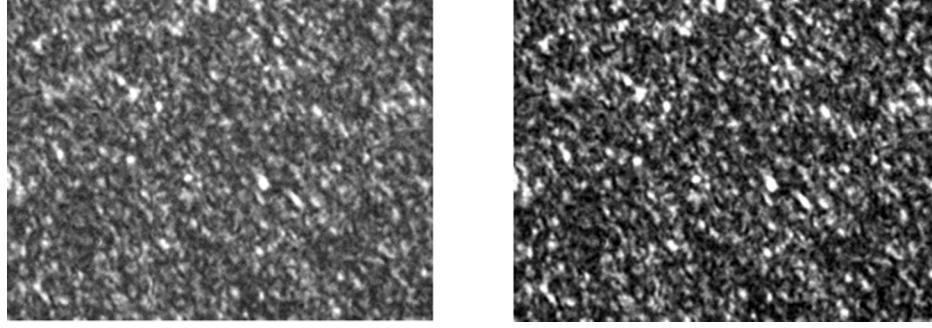
This polarization speckle imaging system was designed such that the speckles were uniformly distributed over a  $1024 \times 768$  pixel image. As mentioned in the introduction, speckle images feature global statistical properties. The speckle image is a sampled portion of a greater speckle interference field, and as such, its statistical properties can vary depending on the sampling size. For example, when calculating the mean intensity of a speckle pattern, we can expect an error on the order of  $\sqrt{N}^{-1}$  from the true statistical value of the speckle field, where  $N$  is the number of speckles within the sampled area. In cropping and patching these speckle patterns, we take this statistical variation into consideration. Although it is yet unclear if speckle images as inputs to a neural network adhere to the same error properties, we would ideally have a mean intensity error of 10% or less, as we have followed in previous work [28].

To determine the average speckle width of our images, we performed an autocorrelation of each column and row of our speckle images. The autocorrelation calculation was done through a shift and multiplication algorithm which operates as described in Eq. (4). This is a non-normalized discrete autocorrelation where  $c(m)$  is the resultant autocorrelation vector, and  $R_{II}(m)$  is the autocorrelation operation with the pixel intensity input vector  $I$ , with  $M$  elements. We estimated the speckle width as the half-width, defined as the distance from peak to base, for each autocorrelated column or row. An averaging of all these values

**Table 1**

The Number of Images Collected From the Skin Care Centre of Vancouver General Hospital for the Analysis.

Malignant Class		Benign Class	
Type	Number of images	Type	Number of images
MM	149	Nevus	285
BCC	168	SK	354
SCC	98	Total	639
AK	33		
Total	448		



**Fig. 3.** Comparison between polarization speckle ratio D before (left) and after (right) gamma correction.

gives us an approximate speckle width of 27 pixels across both red and blue light. As the standard input size of our deep neural architectures is a  $224 \times 224$  pixel image, we can approximate that each speckle image patch contains approximately 69 speckles, which would result in a mean intensity error of 12%, slightly higher than the ideal situation, but the margin was considered acceptable.

$$c(m) = \hat{R}_H(M - m), m = 1, 2, \dots, M,$$

$$\hat{R}_H(m) = \sum_{j=0}^{M-m-1} I_{j+m} I_j \quad (4)$$

### 3.3. Deep residual neural network

In this study, we aim to solve a binary classification problem, i.e., classifying skin lesion images into cancer and benign classes.

We employed a deep residual neural network with 101 layers (ResNet-101) as the backbone of our cancer detection architecture, which is based on plain CNN networks with deeper hierarchical architecture and residual blocks. Over recent years, ResNet displayed robust and impressive performance in the medical imaging area due to its high effectiveness. Differently from traditional deep convolutional neural networks, ResNet utilizes residual blocks to train a model with deep architecture while solving the degradation problem. Therefore, more layers are added in ResNet to make full use of input data and enhance the model's performance without an overfitting problem. This specific structure of ResNet can speed up the convergence of deep layers and keep a high precision by largely increasing the depth of the network [20]. In general, a deep residual network should be composed of multiple residual blocks where each contains several stacked convolutional layers. Each convolutional layer consists of rectified linear unit (ReLU) layers and batch normalization (BN) layers. The mechanism of residual block is formulated as:

$$h_{(l+1)} = \text{ReLU}(h_l + w_l) \quad (5)$$

where  $h_l$  and  $h_{(l+1)}$  represent the input and the output of the  $l^{\text{th}}$  residual block respectively. ReLU is the rectified linear unit function. However, the input of the residual block  $h_l$  and the residual mapping function may not be in the same dimension. Therefore, the residual block equation can be reformulated as:

$$h_{(l+1)} = \text{ReLU}(A_l h_l + f(h_l, w_l)) \quad (6)$$

where  $f(*)$  is the residual mapping function while  $w_l$  is the parameters of the residual block. When  $f(h_l, w_l)$  and  $h_l$  are not in the same dimension, a linear mapping  $A_l$  is leveraged to make their dimensions uniform.

The key component of the residual block is the convolutional layer, which is composed of a series of neurons. Each neuron in the convolutional layer contains a corresponding learnable weight and bias. These weights are step-wise reeducated and tuned during the training phase. These convolutional layers can obtain feature maps from the previous layers as their input to keep their spatial information. Here, we suppose

$y_j^i$  is the  $j^{\text{th}}$  feature map in the  $i^{\text{th}}$  layer and  $y_m^{i-1}$  is the  $m^{\text{th}}$  feature map in the  $(i-1)^{\text{th}}$  layer, then  $y_j^i$  can be formulated as:

$$y_j^i = \sigma \left( \sum_{m=1}^M W_{jm}^i * y_m^{i-1} + b_j^i \right) \quad (7)$$

where  $W_{jm}^i$  is the weight of the  $m^{\text{th}}$  feature map in the previous layer,  $b_j^i$  is the  $j^{\text{th}}$  bias in the  $i^{\text{th}}$  layer, and  $\sigma(*)$  is the non-linear activation function of the convolutional layer. Here, we use  $\text{ReLU}(x) = \max(x, 0)$  as the activation function in our architecture.

The detailed specification of Resnet-101 architecture we used in this study is shown in Table 2. The convolutional units contributed to a total of 101 layers. In this architecture, the residual block is computed and optimized using  $1 \times 1 + 3 \times 3 + 1 \times 1$  instead of traditional  $3 \times 3$  convolutional layers. The middle layer of these convolutional layers is transformed from  $3 \times 3$  to  $1 \times 1$  and then restored by another  $1 \times 1$  layer to keep its accuracy and reduce the complexity of computation. The whole structure is linked by a function that can be mathematically represented as:

$$h_L = h_l + \sum_{i=1}^{L-1} f(h_i, w_i) \quad (8)$$

where  $h_L$  is the input of the  $L^{\text{th}}$  residual unit, and it is calculated as the sum of the inputs of a shallow residual unit and all the middle complex mappings.

**Table 2**  
The Deep Learning Architecture Used In Our Study.

Layer Name	No. of Residual Blocks	No. of Unites
Convolutional Layer1		$[7 \times 7, \text{stride } 2, \text{channel } 64] \times 1$ $[3 \times 3, \text{max pool}, \text{stride } 2] \times 1$
Convolutional Layer2	ResBlock1-3	$[1 \times 1, \text{channel } 64] \times 3$ $[3 \times 3, \text{channel } 64] \times 3$ $[1 \times 1, \text{channel } 256] \times 3$
Convolutional Layer3	ResBlock4-7	$[1 \times 1, \text{channel } 128] \times 4$ $[3 \times 3, \text{channel } 128] \times 4$ $[1 \times 1, \text{channel } 512] \times 4$
Convolutional Layer4	ResBlock8-30	$[1 \times 1, \text{channel } 256] \times 23$ $[3 \times 3, \text{channel } 256] \times 23$ $[1 \times 1, \text{channel } 1024] \times 23$
Convolutional Layer5	ResBlock31-33	$[1 \times 1, \text{channel } 512] \times 3$ $[3 \times 3, \text{channel } 512] \times 3$ $[1 \times 1, \text{channel } 2048] \times 3$
Output		Average pool, 2d-fc, Soft-back, Equally voting

### 3.4. Conventional machine learning methods

#### 3.4.1. Support vector machine

As one of the most popular supervised classifiers, SVM is generally deployed to separate different classes by finding the optimum hyperplane between them. We can consider that  $x_i (i = 1, \dots, M)$  belongs to either class A or class B, and  $y_i$  is its matched label.  $y_i = 1$  means class A and  $y_i = -1$  means class B. The hyperplane can be described as:

$$H_x = (w \cdot x) + b \quad (9)$$

where  $w$  is an  $m$ -dimensional vector,  $b$  is a scalar. The hyperplane  $H_x = 0$  with the maximum margin is called the optimal separating hyperplane.

Although the original SVM is a linear classifier, it can be applied for non-linear problems by using kernel functions which nonlinearly maps the feature vector to a new space. In our study, we use the SVM with a radial basis function (RBF) kernel function [29] to perform the cancer vs benign classification.

#### 3.4.2. K nearest neighbors

KNN is one of the simplest and most popular classifiers in medical image processing. The new input value is classified by the majority voting mechanism based on its neighbors. The new input value will be assigned to the class that is most common amongst its  $K$  nearest by a distance metric. In this study, we used the Euclidean distance [30] as follows:

$$d_{st}^2 = (x_s - y_t) \cdot (x_s - y_t)' \quad (10)$$

where  $x_s$  and  $y_t$  are the vectors of two input cases.

The advantage of KNN is that it is extremely simple and does not need an explicit training step. Thus, KNN is effective for applications with a large training set. However, since the distance computation to all samples from the training set is required, its computation is heavy and a lot of processing power is needed.

#### 3.4.3. Random forests

RF is a popular classification technique, which consists of an ensemble of random decision trees. A prediction made through a random forest method is the average of predictions over its constituent random trees. Over the training phase, a random forest model is constructed by training random trees on random subsets of data. All the subsets are trained as individual decision trees in a similar fashion. It is considered well suited for classification tasks with numerous features.

### 3.5. Deep neural network training

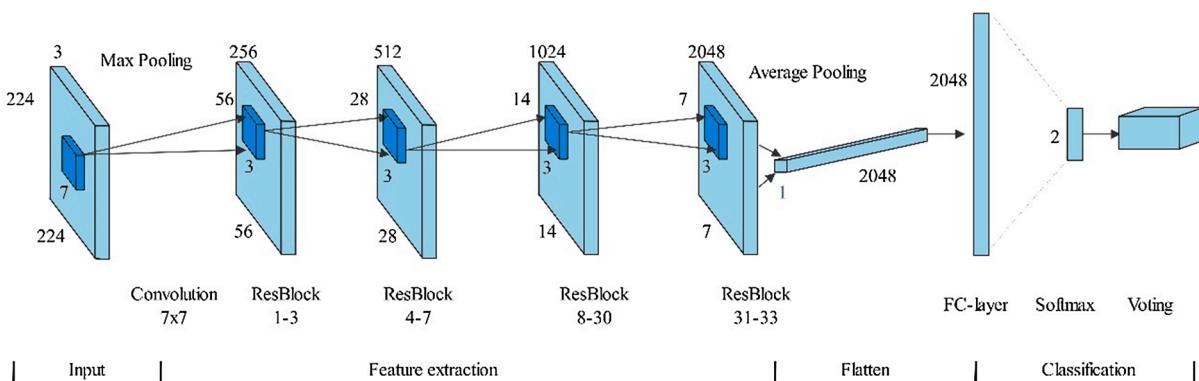
In this study, we added an equally-weighted voting layer after the softmax layer of ResNet in order to ensure the final classification was

based on the entire speckle image. Specifically, a speckle image was divided into patches and each patch went through the deep network during the validation. Then, the decision of the lesion is based on the majority classification status of all the corresponding patches. In other words, the results shown in the next section regard a speckle image as a unit. The whole structure we used in this study is shown in Fig. 4. The well-labelled 100,000 patches were split into 5 folders for further cross-validation steps and wrapped as the input of a ResNet architecture pre-trained on the ImageNet database [31]. Here, all the patches from an input speckle image was kept in either the training or the validation datasets. All validation data were unseen and not used in the training. Then we fine-tuned the parameters of the model using depolarization ratio  $D$ , and the learning rates were step-wise reeducated to find the proper checkpoint for next-step validation. We used Adam [32] as the optimizer to minimize the categorical cross entropy, which computes the dissimilarity of the approximate output distribution over the training steps. For the set-up parameters, we fixed the learning rates to be 0.01, 0.001 and 0.0001 with different steps, the batch size to be 16 and the weight decay to be 0.00004. For each experiment in the cross validation, there were around 100,000 steps in total. The momentum factor was set to 0.9 to avoid loss of function to trap local minima and move global minima. It is worthwhile to note that, unlike general transfer learning based implementation works, we not only trained the last two layers, but fine-tuned all the layers by using back propagation after we fixed the neuron number to suit out specific binary classification tasks. Additionally, we used fixed learning rates instead of placing a certain decay type, which helped us further fine-tune to explore the optimal checkpoints based on the real-time adjustable learning rates.

The deep residual network is implemented using TensorFlow [33] and computed on a Linux Ubuntu computer with an Intel Core i7-@2.6 GHz and two GTX 1080Ti card.

### 3.6. Conventional machine learning features

To further explore the utility of machine learning strategies, we compared the deep learning model's performance with conventional machine learning methods. Unlike deep learning, which is able to recognize hierarchical feature representations in the data, conventional machine learning methods usually require an additional feature extraction step. In our experiment, we used a principal component analysis (PCA) [34] based feature extraction prior to proceeding to the classification step, where the top 10 principal components were extracted and fed into the classifier. PCA was executed on the entire data set before the 5-fold cross-validation structure was set up. The main purpose of this strategy is to avoid data pollution during the training steps and therefore make the results trustable.



**Fig. 4.** The deep learning architecture based on ResNet with 101 layers. Each block within the feature extraction stage denotes a feature map. An equally-weighted voting layer for patch predictions is appended after the softmax layer.

## 4. Results

In this section, we evaluate the performance of the proposed deep learning enabled system and three machine learning methods for classifying skin lesions into malignant and benign classes in a 5-fold cross-validation. Table 3 shows the average performance of depolarization ratio  $D$  with the red laser source using four different classifiers. The ResNet-based deep learning method achieved the highest average accuracy of 82%, suggesting its superiority over traditional machine learning methods for skin cancer classification. In addition, the conventional machine learning methods exhibited inferior performance in sensitivity (SVM: 61%, KNN: 49%, RF: 45%) and specificity (SVM: 58%, KNN: 71% and RF: 74%) compared to the deep learning method (ResNet sensitivity: 78% and ResNet specificity: 85%), indicating the latter approach would have fewer false-negatives and false-positives than the traditional machine learning methods. Similarly, the standard deviation in the 5-fold cross validation indicates that deep learning method is most stable among the other three classification methods.

## 5. Sensitivity analyses

In this section, we will report two additional analyses, substantiating that the deep learning approach and the depolarization ratio are suitable for classifying malignant and benign skin lesions. The first analysis highlights the advantage of depolarization ratio  $D$  by comparing its performance with the raw co-polarized speckle image. Here, we further compared its performance with conventional machine learning methods and our previous statistical methods. The second analysis will demonstrate that the PCA-based features are an excellent choice for speckle image analyses with SVM, KNN and RF. In particular, the features achieve excellent results for differentiating MM and SK, although this approach obtained inferior results to the deep learning approach in classifying malignant vs. benign.

### 5.1. Comparison between depolarization ratio and raw speckle images

To further identify the advantages of depolarization ratio ( $D$ ) in capturing clinical information from polarization speckle patterns, we repeat the same deep learning experiment using the red light co-polarized raw speckle images  $I_{\parallel}$  after black image correction as input (Results are shown in Table 4). The red speckle images are used because they encode both surface and volume scattering, as red light could penetrate more deeply into the skin than blue. With the raw speckle images, the accuracy of differentiating malignant and benign skin lesions was decreased from 82% to 78% in our deep learning enabled structure. It also should be noted that changing the input from depolarization ratio to raw speckle images affected sensitivity more than other evaluation matrices, which can be regarded as more crucial than specificity for clinical skin cancer detection.

### 5.2. PCA-based machine learning could capture skin diseases' information

In our previous study [2], we have demonstrated that depolarization

**Table 3**

Comparisons of the performance of classifying malignant skin lesions and benign skin lesions (mean value and standard deviation) on depolarization ratio  $D$  images with red laser light source among ResNet, SVM, KNN, and RF. The bold font indicates the best performance.

Methods	Accuracy(std)	Sensitivity(std)	Specificity(std)
ResNet	<b>0.82(0.01)</b>	<b>0.78(0.03)</b>	<b>0.85(0.01)</b>
SVM	0.65(0.03)	0.61(0.03)	0.68(0.02)
KNN	0.62(0.02)	0.49(0.04)	0.71(0.06)
RF	0.61(0.02)	0.46(0.05)	0.74(0.08)

**Table 4**

Comparisons of the performance of classifying malignant skin lesions and benign skin lesions (mean value and standard deviation) on co-polarized raw speckle images with red laser light source among ResNet, SVM, KNN, and RF. The bold font indicates the best performance.

Methods	Accuracy(std)	Sensitivity(std)	Specificity(std)
ResNet	<b>0.78(0.02)</b>	<b>0.70(0.03)</b>	<b>0.83(0.04)</b>
SVM	0.60(0.02)	0.54(0.03)	0.63(0.02)
KNN	0.60(0.02)	0.42(0.04)	0.71(0.08)
RF	0.60(0.02)	0.41(0.07)	0.71(0.07)

ratio was able to help classify malignant melanoma (MM) and seborrheic keratosis (SK) with a ROC accuracy (AUC) of 86.5% by using fourth moment analysis. As introduced above, the deep learning enabled structure outperformed other conventional machine learning methods. However, since deep learning and conventional machine learning methods have different principles, it was still necessary to identify whether the machine learning methods described in Section 3.4 are suitable comparison groups. In this study, we applied the 10 PCA components of depolarization ratio  $D$  obtained from both the red and blue light sources to a SVM, KNN and RF classifier. Tables 5 and 6 show the classification results of MM and SK. With the blue light derived depolarization ratio, RF achieved the highest prediction accuracy of 90%, while SVM's sensitivity reached 88%, which indicates that the machine learning methods we selected could capture the skin diseases' information from speckle patterns. It was also interesting to note that the overall performance of blue light was better than red light, which was consistent with our previous study [2]. Blue light penetrates skin less than red light and is more affected by surface morphology. MM is usually considered as a 'smooth' lesion in comparison to the 'rough' SK. We assume that lesion roughness contributes to a higher quality of MM vs SK differentiation using blue light.

## 6. Discussion

### 6.1. Discussion on light sources

An important point that should be noted in this study is the laser light source. In fact, different types of lasers could produce polarization speckle images with unique properties, resulting from different combinations of absorption and scattering in different tissues. Our imaging system, as shown in Fig. 1, was equipped with two different diode lasers: red (663 nm) and blue (405 nm). According to the optical properties of human skin, the depth of light penetration is about 100  $\mu\text{m}$  for the blue laser, while the red one can penetrate as far as 1000  $\mu\text{m}$  [35]. Since the average thickness of the epidermis is about 100  $\mu\text{m}$  [35], we hypothesized that the red polarization speckle could encode information for both the dermo-epidermal junction and the superficial dermis layer. Therefore, we selected red laser derived speckle images in our study as shown in Section 4. However, it is observed that blue light is superior for MM vs SK classification, which is likely a result of the lesions' differing surface roughness. Furthermore, there is great interest in the performance of combining both red and blue light sources as multi-channels to see if this system could be further improved. By adding an extra layer in the deep learning structure, we completed the feature fusion by concatenating the

**Table 5**

Comparisons of the performance of classifying malignant melanoma and seborrheic keratosis (mean value and standard deviation) on red light derived depolarization ratio  $D$  images using SVM, KNN, and RF with 10 PCA components. The bold font indicates the best performance.

Methods	Accuracy(std)	Sensitivity(std)	Specificity(std)
SVM	<b>0.75 (0.07)</b>	<b>0.70(0.01)</b>	0.78(0.02)
KNN	0.70(0.01)	0.39(0.02)	0.84(0.01)
RF	0.75(<0.01)	0.36(<0.01)	<b>0.92(&lt;0.01)</b>

**Table 6**

Comparisons of the performance of classifying malignant melanoma and seborrheic keratosis (mean value and standard deviation) on blue light derived depolarization ratio D images using SVM, KNN, and RF with 10 PCA components. The bold font indicates the best performance.

Methods	Accuracy(std)	Sensitivity(std)	Specificity(std)
SVM	0.86 (<0.01)	<b>0.88(0.01)</b>	0.86(<0.01)
KNN	0.80(<0.01)	0.70(0.02)	0.85(<0.01)
RF	<b>0.90(&lt;0.01)</b>	0.73(0.02)	<b>0.98(&lt;0.01)</b>

red channel and blue channel together for skin cancer detection. Here, we give a comparison of performance in skin cancer detection between different light sources including red, blue, and their combination based on the ResNet\_101. As shown in Fig. 5, the overall performance of red light is better than blue light and the multi-channels. This is a very interesting finding which, combined with the result of our PCA-machine learning study, suggests that the speckle image features that are most applicable in cancer vs. benign classification are generated through volume scattering as opposed to surface scattering, due to the penetration depth difference of red and blue light. A future study is needed to further understand the extent and significance of these differences.

## 6.2. Discussion on classifiers

From the classifier perspective, we evaluated many popular and the state-of-the-art deep neural networks in our preliminary study, including VGG16, Inception-V3, ResNet with different numbers of layers and EfficientNet. As shown in Fig. 6, after having each of these networks perform the same cancer vs. benign classification task on our speckle image dataset we found ResNet had the greatest accuracy, although the differences in performance are small enough (within 5%) to be negligible. Unlike other imaging modalities, speckle images consist of huge quantities of global polarization features, to this end, a patch-based approach was applied. ResNet is preferred in our study since it is able to keep learning abundant information with a deeper structure, as its residual block leaves it exempt from the overfitting problem. With regard to the choice of layers, we tried different numbers of layers and found the ResNet with 101 layers is optimal. The classification performance stopped improving with more than 101 layers, suggesting that the ResNet architecture with 101 layers is sufficient to recognize global

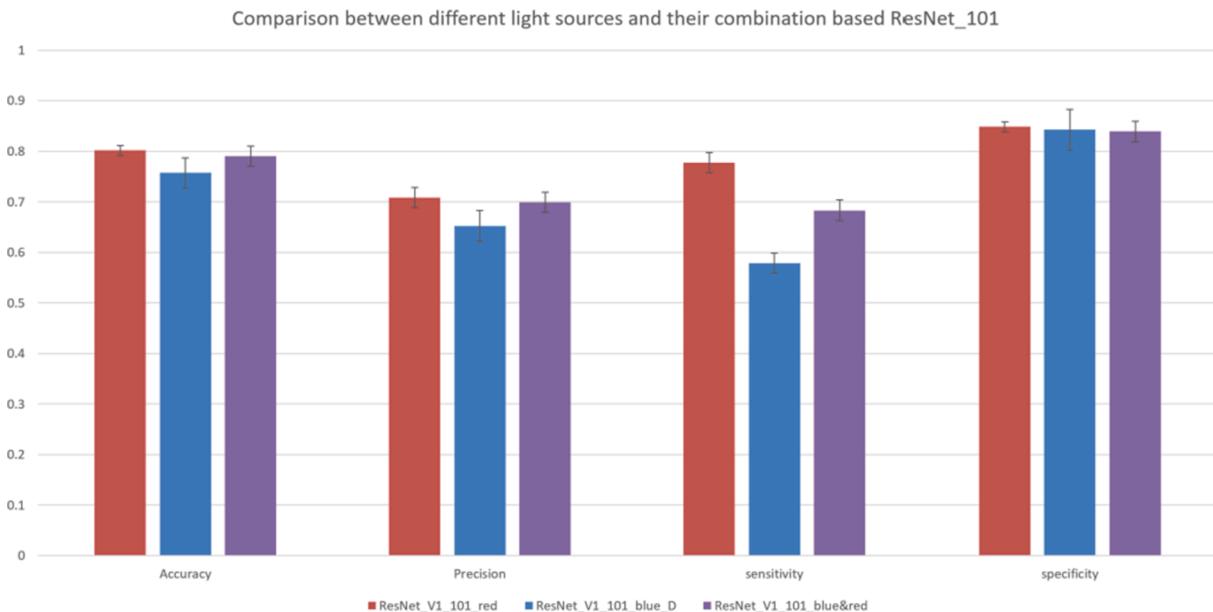
polarization features from speckle images. Additionally, we investigated the performance of ResNet\_v2\_101, however, no obvious difference was found between different versions of the ResNet with the same hyper-parameters.

## 6.3. Discussion on pre-training and training from scratch

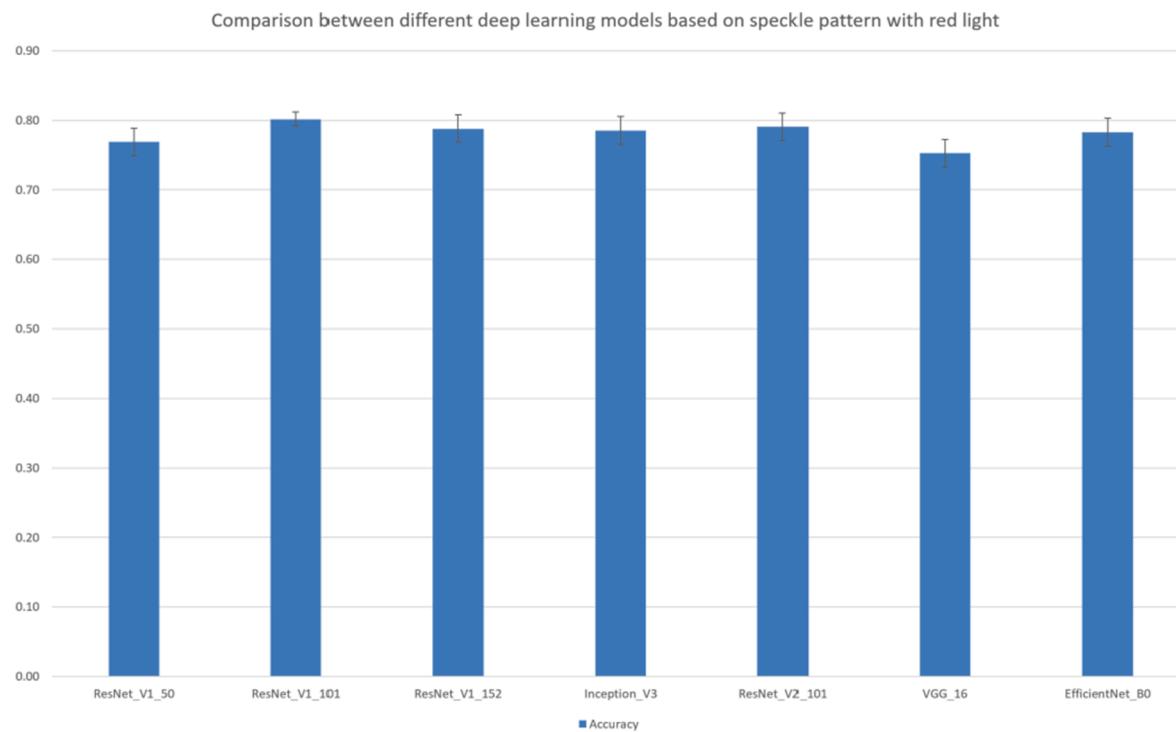
It is natural to question the validity of applying ImageNet pre-trained checkpoints to speckle patterns. Therefore, we also completed a comparison between training based on the pre-trained checkpoints and training from scratch directly on speckle patterns. As shown in Fig. 7, our results indicated that employing the pre-trained checkpoints substantially improved the performance over training from scratch. We believe this is caused by the limited size of our speckle dataset and training directly from scratch generally requires a huge training dataset. There could be further improvement for speckle-based classification if a large enough speckle dataset is acquired.

## 6.4. Discussion on speckle and color images

As mentioned in the introduction section of this paper, color images are regarded as the primary image modality for skin cancer detection. Based on this practical clinical reality, it is worthwhile to compare our speckle pattern's performance with general images. During the data collection phase, we collected clinical color images as well as the speckle images for each lesion. The paired sample images are shown in Fig. 8. We used ResNet\_v1\_101, pre-trained by ImageNet, as the backbone and split the datasets into training and test based on the same list for both speckle images and clinical images. We found that the accuracy performance of speckle based skin cancer detection (80–82%) is comparable to the color clinical images (81–83%). Studies [36] of color clinical images shows classification performance benefit from a large and properly pre-trained network with ImageNet. Our experiment demonstrated that the pre-trained checkpoints could also improve the speckle pattern based performance which leads to the conclusion that the pre-trained dataset from the ImageNet is better than using small speckle dataset. What's more, since the pre-trained model based on ImageNet can benefit both clinical color images and speckle pattern, we believe there is a great potential for speckle based classification. Determining how to leverage the knowledge from both clinical and speckle images to



**Fig. 5.** Comparison of skin cancer detection performance between different light sources and their combination based on ResNet\_101. Red bars refer to red light, blue bars refer to blue light, and the purple bars represent their combination.



**Fig. 6.** Comparison of different state-of-the-art deep learning structures in skin cancer detection based on red light's speckle images.



**Fig. 7.** Comparison between deep learning models with pre-trained checkpoints and without pre-trained checkpoints based on red light speckle patterns and ResNet\_101.

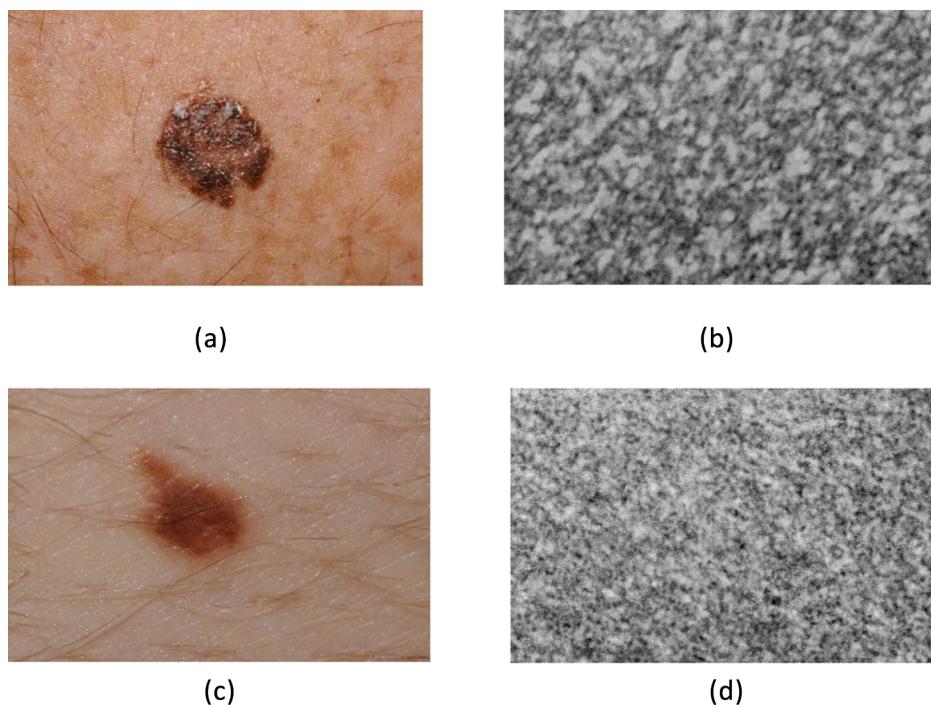
benefit skin cancer detection is an important direction for our future study.

## 7. Conclusion

In this study, we combined polarization speckle imaging and machine learning strategies, including deep learning and conventional machine learning, to obtain a better classification performance for skin cancer detection. We demonstrated the superiority of ResNet-based deep

learning over traditional machine learning methods such as SVM, KNN and RF along with our previous statistical analysis. We improved the performance in classifying benign and malignant skin lesions by 20%. Additional analyses identified that depolarization ratio outperformed raw speckle images in capturing the clinical information in speckle patterns, revealing the importance of the polarization aspect of speckle.

To the best of our knowledge, we are the first group to apply machine learning to polarization speckle images for skin cancer detection. By embracing the stochastic but statistically-uniform nature of polarization



**Fig. 8.** Example paired clinical color images and red light based speckle polarization ratio (D). The first row is cancer images, while the second row represents the benign skin lesion.

speckle patterns, the deep learning enabled system proposed in this paper appears to extract otherwise hidden information from these polarization speckle images. This is an important step forward in automating polarization speckle techniques. In clinical practices, the speed at which a device operates is an important factor. A single speckle image could be acquired and processed by the trained ResNet-101 system within 100 ms. This speed opens up the possibility of real-time skin cancer scanning and detection, which is critical for *in vivo* examination in busy clinics.

#### CRediT authorship contribution statement

**Yuheng Wang:** Methodology, Investigation, Software, Validation, Formal analysis, Writing - original draft. **Daniel C. Louie:** Investigation, Formal analysis, Writing - review & editing. **Jiayue Cai:** Methodology, Investigation, Validation, Software. **Lioudmila Tchvialeva:** Formal analysis, Investigation, Validation, Writing - review & editing. **Harvey Lui:** Resources, Data curation. **Z. Jane Wang:** Methodology, Funding acquisition. **Tim K. Lee:** Conceptualization, Project administration, Supervision, Funding acquisition, Writing - review & editing.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

This work was supported in part by grants from the Canadian Dermatology Foundation, Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, UBC & VGH Hospital Foundation, and UBC Faculty of Medicine.

#### References

- [1] W. Wang, S.G. Hanson, M. Takeda, Statistics of polarization speckle: theory versus experiment. Ninth International Conference on Correlation Optics, International Society for Optics and Photonics, 2009.
- [2] L. Tchvialeva, et al., Polarization speckle imaging as a potential technique for *in vivo* skin cancer detection, *J. Biomed. Opt.* 18 (6) (2012) 061211.
- [3] K. Leonhardt, H.J. Tiziani, Removing ambiguities in surface roughness measurement, *Optica Acta: Int. J. Opt.* 29 (4) (1982) 493–499.
- [4] C. Hun, J.-M. Caussignac, M.M. Bruynooghe, Speckle techniques for pavement surface analysis. *Speckle Metrology 2003*, International Society for Optics and Photonics, 2003.
- [5] T.K. Lee, et al., In-Vivo Skin Roughness Measurement by Laser Speckle, in: *Fringe 2013*, Springer, 2014, pp. 933–936.
- [6] P. Miao, et al., Entropy analysis reveals a simple linear relation between laser speckle and blood flow, *Opt. Lett.* 39 (13) (2014) 3907–3910.
- [7] J. Briers, Speckle techniques, in: *Optical measurement methods in biomechanics*, Springer, 1997, pp. 76–98.
- [8] W. Heeman, et al., Clinical applications of laser speckle contrast imaging: a review, *J. Biomed. Opt.* 24 (8) (2019), 080901.
- [9] L. Tchvialeva, et al., Skin roughness assessment, *New Develop. Biomed. Eng.* (2010) 341–358.
- [10] J.R. Mourant, et al., Predictions and measurements of scattering and absorption over broad wavelength ranges in tissue phantoms, *Appl. Opt.* 36 (4) (1997) 949–957.
- [11] Cancer Facts and Figures. American Cancer Society, 2020 (accessed January 8, 2020).
- [12] A. Esteva, et al., Dermatologist-level classification of skin cancer with deep neural networks, *Nature* 542 (7639) (2017) 115–118.
- [13] N. Gessert, et al., Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data, *MethodsX* (2020) 100864.
- [14] T. Akram, et al., A multilevel features selection framework for skin lesion classification, *Hum.-centric Comput. Inform. Sci.* 10 (2020) 1–26.
- [15] J. Kawahara, et al., Seven-point checklist and skin lesion classification using multitask multimodal neural nets, *IEEE J. Biomed. Health. Inf.* 23 (2) (2018) 538–546.
- [16] A. Hekler, et al., Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images, *Eur. J. Cancer* 118 (2019) 91–96.
- [17] N. Klebanov, et al., Clinical spectrum of cutaneous melanoma morphology, *J. Am. Acad. Dermatol.* 80 (1) (2019) 178–188, e3.
- [18] K. Matsunaga et al., Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble, 2017, arXiv preprint arXiv: 1703.03108.
- [19] T. Lee, et al., *Polarization Speckles and Skin Applications*, in: *Imaging in Dermatology*, Elsevier, 2016, pp. 77–87.
- [20] K. He, et al., Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

- [21] C. Szegedy, et al., Inception-v4, inception-resnet and the impact of residual connections on learning. Thirty-first AAAI conference on artificial intelligence, 2017.
- [22] J.A. Suykens, J. Vandewalle, Least squares support vector machine classifiers, *Neural Process. Lett.* 9 (3) (1999) 293–300.
- [23] L.E. Peterson, K-nearest neighbor, *Scholarpedia* 4 (2) (2009) 1883.
- [24] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32.
- [25] G. Blanchet, M. Charbit, Digital signal and image processing using MATLAB, vol. 4, Wiley Online Library, 2006.
- [26] A. Jung, imgaug, 2017. URL: <https://github.com/aleju/imgaug> (visited on 04/20/2017).
- [27] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.
- [28] L. Tchvialeva, I. Markhvida, T. Lee, Error analysis for polychromatic speckle contrast measurements, *Opt. Lasers Eng.* 49 (12) (2011) 1397–1401.
- [29] S. Han, C. Qubo, H. Meng, Parameter selection in SVM with RBF kernel function. *World Automation Congress 2012*, IEEE, 2012.
- [30] M. Norouzi, D.J. Fleet, R.R. Salakhutdinov, Hamming distance metric learning. *Advances in neural information processing systems*, 2012.
- [31] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 2012.
- [32] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014. arXiv preprint arXiv:1412.6980.
- [33] M. Abadi, et al., Tensorflow: A system for large-scale machine learning. 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), 2016.
- [34] D. Nandi, et al., Principal component analysis in medical image processing: a study, *Int. J. Image Min.* 1 (1) (2015) 65–86.
- [35] E.V. Salomatina, et al., Optical properties of normal and cancerous human skin in the visible and near-infrared spectral range, *J. Biomed. Opt.* 11 (6) (2006), 064026.
- [36] J. Yosinski, et al., How transferable are features in deep neural networks?. *Advances in neural information processing systems*, 2014.