Automatic System for Skin Cancer Detection

Minh Vu-Hoang[1], Duc-Loc Nguyen-Vo[1], and Khoa Nguyen-Dang[1]

[1]Advanced Program in Computer Science, Faculty of Information Technology, University of Science, VNU-HCM

***Abstract*—Skin cancer is the most commonly diagnosed cancer. One person dies of melanoma every hour. The patients need to receive treatment when the cancerous tissue is not widespread. The demand for early detecting skin cancer is rising. Therefore, we present our experiments and compare the results of 2 approaches: using image processing, and using convolutional neural network classification. We achieve the accuracy of ...% on a balanced data set. This model can be used in a cost-effective system to detect malignant skin tumor at domestic level.**

I. Introduction

There are approximately 91,270 new cases with 9,320 deaths due to melanoma in the US in 2018[1]. One in five Americans will develop skin cancer by the age of 70[2]. It costs the US about $8.1 billion every year for treatment[3]. Fortunately, most cases of skin cancer can be cured if we can detect and treat them early[4]. Melanoma skin cancer has the 5-year relative survival rate is 99% for localized stage, 63% for regional stage, and 20% for distant-stage[4]. Skin lesion diagnosis often requires skin biopsy. A skin biopsy removes skin samples from the skin surface and sends that to a laboratory for examination. This is a complicated procedure. Result may take several weeks or months and the patient can be in a higher risk of the disease. We need a more convenient way to predict skin cancer at early stage without the help of complicated equipment. One solution is to make prediction base on color image of the lesion. With the popularity of smartphones, taking a color image is very simple. The aim of this project is to provide a automatic system at domestic level for predicting cancerous tumor given the skin image. The system can help in self-examination for unusual skin lesions or can be used to prioritize cases in skin biopsy.

There are several studies on this subject. Most of them fall into two categories. The first approach is learning from the physical features of the tumor[5, 6]. We apply image processing technique to extract the features of interest from the image. Then, those features are used for classification. The second approach is to train a deep convolutional neural network directly on the input images[3].

II. Background

In this section, we will discuss in more detail about existing methods for this problem.

The first approach is to learn features of the lesion. The input images go through a preprocessing step to reduce noise if necessary. After that, we apply image processing algorithm such as automatic threshold[7] or partial differential equations based algorithm[8] or neural network[9] to get the binary segmentation of the given image. Having the original image together with its segmentation, we can extract useful physical features from the sample. One popular choice of features is the ABCDE rule. It is originally proposed by Friedman et al. [10] and suggested by dermatologists to self-exam skin lesion at home. The ABCDE rule consists of five criteria to identify whether the tumour is malignant: A for asymmetry (one half of the mole does not match the other half), B for border irregularity (the edges are ragged or blurred), C for color (the color is not uniform), D for diameter (the diameter is greater than 6mm), and E for evolution (the change of the lesion over time). The features are fed into a neural network for classification.

The second approach is training an end-to-end deep convolutional neural network using only the images and labels as input. Convolutional neural network is a special architecture of artificial neural networks, proposed by LeCun et al.[11]. It is effective and widely used for image classification. Using fully connected neural networks may lose interaction between pixels. CNN has convolutional layers, which can be considered as automatic feature extractors. They make use of adjacent pixels information to create output and feed into the next fully-connected layers. Convolutional neural networks become more complex and give higher accuracy.

III. Proposed Method

A. Data preprocessing

We remove noise using median blur with the kernel window of size 3 by 3. A kernel window is centered on each pixel and the value of the central pixel is replaced with the median value of all pixels under the kernel window. Because of the memory limit, we resize each image to 300×225 for method III-B and 200×150 for method III-C. We split 80% dataset for training and 20% for testing.

B. Method 1: Image processing

1) Overview: Figure 1 shows 3 stages of this method. After preprocessing step, we apply threshold to find the lesion segmentation. The color image and the binary image are used to extract features of the lesion. Then we use a fully connected neural network to classify images base on these features.

2) Lesion segmentation: Our main idea is to use image thresholding. Every image is converted to grayscale. We choose a value $threshold$ and then convert the grayscale image to binary image using the equation 1. If a pixel has

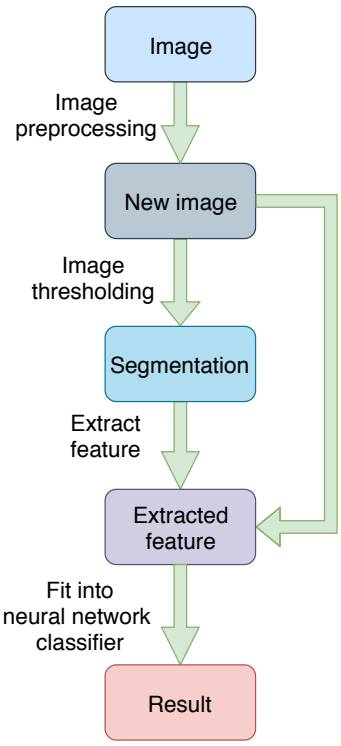


Fig. 1: Image processing method overview

value smaller than $threshold$, then it is black. Otherwise, it is white.

$$destination(x,y) = \begin{cases} 255, & \text{if } source(x,y) > threshold \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Choosing $threshold$ value is a critical step to get an accurate contour of the lesion. For each image, we choose the $threshold$ value slightly smaller than the mean of all pixels

$$threshold = 0.8 \times \frac{\sum_{x=0}^{width-1} \sum_{y=0}^{height-1} pixel(x,y)}{width \times height} + 0.2 \times \min_{x \in \{0...width-1\},\, y \in \{0...height-1\}} pixel(x,y) \tag{2}$$

3) Feature extraction: We adapt the idea of Ercal et al.[6] to choose 4 following features to represent a lesion: circularity index, color mean, color variance, and relative color between the lesion and the skin background.

The circularity index represents how likely the lesion is a circle and it is computed by equation 3

$$circularity\ index = \frac{4 \times \pi \times a}{p^2} \tag{3}$$

where a is the area and p is the perimeter of the lesion. We can obtain a by counting black pixels in the binary image and p by counting black pixels which is next to a white one.

The second feature is the mean color of pixels inside the lesion. Previous studies in diagnosing melanoma[12] shows that spherical color space coordinates give better results than the RGB color space. We transform from RGB pixel to spherical coordinate using the equations 4.

$$\begin{aligned} L &= \sqrt{R^2 + G^2 + B^2} \\ \alpha &= \arccos \frac{B}{L} \\ \beta &= \arccos \frac{R}{L \times \sin \alpha} \end{aligned} \tag{4}$$

Malignant tumor color is not uniform. Therefore, we add the color variance to our feature set. We calculate the color variance on R, G, and B color plane using equation 5.

$$\begin{aligned} Variance_R &= \frac{\sum_{pixel\ \in\ lesion} (R_{pixel} - R_{mean})^2}{area} \\ Variance_G &= \frac{\sum_{pixel\ \in\ lesion} (G_{pixel} - G_{mean})^2}{area} \\ Variance_B &= \frac{\sum_{pixel\ \in\ lesion} (B_{pixel} - B_{mean})^2}{area} \end{aligned} \tag{5}$$

Relative color is added to the feature list to equalize any variations caused by illumination, and to equalize variations in normal skin color between individuals[12]. As shown in equation 6, relative color on a color plane is defined as the subtraction of the normalized value of that color in the lesion area and the normalized value of that color in the skin background.

$$\begin{aligned} Relative_R &= \frac{R_l}{R_l + G_l + B_l} - \frac{R_{bg}}{R_{bg} + G_{bg} + B_{bg}} \\ Relative_G &= \frac{G_l}{R_l + G_l + B_l} - \frac{G_{bg}}{R_{bg} + G_{bg} + B_{bg}} \\ Relative_B &= \frac{B_l}{R_l + G_l + B_l} - \frac{B_{bg}}{R_{bg} + G_{bg} + B_{bg}} \end{aligned} \tag{6}$$

where R_l, G_l, B_l denote the red, green, blue mean value of pixels inside the lesion, and R_{bg}, G_{bg}, B_{bg} denote the red, green, blue mean value of pixels of the skin background.

4) Classification: The features obtained in III-B3 is fed to a small fully connected neural network for classification. The network architecture is shown in the table I

Layer	Number of nodes	Activation function
Input layer	10	None
Dense layer	16	Relu
Dense layer	32	Relu
Dense layer	32	Relu
Output layer	7	Softmax

TABLE I: Method 1 network architecture

C. Method 2: Convolutional neural network classification

1) Overview: On the training set, we perform many image augmentation techniques: rotation, flipping, shifting. We train a CNN model on the training images. Then, the model is evaluated on the test set to give the result.

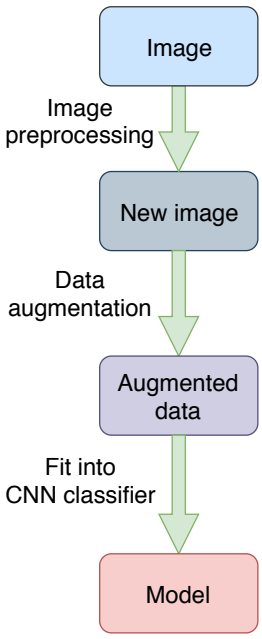


Fig. 2: CNN approach overview

2) Data augmentation: To prevent overfitting, we rotate, flip, and shift the images. The angle of rotation is randomly chosen from 0 to 180. The image is flipped and shifted in vertical and horizontal direction. The shifting amount is at most 10 percent of the image dimension. The pixels outside the input image boundary are filled with the color of the nearest pixel.

3) Training CNN model: After data augmentation, we fit the image created in subsection III-C2 into a CNN model.

We do the experiments on 4 architecture: ResNet50, ResNet152[13], DenseNet121, and DenseNet201[14]. The ResNet152 architecture proposed by He et al.[13] achieved top-5 error rate of 3.57% on ImageNet dataset consisting of around 22000 images from 1000 categories. In ResNet, identity mapping is proposed to help in solving vanishing gradient problem as the network become deeper. DenseNet is an improved version of ResNet. Each layer receives inputs from all preceding layers and passes its feature maps to all subsequent layers. The network becomes thinner and gives higher computational and memory efficiency.

We use the same configuration for each model. A max pooling layer is added after the base network and followed by a fully-connected layer for classification.

We train each model 200 epochs with learning rate 1×10^{-3} and 50 epochs with learning rate 3×10^{-4}.

IV. Experiments and Results

A. Dataset

We use the HAM10000 dataset[15] consisting of 10015 images from 7 types of skin lesion: actinic keratoses, basal cell carcinoma, benign keratosis, dermatofibroma, melanoma, melanocytic nevi, and vascular skin lesions. The lesions appear on many parts of the body including face, chest, back, trunk, abdomen, scalp, genital, neck, hand, foot, acral, upper extremity, and lower extremity. The labels are confirmed by histopathology, follow-up examination, expert consensus, and in-vivo confocal microscopy. Each image has the size 600×450 and only consists the lesion with a skin background. The number of images of each class is show in table II.

Type of lesion	Number of examples
Actinic keratoses	327
Basal cell carcinoma	514
Benign keratosis	1099
Dermatofibroma	115
Melanoma	1113
Melanocytic nevi	6705
Vascular skin lesions	142

TABLE II: Number of images of each class

B. Experimental Design

We run our code on Google Colaboratory. In method III-B, we use OpenCV 3.4.3 to perform image processing. The image augmentation in method III-C is achieved by using ImageDataGenerator class in keras. It returns an iterator to images. For each call, it will generate a batch of images with real-time data augmentation. We use keras 2.2.4 with tensorflow backend to build our models. As we mention in IV-A, our dataset is highly imbalanced. The number of examples of Melanocytic nevi is nearly 60 times the number of Dermatofibroma. Then, we assign the class weights to balance our training set. The class weight is used for weighting the loss function during training. The loss of each image is multiplied by the class weight. Therefore, we choose the weight of each class is the multiplicative inverse of the number of examples of that class in the training set. For each class i:

$$CW_i = \frac{1}{N_i} \tag{7}$$

where CW_i is the class weight of class i and N_i is the number of examples of class i in the training set.

For each model, we calculate the balanced accuracy on the test set and plot its confusion matrix.

C. Results

Balanced accuracy for each experiment is shown in table III

Method	Balanced accuracy
Image processing method	0.4609
ResNet50	0.6806
ResNet152	0.6670
DenseNet121	0.6939
DenseNet201	0.7154

TABLE III: Balanced accuracy

Figure 3 is the confusion matrix obtained in method III-B.

Figure 4 shows the confusion matrix of CNN classification with ResNet50, ResNet152, DenseNet121, and DenseNet201.

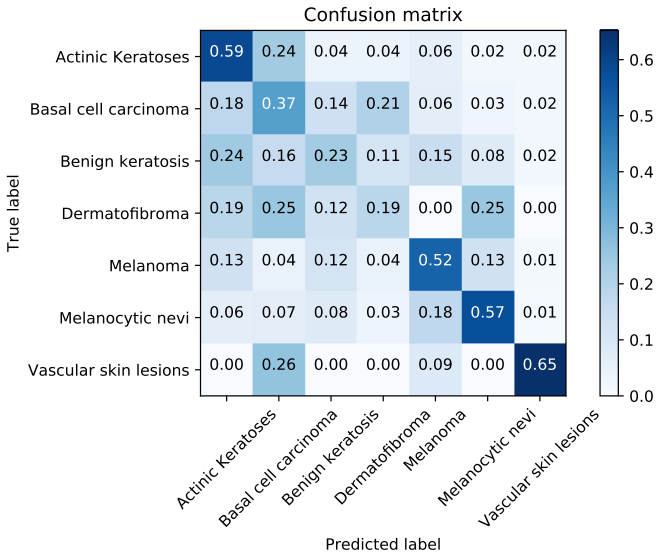


Fig. 3: Confusion matrix of image processing method

V. Conclusion

The image processing based method gives poor result especially for 3 classes Basal cell carcinoma, Benign keratosis, and Dermatobroma. It is not sufficient to choose 4 features to represent lesions coming from 7 classes. Overall, CNN classification with DenseNet201 perform the best in our experiments. The two type of cancerous lesions are basal cell carcinoma and melanoma have the f1 score 0.71 and 0.64 respectively. The class that have the highest f1 score is vascular skin lesion with 0.96. It is reasonable as this type of lesion is related to growth formed from blood vessels and often have pink color whereas other type of skin lesions are darker.

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 1, pp. 7–30, 2018.

[2] R. S. Stern, "Prevalence of a History of Skin Cancer in 2007: Results of an Incidence-Based Model," *Archives of Dermatology*, vol. 146, no. 3, pp. 279–282, 03 2010.

[3] A. Esteva, B. Kuprel, R. Novoa, J. Ko, S. M Swetter, H. M Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, 01 2017.

[4] A. C. Society, "Cancer facts and figures 2018."

[5] "Computer aided melanoma skin cancer detection using image processing," *Procedia Computer Science*, vol. 48, pp. 735 – 740, 2015, international Conference on Computer, Communication and Convergence (ICCC 2015).

[6] F. Ercal, A. Chawla, W. V. Stoecker, , and R. H. Moss, "Neural network diagnosis of malignant melanoma from color images," *IEEE Transactions on Biomedical Engineering*, vol. 41, no. 9, pp. 837–845, Sep. 1994.

[7] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, Jan 1979.

[8] D. H. Chung and G. Sapiro, "Segmenting skin lesions with partial differential equations based image processing algorithms," in *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, vol. 3, Sep. 2000, pp. 404–407 vol.3.

[9] S. Vesal, N. Ravikumar, and A. K. Maier, "Skinnet: A deep learning framework for skin lesion segmentation," *CoRR*, vol. abs/1806.09522, 2018.

[10] R. J. Friedman, D. S. Rigel, and A. W. Kopf, "Early detection of malignant melanoma: The role of physician examination and self-examination of the skin," *CA: A Cancer Journal for Clinicians*, vol. 35, no. 3, pp. 130–151, 6 1985.

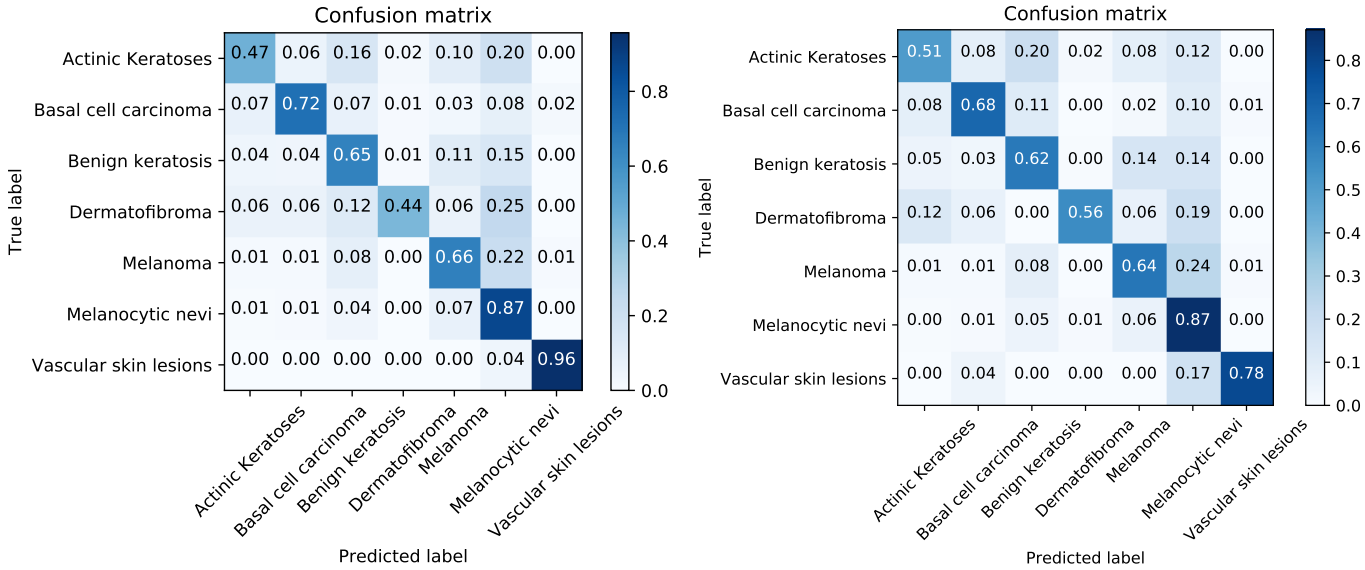
[11] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[12] S. E. Umbaugh, "Computer vision in medicine: Color metrics and image segmentation methods for skin cancer diagnosis," Ph.D. dissertation, Rolla, MO, USA, 1990, uMI Order No. GAX90-32115.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.

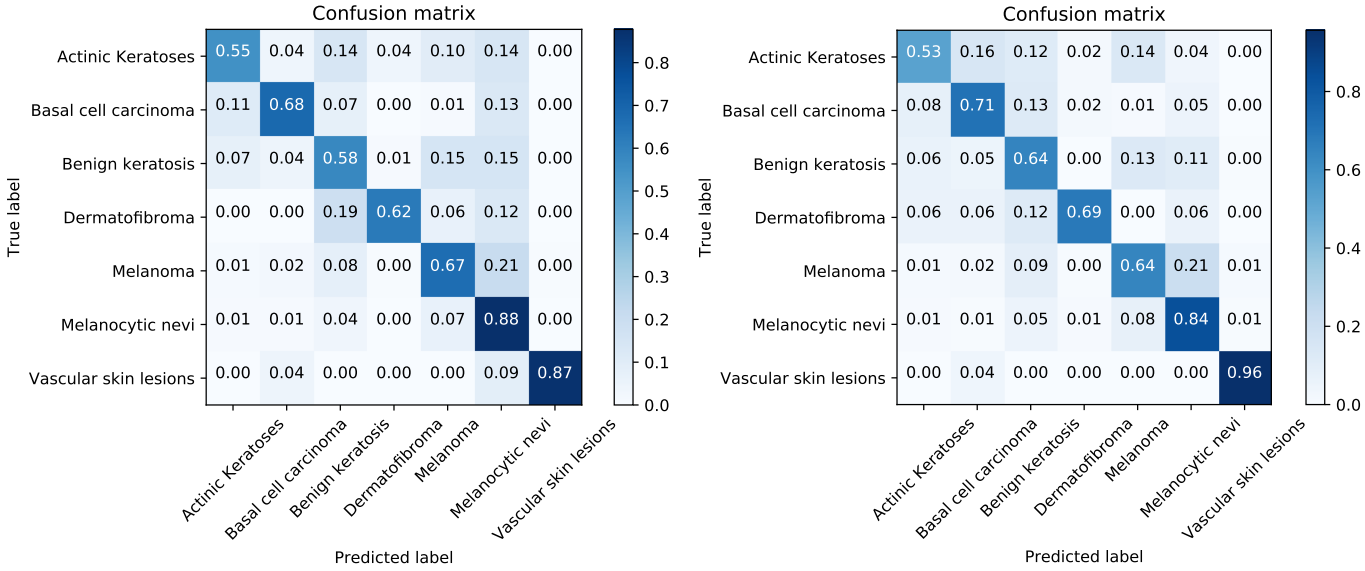
[14] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[15] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, 03 2018.



(a) ResNet50

(b) ResNet152



(c) DenseNet121

(d) DenseNet201

Fig. 4: Confusion matrix of ResNet50, ResNet152, DenseNet121, and DenseNet201