International Journal of Innovative
Computing, Information and Control
Volume **16**, Number **5**, October **2020**

ICIC International ©2020 ISSN 1349-4198
pp. 1821–1827

DEVELOPMENT OF A SKIN TEXTURE EVALUATION SYSTEM USING A CONVOLUTIONAL NEURAL NETWORK

Moeka Nagano and Tadanori Fukami

Graduate School of Science and Engineering
Yamagata University
4-3-16 Jonan, Yonezawa, Yamagata 992-8510, Japan
fukami@yz.yamagata-u.ac.jp

Received March 2020; revised June 2020

Abstract. *Currently, various methods are used to evaluate the beauty of skin, but problems arise with evaluation accuracy and the time required for evaluation. Previous research has proposed a method in which researchers select features and evaluate skin texture using machine learning with a three-layer neural network. However, results of this research were discrepant from visual evaluation by experts. In this paper, we describe the development of a convolutional neural network that acquires skin features through learning. We used images captured with a microscope as input data and visual evaluation scores as training data. To evaluate performance, we examined the correlation coefficient between the estimated evaluation score by the convolutional neural network and the visual evaluation score, as well as the matching rate of all data scores. When data from a single subject were used for learning but were not used as test data (an "open set"), the correlation coefficient and matching rate were 0.903 and 73.2%, respectively. Alternately, when data from a single subject were used for both learning and test data (a "closed set"), the correlation coefficient and matching rate were 0.922 and 77.8%, respectively. In each case, these values were significantly higher than those of previous studies.*
Keywords: Skin texture, Evaluation system, Convolutional neural network (CNN), Deep learning

1. **Introduction.** Recent years have seen an abundance of information on beauty and increasing interest in appearance, especially the condition of skin. In particular, information on skin care and skin cleaning product development is expanding. Understanding their skin's condition is a key concern for consumers choosing a product. Therefore, developing an evaluation system that can easily determine skin condition is important. Here, we focus on skin texture as it strongly influences the perception of skin beauty. Skin is considered to be well-textured when it satisfies the following characteristics [1]: creases in the skin (called sulcus cutis) are of a moderate depth, narrow, and linear, and any skin protrusions (called crista cutis) are small, the area between them is uneven, and they are regular in shape (Figure 1).

For skin evaluation, software programs exist that can evaluate the condition of skin from images taken by a smartphone camera. However, the image processing algorithms used in such programs are proprietary, and evaluation is susceptible to external factors such as illumination; hence the evaluation result is not stable. Thus, a skin evaluation method that addresses such problems is needed.

There have been many reports on the detection of skin cancers such as melanoma [2, 3, 4] in the context of evaluating human skin conditions, as well as in fish species classification [5, 6], but there are few studies on the evaluation of skin for beauty. Kobayashi et

DOI: 10.24507/ijicic.16.05.1821

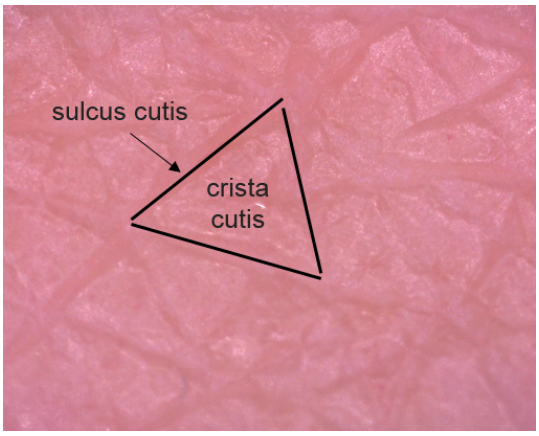


FIGURE 1. Skin texture

al. [7] and Takemae et al. [8] have attempted automatic evaluation using a three-layer neural network using features extracted from skin texture images as input signals and visual evaluation scores by experts as training data. Kobayashi et al. performed their evaluation using the image features of sulcus cutis and used an empirically obtained threshold value for binarization during sulcus cutis extraction [7]. Takemae et al. used 58 image features that were subjectively selected to evaluate the skin texture [8]. In these papers, the selection and extraction of the features are based on experience and subjectivity. Therefore, our goal is to evaluate skin texture using deep learning, which can automatically obtain the optimal features by learning. In face recognition systems, research has been conducted to implement a high-accuracy system for mobile phones [9]. We also aim to ultimately develop an evaluation system for skin texture that can be implemented on mobile phones and is more accurate than conventional methods.

In this study, we describe the methods for image acquisition, preprocessing for learning (including contrast correction and image augmentation), visual evaluation, convolutional neural network (CNN) construction, and results evaluation in Section 2. Next, the results are shown in Section 3, the discussion is given in Section 4, and the conclusion is stated in the final section.

2. **Methods.**

2.1. **Image acquisition.** The images used in this study were obtained from a subject's forehead, both cheeks, the front and back of both wrists, and the lower jaw under conditions in which the skin texture could be seen. A microscope (Dino-Lite Digital Microscope AM4515T), set at a magnification of 70×, was brought into close contact with the skin surface to acquire an image. We acquired 184 images from 23 subjects (19 men and 4 women, age: 22.2 ± 0.74 years old) and 75 images with less body hair were selected. Each image size was 1280 pixels × 1024 pixels, and down-sampling was performed to convert each image into 160- × 128-pixel resolution, at which skin texture can be visually recognized. The measurement process was approved by the Ethics Committee of the Faculty of Education, Art and Science, Yamagata University.

2.2. **Contrast correction.** We confirmed that the sulcus cutis and crista cutis forming skin texture can be visually observed. However, to further emphasize differences between them, contrast correction was performed, with each image being converted from RGB to YCbCr for a down-sampled image. Thereafter, the histogram of the Y component representing the luminance value was flattened and the YCbCr image was returned to an RGB image. The process is shown in Figure 2.

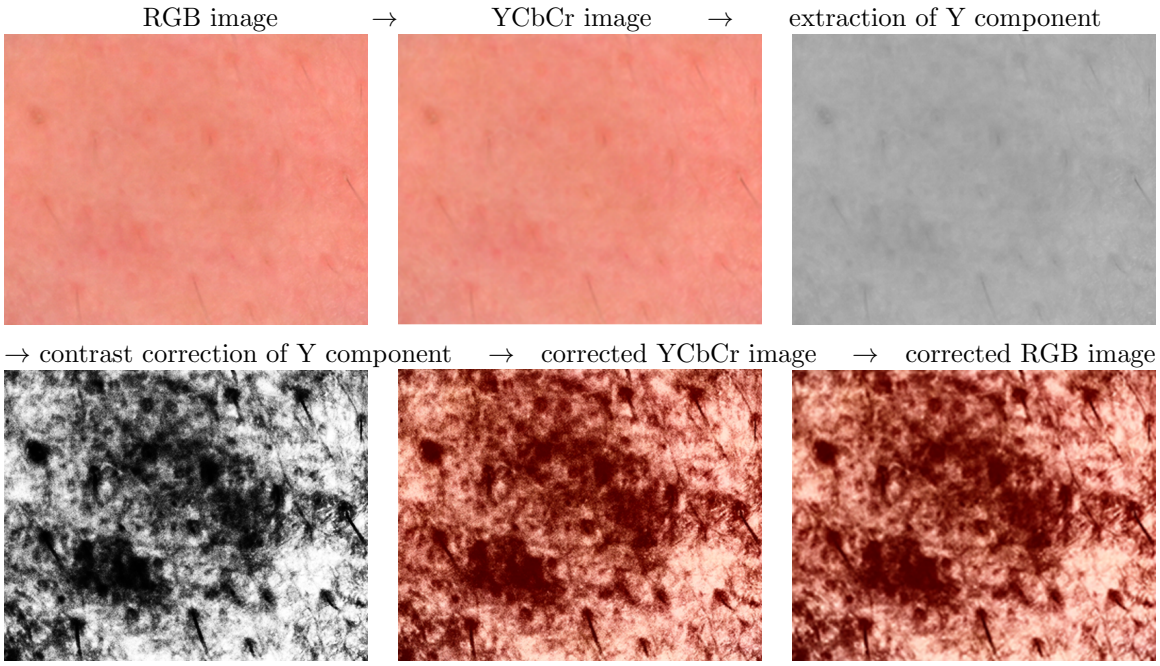


FIGURE 2. Generation of images for learning

2.3. **Image augmentation.** Because the total number of images was relatively small for applying deep learning techniques, we used image augmentation to increase the number of data. For this, we generated additional images by flipping the original images horizontally and inverting them vertically. After that, each image was divided into four parts so that the height and width of each resultant image was half the original size; ultimately, the image size used for learning was 80 pixels × 64 pixels. A total of 900 images were used in this study: 75 images × 4 divisions × 3 (the original image and two types of augmented images).

2.4. **Visual evaluation.** Ten evaluators evaluated the quality of the skin texture on a five-point scale, comparing the image with a reference image presented side-by-side on a display as shown in Figure 3. A score of three was considered an intermediate-level image. An intermediate-level texture image was set as a reference image. Scores of one and five meant much worse and much better than the intermediate level, respectively. The conditions for perfect skin texture were the following six points: sulcus cutis are straight, fine and of moderate depth, with small crista cutis, unevenness between sulcus cutis and

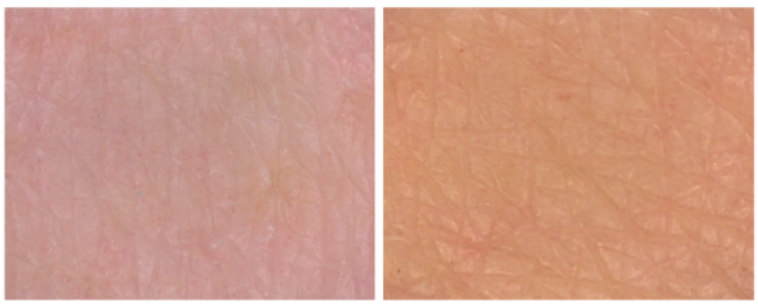


FIGURE 3. Visual evaluation (The reference image is an image equivalent to evaluation score 3.)

crista cutis, and crista cutis that are regular in shape. A final visual evaluation score was obtained as the average of values from the ten evaluators.

2.5. **CNN construction for deep learning.** Using the processed image data and the visual evaluation score as training data, we constructed a network as shown in Figure 4 to perform the learning. Because the estimated evaluation score is continuous in value, this network employs a regression model that outputs an estimated evaluation score of image data. In view of the limited amount of training data in this study, we constructed a CNN with a relatively shallow layer. This CNN has an input layer for the input of color images with three channels (RGB) and additional two sets of a convolutional layer and a pooling layer that are connected in series. In the convolutional layer, the number of filters, kernel size, and stride are set to 32, 3×3, and 1×1, respectively, and the size is maintained by zero padding. Furthermore, in the pooling layer, max pooling is performed over a 2×2 region. Then, the data are unfolded and connected to the fully connected layer and the output layer. To prevent overfitting, dropout at a rate of 0.25 is performed between the convolutional and pooling layers and dropout at a rate of 0.5 is performed between the flatten layer and fully connected layer.

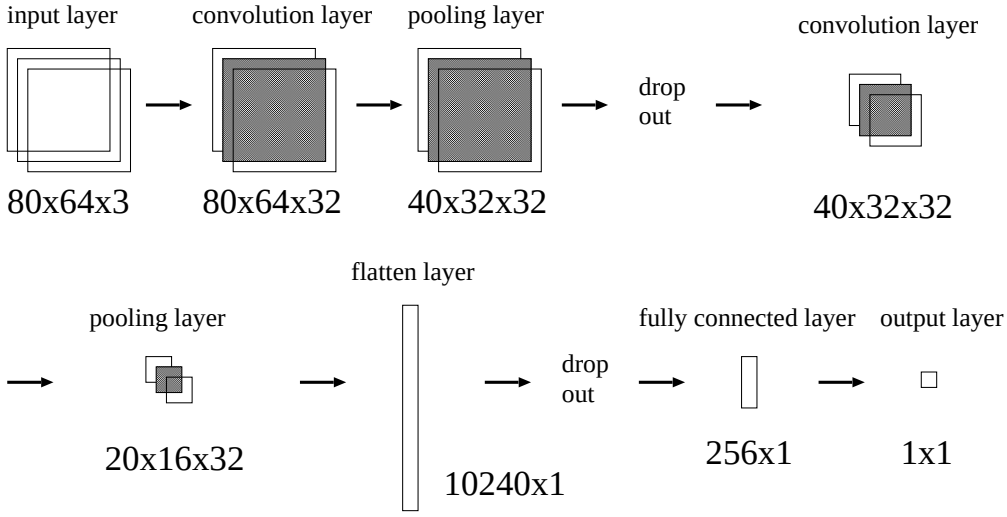


FIGURE 4. Convolutional neural network for skin texture evaluation

2.6. **Evaluation of estimated scores in our system.** In this study, five-fold cross-validation was performed to evaluate performance. This means that 80% of the image data was used for training data, with the remaining 20% used for test data. Two types of data allocation were performed. One type was assignments in which an image derived from a subject used in training was not used as a test datum. The other type was assignments in which an image derived from a subject used in training was allowed as a test datum. We defined datasets allocated in these ways as open and closed sets, respectively.

Evaluation was performed based on two indices used in previous studies: 1) the correlation coefficient between the visual evaluation score and the estimated evaluation score; and 2) the matching rate of the visual evaluation score and the estimated evaluation score. However, the visual evaluation score is a discrete value, whereas the estimated evaluation score is a continuous value because a regression model is used in this study. Therefore, we rounded the estimated evaluation score to the first decimal place and then converted it to an integer to properly compare it with the visual evaluation score.

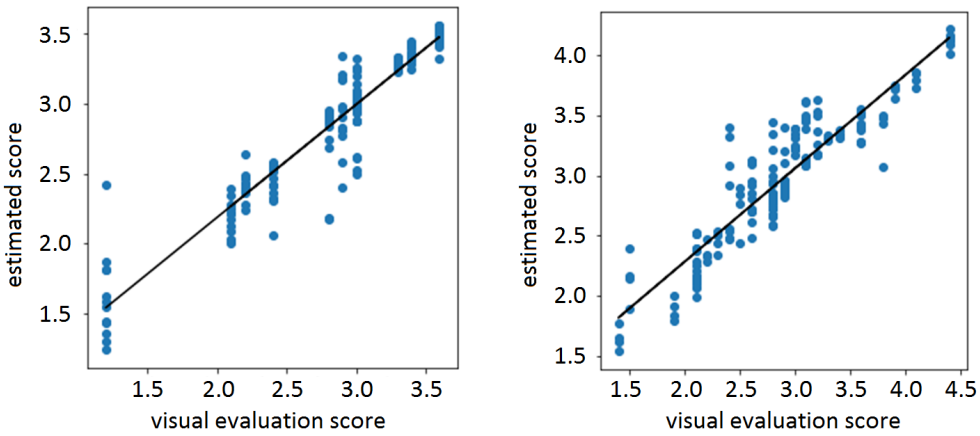
We evaluated two data sets, an open set and a closed set, separately. In the closed set, 900 images were divided into five groups of 180 images without considering subjects,

four groups were used as training data, and the remaining one group was used as test data for evaluation. This procedure was repeated for all five choices, and the performance score from the five runs is then averaged. However, in the open set, images of the same subject used for training are not used as test data. Therefore, although the number of images of each subject is different, each group has approximately 180 images with which to approach the conditions of the closed set. The test was divided into five groups and the same test was performed five times as in the closed set.

3. **Results.** Table 1 shows the correlation coefficient between the visual evaluation score and the estimated score in open and closed sets. Because the evaluation is based on five-fold cross-validation, it is expressed as the average value and standard deviation of five tests. Figure 5 shows an example of plots in a test out of five times. In both open and closed sets, the correlation coefficient showed a high correlation, greater than 0.9.

TABLE 1. Correlation coefficient between estimated score and visual evaluation score

Data set	Correlation coefficient
Open set	0.922 ± 0.0417
Closed set	0.903 ± 0.0265



Open set Closed set

FIGURE 5. Examples of the relationship between estimated score and visual evaluation score

In the two data sets, Table 2 shows the relationship between the value obtained by subtracting the visual evaluation score from the estimated evaluation score and the matching rate of two scores. The estimated score by CNN and the visual evaluation score coincided for more than 70% of all data in both open and closed sets.

TABLE 2. The ratio (%) of the difference between an estimated score and visual evaluation score

	Estimated score – Visual evaluation score			
Data set	-1	0	1	2
Open set	9.87 ± 3.59	$\mathbf{73.2 \pm 10.3}$	16.9 ± 11.2	N/A
Closed set	7.61 ± 2.40	$\mathbf{77.8 \pm 7.87}$	13.9 ± 5.62	1.14 ± 0.57

4. **Discussion.** The correlation coefficients and the differences between the visual and the estimated scores were better than those in the results of previous studies [7, 8]. Simple comparisons are difficult because the data and visual evaluators are different, but the correlation coefficient from our method was 0.922 for the open set, which surpassed the 0.7 of Kobayashi's study [7]. The matching rate between the evaluation score and the estimated evaluation score was 53.4% in Takemae's study [8], whereas our matching rate was 73.2% for the open set.

In previous research, skin texture was evaluated by a three-layer neural network using image features subjectively selected by the researcher such that estimation performance depended on the features selected. Conversely, the deep learning used in this study acquired optimal image features for evaluating skin texture through machine learning. Furthermore, in previous studies, features were acquired after extracting textures by image binarization. In binarization, the threshold value setting is important, but choosing an optimum value is difficult and affects accuracy. However, in our study, texture enhancement was performed using contrast correction as preprocessing; the results show that this correction of variation in luminance among images was effective for feature extraction in deep learning [10, 11].

As shown in Figure 5, most of the plots fit a regression line, but some data plots deviate from the line. Checking the corresponding data, we found relatively large variability across the visual evaluation scores. Presumably, those skin texture images are difficult to evaluate because that form of evaluation depends on human evaluators. In Tables 1 and 2, we have obtained nearly identical results in both open and closed sets. This result indicates that the CNN for evaluating skin texture has obtained a stable result regardless of the subject from which the images were taken. This suggests that variation among body parts is greater than the variation among subjects. As for the relationship between body parts and estimation accuracy, it will be necessary to investigate how the body part affects the learning accuracy in the future. Moreover, because most of the images had visual evaluation scores of two, three, and four, it will be necessary to collect images with extremely good or bad scores.

In this study, the number of men and women in the data are unbalanced. We focused on increasing the number of subjects we recruited, and consequently, there was a large difference in the data we acquired because many male subjects contributed to this experiment. Regarding the difference in skin between men and women, Rahrovan et al. found that the skin parameters of hydration, transepidermal water loss, sebum, microcirculation, pigmentation, and thickness are generally higher in men, but skin pH is higher in women [12]. It is unclear how these parameters affect the skin texture as there are no reports that analyze this issue quantitatively or statistically. However, to realize a more practical system, it will be necessary to reduce the bias in the data by using equal numbers of male and female participants.

5. **Conclusion.** In this study, we developed a CNN-based deep learning to evaluate skin texture. The learning was accomplished using preprocessed images that included contrast correction along with an average evaluation score from visual evaluators as training data. As a result, the correlation coefficients between the estimated evaluation score by the CNN and the visual evaluation score were 0.922 for the open set and 0.903 for the closed set. Matching rates of the two scores were 73.2% and 77.8%, respectively. Our results were better than those of the three-layer neural network used in previous studies. We used a CNN network model because the number of data used was relatively small. However, by increasing the data through augmentation, we were able to realize an evaluation system with highly accurate performance.

A limitation of this study was that it excluded images that included body hair. Future research will need to confirm the extent to which body hair in an image affects the system's performance and to establish a processing method for eliminating body hair in images. Furthermore, in the image acquisition, it will be necessary to eliminate the bias in the amount of data obtained from men and women.

Acknowledgement. We thank Edanz Group (https://en-author-services.edanzgroup.com/) for editing a draft of this manuscript.

REFERENCES

[1] K. Sakashita and H. Takahashi, Skin texture analysis by skin groove extraction, *ITE Technical Report*, vol.39, no.14, pp.133-136, 2015.

[2] N. Zhang, Y. X. Cai, Y. Y. Wang, Y. T. Tian, X. L. Wang et al., Skin cancer diagnosis based on optimized convolutional neural network, *Artificial Intelligence in Medicine*, vol.102, DOI: 10.1016/j.artmed.2019.101756, 2020.

[3] M. A. Al-Masni, D. H. Kim and T. S. Kim, Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification, *Computer Methods and Programs in Biomedicine*, vol.190, DOI: 10.1016/j.cmpb.2020.105351, 2020.

[4] J. K. Winkler, K. Sies, C. Fink, F. Toberer, A. Enk et al., Melanoma recognition by a deep learning convolutional neural network-performance in different melanoma subtypes and localisations, *European Journal of Cancer*, vol.127, pp.21-29, 2020.

[5] D. Rathi, S. Jain and S. Indu, Underwater fish species classification using convolutional neural network and deep learning, *Proc. of the 9th International Conference on Advances in Pattern Recognition (ICAPR)*, Bangalore, pp.1-6, 2017.

[6] P. Hridayami, I K. G. D. Putra and K. S. Wibawa, Fish species recognition using VGG16 deep convolutional neural network, *Journal of Computing Science and Engineering*, vol.13, no.3, pp.124-130, 2019.

[7] H. Kobayashi, T. Hashimoto, K. Yamazaki and Y. Hirai, Proposal of quantitative index of skin texture by the image processing and its practical application, *Transactions of the Japan Society of Mechanical Engineers Series C*, vol.76, no.9, pp.922-929, 2010.

[8] Y. Takemae, H. Saito and S. Ozawa, The evaluating system of human skin surface condition by image processing, *Transactions of the Society of Instrument and Control Engineers*, vol.37, no.11, pp.1097-1103, 2001.

[9] A. Chowanda and R. Sutoyo, Convolutional neural network for face recognition in mobile phones, *ICIC Express Letters*, vol.13, no.7, pp.569-574, 2019.

[10] T. Cogan, M. Cogan and L. Tamil, MAPGI: Accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning, *Computers in Biology and Medicine*, vol.111, DOI: 10.1016/j.compbiomed.2019.103351, 2019.

[11] P. Liskowski and K. Krawiec, Segmenting retinal blood vessels with deep neural networks, *IEEE Trans. Medical Imaging*, vol.35, no.11, pp.2369-2380, 2016.

[12] S. Rahrovan, F. Fanian, P. Mehryan, P. Humbert and A. Firooz, Male versus female skin: What dermatologists and cosmeticians should know, *International Journal of Women's Dermatology*, vol.4, pp.122-130, 2018.