



Evaluation of Kampo disease states using facial images

Futa Matsushita¹ · Kaoru Kiyomitsu² · Keiko Ogawa³ · Norimichi Tsumura²

Received: 29 March 2018 / Accepted: 8 June 2018 / Published online: 22 September 2018
© The Author(s) 2018

Abstract

In the traditional Japanese (Kampo) medicine, diagnosis is based on the observation of facial color and expression. Disease states that are diagnosable from facial characteristics include blood stagnation, blood deficiency, and yin deficiency. However, these diagnoses are subjective and require lots of experience. In this paper, we aimed to quantify facial diagnosis by creating a system to evaluate the severity of disease states using feature values obtained from facial images. A Kampo physician evaluated the facial images and rated them from 1 to 5 according to severity of disease states. We verified the accuracy of this system using the mean squared error calculated from the difference between the physician evaluation scores and the system estimates. The mean squared error was close to zero, indicating that the system has high accuracy. The selection of feature values using this system corresponded with those facial regions used by the Kampo physician in diagnosing the disease states.

Keywords Facial image · Kampo medicine · Diagnosis · Disease state

1 Introduction

In most Western countries, the standard medical approach consists of Western medicine, which is based on the assumption that disease is caused by an underlying pathology that must be removed for health to be re-established. An alternative medical system has developed in Asia. Oriental medicine is much older than biomedicine and is based on the assumption that the “heart (mind)” is inseparable from the “body” and that disease can be cured by correcting physical imbalances. Japanese Kampo medicine has developed independently of other types of Oriental medicine. Physicians with a national qualification in Japan can prescribe Kampo medicine and combine both Kampo and Western treatments [1].

Kampo medicine uses four diagnostic methods: visual observation, auscultation and olfaction examinations, inquiry, and palpation. Diagnosis from visual observation

is based on the observation of aspects such as face color, skin condition, and tongue color. Auscultation and olfaction examinations involve listening to sounds emitted from the body and smelling body odor. Inquiry comprises asking the patient questions. Palpation is conducted by touching the body.

Visual observation is an important diagnostic method in Kampo medicine. Facial diagnosis is based on the observation of skin color and skin condition. For example, a patient with reddish skin color is likely to have heart disease, because the redness indicates expanded capillaries and lack of blood circulation.

Kampo disease states diagnosable by facial observation include “blood stagnation,” “blood deficiency,” and “yin deficiency.” Blood stagnation is caused by poor blood circulation. This disease state is indicated by facial reddening caused by pigmentation. Blood deficiency is caused by lack of blood. This disease state is indicated by a pale white skin color and dry skin, which is caused by reduced blood production and by blood stagnation. Yin deficiency is indicated by facial skin that lacks moisture. This disease state can be diagnosed from the luster and elasticity of the face.

There are two problems in facial diagnosis of these disease states. First, these diagnostic methods require much experience and are based on subjective judgment rather than quantitative measures. Second, the Oriental medical approach involves follow-up of disease states, but this

✉ Futa Matsushita
f.matsushita1231@chiba-u.jp

¹ Department of Information and Imaging System, Chiba University, Chiba, Japan

² Graduate School of Advanced Integration Science, Chiba University, Chiba, Japan

³ Department of Japanese-Traditional (Kampo) Medicine, Kanazawa University Hospital, Kanazawa, Japan

requires access to the same physician who made the original Kampo diagnosis. Patients must go to hospital to access the same physician for the follow-up.

To solve these problems, we aimed to quantify diagnosis by capturing facial images using a general RGB camera and extracting feature values from the facial images. We asked a Kampo physician to evaluate the severity of disease states from facial images. We then created an evaluation system using quantified feature values obtained from facial images and the physician evaluation scores. We assessed the performance of the evaluation system by verifying the system’s precision.

Finally, we used feature selection to investigate which feature values were most effective in the evaluation of disease states.

2 Generation of facial image dataset

In this section, we describe the method of generating the facial image data set. A large number of facial images were needed to create the evaluation system. However, our research indicated that the number of captured facial images was insufficient to construct the system using machine learning algorithms. Therefore, we increased the number of facial images using the methods described in this section.

As mentioned in Sect. 1, blood stagnation and blood deficiency are related to problems with blood production and flow, and yin deficiency is indicated by facial luster and elasticity. We assumed that the state of the blood is indicated by hemoglobin pigmentation and that facial luster and elasticity are related to facial skin gloss (i.e., the lightness of the face). Based on these assumptions, we increased the number of facial images by changing hemoglobin pigmentation and gloss connected with three disease states.

2.1 Changes in hemoglobin pigmentation of facial images

We extracted hemoglobin pigments from facial images using the independent component analysis method proposed by Tsumura et al. [2]. We generated several facial images by changing the hemoglobin pigments. Figure 1 shows an overview of this method. The v^{log} which is density distribution signal is obtained by converting the RGB color space to the logarithmic space. The v^{log} is signified by the weighted linear combination of the three vectors σ_m , σ_h , and I with the bias vector e^{log} . The relations among the observed signal and three independent signals are shown in Fig. 1. The hemoglobin and melanin pigmentation densities are obtained by reprojecting onto the σ_m and σ_h pigmentation density vectors after projecting the v^{log} onto the skin color plane in parallel with the I shading vector. To change the hemoglobin

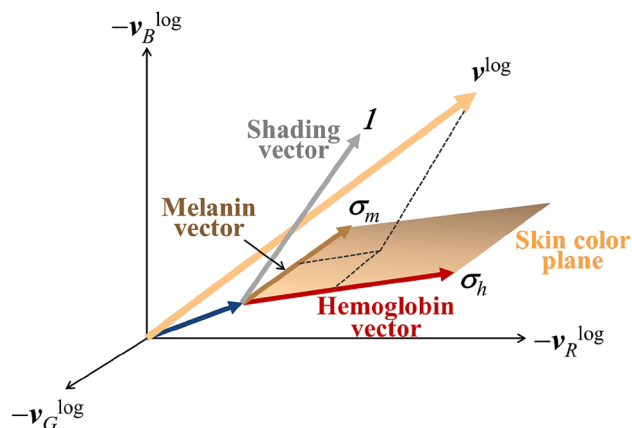


Fig. 1 Diagram of independent component analysis

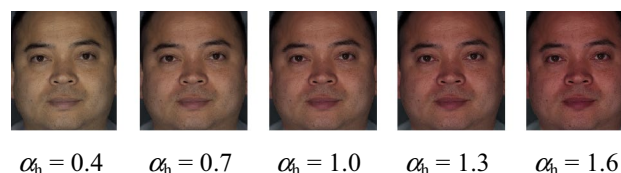


Fig. 2 Facial images produced by changing the hemoglobin component

component, we separated the facial image into hemoglobin, melanin, and shading components. We changed the hemoglobin components multiplied by the coefficient α_h ($\alpha_h = 0.4, 0.7, 1.0, 1.3, 1.6$) and reconstructed the facial images using the changed hemoglobin component and the other components. Figure 2 shows facial images with various amounts of hemoglobin. The facial image with $\alpha_h = 1.0$ represents the original facial image.

2.2 Changing the gloss of facial images

We used the extraction of surface reflection from facial images [3]. We generated several facial images by changing the gloss. The surface reflection represents the facial gloss and is acquired by subtracting the facial image including only internal reflection from the facial image including both surface and internal reflections. The surface reflection retains the same polarized nature as the incident light. The internal reflection becomes unpolarized by scattering within the skin layers. Figure 3 shows the method of extraction of surface reflection using polarization. A facial image is captured by attaching the polarizing plates in front of the camera and light source. The incident light passes through the p -polarizing plate, which lets only p -polarized light pass. The surface reflection remains polarized, whereas the internal reflection becomes unpolarized. Therefore, both the

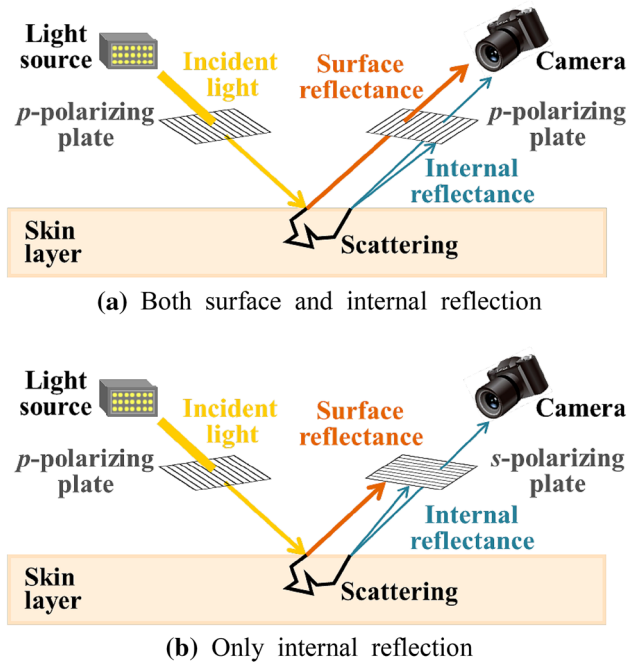


Fig. 3 Diagram of the extraction of surface reflection from skin

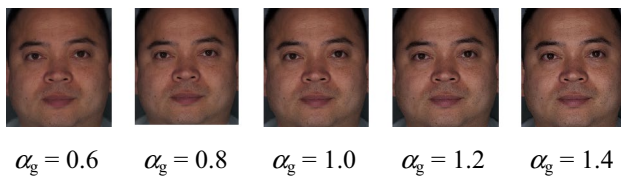


Fig. 4 Facial images produced by changing the gloss component

surface and internal reflection p -polarized light pass through the p -polarizing plate shown in Fig. 3a, and only the s -polarized internal reflection light passes through the s -polarizing plate shown in Fig. 3b. We obtained a gloss-only image using this method. We used a gloss component multiplied by the coefficient α_g ($\alpha_g = 0.6, 0.8, 1.0, 1.2, 1.4$) to synthesize facial images with different gloss. This indicated adding an image for which only the gloss had changed to a facial image that included only internal reflection. Figure 4 shows facial images with various levels of gloss. The facial image with $\alpha_h = 1.0$ represents the original facial image including both surface and internal reflections.

2.3 Generation of facial image data set by changing hemoglobin and gloss components

We generated the facial images by synthesizing a facial image for which only the gloss had changed with a facial image for which the hemoglobin component had changed.

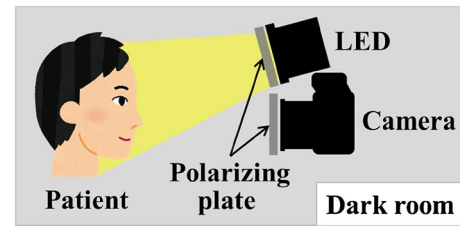


Fig. 5 Setting used to photograph facial images

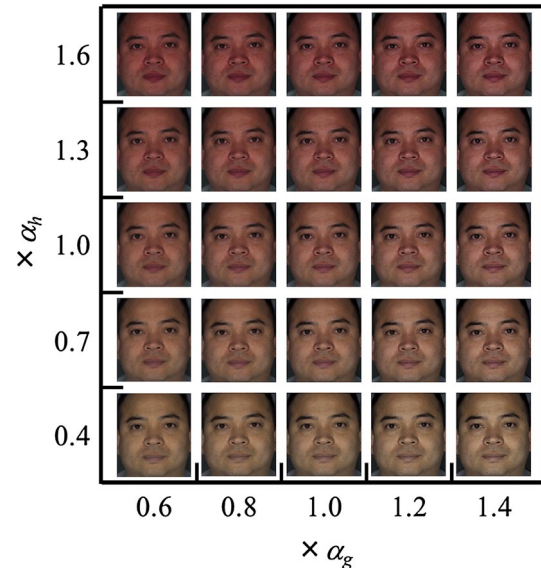


Fig. 6 Five-step construction of facial images by changing hemoglobin and gloss

Figure 5 shows the setting used to capture the facial images. The facial images were taken using an RGB camera (Nikon D5100) with LED light in a dark room. The polarizing plates were positioned in front of the camera and the light. Using the method described in Sect. 2.2, we obtained facial images with surface reflection and facial images without surface reflection. We captured the facial images of seven patients in Kanazawa University Hospital: six females aged in their 20s to their 40s and a male in his 50s. We generated 25 facial images by changing the hemoglobin in five steps ($\alpha_h = 0.4, 0.7, 1.0, 1.3, 1.6$) and changing the gloss in five steps ($\alpha_g = 0.6, 0.8, 1.0, 1.2, 1.4$) (Fig. 6). These ranges and intervals were determined empirically enough to discriminate changes in hemoglobin and gloss. We asked the Kampo physician whether these facial images are natural. As a result, these facial images judged as natural images. In this way, we constructed a data set of 175 facial images for seven patients by changing each component.

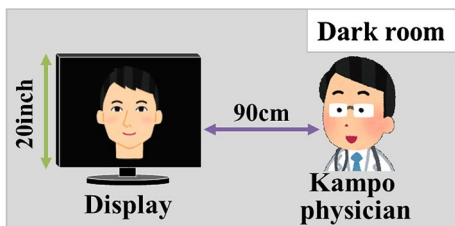


Fig. 7 Kampo diagnosis evaluation setting evaluations of disease states from facial images by a Kampo physician and extracted quantified feature values from facial images

3 Preparation for constructing the evaluation system

To create the evaluation system, we obtained.

3.1 Evaluation of disease states by Kampo Physician

We asked the Kampo physician to evaluate disease states from the facial images. Figure 7 shows the experimental setting in which the physician diagnosed the disease states. The diagnosis was conducted in a dark room, because the result of evaluation scores would change due to surrounding ambient light. A 20-inch display was used with a viewing distance of approximately 90 cm, which was three times the height of the display. The physician used an absolute evaluation method to diagnose three disease states: blood stagnation, blood deficiency, and yin deficiency for each facial image. The resultant scores represented the severity of disease states, from 1 (no disease state) to 5 (severe disease state) with increments of 1. The evaluation time was unlimited. The evaluation interval was 5 s, because a black screen was used to ensure a neutral interval before the next evaluation.

Figure 8 shows the results of the evaluation of disease states from facial images. Figure 8a, c, e shows the average evaluation scores according to changes in hemoglobin pigmentation. Figure 8b, d, f shows the average evaluation scores according to changes in gloss. The blood stagnation scores tended to increase as the skin color became more red from an increase in hemoglobin pigmentation (Fig. 8a), whereas blood stagnation scores remained constant as gloss changed (Fig. 8b). Blood deficiency scores increased as the skin color became pale white by decreasing hemoglobin pigmentation (Fig. 8c), and slightly increased as the skin became dry as gloss decreased (Fig. 8d). The yin deficiency scores increased with the reduction in facial luster and elasticity from decreasing gloss (Fig. 8f), whereas yin deficiency scores remained constant as hemoglobin changed (Fig. 8e). These results are in accord with the theory underlying Kampo diagnosis (described in the Sect. 1).

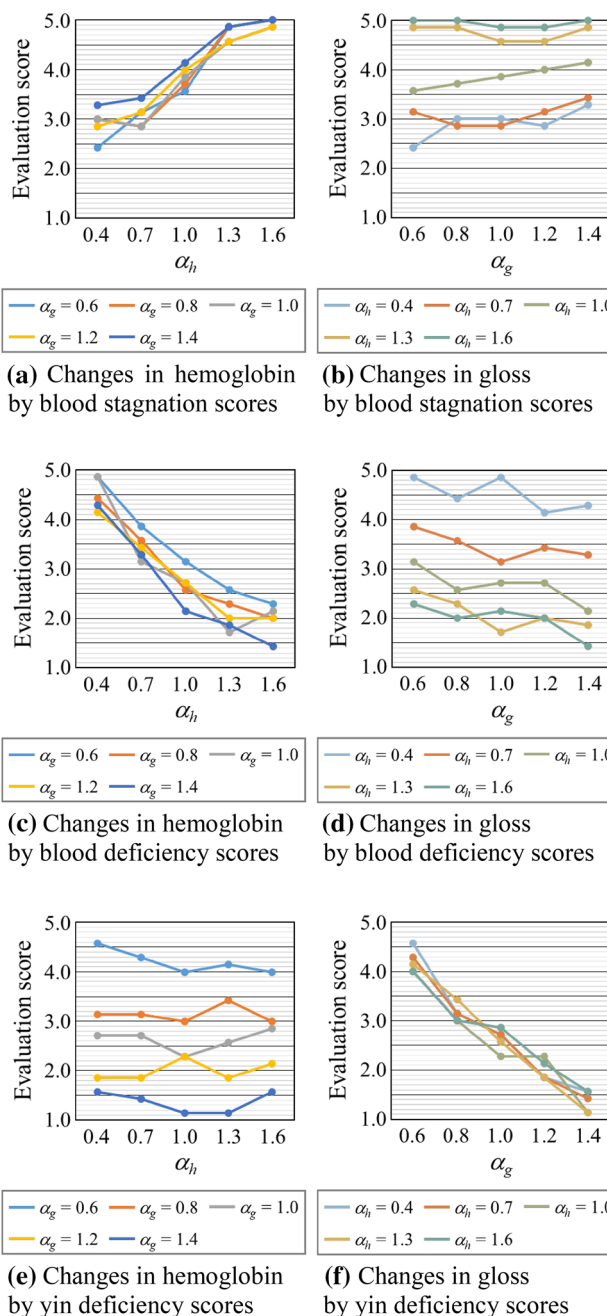


Fig. 8 Changes in Kampo physician evaluation scores according to facial image changes

3.2 Extraction of feature values from facial images

Feature values were extracted from facial images to create the evaluation system. Figure 9 shows a hemoglobin density image and a gloss image obtained using pigmentation component separation and extraction of surface reflection, respectively. We used five regions from these images: the forehead (100 px × 100 px), the areas below the left and right eyes (100 px × 50 px), and the left and right cheeks

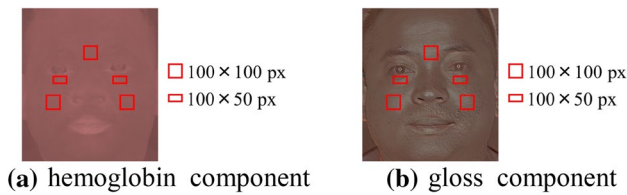


Fig. 9 Extraction of feature values from facial image areas (red rectangles). (Color figure online)

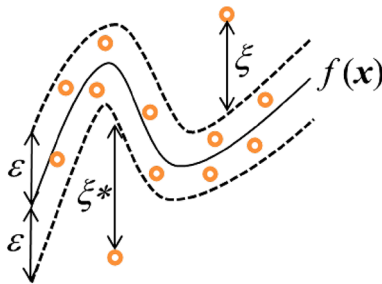


Fig. 10 Diagram showing support vector regression

(100 px × 100 px); we also used the whole facial area (Fig. 9), producing a total of six regions. We set ROI manually and extracted these regions for each facial image. We calculated five measures for these six regions (the average, standard deviation, maximum, minimum, and range) for each hemoglobin density and gloss image. Finally, we used 60 feature values to construct the evaluation system.

4 Estimation of evaluation scores using support vector regression

We created the evaluation system using support vector regression (SVR) and assessed the system using the estimated error.

4.1 Support vector regression

SVR is a method based on support vector machine methods. In recent years, SVR has attracted attention as a tool for classification and regression problems [4]. We describe here the SVR method in some detail. In this study, we used the ϵ -SVR algorithm. An overview of the ϵ -SVR is shown in Fig. 10. In ϵ -SVR, we find a regression function $f(x)$ by permitting the range of ϵ for each training data x . We will first explain linear SVR. x_n is a multivariate data set of N observations containing the corresponding response variable y_n . We define the linear regression function as follows:

$$f(x) = x^t \beta + b, \quad (1)$$

where $\beta \in R^n$ and b is a real number. To find the function $f(x)$, we flattened this function as much as possible by minimizing norm value ($\beta^t \beta$). In addition, we introduced slack variables ξ_i and ξ_i^* , for each training data. These variables allow the existence of regression errors to the value of ξ_i and ξ_i^* with satisfying the necessary condition. This leads to an optimization problem, as follows:

$$\text{Minimize } J(\beta) = \frac{1}{2} \beta^t \beta + C \sum_{i=1}^n (\xi_i \xi_i^*), \quad (2)$$

$$\text{s.t. } \begin{aligned} (\beta^t x_i + b) - y_i &\leq \epsilon + \xi_i \\ y_i - (\beta^t x_i + b) &\leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned} \quad (3)$$

The constant $C (> 0)$ controls how much observations outside the range of ϵ are tolerated. The optimization problem in Eq. 2 can be easily solved using the Lagrange multiplier method. By finding β using this method, we derive the function $f(x)$ as follows:

$$f(x) = \sum_{i=1}^n (\lambda_i - \lambda_i^*) x_i^t x + b, \quad (4)$$

where λ_n and λ_n^* are Lagrange multipliers for each training data x_n .

We previously considered linear SVR; however, some regression problems cannot be solved using this method. Therefore, we introduced nonlinear SVR model. To expand the linear SVR to the nonlinear SVR, we replaced the observation x with the function $\varphi(x) = (\varphi_1(x), \varphi_2(x), \dots, \varphi_d(x))^t$, where $\varphi_i(x)$ is a nonlinear function that the output is scalar. This allows nonlinear regression to be treated as linear regression. In addition, by introducing kernel function $K(x_i, x_j) = \varphi(x_i)^t \varphi(x_j)$, this function can be obtained directly without calculating φ . Using this method, nonlinear SVR produces the regression function $f(x)$ as follows:

$$f(x) = \sum_{i=1}^n (\lambda_i - \lambda_i^*) K(x_i, x) + b. \quad (5)$$

4.2 Verification of estimate accuracy

We verified the accuracy of the evaluation system. First, 60 feature values were normalized to 0–1, as significant differences in the range of feature values would reduce accuracy. Then, we constructed an SVR model to diagnose disease states using these feature values extracted from facial images and the Kampo doctor's evaluation scores.

We used the 60 feature values as explanatory variables and the evaluated scores as objective variables. In this

research, we used various kernel functions, including linear, Gaussian, 2-dimensional, and 3-dimensional kernels.

We used the mean squared error (MSE) to measure the difference between the SVR-estimated values and the Kampo doctor’s evaluation scores; this produced an index of the predictive accuracy. The MSE approaches 0, indicating high accuracy.

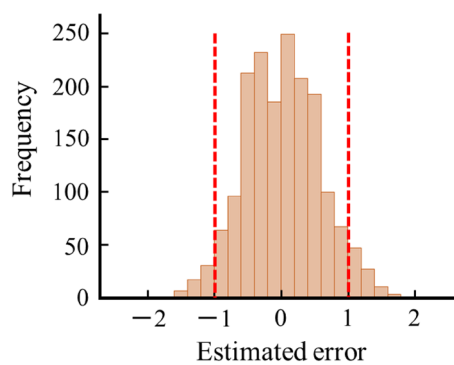
We calculated the accuracy of the evaluation system using *k*-fold cross validation. Table 1 shows the result of averaged *MSE* for each kernel function by calculating the fivefold cross validation ten times. The 2-dimensional kernel produced the best accuracy for blood stagnation and yin deficiency evaluation. The linear kernel produced the best accuracy for blood deficiency.

We obtained the estimated errors by subtracting estimated values from evaluation scores. Figure 11 shows histograms of estimated errors for each disease state.

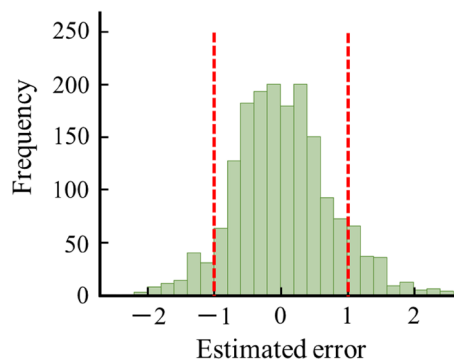
4.3 Feature selection

We selected feature values as follows. If there are more feature values than estimated classes or samples, accuracy can be increased by removing feature values that represent weak correlations. Therefore, we selected those feature values that would increase the accuracy of the evaluation system.

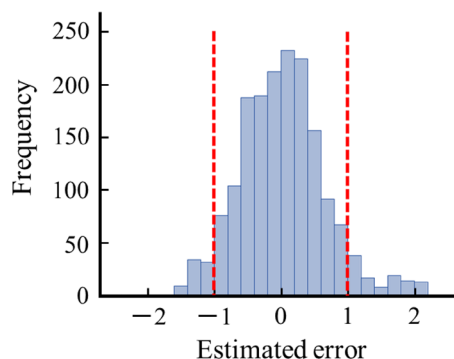
We used a wrapper method to select the features [5]. In this method, the model is practically constructed using a subset of feature values, and the optimal combination of feature values is chosen by verifying accuracy with this model. The wrapper method requires many calculations, but is more accurate than other methods. We used an individual optimization approach based on a wrapper method to select feature values. This method evaluates the feature values individually and selects them based on this evaluation. As shown in Fig. 12, we evaluated the feature values individually by removing each value one by one. Removal of feature values that have a low MSE had a large influence on the evaluation system. We ranked feature values using this method. The top five feature values selected from 60 feature values for each disease state are shown in Tables 2, 3, and 4. The MSE using the top 15 selected feature values are shown in Table 5. The



(a) Blood stagnation



(b) Blood deficiency



(c) Yin deficiency

Fig. 11 Histograms of estimated error

Table 1 Mean squared errors from various kernel functions

Kernel	Disease states		
	Blood stag-nation	Blood defi-ciency	Yin deficiency
Linear	0.34	0.52	0.44
Gaussian	0.42	0.66	0.62
2-Dimensional	0.32	0.52	0.42
3-Dimensional	0.37	0.66	0.48

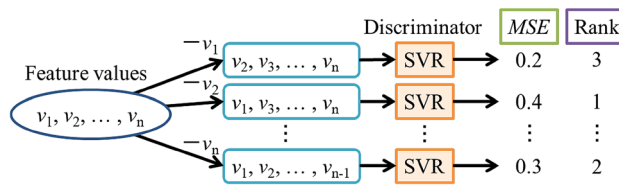


Fig. 12 Individual optimization method

Table 2 Ranked feature values for blood stagnation

Rank	Hemo/gloss	Value	Region
1	Hemoglobin	Range	Forehead
2	Hemoglobin	Standard deviation	Left eye below
3	Hemoglobin	Range	Right cheek
4	Hemoglobin	Range	Right eye below
5	Hemoglobin	Minimum	All

Table 3 Ranked feature values for blood deficiency

Rank	Hemo/gloss	Value	Region
1	Hemoglobin	Range	Forehead
2	Gloss	Range	Left cheek
3	Hemoglobin	Minimum	Forehead
4	Hemoglobin	Average	Left cheek
5	Hemoglobin	Average	Right cheek

Table 4 Ranked feature values for yin deficiency

Rank	Hemo/gloss	Value	Region
1	Gloss	Standard deviation	Right eye below
2	Gloss	Range	Right eye below
3	Gloss	Average	Right cheek
4	Gloss	Range	Right cheek
5	Gloss	Maximum	Right eye below

Table 5 Mean squared errors of the top 15 selected feature values

	Disease state		
	Blood stag- nation	Blood deficiency	Yin deficiency
Before feature selection	0.32	0.52	0.42
After feature selection	0.31	0.51	0.40

resultant accuracy after feature selection was slightly better than the original accuracy prior to feature selection.

5 Discussion

In this study, we constructed an evaluation system reflecting the diagnosis of one physician. It is significant to quantify the diagnosis of a physician who has many knowledge and experience on Kampo medicine.

In the conventional Kampo diagnosis, Kampo physicians subjectively diagnose patients based on experience. Therefore, the results of diagnosis differ depending on Kampo physician. In addition, differences between

patients are not clear, because Kampo diagnosis has not been quantified in the past.

Kampo physician said that physicians judge the disease state based on various information such as walking, tone, and size of voice in case of face-to-face observation. On the other hand, it is diagnosed only by observing the region where the complexion or gloss appears on electrical display. These disease states can be diagnosed through electronic display depending on method of acquiring information. We assume that more accurate diagnosis can be done on electronic display by correcting facial images captured in various environment.

Table 1 shows that the MSE for blood stagnation (0.32) and yin deficiency (0.42) were the most accurate, and these were calculated using the 2-dimensional kernel function. The MSE of blood deficiency (0.52) was the most accurate, and was calculated using the linear and 2-dimensional kernel functions. Therefore, the most accurate estimates of all three disease states were produced with the 2-dimensional kernel function. In addition, we calculate the standard deviations of each disease states in case of 2-dimensional kernel function. The standard deviations are 0.56 (blood stagnation), 0.71 (blood deficiency), and 0.64 (yin deficiency).

We used an evaluation score range of 1 (no disease state)–5 (severe disease state) with increments of (1) based on this evaluation criteria; the values adjacent are highly relevant. For example, we assume that it is little problem to estimate three when the correct value is (2) we considered that values adjacent to the estimated score represented sufficient accuracy of estimation. As mentioned above, the findings showed that the estimates of the evaluation scores were accurate, because the MSE and the standard deviations were lower than 1.0 and most of the estimated error was within the range of 1.0, as shown in Fig. 11.

We ranked and selected effective feature values using an individual optimization method. We calculated the MSE using the top 15 selected feature values. The resultant accuracy after feature selection was slightly better than the original accuracy prior to feature selection. Hemoglobin was the most important feature for blood stagnation and blood deficiency diagnosis, and gloss was the most important feature for yin deficiency diagnosis. The most effective feature value was the range of hemoglobin in the forehead for blood stagnation and blood deficiency and the standard deviation of gloss below the right eye for yin deficiency. The forehead is an important area for Kampo facial diagnosis of the disease states of blood stagnation and blood deficiency, because the dermis of this region is thin. The range which is difference between the maximum and the minimum affects the appearance of blood. In other words, the larger the difference in hemoglobin amount is, the clearer the appearance of blood becomes.

The facial regions in which gloss is most evident were selected for yin deficiency. We assume that the larger the value variation, the larger the glossy range is. Therefore, the gloss looks outstanding. We concluded that our findings were in accord with the knowledge made by a Kampo medicine.

6 Conclusion and future work

In this study, we estimated the severity of disease states using facial images captured using an RGB camera. We concluded that physician evaluation scores could be accurately estimated by the system, as the MSE was close to zero. We ranked effective feature values using feature selection and the system estimates of disease states were in accord with the perceptions of the Kampo doctor.

We asked the Kampo doctor to evaluate images generated from only seven facial images, and facial images were taken under strictly controlled lighting conditions. In future studies, we plan to increase the number of original facial images of patients to verify how many original faces are needed to create the system, and to create a system that can be used under more common lighting conditions. We would also like to automate the process of capturing facial images to diagnose disease states.

Acknowledgements We thank Diane Williams, Ph.D., from Edanz Group (<http://www.edanzediting.com/ac>) for editing a draft of this manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Sato Y et al (2005) Introduction to Kampo: Japanese traditional medicine. The Japan Society for Oriental Medicine, Tokyo
2. Norimichi Tsumura N, Ojima K, Sato et al (2003) Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin. *ACM Trans Graphics* 22(3):770–779
3. Ojima N, Minami T, Kawai M (1997) Transmittance measurement of cosmetic layer applied on skin by using processing. *Proceeding of The 3rd scientific conference of the Asian Societies of Cosmetic Scientists*, vol 114
4. Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14(3):199–222
5. Leslie KP (2003) An Introduction to variable and feature selection. *J Mach Learning Res* 3:1157–1182