https://doi.org/10.2352/issn.2169-2629.2020.28.35

Using images of the tongue for diagnostic assistance in Kampo Medicine

Reimei Koike[1], Keiko Ogawa-Ochiai[2], Hongyang Li[2] , and Norimichi Tsumura[3]
1) Graduate School of Science and Engineering, Division of Creative Engineering, CHIBA, JAPAN
2) Department of Japanese Traditional (Kampo) Medicine, Kanazawa University Hospital, KANAZAWA, JAPAN
3) Graduate School of Engineering, Chiba University, CHIBA, JAPAN

Abstract

In this research, we propose a method to assess images of the *tongue captured using a polarized light camera for diagnostic use in Kampo Medicine. The polarized light camera is used to simultaneously capture glossy and non-glossy images of the tongue. Data augmentation was performed by modulating the color and gloss, through which the number of images was increased from 11 to 275. A diagnostic assistance module was built to evaluate a given disease by learning a specialist's assessment of it along with feature values obtained from the captured image using a machine learning technique. The resulting mean absolute error of the assessment of five diseases was sufficiently small for it to be accurate.*

Introduction

The general medical care we receive in hospitals worldwide is based on Western medicine. However, in China and other East Asian countries, the field of traditional medicine—called Kampo medicine in Japan—has developed over 1,000 years. Western medicine and Kampo medicine are very different in their treatment methods. In Kampo medicine, the method of understanding the pathology of the organism using the three elements of "qi," "blood," and "fluid" is called Qi, Blood, and Fluid Theory [1]. A healthy state is one in which these three circulate normally and in sufficient quantities: "Qi" represents the energy of the entire body, "blood" represents the blood and its flow through the body, and "fluid" represents water other than the blood flowing through the body, such as lymph and digestive fluids. The conditions in which qi, blood, and fluid are "insufficient" are called "qi deficiency," "blood deficiency," and "yin deficiency," respectively, and those in which in which qi, blood, and fluid are "stagnated" are called "qi stagnation," "blood stagnation," and "fluid stagnation," respectively.

There are four ways for physicians practicing Kampo medicine to diagnose diseases: inspection, inquiry, palpation, and aural and olfactory examinations [1]. Inspection is among the most important diagnostic methods in Kampo medicine that is rarely applied to Western medicine. One reason for this is that it is not possible to make a quantitative diagnosis because inspection is subjective, and is based on the experience of the physician. To solve this problem, a quantitative method of descriptive diagnosis has been formulated using image processing in recent years. By allowing a computer to perform an advisory diagnosis based on images of the face and tongue, it is possible to quantify the diagnosis without prior experience and identify the relevant disease.

To computerize descriptive diagnosis, Matsushita *et al*. proposed a method to assess diseases using facial images of patients [2]. However, among the six conditions mentioned above, only three—"blood deficiency, blood stagnation, yin deficiency" —can be diagnosed using images of the face. However, Kampo medicine can help diagnose five of the six conditions—"qi stagnation," "qi deficiency," "blood deficiency," "blood stagnation," "yin deficiency," and "water stagnation"—by observing the color of the patient's tongue, his/her water content, shape of the tongue, and condition of the tongue coating. Table 1 lists diseases that can be diagnosed from the face and tongue of the patient using Kampo medicine.

Table 1. Diagnostic method for each of the six diseases in Kampo medicine

	Qi	Blood	Fluid
deficiency	**tongue**	**face tongue**	**face tongue**
stagnation	**×**	**face tongue**	**tongue**

Because the tongue is close to the internal organs of the body with exposed mucous membranes, it has been regarded as an important indicator of many diseases in Kampo medicine. However, because it is highly glossy under normal lighting conditions owing to its moist surface, it is difficult to obtain accurate information on the color of the tongue. However, as mentioned above, physicians in Kampo medicine make diagnoses based on observations of gloss, in part, and assessments made without incorporating gloss-related information are inaccurate. To inspect the tongue using a computer, it is necessary to use an imaging method that can accurately acquire color-related information after acquiring gloss-related information.

Nakaguchi *et al*. [3], by combining the integrating sphere and directional illumination, took images of the tongue with and without gloss. However, because the integrating sphere is large and expensive, it is not appropriate for use in personal diagnosis.

In this study, we use a polarizing camera to photograph the tongue. Such a camera can capture glossy and non-glossy images at the same time, and is more compact and cheaper than an integrating sphere. We then assess a patient's diseases using these images. A system of quantitative evaluation is constructed by learning the values and features obtained from images of the tongue through machine learning. The accuracy of the evaluation system is also verified.

Acquiring glossy and non-glossy images using polarized light

Incident light from a source is reflected from the surface of the object. In this case, the reflection of the light can be classified into two types: surface reflection, where the angles of incidence and reflection of light are the same with respect to the surface of reflection according to the laws of reflection, and internal diffuse reflection, which is caused by the scattering of the incident light inside the object. The human eye and a camera can perceive the color and luster of an object by capturing the reflected light. In general, an object appears colored owing to internal diffuse reflection, and luster is caused by surface reflection.

We explain the nature of polarization. Light, which is an electromagnetic wave, oscillates in a direction perpendicular to its direction of travel, and normal light is in a state called unpolarized or partially polarized, which is a mixture of light oscillating in all directions. Polarization, conversely, refers to light that oscillates only in a certain direction. By passing ordinary unpolarized light through an optical filter called a polarizing plate, we can obtain polarized light in a direction identical to that of the polarizing plate. This property of polarization has important links to the surface reflection and internal diffusion described above. It is reflected while the polarization of the incident light is maintained; when the incident light is reflected off the surface of the object, it is linearly polarized in the direction of the polarizing plate, as in case of incident light. Conversely, internal diffuse light is unpolarized even if the incident light is linearly polarized when internally reflected, because the incident light is reflected in various directions and the light is scattered. Using this property, we explain how to obtain glossy and non-glossy images below.

Figure 1 shows how to obtain an image with enhanced gloss and one with reduced gloss using two orthogonal polarizing plates. S indicates the surface-reflected light, D indicates diffuse reflected light, and the subscripts p and s indicate the directions of orthogonal polarization. Because S in Figure 1(a) maintains the same polarization as the incident light, the relation between the intensity of S and that of the p-polarization component S_p of S is as below. I_x indicates the intensity of x:

$$I_S = I_{S_p}. \tag{1}$$

The diffuse component D in Figure 1(a) is the loss of polarization of the incident light. The ratio of intensity of the p-polarization component D_p to that of the s-polarization component D_s in D is 1:1, and can be expressed as follows:

$$I_D = I_{D_p} + I_{D_s} = 2I_{D_p} = 2I_{D_s}. \tag{2}$$

Consider the case where a second polarizing plate is placed between the object and the camera. Figure 1(b) shows this case. When light passes through the p-polarizing plate, the component of s-polarization in the vertical direction cannot pass through the p-polarizing plate. Thus, the internally reflected light D, which contains D_s, has half the intensity but S_p, which has only the p-polarization component, can pass through the p-polarizing plate without attenuation. Therefore, in Figure 1(b), we obtain an image with enhanced gloss. Similarly, Figure 1(c) shows the case where an

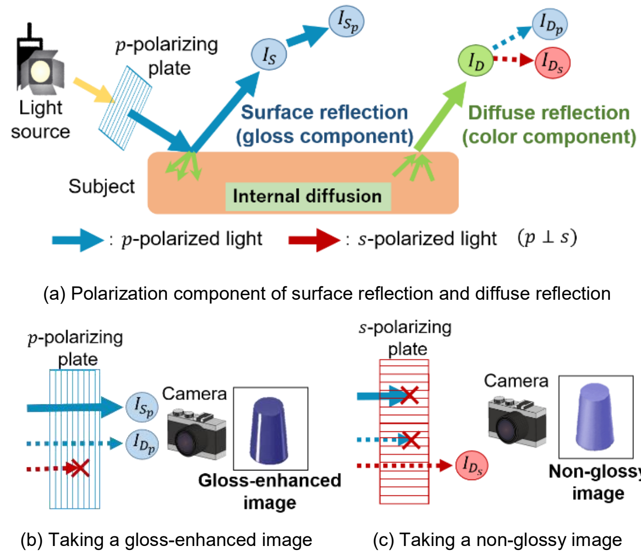


Figure 1. *Acquisition of surface and internal reflection using polarizing plate.*

s-polarizing plate is placed between the camera and the object; because this plate is in a direction perpendicular to the p-polarizing plate, the p-polarization component of the reflected light cannot pass through the s-polarizing plate. Therefore, in Figure 1(c), the image with the gloss removed is obtained. The gloss component can be obtained by subtracting the non-glossy image from the gloss-enhanced image.

A polarized camera (Lucid, PHX050S-Q) was used for photographic work in this study. A polarization camera can capture four color polarization images at polarization angles of 0°, 45°, 90°, and 135° in a single shot; a 0° polarizing plate was placed between a source of illumination and the object. The polarization camera acquired a gloss-enhanced image from a 0° polarization image and a non-glossy image from a 90° polarization image simultaneously. The polarization camera does not perform color calibration such as white balance adjustment.

Data augmentation by color and gloss modulation

In this section, we describe a method of data enhancement by applying color and gloss modulation to a captured face image. The purpose of this study is to evaluate diseases using machine learning, but the number of images of the tongue captured was small and insufficient for learning. Therefore, we performed physiologically corrected modulation for each sample to produce more training data.

We first explain the color modulation technique. The tongue is the exposed biological mucosa, and its color is determined by the amount of hemoglobin in the blood. Therefore, it is possible to create images with different colors of the tongue by modulating the hemoglobin components in them. To extract components of hemoglobin from non-glossy tongue images, we use the method of separating the components of pigmentation proposed by Tsumura *et al*. [4]. The color-modulated images of the tongue can be obtained by multiplying the hemoglobin component by the color modulation factor α_i in log space, and re-synthesizing the components of melanin and shading by the reverse processing of the separation of the pigment component. Figures 2(a)–(e) show color-modulated images with $[\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5]$= [0.6,0.8,1.0,1.2,1.4], respectively.

We now explain the method of gloss modulation. Both glossy and non-glossy images are captured simultaneously. The glossy component can thus be easily extracted by subtracting the non-glossy image from the glossy image. By multiplying the extracted components of gloss with the gloss modulation factor β_j, as well as color modulation, and re-synthesizing the image with no gloss, the image of the tongue with modulated gloss can be obtained. Figure 3(a)–(e) show color-modulated images for $[\beta_1, \beta_2, \beta_3, \beta_4, \beta_5] = [0.0, 0.3, 0.6, 1.0, 1.5]$, respectively. As shown in Figure 4, by combining these two modulation methods, the number of original images of the tongue was augmented by a factor of 25.

Separating tongue coating from tongue body

The surface of the tongue can be categorized into coating and body. The coating of the tongue refers to the part coated with food scum, mucosal cells, and bacteria on tissues called filamentous papillae, and in the gaps between protrusions on the surface of the tongue. The body of the tongue refers to the part that is not covered by the coating of the tongue [5]. The lichen and body of the tongue have trends of colors of white, yellow, gray, and black, and light-red, red, crimson, and purple, respectively [6]. Because the lichen of the tongue and its body show different changes in color, it is necessary to distinguish between them when evaluating tongue color.

Jiatuo *et al*. focused on the fact that the tongue has a large proportion of red and the R component has a protruding tendency in terms of color, whereas the lichen of the tongue does not have a protruding R component when considered in RGB [7]. The method used by them is shown below. I for all pixels in the tongue image is calculated by the following equation:

$$I = R - \frac{G+B}{2}. \quad (3)$$

All pixels are classified as belonging to the body or coating of the tongue according to Eq. (4). However, let I_{body}, be the mean value of I classified as belonging to the body of the tongue, $Mean(I_{body})$ be that of I classified as belonging to the tongue body, and $Mean(I_{coating})$ be the mean value of I classified as belonging to the tongue coating. I_m is determined to maximize Eq. (5):

$$I = \begin{cases} I_{body} & (\text{if } I \geq I_m) \\ I_{coating} & (\text{if } I < I_m) \end{cases}, \quad (4)$$

$$Maximize(I_m): \left|Mean(I_{body}) - Mean(I_{coating})\right|. \quad (5)$$

Yellow tongue coating can be incorrectly classified owing to its high value of I, although the above methods can correctly distinguish most tongues. The coating of the tongue was distinguished here from its body by changing Eq. (4) as below. As shown in Eq. (6), by increasing the ratio of G to reduce I, and by decreasing the ratio of B to subtract I, we can classify yellow coating correctly while maintaining the ability to classify other colors.

$$I = R - \frac{1.5G+0.5B}{2}. \quad (6)$$

(a)$\alpha_1 = 0.6$ (b)$\alpha_2 = 0.8$ (c)$\alpha_3 = 1.0$ (d)$\alpha_4 = 1.2$ (e)$\alpha_5 = 1.4$

Figure 2. *Five levels of color modulation using modulation coefficient* α_i

(a)$\beta_1 = 0.0$ (b)$\beta_2 = 0.3$ (c)$\beta_3 = 0.6$ (d)$\beta_4 = 1.0$ (e)$\beta_5 = 1.5$

Figure 3. *Five levels of gloss modulation using modulation coefficient* β_i

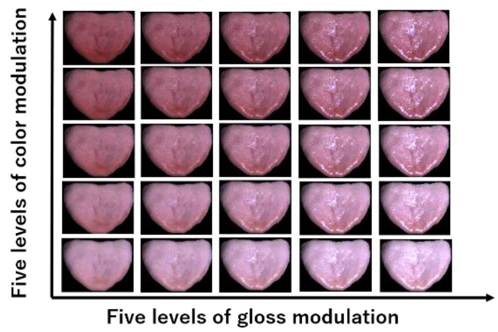
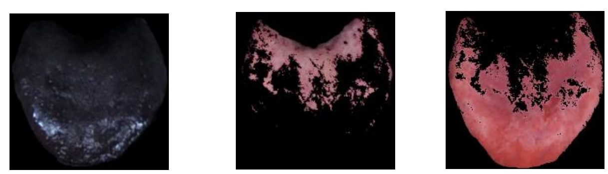


Figure 4. *Data augmentation by color and gloss modulation.*



(a) Glossy image **(b) Tongue coating** **(c) Tongue body**

Figure 5. *Separation tongue into three components,*
(a) Glossy image, (b) tongue coating, (c)tongue body.

Extracting feature values from tongue images

In this section, we explain how to extract feature values from images of the tongue for machine learning. As mentioned above, from the two images with and without gloss, three images can be obtained—(a) the gloss component, (b) image of tongue coating, and (c) image of tongue body—as shown in Figure 5. From (a), we obtained the mean, standard deviation, maximum, minimum, and median of vales of L^* in $L^*a^*b^*$ color space of the glossy area. From (b) and (c), we obtained the mean, standard deviation, maximum, minimum, and median of a^* and b^* in the $L^*a^*b^*$ color space of the area of tongue coating. We thus obtained a total of 27 color and gloss features.

Environment for capturing tongue images

Figure 6 shows the photographic environment in which the learning data were collected in this study. A polarizing camera (Lucid, PHX050S-Q) was used as photographic equipment. The resolution per polarized image was 1224×1024 [px] and the bit depth was 16 [bits]. The exposure time was 8000 [µs] and the distance between the camera and the subject was 73 cm. The images were taken in a dark room and the light source was an LED (Neewer, PT-176S) with a polarizing plate. We photographed 11 healthy subjects for the experiment.

Subjective evaluation of diseases by Kampo doctor using tongue images

In this section, we describe how a Kampo doctor evaluated each tongue image to diagnose the subjects. The evaluation was performed in a darkened room, with 80 [cm] between the display and the evaluator, and a display size of 20.1 [in]. The screen was displayed in device-independent sRGB. The evaluator assessed the five conditions that can be diagnosed from images of the tongue—qi deficiency, blood deficiency, blood stagnation, yin deficiency, and fluid stagnation—in five levels of 0, 1, 2, 3, and 4. The smaller the value was, the more severe the condition was. The evaluator was a doctor specializing in Chinese herbal medicine. Note that in this study, the data were augmented by modulation, which produced some images of tongues that were unrepresentative. Therefore, tongues judged to be thus, based on the experience of the physician, were excluded from the evaluation.

The display showed images of the tongue randomly extracted from a set of unexamined images. The rater evaluate the images, and the display turned dark when the evaluation ended. Four seconds later, the randomly selected images were displayed again. This was repeated until the entire image of the tongue had been assessed. To prevent a decline in evaluation capability due to fatigue, breaks for the Kampo doctor were allowed as appropriate. Interruptions in the experiment were timed to complete the evaluation of all five conditions for the immediately preceding sample, and the display was darkened during the interruption.

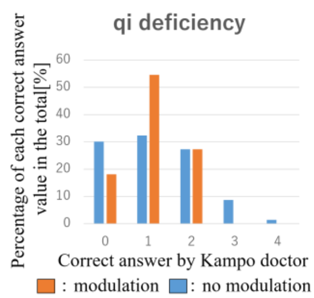
Figure 7 shows the percentage of correct answers by the specialist when data modulation was not performed for each condition, and the percentage when it was performed. We refer to the specialist's evaluation as the correct value here. Because the data had been obtained mainly from healthy subjects, the percentage of correct answers was biased before data modulation, but this bias decreased after modulation for many conditions.

We used a dataset of 275 pairs obtained by modulating the images of the tongues of 11 subjects for machine learning, and to assess its accuracy.

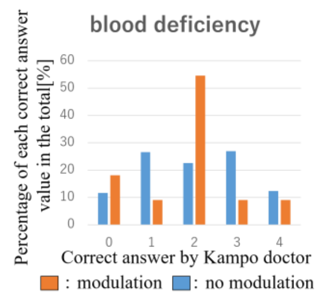
Feature Selection

Support vector regression (SVR) [8] was used for learning. In machine learning, training with ineffective features may reduce the accuracy of estimation compared with training using only the optimal features. This is because generalization performance suffers from over-learning, and the selection of optimal features plays an important role in improving the diagnostic accuracy of machine learning. We used a feature selection method called step forward selection (SFS) to choose the best combination of features to evaluate each disease. In the SFS, features are selected by the following procedure:

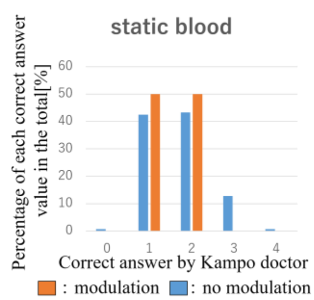
Figure 6. *Experimental environment.*



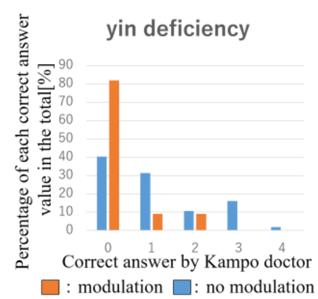
(a) Qi deficiency



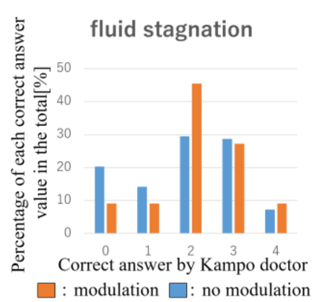
(b) Blood deficiency



(c) Blood stagnation



(d) Yin deficiency



(e) Fluid stagnation

Figure 7. *For each disease, the percentage of correct answers given by the Kampo doctor: no modulation (blue bar); with modulation (orange bar).*

It learned and evaluated the images one by one independently using all features, and chose the ones with the highest evaluation accuracy. It then added one of the remaining features to the selected feature, and learned and evaluated it. This process was repeated for all remaining features. It continued adding features until the evaluation accuracy no longer rose.

The parameters of the SVR used for training were C=100, γ=0.1, and ε=0.1. Features that were determined to be optimal for the evaluation of each disease are shown in Figure 8. Table 2 shows the mean absolute error (MAE) between the values of the correct answers by the Kampo doctor and the estimated values obtained by SVR before and after feature selection. It is evident that evaluation accuracy increased when feature selection was used. Thus, the feature selection was appropriate .

Verifying Accuracy of the Evaluation System

Table 3 shows the MAE of the correct and estimated values before and after data modulation. Each disease was diagnosed using features selected through the SFS and the parameters were optimized by grid search. From the table, it is clear that data modulation improved the accuracy of diagnosis of all diseases.

Discussion

The correct answers were assigned a value from 0 to 4, with adjacent values having close meanings. Therefore, this method yielded good diagnostic performance when the average error was within one.

The features chosen by feature selection were closely related to the aspects of evaluation attended to by the Kampo doctor. Qi deficiency appeared as glossy, colored, and shaped; blood deficiency appeared as colored and shaped; blood stagnation was mainly manifested in terms of its color; yin deficiency was mainly manifested as glossy; and fluid stagnation was mainly reflected in shape. It is clear from this that features that were close to the aspects attended to by the Kampo doctor had been selected by the proposed method. However, we did not use the features of form, for example, size and tooth marks, which means that a small number of water stagnation, which is diagnosed only from form, could not determine the dominant features, and many features were selected.

This was also evident in the final diagnostic results. The diagnostic accuracies for stagnation and hemorrhage, where shape is important, were worse than those for the other diseases. These results show that the addition of new features related to shape can improve diagnostic accuracy.

We did not use many subjects in this study, and increased the number of data items by modulating the data. However, the distribution of data for blood stagnation was biased, as shown in Figure 7. This might have yielded the high accuracy of evaluation of blood stagnation. More scattered data need to be collected.

Conclusion

This study developed an evaluation system that considers the gloss and color components of images of the tongue taken by a polarizing camera to diagnose diseases according to the principles of Kampo. The results of this study are as below.

The number of samples was increased to 275 by modulating color and gloss, as only 11 original images were used in this study. A total of 27 gloss- and color-related features were obtained

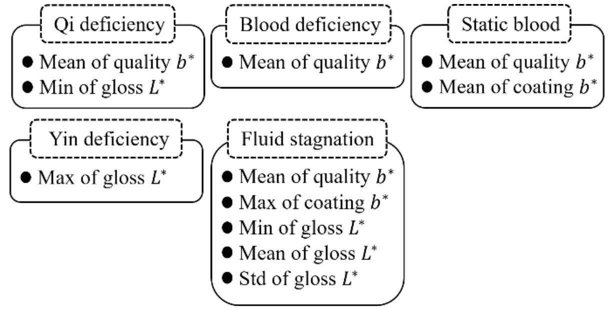


***Figure 8.** For each disease, features selected by SFS.*

Table 2. In each disease, MAE between the correct answer value by the Kampo doctor and the estimated value by SVR before and after the feature selection.

		Before feature selection	After feature selection by SFS
Diseases	Qi deficiency	**0.84**	**0.64**
	Blood deficiency	**1.00**	**0.78**
	Blood stagnation	**0.62**	**0.48**
	Yin deficiency	**0.91**	**0.56**
	Fluid stagnation	**1.01**	**0.88**

Table 3. In each disease, MAE between the correct answer value by the Kampo doctor and the estimated value by SVR with no modulation and modulation

		No modulation	Modulation
Diseases	Qi deficiency	**0.69**	**0.63**
	Blood deficiency	**1.00**	**0.78**
	Blood stagnation	**0.52**	**0.43**
	Yin deficiency	**0.93**	**0.56**
	Fluid stagnation	**0.92**	**0.87**

and combined with subjective evaluation by a Kampo physician to generate the training data. The MAE of SVR was 0.63 for qi deficiency 0.56 for yin deficiency, 0.87 for water stagnation, 0.78 for blood deficiency, and 0.43 for blood stagnation when its parameters were optimized by grid search. The proposed method was accurate because the error in its results was smaller than one because the neighboring assessment values had very close meanings.

The results of feature selection were consistent with the evaluation scores assigned by the Kampo doctor, indicating their correctness. Because the diagnosis of water stagnation required a large number of features, no useful set of features was available for it compared with the other conditions. Adding the shape of the tongue, which was not used here, as a feature can improve the accuracy of the diagnosis.

In future work, more images of the tongue should be collected and re-evaluated because some of the pathological conditions were biased, as described above. The results of a single diagnosis made by one specialist were used here as the correct answer, and conflicts in assessment values between specialists and diagnostic errors were not considered. By taking the mean of multiple evaluations by multiple specialists as the correct answer, we can create a more general evaluator with less subjectivity.

Acknowledgment

We thank Saad Anis, PhD, from Edanz Group (https://en-author-services.edanzgroup.com/) for editing a draft of this manuscript.

References

[1] Y. Sato *et al*., “Introduction to Kampo: Japanese Traditional Medicine,” The Japan Society for Oriental Medicine, 132(5), 260-264(2008)

[2] F. Matsushita, *et al.,* “System for Evaluating Pathophysiology using Facial Image,” Color and Imaging Conference, 25, 274-279 (6), (2017).

[3] T. Nakaguchi, *et al.*, “Proposal for a New Noncontact Method for Measuring Tongue Moisture to Assist in Tongue Diagnosis and Development of the Tongue Image Analyzing System, Which Can Separately Record the Gloss Components of the Tongue,” BioMed Research International*,* Article ID 249609, 10 pages (2015).

[4] N. Tsumura, *et al*., “Image-based Skin Color and Quality Analysis/Synthesis by Extracting Hemoglobin and Melanin Information in the Skin,” ACM Transactions on Graphics, 22(3), 770-779 (2003).

[5] M. M Danser, *et al.*, “Tongue Coating and Tongue Brushing: A Literature Review," International Journal of Dental Hygiene, 1(3), 151-158 (2003).

[6] J. Zhang *et al*., "Diagnostic Method of Diabetes Based on Support Vector Machine and Tongue Images," BioMed Research International, 2017, Article ID 7961494, 9 pages (2017).

[7] J. Xu, *et al*., “The Region Partition of Quality and Coating for Tongue Image based on Color Image Segmentation Method,” 2008 IEEE International Symposium on IT in Medicine and Education, Xiamen, 2008, pp. 817-821 (2008)

[8] C. J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, 2, 121-167 (1998).