Original article

Extraction of tongue coating area from tongue image for automated tongue diagnosis

Gou Nambu (School of Science and Technology, Keio University),
Takao Namiki (Graduate School of Medicine, Chiba University),
Toshiya Nakaguchi (Center for Frontier Medical Engineering, Chiba University),
*Toshiyuki Tanaka (School of Science and Technology, Keio University)

Abstract

Purpose: To automate tongue diagnosis, we propose a method utilizing machine learning to extract the tongue coating area from tongue images captured using a tongue image analysis system (TIAS).

Subjects and Methods: Tongue images were captured using a TIAS and a fluorescence imaging system from 11 participants (20 to 24 years old), and only the tongue coating area was extracted from the images. For extraction, the TIAS and fluorescence images were segmented into superpixels, machine learning was performed based on the features of the corresponding superpixels to obtain information regarding the presence of tongue coating, and the coating areas were extracted from the TIAS images. Furthermore, cross-validation of the leave-on-out and comparison of the performances of a support vector machine (SVM) and random forest (RF) were performed.

Results: Two machine learning classifiers were built for tongue coating extraction. With the use of these classifiers, which utilized the SVM and RF for learning the data, the percentage of correct responses was approximately 86% for both the classifiers, and this accuracy is similar to the those obtained in previous studies.

Conclusion: We proposed a tongue coating discrimination method utilizing feature analysis with an accuracy equal to or better than those of previous studies. Our proposed method is superior to the conventional method because it can analyze both the tongue as well as its peripheries. However, its accuracy is low for cases involving thin tongue coating (white coating, etc.), and the accurate extraction for cases involving white coating is difficult using our method, which forms a future direction of research.

Keywords: Machine learning, Tongue diagnosis, Eastern medicine, Support vector machine (SVM), Random Forrest

Introduction

In recent years, the need for preventive medical care before the onset of illness has increased as a measure to mitigate the decrease in the working population owing to the declining birthrates, aging society, and short life expectancy relative to the average life expectancy. In contrast to Western medicine, which has a dualistic view of diseases or health, Eastern medicine is based on the premise that diseases and health are constantly evolving (Hori, 2014). Kampo medicine can be considered as a suitable way of providing medical care for presymptomatic diagnosis, which lies in between disease and health. In tongue diagnosis, which is one of the diagnostic methods performed in Kampo medicine, a patient's current health condition is diagnosed based on the color and shape of the tongue and coating as well as the amount of moisture on the tongue. However, this diagnosis depends significantly on the physician's ability and the diagnosis environment. Therefore, it is necessary to establish universal diagnostic criteria by performing quantitative and objective analyses through image analysis. This study aims to automate tongue diagnosis and proposes a

Accepted on 4 May, 2020
*: Corresponding Author

method for the extraction of tongue coating area—the target tongue diagnosis—from tongue images captured using a light-emitting diode (LED) light source and a digital camera.

When capturing an image of the tongue using a normal digital camera, a portion of the image may become overexposed owing to specular reflection from tongue caused by the moisture present on it. Furthermore, the movements of the patient and the photographer during the capture may cause unexpected blurring. The tongue image analysis system (TIAS) (Nakaguchi, 2015) was proposed to solve the aforementioned issues. The TIAS comprises an integrated sphere that allows for the uniform illumination of LED white light on the tongue surface, thereby suppressing specular reflections caused by the moisture on the tongue. TIAS imaging therefore appears to be an effective method for the automation of tongue diagnosis. In addition, researchers (Ota, 2018) have reported that bacteria present the tongue coating emit fluorescence when exposed to visible light at a wavelength of approximately 405 nm. Utilizing this property, we performed florescence photography of the tongue surface using visible light illumination and analyzed its fluorescence intensity to obtain an accurate extraction of the tongue coating area. However, tongue coating extraction using visible light requires a fluorescence imaging device.

To solve this problem, the extraction of the tongue coating area was attempted using an image analysis method that utilizes machine learning to capture images using white LED illumination. To perform machine learning, the teaching data must contain images of the tongue coating area captured from the same subject using TIAS, and each image should be a 405 nm visible light image. Furthermore, as each tongue coating area must be aligned, the thin-plate spline method (Fred, 1989) was utilized to collect these data. Using this method, image extraction could be performed with an accuracy of over 80%. In the preprocessing step of this method, a rectangular area was cropped from the central part of the tongue image, which was then segmented into small areas. Since these segmented areas were square-shaped, information regarding the tongue coating shape was lost. Moreover, since only a part the center of the tongue was analyzed, extraction of the tongue coating in the peripheries of the tongue was challenging. In addition to the central part, information from the peripheries of the tongue is required for tongue diagnosis, as this method is a comprehensive in nature and is based on not only the characteristics (area, thickness, distribution, etc.) of the coating, but also the presence of coating.

Therefore, this study purposes an automated method for the extraction of the tongue coating area including tongue peripheries by segmenting an entire tongue into superpixels (Achanta, 2010) and then determining the presence of tongue coating in each superpixel. Sub-sequently, we aim to develop a system suitable for clinical use. To calculate the number of features from each superpixel, we used statistics, which are considered effective in discriminating between three types of color spaces (RGB, CIE-L*a*b*, HSV), the profile features when crossing the tongue images, and the number of co-occurrence matrix features (gray-level co-occurrence matrix, GLCM features) (Haralick, 1973; Haralick, 1992).

We compared the performance of methods based on the differences between machine learning methods in cases involving the support vector machine (SVM) (Fan, 2005) and random forest (Breiman, 2001; Breiman, 1996) as post-feature analysis machine learning methods. Since this study uses images illuminated with visible light, only those images that have been approved by the Chiba University Ethics Committee (approval numbers 2018-1234, 2018) were used.

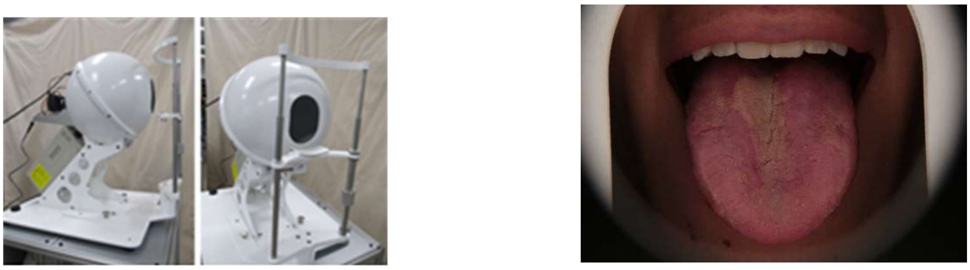
Research data and methods

1. Subject and learning data

In this study, the tongue coating area was extracted using machine learning. The learning data were gathered from images captured using the TIAS and a fluorescence imaging system, as shown in Figure 1(a). The TIAS captures images of the tongue using white LEDs and a digital camera (Canon EOS Kiss X7, resolution 5,184 × 3,456 pixels, RGB, eight gradations each). It allows the uniform illumination of the tongue and produces blur-free images by restraining the participant’s movement using a face-fixing device. It can capture images of the entire tongue. Figure 1(b) shows an example of a tongue image captured using the TIAS.

Figure 2(a) depicts the fluorescence imaging system. Using this system, accurate position information of the tongue coating can be obtained using the fluorescence of the tongue coating, which is acquired when the tongue area is illuminated by visible light (405 nm). Figure 2(b) shows an example of a tongue image captured using this system Figure 3 shows an example of the tongue coating area extracted from a fluorescent image by thresholding.

The tongue images were captured from 11 participants (20 to 24 years old).



（a） Tongue image analysis system (TIAS) （b） Image captured using TIAS

Figure 1: Tongue (TIAS) and a sample image analyzing system

（a） Fluorescence imaging device （b） Fluorescence image

Figure 2: Fluorescence imaging device and a sample image

Figure 3: Tongue coating area extracted by thresholding

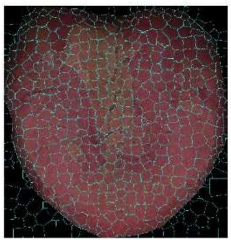


Figure 4: Tongue coating area image segmented into superpixels

2. Superpixel division of TIAS image and fluorescence image

To extract the tongue coating area from a TIAS image, the image was segmented into small areas, and machine learning was used to determine whether these small areas contained tongue coating. However, during segmentation, mixing of the parts of the tongue coating and the tongue itself should be avoided. To achieve this, the area was segmented using a superpixel algorithm (Achanta, 2012). It is necessary to specify the number of area divisions in the superpixel algorithm; based on our experience, we segmented the image into 600 areas. A disadvantage of this method is that many superpixels can be assigned to the background as several background parts exist in the original image captured using the TIAS. Therefore, the original image was trimmed to identify the smallest rectangular area containing the tongue area. As such, most of the superpixels can be assigned to the divisions within the tongue area. Figure 4 shows a trimmed image segmented into superpixels. Because the position of the tongue differed between the TIAS and fluorescence images, the tongue in both the images must be aligned. The boundary of the superpixels obtained from the TIAS image corresponds to the tongue coating extraction image acquired from the fluorescence image. Hence, the same superpixels can be allocated to the same position in the TIAS and fluorescence images.

3. Extraction of tongue coating position using fluorescence image

In the superpixels located at the same position in the TIAS and fluorescence images, the image features from the TIAS image and the information regarding the presence of tongue coating from the fluorescence image are required as machine learning data. Next, we discuss the method of identifying the tongue coating area in the fluorescence image.

First, small areas of the tongue that were to be used for analysis were manually cropped from the fluorescence image (Figure 2 (b)). As shown in the figure, the fluorescence image is red, which means that the R component image in the RGB color space was used to extract the tongue coating position from the image. The increase in the density of the R component in the tongue coating area allowed the extraction of the tongue coating area by thresholding. However, different thresholds were set for each image as individual differences occurred in the emission intensity of the tongue coating caused by fluorescence. Pixels that were below the threshold were not considered as tongue coating even if they had high relative densities. In this study, pixels that satisfied both the criteria were considered as the tongue coating. As the first criterion, pixels that reached the threshold obtained from Otsu's binarization method were considered as the tongue coating. In the second criterion, pixels with a density of 93 or higher (maximum density: 255) were considered as the tongue coating, by referring to the criteria determined by the physician. Because the possibility of erroneously determining the presence of tongue coating in images without coating increases if the threshold value is obtained using Otsu's binarization method, a fixed threshold was used in such cases. In the algorithm, the threshold was first obtained using Otsu's binarization method. If it is 93 or higher, the value obtained through the abovementioned method was set as the threshold for tongue coating discrimination. By contrast, if the value is below 93, a fixed threshold of 93 was used for tongue coating discrimination. Through this process, we prevented the erroneous determination of the presence of tongue coating in tongues that did not have tongue coating (Refer to Figure 3). The tongue coating area thus obtained is necessary for performed machine learning, and this will be discussed in the following sections. Moreover, the validity of the areas obtained by thresholding where the tongue coating was determined was confirmed by a specialist.

4. Feature extraction from TIAS image

(1) Color features and grayscale statistics

The component image in some color spaces (RGB, CIE-L*a*b*, HSV) and the grayscale image were used to obtain features from the TIAS image. Because the main color of the tongue is a reddish component, the density of the R component image in the RGB color space and the density of the a* component image in the CIE-L*a*b* color space were used to calculate the color features. Using the k-means method (Onoda, 2011) in the HSV color space (Burger, 2016), the density of the S component, which exhibited the strongest relationship with the tongue coating in the preliminary experiment, was used to calculate the features. In addition, the density of the grayscale image was used to calculate the textural features. The superpixel algorithm was applied on the TIAS image, whereas the mean, variance, skewness, kurtosis, and mode of the density in each obtained superpixel were used as features; a total of 20 features (four components × five features) were used.

(2) Co-occurrence matrix features

The TIAS image was converted to grayscale to calculate the textural features. The co-occurrence matrix features (GLCM feature value) (Tanaka, 2019) were used as textural features in the grayscale image. Let the relative position of pixels i and j be $\delta = \langle d, \theta \rangle$, each pixel value be L_i, L_j, and $P_\delta(L_i, L_j)$ be the co-occurrence matrix converted to probability from the co-occurrence matrix, which is the frequency of occurrence of the pixel value pair (L_i, L_j). Four features (contrast, correlation, energy, and inverse difference moment) that are considered to be advantageous in the classify-cation of tongue coating were used, amongst the 14 features that are popularly known as co-occurrence matrix features. These four features can be calculated using the following equations:

(a) Contrast (CNT):

$$\mathrm{CNT} = \sum_{k=0}^{L-1} k^2 P_{x-y}(k)$$

$$P_{x-y}(k) = \sum_{L_i=0}^{L-1} \sum_{L_j=0}^{L-1} P_\delta(L_i, L_j)$$

(b) Correlation (CRR):

$$CRR = \frac{\sum_{L_i=0}^{L-1} \sum_{L_j=0}^{L-1} P_\delta(L_i, L_j) - \mu_x \mu_y}{\delta_x \delta_y}$$

$$\mu_x = \sum_{L_i=0}^{L-1} L_i P_x(L_i), \quad \mu_y = \sum_{L_j=0}^{L-1} L_i P_y(L_j)$$

$$\delta_x = \sum_{L_i=0}^{L-1} (L_i - \mu_x)^2 P_x(L_i), \quad \delta_y = \sum_{L_j=0}^{L-1} (L_j - \mu_y)^2 P_y(L_j)$$

(c) Energy (ENR)

$$ENR = \sum_{L_i=0}^{L-1} \sum_{L_j=0}^{L-1} \{P_\delta(L_i, L_j)\}^2$$

(d) Inverse difference moment (IDM):

$$IDM = \sum_{L_i=0}^{L-1} \sum_{L_j=0}^{L-1} \frac{1}{1 + (L_i - L_j)^2} P_\delta(L_i, L_j)$$

These features were defined to be calculated for a rectangular area. Since the superpixels, which were segmented for feature analysis of the tongue area, did not always form rectangles, the features were calculated for a rectangular area circumscribing each superpixel.

(3) Profile features

The tongue coating area was extracted from the entire image even for images captured using the TIAS. In this section, we describe the method of digitizing the change in density in the tongue image as a feature value. A profile is a graph that shows the changes in density when horizontally crossing (from left to right) the tongue area in each color image. Figure 5 shows the relationship between the a* component image profile and the tongue coating position. Here, the profile is presented based on the solid blue line. As shown in the figure, the part where the density changes in the center of the profile corresponds with the position of the tongue coating in the tongue image. The a* and b* components in the CIE-L*a*b* color space corresponded well with the S component in the HSV color space between the profile and the tongue coating position; therefore, these color component images were used in the profile extraction.

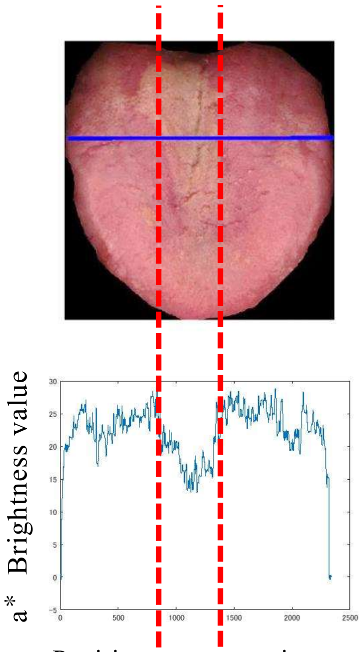


Figure 5: Relationship between tongue position and density of a* component image in profile

When calculating the features from the profile, the tongue image was raster scanned, and the tongue coating area was determined based on the maximum and minimum values of each profile. As shown in Figure 5, the profile exhibited several fluctuations in the density; therefore, it was difficult to obtain the maximum and minimum values from the difference in the density of adjacent pixels owing to the significant amount of noise. Hence, noise removal was performed on each line profile using a moving average filter. In smooth profile data with not much noise, the maximum value was always located within the tongue coating area, whereas the minimum value was in the area without tongue coating. Pixels with other values were classified into pixels with and without tongue coating based on the intermediate value between the maximum and minimum values. All line profiles were subjected to the same processing, and the tongue extraction image was displayed two dimensionally. Figure 6(a) shows the tongue area extracted from the fluorescence image, and Figure 6(b) shows the tongue coating detected by the profile feature of the TIAS image. A mock tongue area was extracted from the TIAS through a series of these processes.

5. Creation of classifier

When extracting the tongue coating area from the TIAS image, the image was segmented into small areas, and the tongue coating area was extracted by determining whether each of the small area was a tongue coating area. In this study, machine learning was performed using the image features in the superpixels of the TIAS image and the information on the presence of tongue coating inside the superpixels of the corresponding fluorescence image as the teaching data. First, the abovementioned features were calculated in each superpixel of the TIAS image.

Next, areas were classified as tongue coating areas if several pixels with tongue coating existed inside each superpixel of the fluorescence image. The areas were classified as otherwise if only a few of the abovementioned pixels existed. Furthermore, the tongue coating image was a binary image, hence the tongue coating appeared white (density: 255), whereas the other parts

(a) Correct label

(b) Tongue coating detection with profile features

Figure 6: Detection results using correct labels and profile features

appeared black (density: 0). When learning was performed on the discriminator, the learning data were created by removing the black background, which was noise. Figure 7(a) shows the tongue coating image obtained by converting the fluorescence image into a binary image, whereas Figure 7(b) shows an example of learning data created using superpixel information. Additionally, to prevent the deviation of the number of data with and without tongue coating in the learning data, learning was performed using data that were randomly extracted from a small number of data.

The machine learning in this study involved discrimination using an SVM and random forest. Performance comparison was performed through leave-one-out cross-validation using the discriminator, which has undergone learning. As 11 participants were involved, 11 verification patterns were performed, with one person as the test data and the others as the learning data. Subsequently, the average of these was obtained as the result.

Results

Figure 8(b) shows an example of the learning data, wherein the presence of tongue coating on a TIAS image was detected using SVM and random forest. As shown, we successfully extracted the tongue position in the TIAS image The mean of the area under curve obtained from the receiver operating curve was 0.939 when random forest was used, whereas it was 0.935 when SVM was used, thereby indicating that both the method showed similar performance. Table 1 presents the results wherein tongue discrimination was performed on the TIAS image using an SVM. The discrimination rate of the area with tongue coating was 89.4%, whereas that without tongue coating was 84.4%.

Table 1: Results of discrimination of presence or absence of tongue coating using SVM

		SVM(%)	
		With coating	Without coating
Learning Data	With coating	89.4	10.6
	Without coating	15.6	84.4

Table 2: Results of discrimination of presence or absence of tongue coating using RF

		Random Forest(%)	
		With coating	Without coating
Learning Data	With coating	88.3	11.7
	Without coating	15.7	84.3

These results indicate that we successfully obtained sufficient discrimination accuracy. Moreover, in the discrimination using the SVM and random forest, the discrimination rates were 88.3% and 84.3%, respectively. These results show that the discrimination rate was similar to that of the SVM. We discovered that similar results can be obtained from the SVM and random forest, provided that the quality of the learning data is maintained.

Discussion

In a previous study (Ota, 2018), a rectangular area was cropped from the central part of the tongue, the area was segmented into small areas, and random forest was used to determine whether area indicated the presence of tongue coating. Because all these areas were square-shaped, information regarding the shape of the tongue coating was lost.

Therefore, to address the aforementioned issue, the entire image was segmented into superpixels, and the presence or absence of tongue coating in these superpixels was automatically determined. The results of this study exceeded the average area under the curve (AUC) (0.9) and the percentage of correct responses (85.8%) in the determination of the presence or absence of tongue coating and in the position of tongue coating area in the images. Therefore, it can be said the profile features used in the extraction of the tongue coating area from the entire image were effective. Because noise removal, which is necessary for calculating profile features, was not sufficiently investigated, the smoothing method is expected to further improve the performance of tongue coating extraction.

In cross-validation through machine learning, the percentage of correct responses for tongue coating extraction decreased when data from participants with thin tongue coating were used as test data. In cases involving pale-colored white coating, the difference in the density between the white coating and other parts decreased, which in turn lead to a decrease in the discrimination rate. The GLCM features used in this study did not yield sufficient results for cases with small differences in density; therefore, features that are effective for discriminating the difference in the density between the white coating and other parts must be analyzed. Pale-colored tongue coating is often associated with healthy people; therefore, it is not a significant clinical problem if white coating cannot be detected. However, we believe that the extraction of the shape of the white coating area is an important topic for future studies.

Furthermore, for determining the presence or absence of tongue coating, the correspondence between the image features of a superpixel in a TIAS image and the information regarding the presence or absence of tongue coating in the superpixel of the fluorescence image is crucial. During this process, the total number of superpixels may affect the final result. The value (600) used in this study was selected based on experience; however, the number of superpixels must be determined based on the calculated cost. In addition, methods for adaptive decision-making must be considered.

Conclusion

A method for the automatic extraction of the tongue coating area using TIAS was proposed for tongue diagnosis. The tongue coating area obtained from the TIAS image and the information regarding the presence or absence of tongue coating obtained from the fluorescence image were associated with each other. The extraction of the images was performed using a classifier obtained using machine learning. We successfully automated the entire system using superpixel divisions to associate the TIAS image with the fluorescence image. Compared with the previous methods, the proposed method can extract the tongue coating area from the entire tongue, rendering it an effective system for clinical applications.

Moreover, to improve the accuracy of the system, more participants should be recruited. If Type II error α = 0.05, detection power = 0.95, and effect ρ = 0.3, the number of participants will be 134. When combined with our current data, at least 123 participants are required.

In addition, extraction of tongue coating thickness from TIAS images is required. The thickness of the tongue coating is one of the important criteria for diagnosing diseases. Based on the findings obtained this study, we confirm that the thickness of the tongue coating can be determined via machine learning using features related to the thickness of the coating.

References

Achanta R, Shaji A, Smith K, Lucchi A, Fua P, and Susstrunk S, 2012. SLIC superpixels Compared to state-of-the-art superpixel methods," IEEE Trans. Pat. Anal. Mach. Intel., 34(1), 1-8.

Breiman L, et al, 2001. Random forests. Machine Learning 45: 5-32.

Breiman L. 1996. Bagging predictors. Machine Learning 26, 123-40.

Burger W, Burge MJ, 2016. Digital image processing: an algorithmic introduction using java (texts in computer science) 2nd edition, Springer.

Fan RE, Chen PH, and Lin CJ, 2005. Working set selection using the second order information for training SVM. Journal of Machine Learning Research 6, 1889-1918.

Fred L, Bookstein, 1989. Principle warps: thin-plate spline and the decomposition of deformations." IEEE Trans. on Pattern Analysis and Machine Intelligence 11(6), 567-85.

Haralick RM, Shanmugan K, and Dinstein I, 1973. Textural features for image Classification. IEEE Trans. on Systems, Man, and Cybernetics SMC-3, 610-21.

Haralick RM, and Shapiro LG, 1992. Computer and Robot Vision, Vol.1, Addison-Wesley.

Hori T, , 2014. A comparative study of eastern and western medicine. Chuo Gakuin University Human and Natural Sciences, 38, 27-49 (in Japanese).

Nakaguchi T et al, 2015. Proposal for a new non-contact 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 2015, 10.

Ota Y, et al, 2018. Tongue coating analysis via machine learning using texture and color features. In Int'l Journal of Computer Assisted Radiology and Surgery (CARS2018), 212-213.

Onoda T, Sakai M, and Yamada S, 2011. Experimental comparison of clustering results for k-means by using different seeding methods. The 25th annual conference of the Japanese Society for Artificial Intelligence, 1-4. (in Japanese)

Tanaka T, 2019. Fundamentals of image information processing, Colona Publishing Co., Ltd., 109-112. (in Japanese)