Research paper

Quantification of tongue colour using machine learning in Kampo medicine

Tadaaki Kawanabe a, Nur Diyana Kamarudin b,*, Chia Yee Ooi b, Fuminori Kobayashi b, Xiaoyu Mi c, Mariko Sekine a, Akino Wakasugi a, Hirodi Odaguchi a, Toshihiko Hanawa a

a Oriental Medicine Research Center, Kitasato University, Japan
b Malaysia-Japan International Institute of Technology, Malaysia
c Fujitsu Laboratories Ltd., Japan

ARTICLE INFO

Article history:
Received 6 January 2016
Received in revised form 6 April 2016
Accepted 7 April 2016

Keywords:
Tongue colour diagnosis
Machine learning
K-means clustering
Kampo medicine
Traditional Japanese medicine

ABSTRACT

Introduction: The evaluation of tongue colour has been an important approach to examine human health in Kampo medicine (traditional Japanese medicine) because the change in tongue colour may suggest physical or mental disorders. Several tongue colour quantification methods have been published to objectify clinical information among East Asian countries. However, reliable tongue colour analysis results among Japanese test persons are limited because of a lack of quantitative evaluation of tongue colour. We aimed to use advances in digital imaging processing to quantify and verify clinical data tongue colour diagnosis by characterising differences into tongue features.

Methods: The D501-B tongue colour information acquisition system was used to extract tongue images of 1080 Japanese test subjects. Evaluation of tongue colour, body and coating was performed by 10 experienced Kampo medicine physicians. The acquired images were classified into five tongue body colour categories and six tongue coating colour categories based on evaluations from 10 physicians with extensive Kampo medicine experience. K-means clustering algorithm was applied as a machine learning (the study of pattern recognition by computational learning) method to the acquired images to quantify tongue body and coating colour information.

Results: Tongue body (n = 550) and tongue coating (n = 516) colour samples were classified and analysed. Clusters consisting of five tongue body colour categories and six tongue coating colour categories were experimentally described in the CIELAB colour space. Statistical differences were evident among the clinically primary tongue colours.

Conclusions: Clinically important tongue colour differences in Kampo medicine can be visualised by applying machine learning to tongue images taken under stable conditions. This has implications for developing globally unified, reliable tongue colour diagnostic criteria which could be used to explore the relevance between clinical status and tongue colour.

© 2016 The Author(s). Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Tongue diagnosis is a simple and non-invasive clinical technique in traditional medicine. The technique has been used by traditional medicine practitioners in both East Asian and Western countries. The basic idea of tongue diagnosis can be found in ancient Chinese clinical textbooks such as ‘Nan Jing’, ‘Shang Han Lun’ and ‘Jingui Yaolue’. It is thought that the essence of patients’ disorders is indicated by the tongue condition. Physicians evaluate tongue colour, shape and coating visually and adopt the findings as an important basis for the choice of treatment. Specific findings indicate patients’ disharmony. For example, tongue colour provides useful diagnostic information such as blood congestion, water imbalance and psychological problems.

In Kampo medicine as well as in traditional Chinese medicine, clinically important tongue colour information is obtained from the tongue body area and tongue coating area. The tongue body colours are usually classified into at least five subtypes (light white, light red, red, deep red and purple) by observing the bilateral edges of the tongue surface (non-coating area). Tongue coating colours are classified into six subtypes (white, white–yellow, yellow, brown, grey and black) [1,2]. Colour samples are shown in Figs. 1 and 2.

* This issue belongs to the Special Issue on Methodological Challenges in Whole Systems Research.
* Corresponding author.
E-mail address: east.diyana@outlook.com (N.D. Kamarudin).

http://dx.doi.org/10.1016/j.eujim.2016.04.002
1876-3820 © 2016 The Author(s). Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
In practice, evaluation results are influenced by several factors such as the medical practitioner’s tactile sense, colour sensitivity or viewing environment as well as their interpretation tendencies based on experience or other clinical information. In addition, clinical practices vary among physicians in different countries. Thus, a method to objectify tongue colour diagnosis using optical methods under stable conditions is desirable [3–5]. A colour quantification method that can differentiate colours reliably will advance the field of traditional medicine.

Objectification and standardisation in traditional medicine have received increasing attention worldwide, particularly in East Asian countries. Advances in digital image processing with feature pattern recognition approaches have contributed to the objectification of tongue diagnosis and led to the development of various tongue image analysis systems [5–13].

Evaluation of quantitative clinical data is essential for the acquisition of accurate tongue image information. However, some elements in tongue image analysis systems, such as illumination conditions and colour correction procedures, affect tongue feature evaluation results. Consequently, sustaining the accuracy and reproducibility of tongue image information is difficult [5–11,14–19]. Overcoming these difficulties will benefit further development of objectification in clinical practice, basic research and education.

To address these difficulties, an information acquisition system for tongue diagnosis, DS01-B, was developed in China in 2011 (Fig. 3). DS01-B is a medical optical instrument that simulates stable natural illumination with no light reflection or shadow on the tongue surface and demonstrates high colour reproduction performance. The components implemented in DS01-B are alight-emitting diode light source to provide uniform illumination and specular reflection reduction on the tongue surface and a charge-coupled device (CCD) camera (Nikon D70®) to capture high-resolution tongue images. DS01-B removes ambiguity and maintains colour value accuracy to provide reliable evaluation of tongue colour information [12].

Fig. 1. Tongue body colour samples ([left-right] light white, light red, red, deep red and purple). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 2. Tongue coating colour samples ([left-right] white, white-yellow, yellow, brown, grey and black). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 3. Outward appearance of DS01-B.

Fig. 4. Clipping of entire tongue (left, before processing; right, after processing).
Analysis software or algorithms to quantify the colour information of specific regions of the tongue surface are also important. DS01-B implements a precise clipping algorithm that extracts the entire tongue area and eliminates the perioral area from acquired images (Fig. 4).

However, DS01-B does not implement a clinically reasonable tongue body and coating separation algorithm. We examined several colour quantification methods for acquired tongue images. Previously reported methods are based on algorithms with predefined thresholding, colour extraction or machine learning methods [11,20–22]. Although machine learning results are often unreasonable because of significant variation of tongue features, machine learning has been reported to be one of the most credible methods. Consequently, we applied a machine learning process to correspond to the feature differences of each tongue image for detecting tongue body and coating.

Today, one of the challenges in biomedical research and clinical practice is the excesses of valuable medical information which needs to be analyzed and optimized to improve the efficiency of diagnosis while also reducing the cost of computerized devices. Because of this abundance of non-analysed data that cannot be handled manually by the medical doctors or researchers, there is an urgent need for integrative and interactive machine which can provide learning solutions [23]. Machine learning explores the study of computer science which involves data mining technique and the development of intelligence algorithms which has the ability to learn from the vast training inputs and make predictions based on these learning data for classification by optimisation. Therefore, we tested a conventional mathematical machine learning method that can process characteristic differences of tongue features, and the results obtained from a CIELAB-based K-means method had high correlation to physicians. K-means clustering algorithm is one of unsupervised machine learning techniques where the predictions for the test outputs do not depend on the learning algorithm (training data) but on the hidden patterns of the data itself or a feature-based learning. K-means is a popular and distance-based clustering algorithm used in data mining to partition data into a pre-determined number of clusters.

After analysing data similarity, each data item belongs to the cluster with the nearest mean. Using the K-means clustering algorithm, we can successfully obtain well-separated tongue body and coating images (Fig. 5).

To the best of our knowledge, to date, there have been no reports of tongue colour evaluation using the K-means method for a large number of Japanese test persons. In this study, we present quantified tongue colour data using the DS01-B as an assured instrument and a K-means clustering algorithm for universal mathematical analysis to promote future clinical research.

2. Materials and methods

2.1. Tongue image acquisition procedure and colour reproduction performance

The process to acquire a tongue image using the DS01-B is shown in Fig. 6. Although the DS01-B is presumed to have a reliable automated colour calibration function, we evaluated the colour reproduction of the DS01-B to confirm precision. We used a commercial colour chart (Color Checker Passport® X-rite®) for testing and evaluated the CIELAB colour space value difference

$$
\Delta E = \sqrt{(L_2 - L_1)^2 + (a_2 - a_1)^2 + (b_2 - b_1)^2}
$$

between each measured average value (\(L_2, a_2, b_2\)) and the ground truth value (\(L_1, a_1, b_1\)). The colour difference of each chart was within 8.16 at 2σ of normal distribution. The results indicate that the colour correction of the DS01-B is sufficiently accurate to distinguish difference with the naked eye and acceptable for colour evaluation [24]. The clinical technologists controlled the test persons’ position and tongue protrusion by checking the moving image taken by the CCD camera before photographing. The tongue images were taken within 3 s of protrusion because it is known that the colour of the tongue darkens as a result of blood congestion within a few seconds after protrusion [25]. The acquired colour corrected image is automatically segmented into an entire tongue image and a perioral image (Fig. 4).

2.2. Acquired tongue image analysis (separation algorithm for tongue body and coating)

The tongue image (after clipping) was used as input to the K-means clustering algorithm to extract the tongue body and coating. The standard RGB colour space of a tongue image has been converted to the device-independent CIELAB colour space because of its perceptual uniformity [14,18]. Thus, a distance metric that uses the CIELAB colour value implemented as the distance metric function in the K-means clustering algorithm is more meaningful for similarity assessment using the naked eye. This distance metric is expressed as follows:

$$
\Delta \text{distance metric} = \sqrt{(L_{c1} - C_{c1})^2 + (a_{c1} - C_{c2})^2 + (b_{c1} - C_{c3})^2}
$$

We applied the K-means clustering algorithm to cluster objects into four clusters using the Euclidean distance metric with centroid detection. (C_{c1}, C_{c2}, C_{c3}) is the centroid position for each cluster, and k=4 is the desired number of clusters. We adopted k=4 because the separation of the tongue body and coating of the tongue are evident according to physician assessment.
(L∗, a∗, b∗) is the CIELAB colour value for each pixel of the given tongue image. In image processing, the K-means clustering algorithm is used for image clipping and pattern detection and classification. The K-means method is a vector quantisation method that treats each object in an image as having location in space. The chromaticity elements a∗ and b∗ and luminance L∗ are important parameters for tongue body and coating cluster separation. By labelling each cluster according to its distance metric and colour, the average tongue body and coating colour in the cluster can be determined and analysed. A flowchart of the K-means clustering algorithm is shown in Fig. 7.

We applied the K-means (k=4) clustering method based on CIELAB colour values to detect tongue body and coating for the segmented tongue images using MATLAB. We confirmed all clustered images on an IPS (in-plane switching) monitor (ColorEdge CG246, EIZO®). The major reason for using four clusters comes from the anatomical features of the tongue surface. The tongue surface typically comprises a coating area, a body area (non-coating rim area) and a transition area between the coating and body. Thus, we considered four clusters reasonable for distinguishing these areas and the background. Clustering results with 1) background (black), 2) tongue coating, 3) tongue body and 4) transition areas are shown in Fig. 8. We differentiated the resulting images and selected the tongue coating and body areas to accumulate sample images.

We also evaluated three (k=3) and five (k=5) cluster approaches to verify the appropriateness of the four cluster approach, as shown in Figs. 9 and 10, respectively. For k=3, over-detection of the tongue coating area (including the tongue body area) and insufficient detection of the tongue body area were evident, as shown by 2 and 3 in Fig. 9, respectively. For k=5, excessively separated tongue body images were evident, as shown by 3, 4 and 5 in Fig. 10.

Thus, we applied the four cluster approach because it is the most suitable quantification method for clinical tongue colour diagnosis.

2.3. Analysis of processed images

Tongue image acquisition was performed with prior consent of 1080 Japanese test persons (368 males and 712 females) using the DS01-B at the Oriental Medicine Research Center, Kitasato University between January 2015 and December 2015. The colour evaluation of tongue body and coating was performed by 10 physicians in our institute who have more than 5–30 years of experience with Kampo medicine. The physicians performed their
evaluations under 5500 K colour temperature fluorescent light in an outpatient ward. The processed tongue body and coating images were classified according to the physicians’ evaluations. For appropriate evaluation, we excluded inappropriate samples such as samples with insufficient protrusion. We also excluded samples of wholly coated tongue images (this makes tongue body detection difficult) and very thin or non-coated tongue images (this makes tongue coating detection difficult). After eliminating inappropriate samples, the remaining images were analysed using the four cluster k-means method to quantify the average L*a*b* colour values of the tongue body and coating images. The statistical differences of each accumulated colour value were examined using a Kruskal–Wallis multiple comparison test for major subtype colour cluster analysis of tongue body (light red, red, deep red) and coating (white, white-yellow, yellow, brown) and a Welch’s t-test for whole cluster analysis of tongue body and coating.

3. Results

After processing, typical sample images (550 tongue body and 516 tongue coating images) were extracted and analysed. The average L*a*b* colour values of each tongue body and coating colour group are shown in Tables 1 and 2, respectively. Because of the small sample size (n < 10), the median values of light white, purple, grey and black are also described. Clusters comprising five tongue body colours (light white, light red, red, deep red and purple) and six tongue coating colours (white, white-yellow, yellow, brown, grey and black) are shown in Figs. 11 and 12, respectively. Each cluster centre is shown as *. In tongue body colour analysis, the statistical differences (P < 0.01) can be observed in a* and b* values among light red, red and deep red. In tongue coating colour analysis, statistical differences (P < 0.01) can be observed in L* and b* values among white, white-yellow, yellow and brown.

The average L*a*b* colour values of summed tongue body and coating colour groups are shown in Table 3. Clusters comprising entire tongue body and coating colours are shown in Fig. 13. Each cluster centre is shown as *. The statistical differences (P < 0.01) can be observed only for a* values.

4. Discussion

We have reported the possibility that each cluster derived from clinically different tongue colours of Japanese test persons can be visualised in the CIELAB colour space. The results do not contradict the characteristics of CIELAB colour space because higher a* values indicate richer red, lower b* values indicate richer blue and higher b* values indicate richer yellow components. Tongue colours are
primarily classified into three body colours (light red, red and deep red) and three coating colours (white, white-yellow, yellow); therefore, the other colour samples were limited, and it was difficult to obtain sufficient samples for statistical analysis. Although it was difficult to obtain sufficient sample images for all tongue colours, we were able to demonstrate a partial statistical difference between the primary body and coating colours (Tables 1 and 2). Although overlaps were evident among the tongue body and coating clusters in the CIELAB colour space, the difference of each cluster's centre was evident (Figs. 11 and 12). The difference between 2 clusters comprising entire tongue body and coating colours was also evident (Fig. 13). These results suggest

Table 2

<table>
<thead>
<tr>
<th></th>
<th>L*</th>
<th>a*</th>
<th>b*</th>
</tr>
</thead>
<tbody>
<tr>
<td>White (n = 303)</td>
<td>56.00 ± 3.35**</td>
<td>14.80 ± 2.92</td>
<td>10.38 ± 2.22**</td>
</tr>
<tr>
<td>White-yellow (n = 129)</td>
<td>54.09 ± 3.78**</td>
<td>14.52 ± 2.94</td>
<td>14.28 ± 1.56**</td>
</tr>
<tr>
<td>Yellow (n = 57)</td>
<td>51.95 ± 4.27**</td>
<td>14.65 ± 2.73</td>
<td>19.21 ± 2.00**</td>
</tr>
<tr>
<td>Brown (n = 18)</td>
<td>54.09 ± 3.78**</td>
<td>15.90 ± 1.60</td>
<td>16.85 ± 2.72**</td>
</tr>
<tr>
<td>Grey (n = 7)</td>
<td>39.80 ± 5.62</td>
<td>10.90 ± 1.95</td>
<td>7.56 ± 2.24</td>
</tr>
<tr>
<td>Black (n = 2)</td>
<td>39.72 (35.43–46.28)</td>
<td>10.99 (8.91–11.83)</td>
<td>7.36 (5.96–9.84)</td>
</tr>
</tbody>
</table>

**P < 0.01.

Fig. 11. Distribution of respective average tongue body colours in CIELAB colour space.

Fig. 12. Distribution of respective average tongue coating colours in CIELAB colour space.
that all of the differences depend on $a^*$ or $b^*$ values constantly and the appropriateness of K-means ($k = 4$) clustering for tongue image analysis under the illumination uniformity of DS01-B. Similar analysis has been reported from other countries; however, it is difficult to compare the measured colour data directly because of lower colour reproduction and different analysis methods and clinical practices in different countries [7,13–15].

The most influential difference is diagnostic procedure or physician ability for tongue colour identification [2–5,26]. A tongue colour evaluation method with appropriate viewing conditions has not been established among physicians, researchers or facilities, and globally unified tongue colour diagnostic criteria have not been established to date. As shown in Figs. 11 and 12, physicians evaluate the differences in tongue colour in a very narrow range of CIELAB colour, and it is difficult to form consensus about diagnostic criteria. If we can establish a robust tongue colour quantification method with high reliability, it will become possible to overcome these unsolved problems because we can use a common classification index that is suitable for scientific discussion without forcibly uniting different diagnostic criteria. In addition, the difference in physicians’ abilities to identify colour is another problem because it generally improves with clinical experience and is influenced by a physician’s inherent ability to differentiate colour. At our university, workshops for tongue feature diagnosis have been held annually, and the gaps in evaluation among physicians have been corrected over time [2]. We believe that this activity has led to the significant results of this research.

Another important factor is specification difference among tongue image analysis systems. For example, results can be

![Fig. 12. Distribution of respective average tongue coating colours in CIELAB colour space.](image)

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Average $L^*a^<em>b^</em>$ values of tongue body and coating colour ($**P &lt; 0.01$).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$L^*$</td>
</tr>
<tr>
<td>Tongue body ($n = 550$)</td>
<td>$51.24 \pm 2.85$</td>
</tr>
<tr>
<td>Tongue coating ($n = 516$)</td>
<td>$54.22 \pm 3.28$</td>
</tr>
</tbody>
</table>
influenced by photography conditions such as illumination, device shape or colour correction procedure [5–13]. Appropriate lighting conditions with proper colour correction are considered essential because limited colour information is derived from varying or unstable illumination, and there is an insufficient number of standard colour charts for accurate evaluation [15–19]. To address these problems, various types of tongue image analysis systems with predictable and stable settings have been developed to establish quantitative diagnostic systems, particularly in East Asian countries [5,6,8,11]. On the other hand, the usefulness of high-resolution hyperspectral imaging for tongue colour evaluation has been reported in many research papers because of its absolute colour information and effective colour reproducibility management [27–32]. Compared with hyperspectral imaging, typical tongue image analysis systems can provide tongue images faster and easily but provide limited colour information. Most tongue image analysis systems do not implement a hyperspectral camera; however, our results suggest that a conventional CCD camera with appropriate colour calibration can be reasonably effective for objectification. Nevertheless, several problems are associated with the DS01-B, such as the influence of reflected light from facial skin or stray light that comes through the wide face contacting area as large as the human face. These undesirable influences should be eliminated to achieve good colour reproduction and minimise error. Thus, the specifications of tongue image analysis systems must be enhanced.

Furthermore, the differences among tongue colour analysis methods are another issue. Although many tongue image analysis systems with various types of colour diagnostic functions have been developed, the difference between clinical diagnostic criteria and analysed colour results has seldom been reported. For example, average tongue body colour analysis results are often quite different from physician colour evaluation because of the influence of colour distribution. Even in one tongue image, there often exist various colour patterns in the tongue body area, and the average colour information is insufficient for clinical evaluation.
The difficulty in thresholding different tongue colours quantitatively has also been reported [10,14]. In this study, we extracted tongue body and coating colour data by applying a CIELAB-based K-means clustering algorithm to visualise colour differences to some extent in 3D colour space. Although there are various types of colour combinations between the tongue body and coating, it is not particularly difficult to differentiate each colour with the naked eye because tongue body colours are typically light red, red or deep red and coating colours are typically white, white–yellow or yellow. We consider that this is one reason why K-means clustering worked effectively. However, overlaps among clusters were evident particularly in tongue body colour evaluation, and these overlaps were likely caused by the evaluation method. In future, both the average colour value and the distribution of colours should be evaluated to clarify the most contributing colour information and other characteristic findings such as red spot or purple spot by expanding the clustering algorithm to regional analysis (tips, root and edge) and to avoid divergence between clinical evaluation and quantified results. Thus, more precise colour distribution analysis is required for rigorous evaluation that leads to further degree of standardisation.

Only after the establishment of a clinically matched tongue colour quantification method, we can compare the relevance between clinical information and tongue colour directly. Some studies have reported the relevance between tongue colour and common diseases such as diabetes, hypertension, cardiovascular diseases, gastrointestinal diseases and cancers [7,8,31–34]. Tongue colour and tongue shape changes, such as toothmarks or cracks, are useful clinical information in integrative medicine: thus, additional quantification algorithm appropriate for tongue feature evaluation is required. By promoting assured tongue feature evaluation support system in the future, the quality of clinical evidence of integrative medicine will be improved, and even medical doctors with little knowledge on traditional medicine will be able to easily understand tongue findings at the same standard and conduct clinical research from the perspective of both Western and East Asian medicine. The global accumulation of reliable information may lead to the discovery of correlations between tongue colour and physical or psychological problems. If we can evaluate the risks of diseases from tongue colours, this would be beneficial for predicting particular diseases and promoting human health. We would like to improve the reliability of the quantification algorithm to increase evaluation precision and provide more effective feedback for future clinical research.

5. Conclusion

In this paper, we have discussed the possibility that a conventional mathematical method can appropriately divide a segmented tongue picture into a body and coating parts that are strongly correlated with clinical evaluation. K-means machine learning analysis based on CIELAB colour information is a promising clustering algorithm for tongue colour evaluation. Although the colour data used in this study were obtained from only one institute and it is difficult to generalise our discussion, once we establish an effective tongue colour quantification method, clinical practices in traditional medicine could be visualised. Because diagnostic procedures differ among medical practitioners, particularly in East Asian countries, if we can use a common tongue colour measurement method under stable conditions, clinical research can be promoted worldwide. To improve the accuracy of clinical tongue colour evaluation, we must find a way to quantify contributing colour information rather than the average colour to objectify clinical practices more precisely. There is room for further development of colour reproduction and accurate tongue colour extraction methods, and we would like to continue improving the reliability of tongue evaluation systems.

Acknowledgments

This research was supported by the Japan Society for the Promotion of Science KAKENHI Grant Number 26800420 from the Ministry of Education, Culture, Sports, Science and Technology, Japan and Kitasato University Research Grant for Young Researchers. The authors would like to thank HuiJun Zhou et al. and the Shanghai Daosheng Medical Technology Co., Ltd. for the DS01-B, which we used to obtain samples for our colour analysis and also would like to thank Enago (www.enago.jp) for the English language review. We declare that there is no conflict of interest regarding the publication of this paper.

References


