



Feasibility Study of Google's Teachable Machine in Diagnosis of Tooth-Marked Tongue

Hyunja Jeong[†]

Department of Dental Hygiene, Daegu Health College, Daegu 41453, Korea

Background: A Teachable Machine is a kind of machine learning web-based tool for general persons. In this paper, the feasibility of Google's Teachable Machine (ver. 2.0) was studied in the diagnosis of the tooth-marked tongue.

Methods: For machine learning of tooth-marked tongue diagnosis, a total of 1,250 tongue images were used on Kaggle's web site. Ninety percent of the images were used for the training data set, and the remaining 10% were used for the test data set. Using Google's Teachable Machine (ver. 2.0), machine learning was performed using separated images. To optimize the machine learning parameters, I measured the diagnosis accuracies according to the value of epoch, batch size, and learning rate. After hyper-parameter tuning, the ROC (receiver operating characteristic) analysis method determined the sensitivity (true positive rate, TPR) and specificity (false positive rate, FPR) of the machine learning model to diagnose the tooth-marked tongue.

Results: To evaluate the usefulness of the Teachable Machine in clinical application, I used 634 tooth-marked tongue images and 491 no-marked tongue images for machine learning. When the epoch, batch size, and learning rate as hyper-parameters were 75, 0.0001, and 128, respectively, the accuracy of the tooth-marked tongue's diagnosis was best. The accuracies for the tooth-marked tongue and the no-marked tongue were 92.1% and 72.6%, respectively. And, the sensitivity (TPR) and specificity (FPR) were 0.92 and 0.28, respectively.

Conclusion: These results are more accurate than Li's experimental results calculated with convolution neural network. Google's Teachable Machines show good performance by hyper-parameters tuning in the diagnosis of the tooth-marked tongue. We confirmed that the tool is useful for several clinical applications.

Key Words: Hyper-parameter tuning, Machine learning, Oral health, Teachable Machine, Tooth-marked-tongue

Introduction

The term of machine learning was first used by Arther Samuel (1959) in the late 1950s. As AlphaGo was developed by Google (DeepMind Technologies Limited) in 2014, the interest of the general public in artificial intelligence increased rapidly. Artificial intelligence is currently affecting our lives and is widely used in industries and surrounding environments and in medical fields for diagnosis, treatment, and dentistry¹⁻³. Moreover, platforms with various artificial intelligence libraries such as TensorFlow, Keras, and Pytorch, etc. have been made available, allowing easier implementation of artificial

intelligence⁴. However, artificial intelligence is still a challenging subject to access for ordinary people and beginners. Recently, Google started a web-based artificial intelligence tool service called Teachable Machine for the general public⁵, and version 2.0 is currently available. The Teachable Machine is thought to be insufficient to derive optimal training results due to limited learning parameters that can be adjusted. However, it is thought that meaningful results could be derived from hyper-parameter tuning. The purpose of this study was to evaluate the feasibility of the Teachable Machine web-based artificial intelligence tool in the diagnosis of tooth-marked tongue, in which teeth traces appear on the tongue, among many various findings

Received: October 6, 2020, Revised: October 25, 2020, Accepted: November 5, 2020

eISSN 2233-7679

[†]Correspondence to: Hyunja Jeong, <https://orcid.org/0000-0003-0540-1893>

Department of Dental Hygiene, Daegu Health College, 15 Youngsong-ro, Buk-ku, Daegu 41453, Korea
Tel: +82-53-320-1332, Fax: +82-53-320-1340, E-mail: ywb2014@hanmail.net

Copyright © The Korean Society of Dental Hygiene Science.

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

observed in the oral cavity.

The oral cavity is the first gateway to the digestive system, and it carries out functions such as food intake and mastication. It plays an important role in vocalization, taste, saliva secretion, and auxiliary functions of digestion⁶⁾. Moreover, oral health status is known to be closely related to systemic health. In the past, the relationship between oral disease and systemic health was regarded only as a result of infection by specific bacteria that adhere to the oral cavity⁷⁾. However, nowadays, oral health status is known to have large effects on whole-body health and is used as an index to measure the state of whole-body health^{8,9)}. Among them, oriental medicine diagnoses the whole body's health status by the state of the tongue. An example is tooth-marked tongue. Tooth-marked tongue is a phenomenon in which tooth marks are visible along the edge of the tongue. It is usually accompanied by macroglossia that increases the tongue due to stagnation of recovery (Fig. 1, indicated by the arrows). Macroglossia causes the tongue to press on the teeth, which leads to tooth marks on the tongue after a prolonged period of time¹⁰⁾. In oriental medicine, tooth-marked tongue is thought to be related to spleen abnormalities, and most diagnosis is made by microscopic observations based on experience¹¹⁾.

In this study, we investigated the effects of learning frequency, batch size, and learning rate, which are machine

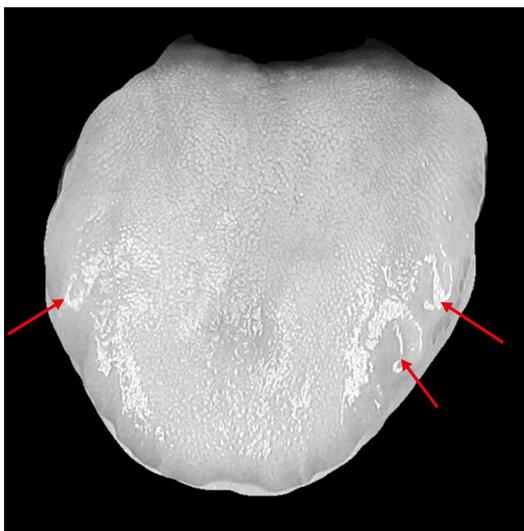


Fig. 1. Photograph of tooth-marked (arrow) tongue.

learning variables of the Teachable Machine provided by Google, on the diagnostic accuracy for a tooth-marked tongue. In addition, we evaluated the feasibility of the artificial model trained under the optimal condition showing the highest diagnostic accuracy for tooth-marked tongue.

Materials and Methods

1. Teachable Machine

Google's Teachable Machine is a web-based artificial intelligence development tool that can build simple artificial intelligence learning models without expertise⁵⁾. It was first introduced in 2017 and has been updated to version 2.0. Artificial intelligence learning model projects related to video, audio, and posture can be performed via the machine.

As shown in Fig. 2, the user interface of the Teachable Machine largely consists of data input, learning, and preview. The trained model can be registered and used on the web through the export function. Training data and test data can be entered through webcam or upload of data files, and multiple classifications of data can be performed by adding classes based on binary classification. It is necessary to adjust various variables according to the video's type and quality during video learning. The artificial intelligence learning model's accuracy can be improved in the advanced learning mode by adjusting three hyper-parameters: epoch, batch size, and learning rate. In the preview mode, the webcam classified results or entered file is displayed by applying the learned model. Diagnostic accuracy of the learned tooth-marked tongue

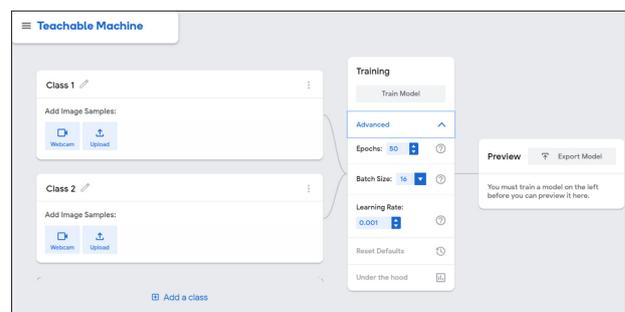


Fig. 2. The graphical user interface of Google's Teachable Machine.

diagnosis artificial intelligence learning model was evaluated by entering test data with known answers. In the export mode, one can output the trained model as a Tensorflow.js model that can be used in a web browsing environment, a Tensorflow model that can be coded using Keras in Python, and a TensorflowLite model that can be used on Android-based mobile.

In this study, machine learning was performed by classifying the training data set into two classes, tooth-marked tongue, and no-marked tongue, to evaluate the feasibility of applying tooth-marked tongue machine learning.

2. Data set

In this study, the feasibility of the Teachable Machine for diagnosis of the tooth-marked tongue was evaluated by using Hanhui's tooth-marked tongue and no-marked tongue data published on the Kaggle site as an example of binary classification¹²⁾. Kaggle is a predictive model and analysis competition platform established in 2010. It is currently being used as a learning platform for data analysis and machine learning¹³⁾. The data consisted of 1,250 tongue images with 704 abnormal images of the tooth-marked tongue and 546 normal images of the no-marked tongue. Ninety percent of all images randomly selected for machine learning were used as the training data set. The remaining 10% was used as the test data set to evaluate the accuracy of the training results. The Teachable Machine did not provide a verification function. Thus, a separate set of verification data was not assigned (Table 1).

3. Hyper-parameter tuning

All image data, including medical images, have their own characteristics depending on the imaging environment, application field, and nature of the acquisition device. Therefore, to appropriately extract each image's character-

istics during machine learning, variables that affect the training results must be set in the optimal condition. The process of increasing the learning accuracy by adjusting variables according to the image data's characteristics of the image data to be learned is called hyper-parameter tuning. It sets the optimal variables to obtain the best learning effects in a learning machine.

The Teachable Machine has three learning parameters, which are epoch, batch size, and learning rate, and machine learning is performed by adjusting these parameters. An increased number of learnings improve accuracy. However, if the number of learnings is set too high, the model is over-optimized for only the training data due to overstaffing. This increases the accuracy of the training data, but it also decreases the data's accuracy. The learning rate is a variable that determines the step size for calculating the loss function. If it is set too high, learning is not performed. If the rate is set too low, a great amount of learning time is consumed, and the model may be trapped at the local minima in the gradient descent method, which is a learning algorithm of machine learning. This may cause a decrease in diagnostic accuracy. In this study, the tooth-marked tongue diagnosis model was calculated by changing the three parameters of an epoch, batch size, and learning rate that affect the learning model's accuracy. Then, test data was entered into the trained model, and the learning accuracy was evaluated to determine the optimal parameter condition. The learning accuracy was determined as the number *n* of cases accurately judged for *N* test data with already known correct answers. The number of wrong answers was calculated as (*N* - *n*).

$$\% \text{ of correct answers} = \frac{n}{N} \times 100 (\%)$$

4. Receiver operating characteristic (ROC)

ROC is a method used to measure sensitivity and specificity in binary classification and is commonly used to measure model performance in machine learning¹⁴⁾. The analysis results were classified as true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN). Then, sensitivity (true positive rate, TPR) and

Table 1. Data Configuration (n=1,250)

Training data		Test data	
No-marked tongue	Tooth-marked tongue	No-marked tongue	Tooth-marked tongue
491	634	55	70

specificity (false positive rate, FPR) were calculated. Test data was entered into the trained model under optimal conditions to measure the accuracy, sensitivity (TPR), and specificity (FPR) and evaluate the model's clinical feasibility.

Results

1. Optimization of hyper-parameters tuning in Teachable Machine

1) Epoch

Artificial intelligence solves overfitting resulting from

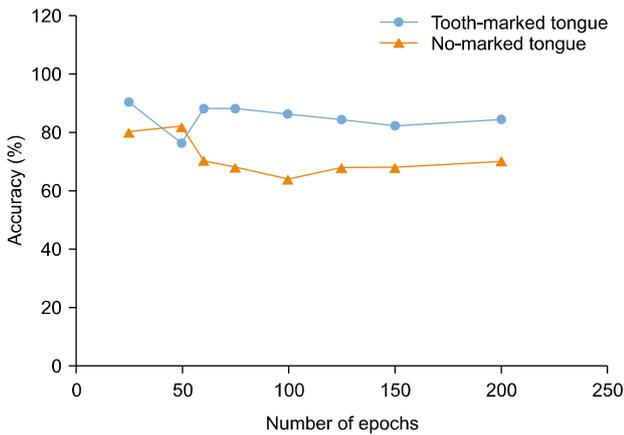


Fig. 3. Diagnostic accuracy of tooth-marked and no-marked tongues as a function of epoch in Teachable Machine learning.

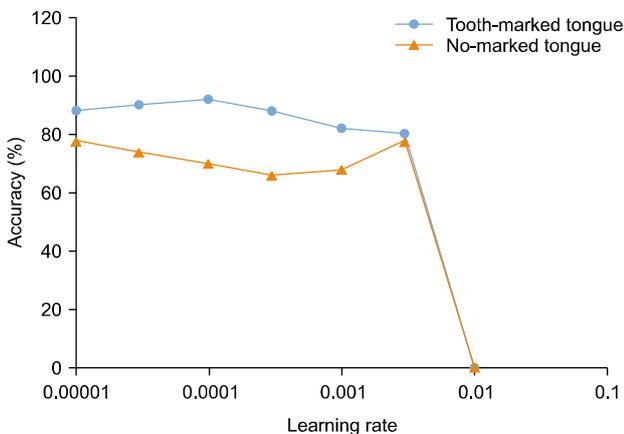


Fig. 4. Diagnostic accuracy of tooth-marked and no-marked tongues as a function of learning rate in Teachable Machine learning.

an excessive number of learning through machine learning on training data and accuracy calculation on verification data at the same time. However, there is no calculation and verification procedure in the Teachable Machine. Thus, the epoch was determined with the test data's diagnostic accuracy while changing the learning number.

Fig. 3 shows the test data's diagnostic accuracy evaluated by the trained model while changing the epoch from 20 to 200. At 50 times or less, where the epoch was insufficient, the accuracy was high and statistical fluctuations were large. At 75 times, the tooth-marked tongue's diagnostic accuracy was 88%, which was the highest. Subsequently, as the epoch increased, the diagnostic decreased due to overfitting, resulting in 82% accuracy at 200 times. On the other hand, at 50 times, the diagnostic accuracy was 82%, which was the highest, and it decreased subsequently due to overfitting. At 200 times, the diagnostic accuracy was 70%. The diagnostic accuracy for abnormality is clinically more meaningful than the diagnostic accuracy for normality. Thus, the epoch was optimized to 75.

2) Learning rate

Fig. 4 shows the test data's diagnostic accuracy evaluated by the trained model while increasing the learning rate by 3-fold from 0.00001 to 0.01. When the learning rate was 0.01 or higher, normal learning for both tooth-marked tongue and no marked tongue could not be performed, resulting in no training results. The tooth-

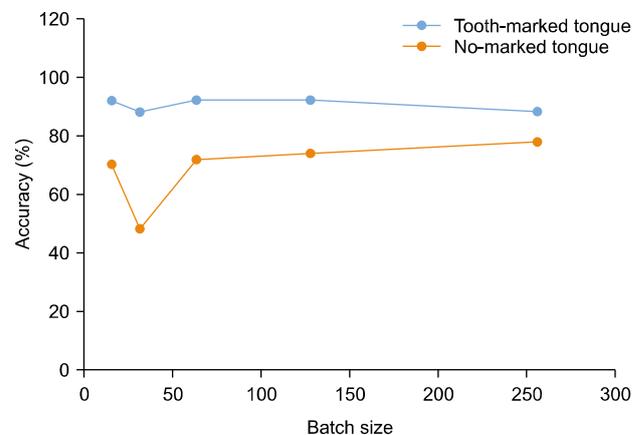


Fig. 5. Diagnostic accuracy of tooth-marked and no-marked tongues as a function of batch size in Teachable Machine learning.

marked tongue's diagnostic accuracy was the highest at 92% when the learning rate was 0.0001. The diagnostic accuracy was the highest at 78% for no marked tongue when the learning rate was 0.003. In consideration of clinical significance, 0.0001, which showed the highest accuracy for the tooth-marked tongue, was determined as the optimal learning rate.

3) Batch size

In machine learning, training is performed by dividing the training data into data sets of a certain size. The divided data set is called the batch size, and the process of learning by dividing the training data set into small batches is called mini-batch gradient descent. Compared to a full batch, which involves learning the entire training data, mini-batch shows a faster computational speed and faster updating of the computed data.

Fig. 5 shows the diagnostic accuracy of the test data evaluated by the trained model while increasing the batch size by 2-fold from 16 to 256. When the batch size was less than 64, the statistical fluctuation of accuracy was large. At the batch size of 128, the diagnostic accuracy for the tooth-marked tongue was 92%, which was the highest. For no-marked tongue, the diagnostic accuracy was the highest at 78% when the batch size was 256. Considering that the accuracy of abnormality has a higher clinical significance, the optimal batch size was determined as 128, at which the highest accuracy for the tooth-marked

tongue was observed.

2. Optimized results of hyper-parameters in Teachable Machine

Fig. 6 shows the screen of the Teachable Machine optimized through hyper-parameter tuning. The epoch, learning rate, and batch size were set as 75, 0.0001, and 128, respectively, and the classes were classified into two types: tooth-marked tongue and no-marked tongue. The training results were 92.1% and 72.6% for tooth-marked tongue and no-marked tongue. Sensitivity (TPR) was 0.92, and specificity (FPR) was 0.28.

Discussion

The oral cavity is the first gateway to the digestive system, and it carries out functions such as food intake and mastication. It plays an important role in vocalization, taste, saliva secretion, and auxiliary functions of digestion⁶⁾. In the oral cavity, the tongue is a muscular organ that plays important roles in pronunciation, mastication, swallowing, and taste. Taste receptors are present in the tongue, and the tongue's state reflects the human body's state. During oral examination in the dental treatment processes, the dorsum in the anterior part of the tongue, tongue root at the posterior end, both sides of the tongue, and the tip of the tongue are rested close to the palate to examine or palpate the lower tongue and floor of the oral cavity. In oriental

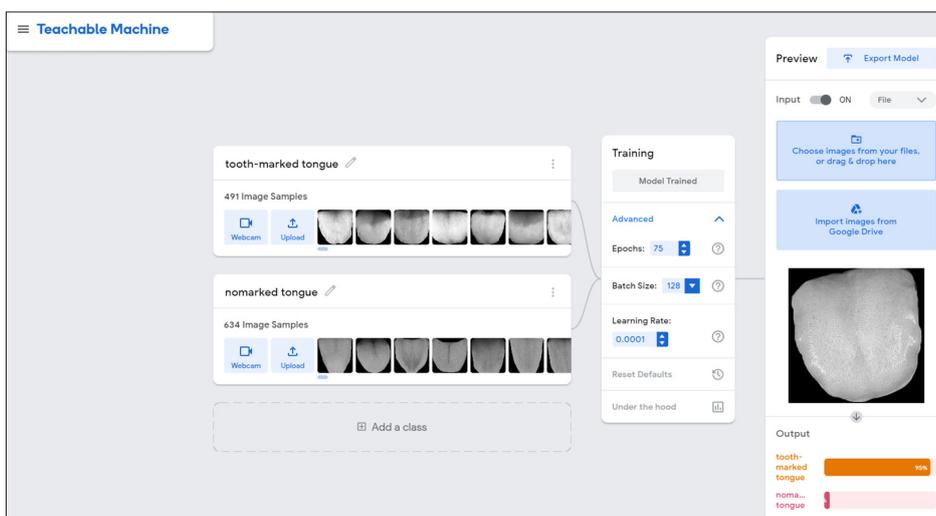


Fig. 6. The studied machine learning model after hyper-parameters tuning using Teachable Machine.

medicine, various diagnoses and prescriptions are performed by referring to the state and color of the tongue and teeth' shapes. In modern medicine, tooth marks are considered as an index to assess whether the pressure balance is well established in the living tissues. In particular, tooth marks appear particularly sensitively in patients with cardiovascular disease, kidney disease, liver disease, and other electrolyte changes¹⁵⁾. Tooth marks result from the excessive amounts of water in the body and hypoproteinemia caused by a lack of nutrition. These lead to enlargement and swelling of the tongue, pressing the tongue against the teeth¹⁶⁾. Furthermore, when the immune function is abnormal in chronic nephritis, thyroiditis, and mammary cancer, macro-glossia, and tooth-marked tongue are observed¹⁵⁾.

As various clinical information can be collected from the tongue's state, various studies related to tongue diagnosis using artificial intelligence have been performed. Ma et al.¹⁷⁾ proposed a new high-level recognition classification method to assess the complex association between disease and tongue characteristics by performing deep learning on tongue images. Furthermore, Tania et al.¹⁸⁾ demonstrated the possibility of non-invasive auxiliary diagnosis by implementing an automatic tongue diagnosis function using machine learning. Li et al.¹¹⁾ performed artificial intelligence on the tooth-marked tongue using a convolutional neural network for 97 and 344 images of the tooth-marked tongue and no-marked tongue, respectively. The accuracy of TP and TN were 69.1% and 76.2%, respectively.

This study's limitations are that the number of tongue image data was limited to 1,250 and that all data were obtained from foreigners. Nonetheless, with an epoch of 75, batch size of 128, and learning rate of 0.0001 for set-up parameters of the Teachable Machine, the diagnostic accuracy for tooth-marked tongue and no-marked tongue was 92.1% and 72.6%, respectively. Sensitivity (TPR) was 0.92, and specificity (FPR) was 0.28. The accuracy of TP that correctly diagnosed the tooth-marked tongue was approximately 23% higher than that in Li et al.'s study¹¹⁾ on the diagnosis of tooth-marked tongue using a convolution artificial intelligence network. Also, the accuracy of TN that was correctly diagnosed with no-

marked tongue was similar. In conclusion, it was observed that the Teachable Machine used for diagnosis of the tooth-marked tongue could learn an artificial intelligence model with sufficient clinical significance. In deep learning artificial intelligence training, it is essential to secure sufficient data. Therefore, it is thought that more quality data would need to be collected to obtain clinically meaningful diagnosis results.

In this study, the following results were obtained for diagnosing tooth-marked tongue through hyper-parameter tuning using the Teachable Machine as a web-based artificial intelligence model learning tool. The optimal conditions for artificial intelligence learning in the Teachable Machine were the epoch of 75, the batch size of 128, and the learning rate of 0.0001. At this time, the diagnostic accuracy of tooth-marked tongue and no-marked tongue was 92.1% and 72.6%, respectively, and the sensitivity (TPR) and specificity (FPR) were 0.92 and 0.28, respectively. Compared to the artificial intelligence learning model for diagnosing tooth-marked tongue trained by the convolution artificial network, the Teachable Machine's learning results were better, confirming the possibility of using the Teachable Machine in dental clinics. To diagnose various cases with artificial intelligence in dental clinics, above all, sufficient clinical data must be secured for each case. If sufficient data can be obtained, it can also be used as an educational tool for Dental Hygiene students.

Notes

Conflict of interest

No potential conflict of interest relevant to this article was reported.

Ethical approval

This study's clinical data already public information on the web and does not include personal information, so IRB is not required.

Author contributions

Conceptualization, Data acquisition, Formal analysis, Funding, Supervision, Writing-original draft, Writing-review & editing: Hyunja Jeong.

ORCID

Hyunja Jeong, <https://orcid.org/0000-0003-0540-1893>

References

1. López-Úbeda P, Díaz-Galiano MC, Martín-Noguerol T, Ureña-López A, Martín-Valdivia MT, Luna A: Detection of unexpected findings in radiology reports: a comparative study of machine learning approaches. *Expert Syst Appl* 160: 113647, 2020.
<https://doi.org/10.1016/j.eswa.2020.113647>
2. Chamunyonga C, Edwards C, Caldwell P, Rutledge P, Burbery J: The impact of artificial intelligence and machine learning in radiation therapy: considerations for future curriculum enhancement. *J Med Imaging Radiat Sci* 51: 214-220, 2020.
<https://doi.org/10.1016/j.jmir.2020.01.008>
3. Abdalla-Aslan R, Yeshua T, Kabla D, Leichter I, Nadler C: An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography. *Oral Surg Oral Med Oral Pathol Oral Radiol* 130: 593-602, 2020.
<https://doi.org/10.1016/j.oooo.2020.05.012>
4. Lee JH, Kim DH, Jeong SN, Choi SH: Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent* 77: 106-111, 2018.
<https://doi.org/10.1016/j.jdent.2018.07.015>
5. Teachable machine: Train a computer to recognize your own images, sounds, & poses. Retrieved September 30, 2020, from <https://teachablemachine.withgoogle.com/> Google(2019, November 1).
6. Jang KH, Jeong HJ: *Oral anatomy*. Komoonsa, Seoul, pp.47-142, 2005.
7. Slavkin HC, Baum BJ: Relationship of dental and oral pathology to systemic illness. *JAMA* 284: 1215-1217, 2000.
<https://doi.org/10.1001/jama.284.10.1215>
8. Shi D, Tang C, Blackley SV, et al.: An annotated dataset of tongue images supporting geriatric disease diagnosis. *Data Brief* 32: 106153, 2020.
<https://doi.org/10.1016/j.dib.2020.106153>
9. Solos I, Liang Y: A historical evaluation of Chinese tongue diagnosis in the treatment of septicemic plague in the pre-antibiotic era, and as a new direction for revolutionary clinical research applications. *J Integr Med* 16: 141-146, 2018.
<https://doi.org/10.1016/j.joim.2018.04.001>
10. Hong SS: *Diagnosis of oriental medicines*. Koonja, Seoul, pp.54-55, 2009.
11. Li X, Zhang Y, Cui Q, Yi X, Zhang Y: Tooth-marked tongue recognition using multiple instance learning and CNN features. *IEEE Trans Cybern* 49: 380-387, 2019.
<https://doi.org/10.1109/TCYB.2017.2772289>
12. Tooth-marked-tongue: Dataset. Retrieved September 30, 2020, from <https://www.kaggle.com/clearhanhui/biyeshheji#> Kaggle(2020, March 11).
13. Kaggle: Tour machine learning and data science community. Retrieved September 30, 2020, from <https://www.kaggle.com/#> Google(2010, April 1).
14. Omar L, Ivrisimtzis I: Using theoretical ROC curves for analysing machine learning binary classifiers. *Pattern Recognit Lett* 128: 447-451, 2019.
<https://doi.org/10.1016/j.patrec.2019.10.004>
15. Cho KH, Kim JS, Hong JP, Eo GS: *Tongue diagnosis for clinicians: integrating oriental and western medicines*. Koonja, Seoul, pp.4-49, 2007.
16. Im YG: *Atlas of clinical diagnosis 2: tongue diagnosis*. Jungdam, Seoul, pp.74-76, 2003.
17. Ma J, Wen G, Wang C, Jiang L: Complexity perception classification method for tongue constitution recognition. *Artif Intell Med* 96: 123-133, 2019.
<https://doi.org/10.1016/j.artmed.2019.03.008>
18. Tania MH, Lwin K, Hossain MA: Advances in automated tongue diagnosis techniques. *Integr Med Res* 8: 42-56, 2019.
<https://doi.org/10.1016/j.imr.2018.03.001>