Published online in advance by J-STAGE

J Prosthodont Res. 2023; **(**): ****–****

OPEN ACCESS JPR Prosthodontic Research

Journal of Prosthodontic Research

Original article

# Developing tongue coating status assessment using image recognition with deep learning

Jumpei Okawa [a], Kazuhiro Hori [a,*], Hiromi Izuno [b], Masayo Fukuda [c], Takako Ujihashi [a,c], Shohei Kodama [a], Tasuku Yoshimoto [a], Rikako Sato [a], Takahiro Ono [a,d]

[a] Division of Comprehensive Prosthodontics, Faculty of Dentistry & Graduate School of Medical and Dental Sciences, Niigata University, Niigata, Japan, [b] Department of Oral Health Sciences, Faculty of Nursing and Health Care, BAIKA Women's University, Ibaraki, Japan, [c] Department of Oral Health Science, Faculty of Health Science, Kobe Tokiwa University, Kobe, Japan, [d] Department of Geriatric Dentistry, Osaka Dental University, Osaka, Japan

## Abstract

**Purpose:** To build an image recognition network to evaluate tongue coating status.

**Methods:** Two image recognition networks were built: one for tongue detection and another for tongue coating classification. Digital tongue photographs were used to develop both networks; images from 251 (178 women, 74.7±6.6 years) and 144 older adults (83 women, 73.8±7.3 years) who volunteered to participate were used for the tongue detection network and coating classification network, respectively. The learning objective of the tongue detection network is to extract a rectangular region that includes the tongue. You-Only-Look-Once (YOLO) v2 was used as the detection network, and transfer learning was performed using ResNet-50. The accuracy was evaluated by calculating the intersection over the union. For tongue coating classification, the rectangular area including the tongue was divided into a grid of 7×7. Five experienced panelists scored the tongue coating in each area using one of five grades, and the tongue coating index (TCI) was calculated. Transfer learning for tongue coating grades was performed using ResNet-18, and the TCI was calculated. Agreement between the panelists and network for the tongue coating grades in each area and TCI was evaluated using the kappa coefficient and intraclass correlation, respectively.

**Results:** The tongue detection network recognized the tongue with a high intersection over union (0.885±0.081). The tongue coating classification network showed high agreement with tongue coating grades and TCI, with a kappa coefficient of 0.826 and an intraclass correlation coefficient of 0.807, respectively.

**Conclusions:** Image recognition enables simple and detailed assessment of tongue coating status.

**Keywords:** Tongue coating, Tongue coating index, Image recognition, Deep learning

Received 19 April 2023, Accepted 29 August 2023, Available online 28 September 2023

## 1. Introduction

Numerous microbes exist within the oral cavity, and the tongue is the site for bacterial accumulation[1]. Increased tongue coating results in the accumulation of oral microbes on the tongue, causing an increase in salivary microbes, and is associated with oral and systemic infections and aspiration pneumonia[2–5]. Since the tongue has functions in mastication and swallowing, decreased tongue motor function in older adults is associated with oral hypofunction, as proposed by the Japanese Society of Gerodontology[6]. In addition, oral hypofunction may lead to frailty and sarcopenia[7–10]. One of the diagnostic criteria for oral hypofunction[6] is poor oral hygiene, which can be assessed by a visual evaluation of the tongue coating[11]. However, a simple, objective, and detailed method for examining the status of tongue coating has not yet been established.

Tongue coating adhesions are common, especially in older adults, and are associated with decreased tongue function[6,9,10,12,13], which necessitates the evaluation of many individuals under a wide range of circumstances, including in-home care, institutional residents, and hospitalized patients. As tongue coating status can be observed on the surface of the tongue, several visual evaluation methods have been reported[11,13–17]. The Tongue Coating Index (TCI) divides the tongue surface into nine areas and visually scores and evaluates the degree of tongue coating[11]. The TCI is associated with the total anaerobic bacteria count and oral function and is used to examine oral hypofunction[6,8–11,18]. However, visual evaluation varies depending on the evaluator, and it is difficult to obtain more detailed scores.

Artificial intelligence for the detection and classification of objects in images has undergone considerable development, and its accuracy has been improved through deep learning. Image recognition technology is frequently applied in the medical and dental fields, with diagnostic support from medical imaging modalities such as computed tomography (CT), magnetic resonance imaging (MRI), dental panoramic radiographs, and intraoral photographs, and measurement of vital signs from facial photographs taken with general

**DOI:** https://doi.org/10.2186/jpr.JPR_D_23_00117

***Corresponding author:** Kazuhiro Hori, Division of Comprehensive Prosthodontics, Faculty of Dentistry & Graduate School of Medical and Dental Sciences, Niigata University, 2-5274 Gakkocho-dori, Chuo-ku, Niigata, Japan.
**E-mail address:** hori@dent.niigata-u.ac.jp

digital cameras are examples[19–24]. A digital camera can be used to obtain photographs of the tongue, as the surface of the tongue can be observed from outside the oral cavity. Previous studies have used a dedicated camera for collecting images of the tongue and assessed in relation to traditional Chinese medicine, diabetes, and non-alcoholic fatty liver disease[25–28]. Additionally, a deep learning model has been developed for diagnosing tongue conditions in traditional Chinese medicine using images obtained with a smartphone camera[29]. However, studies describing the application of deep learning to digital camera images for the evaluation of tongue coating status related to oral hypofunction is not known.

Because tongue coating status reflects oral hygiene and may lead to oral hypofunction and frailty, developing a method for evaluating tongue coating may assist in the early detection of such conditions. We hypothesized that a simple, objective, and detailed evaluation can be established through the fusion of visual evaluation by evaluators using TCI and image recognition using deep learning. To achieve this evaluation, we used two networks: one for tongue detection and the other for the classification of tongue coating by area. Therefore, this study aimed to generate 1) an object detection network that detects the tongue surface from a photograph and 2) a classification network that classifies the degree of tongue coating based on the TCI and verifies its accuracy.

## 2. Materials and Methods

### *2.1. Participants*

In this study, data collection and network generation were performed in two parts: 1) tongue detection network and 2) tongue coating classification network. The tongue detection network and tongue coating classification network included 251 (178 women, 73 men, 74.7 ± 6.6 years) and 144 older adults (83 women, 61 men, 73.8 ± 7.3 years) who were willing to volunteer, respectively. All participants lived independently in the community. Participants with a history of cerebrovascular disorder, dementia, neuromuscular disease, tumors, postoperative disorders with structural changes and symptoms in the tongue, head and neck tumors, and those on medications for oral hypofunction were excluded. This study was conducted in conformance with the standards of the Declaration of Helsinki and was approved by the Ethics Committee of Niigata University (2020-0044). All the participants provided written informed consent to participate in this study.

### *2.2. Image collection*

Tongue images from the participants were acquired using a digital camera (EyeSpecial C-II, SHOFU Inc.) with the tongue in a protruded position. Images were collected from each participant on two different days. Some participants attended only one of the two days of the study; therefore, the total number of images collected was 443 and 215 for the tongue detection and tongue coating networks, respectively. Images were captured using autofocus and flash. The shutter speed, aperture, ISO, and focal length were set automatically. The photographs were imported into MATLAB version R2021a (The MathWorks Inc., Natick, MA, USA) and used to generate networks. The workstation was equipped with an Intel Core i7-9750H CPU @ 2.60 GHz, 32 GB DDR4 RAM 2666 MHz, and an NVIDIA GeForce GTX 1650 GPU with 896 CUDA cores. Windows 11 Pro was the operating system used.

### *2.3. Tongue detection network*

#### *2.3.1. Image dataset*

Manual annotation was performed on 443 images using MATLAB (Computer Vision Toolbox). A rectangular bounding box region of interest was drawn to encompass the entire tongue. The position, width, and height of each bounding box were output as a dataset linked to the image and used to train and validate the network.

#### *2.3.2. Learning*

The You-Only-Look-Once (YOLO) v2 network architecture was selected as the object detection network[30]. Of the 443 annotated images, 60%, 10%, and 30% were used as training data, validation data, and test data, respectively. The validation data were used to adjust the hyperparameters during learning, and the test data were used to evaluate the accuracy of the networks after learning. The following data augmentation was applied to the training data at every epoch: random horizontal flipping, random aspect ratio scaling, and random alteration of the pixel color (contrast and brightness saturation).

Transfer learning was performed for ResNet-50[31], a convolutional neural network, and the feature extraction layer was set to ReLU layer 40, with the subsequent layers replaced with detection subnetworks. The input image size was converted to 224 × 224 pixels. The hyperparameters were set as follows: batch size, 16; number of epochs, 40; and learning rate, 0.001.

#### *2.3.3. Verification*

The input size of the test data was 2080 × 1560 pixels, which was the size of the photographs. If the network detected multiple bounding boxes, the box with the highest confidence score was selected. The confidence score is an index of the extent to which the bounding box includes the tongue surface, and is evaluated using the object detection network itself[32]. The accuracy was evaluated by calculating the intersection over union (IoU) for each result. IoU represents the overlap ratio between the true bounding box and network-estimated bounding box, and the closer the value is to 1, the better the match[33].

In addition, precision-recall curves were created from the precision and recall with IoU thresholds of 0.50 and 0.75, and the average precision was calculated from the area under the precision-recall curve. The formulae for precision and recall are as follows:

$$Presicion = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{All\ Positive\ Instances}$$

True positives represent the number of test data points detected by the network above the set IoU, that overlapped sufficiently with the true bounding box and network-estimated bounding box. False positives represent the number of test data points detected by the network under the set IoU, that overlapped poorly with the true bounding box and the network-estimated bounding box and were determined not to have been detected correctly. “All positive

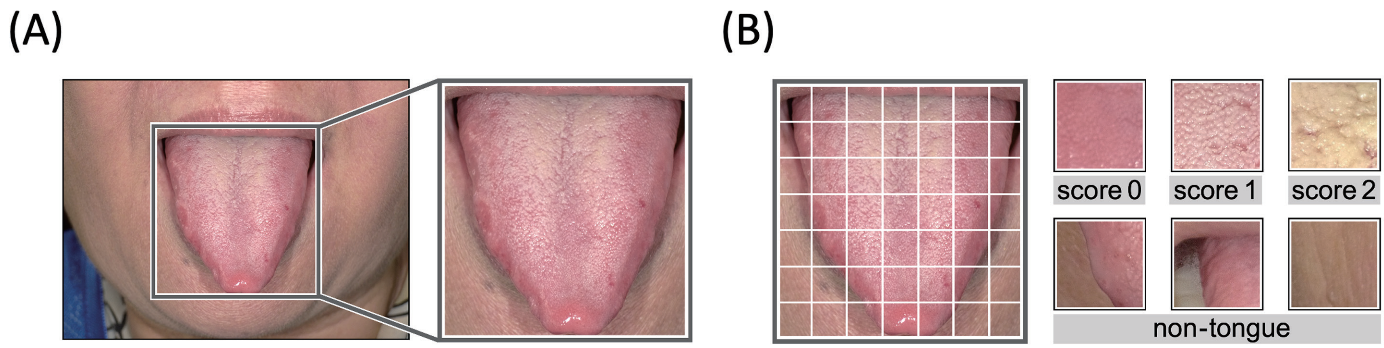


**Fig. 1.** Procedure for labeling the degree of tongue coating. (A): The rectangular area of the photograph containing the tongue cropped using a tongue-detection network. (B): Images divided into 7 × 7 = 49 areas, and sample images showing scoring of the degree of tongue coating.

**Table 1.** Counts for each tongue coating grades and non-tongue labels in the training, validation, and test datasets

| | Tongue coating grade | | | | | | |
|---|---|---|---|---|---|---|---|
| Label | 0 | 1 | 2 | 3 | 4 | non-tongue | Total |
| Training Data | 510 | 635 | 653 | 776 | 527 | 2285 | 5386 |
| Validation Data | 218 | 272 | 280 | 332 | 226 | 979 | 2307 |
| Test Data | 259 | 264 | 266 | 258 | 204 | 1591 | 2842 |

instances" represents the number of all truth bounding boxes in the test data, which was similar to the number of test data (30% of the 433 annotated images).

*2.4. Tongue coating classification network*

*2.4.1. Image dataset and labeling*

The training and validation data consisted of 157 images from 97 participants, and the test data consisted of 58 images from 37 participants. The same participants were not included in either dataset. **Figure 1** shows the procedure for labeling the degree of tongue coating. The tongue detection network was used to detect the rectangular region containing the tongue in each image, which was programmatically divided into 49 areas (7 × 7). The tongue detection network could detect the tongue area in all datasets, and 7693 images (157 images × 49 areas) were generated for the training and validation data and 2842 images (58 images × 49 areas) for the test data.

To label all areas based on the degree of tongue coating, each area was evaluated by a panel of five dentists specializing in geriatric dentistry (years of experience: 2–9 years; median: 6 years). The image presented to the panelists for evaluation was a rectangular area, including the tongue surface, divided into 49 areas, as shown in **Figure 1B**. First, the panelists classified each area as "tongue" or "non-tongue." Non-tongue areas were defined as those where less than two-thirds of the area was occupied by the tongue region and were labeled "non-tongue" if more than one panelist selected that classification. For the areas classified as the tongue, the degree of tongue coating in each area was evaluated according to three grades (scores 0, 1, and 2) based on the TCI. Then, the degree of tongue coating in each area was subdivided into the following five grades according to the average values of the five panelists: score 0, average value < 0.4; score 1, < 0.8; score 2, < 1.2; score 3, < 1.6; and score 4, ≤ 2. Therefore, the areas were labeled with one of six types of labels: five grades of tongue coating and "non-tongue."

*2.4.2. Learning*

For each of the tongue coating grades and non-tongue in the 7693 images, 70% were used as the training data and 30% were used as the validation data. **Table 1** shows the number of areas labeled with each of the five grades of tongue coating and non-tongue in the training, validation, and test data. The following data augmentation was applied to the training data at every epoch: random vertical and horizontal translations, random horizontal flipping, random rotation, and random aspect ratio scaling.

Transfer learning was performed using ResNet-18[31]. The last learning layer (fc1000 layer) and the final classification layer were replaced with new layers adapted to the new dataset. The input image was converted to 224 × 224 pixels. The hyperparameters were set as follows: batch size, 32; number of epochs, 30, and learning rate, 0.001.

*2.4.3. Verification*

To verify the validity of the training data, Fleiss' kappa[34] was used to calculate the inter-rater agreement between the five panelists for the three-grade tongue-coating evaluation in each area.

The agreement between the panelists and the tongue coating classification network was also calculated. The unweighted Cohen's kappa coefficient[35] was calculated for the "tongue" or "non-tongue" classification in each area. A quadratic-weighted Cohen's kappa[36] was calculated for the classification of the five-grade tongue coatings in each area.

For each of the panelists and the tongue coating classification network, the TCI was calculated from the degree of tongue coating in each area as follows:

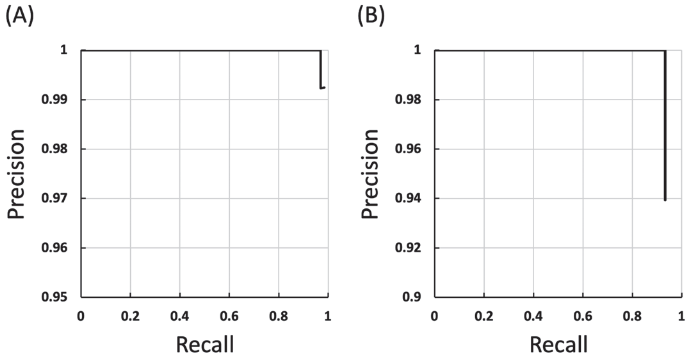


**Fig. 2.** Precision-recall curves with the intersection over union threshold set to 0.5 (A) and 0.75 (B)

$$TCI = \frac{Number\ of\ score1 \times 1 + Number\ of\ score2 \times 2 + Number\ of\ score3 \times 3 + Number\ of\ score4 \times 4}{(Number\ of\ score0-4) \times 4} \times 100[\%]$$

The numerator is the total score obtained, and the denominator is the maximum possible score. The intraclass correlation coefficient (ICC)[37] was calculated for the TCI between the panelist score and tongue coating network.

Fleiss' kappa, Cohen's kappa, and ICC were calculated at a significance level of α = 0.05. All statistical analyses were performed using MATLAB version R2021a.

## 3. Results

### *3.1. Tongue detection network*

The tongue was not detected in 1 of the 133 entries in the test dataset, an image in which the tongue occupied only approximately 10% of the photographs, which was the lowest in the dataset (the second lowest was approximately 16%). The average IoU, which is the ratio between the true bounding box of the tongue and the bounding box that the network estimated for the tongue, was 0.880 (95% confidence interval, 0.866–0.894; standard deviation, 0.081). The average precisions with IoU thresholds of 0.50 and 0.75 were 0.985 and 0.932, respectively (the maximum value was 1) (**Fig. 2**). Examples of the images obtained after applying the tongue detection network are shown in **Figure S1**.

### *3.2. Tongue coating classification network*

In the three-grade evaluation of tongue coating used to generate the training data, the inter-rater agreement between the five panelists was moderate, with a mean kappa coefficient of 0.554 (95% confidence interval: 0.535–0.573, $P < 0.001$; standard deviation 0.121) [38].

**Figure 3** shows the confusion matrix between the panelists' evaluations and the labels given by the network. The agreement between the tongue and non-tongue was almost perfect, with a kappa coefficient of 0.881 (confidence interval, 0.863–0.898; $P < 0.001$)[38]. In the area classified as tongue, the agreement for the five-grade tongue coating was almost perfect, with a kappa coefficient of 0.826 (95% confidence interval 0.774–0.877, $P < 0.001$)[38].

| | | Tongue coating score evaluated by network | | | | | |
|---|---|---|---|---|---|---|---|
| | | tongue area and scores | | | | | non-tongue |
| | | 0 | 1 | 2 | 3 | 4 | |
| Tongue coating score evaluated by panelists: tongue area and scores | 0 | 165 | 54 | 10 | 0 | 0 | 30 |
| | 1 | 39 | 146 | 61 | 2 | 0 | 16 |
| | 2 | 5 | 64 | 149 | 42 | 0 | 6 |
| | 3 | 1 | 18 | 101 | 102 | 24 | 12 |
| | 4 | 0 | 1 | 15 | 86 | 89 | 13 |
| | non-tongue | 53 | 18 | 10 | 5 | 4 | 1501 |

**Fig. 3.** Confusion matrix between the evaluations by the panelists and the labels given by the network. Dark gray backgrounds indicate matched evaluations. The Cohen's kappa between tongue area and non-tongue was 0.881 ($P < 0.001$). The Cohen's kappa between scores 0, 1, 2, 3, and 4 was 0.826 ($P < 0.001$).

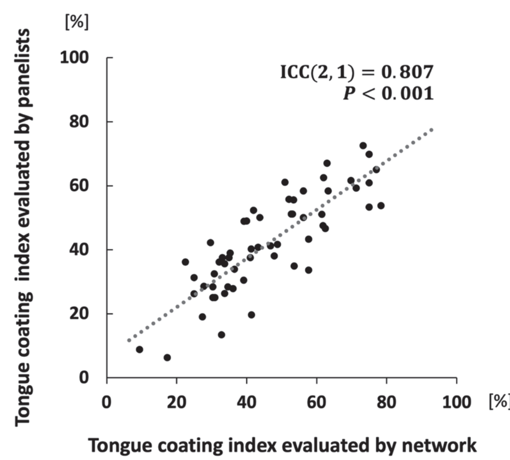


**Fig. 4.** Agreements for the tongue coating index between the panelists and the network ICC(2,1) = 0.807 ($P < 0.001$)

The ICC for the TCI between the panelists and the network was 0.807 (95% confidence interval 0.654–0.890, $P < 0.001$; **Fig. 4**).

## 4. Discussion

In this study, we developed an image recognition network based on deep learning to evaluate tongue coating status from digital photographs and verified its accuracy in comparison with a panel of dentists. The results showed that the object detection network could detect the tongue surface and that the image classification network could evaluate the tongue coating status and provide detailed scoring for each area of the tongue with high accuracy.

### *4.1. Methodological Considerations*

Tongue coating is often observed in older individuals with impaired oral function and reflects poor oral hygiene. Therefore, this study included older adults willing to volunteer, and the panelists evaluating tongue coating were dentists who specialized in geriatric dentistry and were sufficiently trained in the visual assessment of tongue coating adhesion. The inter-rater agreement for the three-grade tongue coating for each area was consistent with that of previous studies[11], although the number of areas differed and non-tongue areas were included. Furthermore, to compensate for small discrepancies in evaluations among the panelists, the average score of the five panelists for tongue coating was calculated and subdivided into five grades. A more detailed evaluation was performed by dividing the rectangular area, including the tongue surface, into 49 areas.

The network architecture and the convolutional neural network were chosen from among the possible networks that were officially supported by the MATLAB software and could operate on this hardware. The important characteristics of networks are speed and accuracy; therefore, we chose networks with high processing speed and precision, according to previous reports[39,40]. In the hyperparameter settings, the batch size was the maximum that the hardware could work with and was in line with the values generally used in deep learning, and the number of epochs was set such that the value of the loss function nearly converged. Tongue images collected using a general camera can be affected by the shape of the tongue and features of the environment in which the photograph is taken, such as the brightness of the location and distance from the subject. Therefore, data augmentation was performed for each epoch to increase the size of the training dataset and provide examples with a range of features to prevent errors and overfitting. Furthermore, two models were generated from images for tongue detection and tongue coating classification. Studies reporting high accuracy in the detection of the tongue and assessment of tongue conditions using deep learning for photographs that include the tongue, validate our study findings[25–29].

### *4.2. Tongue detection network*

The IoU indicates the overlap ratio between the truth and detection ranges and was calculated to verify the accuracy of tongue detection. In general, when calculating the average precision, the IoU threshold is set to 0.5, to account for the uncertainty in the bounding-box annotation[33]. When the IoU threshold is set higher than 0.5, the average precision drops sharply[33]. In this study, the average IoU of the bounding boxes detected by the network was high, and a higher IoU threshold was set to calculate the average precision. As a result, a high average precision was obtained, even at an IoU threshold of 0.75, indicating that this network could detect the tongue with very high accuracy, even from images captured with a standard digital camera. In contrast, the IoU was lower in images, with a smaller area occupied by the tongue surface. YOLO v2 has the lowest computational complexity, faster learning, and detection speed; however, it is not effective in detecting small objects[40]. Using a newer version may result in more accurate detection.

### *4.3. Tongue coating classification network*

The agreement between the tongue or non-tongue and tongue coating grades in each area was almost perfect. The non-tongue areas included the lips, skin, and dentition. Tongue coating is an accumulation that covers the tongue surface and can be observed as a change in color, tone, or roughness; however, the degree and extent of these changes are not regular[17]. Although only one tongue coating classification network was created to classify the tongue, non-tongue, and tongue coating grades, the network could simultaneously recognize various changes on the tongue surface.

Almost all tongue coating grades in each area classified by the network were correct or within one grade of the panel classification. Compared with the conventional visual 3-grade evaluation, the network provides evaluations with detailed grades and areas. In addition, the intraclass correlation verification for the TCI, which is an evaluation of the tongue coating status of the entire tongue[11], showed high agreement between the panelists and the network. TCI is one of the items used in the examination of oral hypofunction, and the image recognition network has the potential to enable anyone to simply examine tongue coating status. Furthermore, since the variability among panelists was suppressed, even panelists without experience in dentistry could evaluate the tongue coating status in detail.

### *4.4. Limitations*

This study focused on older adults living independently in the community. Future research including a wide range of people with severe deterioration in oral hygiene and function, such as patients requiring nursing care and those in the acute phase, is warranted[41,42]. Further verification is required for the application of the networks to patients with a history of tumors and postoperative disorders with structural changes and symptoms of the tongue, who were excluded. In addition, the panelists comprised five dentists; therefore, increasing the number of panelists is needed to examine for errors in the training data. A tongue coating classification network was generated based on the results of the tongue detection network. The decrease in the accuracy of detection of images with a smaller area occupied by the tongue surface might have affected the learning and accuracy of classification. Because of the availability of different types of cameras, further research must be performed using other networks and applications to support various cameras, including smartphones. This will increase the number of images available for training and testing and include diverse study populations and tongue changes. In this study, we performed data augmentation to make the network more robust to variations in photography environments, equipment, overlearning, and patient characteristics[43].

Despite the above limitations, to our knowledge, this is the first method to evaluate the tongue coating status using a conventional digital camera. The findings of this study suggest that the tongue surface can be automatically detected from photographs and that the tongue coating status can be evaluated in more detail by dividing the tongue into numerous areas. Tongue coating affects oral hygiene and can lead to aspiration pneumonia, oral hypofunction, and frailty. The simple and detailed examination of tongue coating status may assist in the maintenance of oral function and the early detection of predisposing conditions. Further research is needed to validate our findings across different participant groups, photography equipment, and photography environments to expand the range of utilization.

## 5. Conclusions

The tongue coating status was evaluated from digital tongue photographs using deep learning-based image recognition technology. The networks created automatically detected the tongue surface with high accuracy and evaluated the tongue-coating grades in high agreement with a panel of dentists. This evaluation method using image recognition technology can perform a simple, objective, and detailed evaluation of the tongue coating status by learning from data collected from diverse participant groups and environments.

## Acknowledgments

This study was supported by JSPS KAKENHI (grant number: 20K18593). We thank Edanz (https://jp.edanz.com/ac) for editing this manuscript.

## Conflict of interest

All authors declare no conflict of interest.

## References

[1] Danser MM, Gómez SM, Van der Weijden GA. Tongue coating and tongue brushing: a literature review. Int J Dent Hyg. 2003;1:151–8. https://doi.org/10.1034/j.1601-5037.2003.00034.x, PMID:16451515

[2] Ryu M, Ueda T, Saito T, Yasui M, Ishihara K, Sakurai K. Oral environmental factors affecting number of microbes in saliva of complete denture wearers. J Oral Rehabil. 2010;37:194–201. https://doi.org/10.1111/j.1365-2842.2009.02042.x, PMID:20050985

[3] Abe S, Ishihara K, Adachi M, Okuda K. Tongue-coating as risk indicator for aspiration pneumonia in edentate elderly. Arch Gerontol Geriatr. 2008;47:267–75. https://doi.org/10.1016/j.archger.2007.08.005, PMID:17913263

[4] Abe S, Ishihara K, Adachi M, Sasaki H, Tanaka K, Okuda K. Professional oral care reduces influenza infection in elderly. Arch Gerontol Geriatr. 2006;43:157–64. https://doi.org/10.1016/j.archger.2005.10.004, PMID:16325937

[5] Zhao Y, Gao X, Guo J, Yu D, Xiao Y, Wang H, *et al. Helicobacter pylori* infection alters gastric and tongue coating microbial communities. Helicobacter. 2019;24:e12567. https://doi.org/10.1111/hel.12567, PMID:30734438

[6] Minakuchi S, Tsuga K, Ikebe K, Ueda T, Tamura F, Nagao K, *et al.* Oral hypofunction in the older population: Position paper of the Japanese Society of Gerodontology in 2016. Gerodontology. 2018;35:317–24. https://doi.org/10.1111/ger.12347, PMID:29882364

[7] Nomoto A, Shimizu A, Ohno T, Tohara H, Hashidume M, Hatano M, *et al.* Poor oral health and anorexia in older rehabilitation patients. Gerodontology. 2022;39:59–66. https://doi.org/10.1111/ger.12600, PMID:34687077

[8] Hatayama C, Hori K, Izuno H, Fukuda M, Sawada M, Ujihashi T, *et al.* Features of masticatory behaviors in older adults with oral hypofunction: A cross-sectional study. J Clin Med. 2022;11:5902. https://doi.org/10.3390/jcm11195902, PMID:36233767

[9] Yoshida M, Hiraoka A, Takeda C, Mori T, Maruyama M, Yoshikawa M, *et al.* Oral hypofunction and its relation to frailty and sarcopenia in community-dwelling older people. Gerodontology. 2022;39:26–32. https://doi.org/10.1111/ger.12603, PMID:34727388

[10] Iwasaki M, Motokawa K, Watanabe Y, Shirobe M, Ohara Y, Edahiro A, *et al.* Oral hypofunction and malnutrition among community—dwelling older adults: evidence from the Otassha study. Gerodontology. 2022;39:17–25. https://doi.org/10.1111/ger.12580, PMID:34212426

[11] Shimizu T, Ueda T, Sakurai K. New method for evaluation of tongue-coating status. J Oral Rehabil. 2007;34:442–7. https://doi.org/10.1111/j.1365-2842.2007.01733.x, PMID:17518979

[12] Avcu N, Kanli A. The prevalence of tongue lesions in 5150 Turkish dental outpatients. Oral Dis. 2003;9:188–95. https://doi.org/10.1034/j.1601-0825.2003.02933.x, PMID:12974518

[13] Mantilla Gómez S, Danser MM, Sipos PM, Rowshani B, Van Der Velden U, Van Der Weijden GA. Tongue coating and salivary bacterial counts in healthy/gingivitis subjects and periodontitis patients. J Clin Periodontol. 2001;28:970–8. https://doi.org/10.1034/j.1600-051x.2001.028010970.x, PMID:11686816

[14] Kojima K. Clinical studies on the coated tongue. Japanese Journal of Oral & Maxillofacial Surgery. 1985;31:1659–78. https://doi.org/10.5794/jjoms.31.1659

[15] Yonezawa T, Michishige F. Literature review on trend, and assessment criteria and their reliability and reproducibility of tongue coating assessment methods. Osaka Med Coll J Nurs Res. 2018;8:116–20.

[16] Chalmers JM, King PL, Spencer AJ, Wright FAC, Carter KD. The oral health assessment tool--validity and reliability. Aust Dent J. 2005;50:191–9. https://doi.org/10.1111/j.1834-7819.2005.tb00360.x, PMID:16238218

[17] Haresaku S, Yamamoto M, Hanioka T. Association of tongue coating with lifestyle and subjective symptoms. J Dent Hlth. 2007;57:662–70. https://doi.org/https://doi.org/10.5834/jdh.57.5_662

[18] Takahashi M, Kishi M. Relationship between dental pathogenic bacteria in tongue coat and clinical oral status. J Dent Hlth. 2006;56:137–47. https://doi.org/10.5834/jdh.56.2_137

[19] Patil O, Wang W, Gao Y, Jin Z. MobiEye: turning your smartphones into a ubiquitous unobtrusive vital sign monitoring system. CCF Transactions on Pervasive Computing and Interaction. 2020;2:97–112. https://doi.org/10.1007/s42486-020-00033-3

[20] Mahdi FP, Motoki K, Kobashi S. Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs. Sci Rep. 2020;10:19261. https://doi.org/10.1038/s41598-020-75887-9, PMID:33159125

[21] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, *et al.* A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88. https://doi.org/10.1016/j.media.2017.07.005, PMID:28778026

[22] Kishimoto T, Goto T, Matsuda T, Iwawaki Y, Ichikawa T. Application of artificial intelligence in the dental field: A literature review. J Prosthodont Res. 2022;66:19–28. https://doi.org/10.2186/jpr.JPR_D_20_00139, PMID:33441504

[23] Sakai T, Li H, Shimada T, Kita S, Iida M, Lee C, *et al.* Development of artificial intelligence model for supporting implant drilling protocol decision making. J Prosthodont Res. 2023;67:360–5. https://doi.org/10.2186/jpr.JPR_D_22_00053, PMID:36002334

[24] Takahashi T, Nozaki K, Gonda T, Ikebe K. A system for designing removable partial dentures using artificial intelligence. Part 1. Classification of partially edentulous arches using a convolutional neural network. J Prosthodont Res. 2021;65:115–8. https://doi.org/10.2186/jpr.JPOR_2019_354, PMID:32938860

[25] Chen Z, Zhang X, Qiu R, Sun Y, Zheng R, Pan H, *et al.* Application of artificial intelligence in tongue diagnosis of traditional Chinese medicine: A review. TMR Modern Herbal Medicine. 2021;4:14. https://doi.org/10.53388/MHM2021B0329001

[26] Jiang T, Guo X, Tu L, Lu Z, Cui J, Ma X, *et al.* Application of computer tongue image analysis technology in the diagnosis of NAFLD. Comput Biol Med. 2021;135:104622. https://doi.org/10.1016/j.compbiomed.2021.104622, PMID:34242868

[27] Qi Z, Tu L, Chen J, Hu X, Xu J, Zhang Z. The classification of tongue colors with standardized acquisition and ICC profile correction in traditional chinese medicine. BioMed Res Int. 2016;2016:1–9. https://doi.org/10.1155/2016/3510807, PMID:28050555

[28] Zhang J, Xu J, Hu X, Chen Q, Tu L, Huang J, *et al.* Diagnostic method of diabetes based on support vector machine and tongue images. BioMed Res Int. 2017;2017:1–9. https://doi.org/10.1155/2017/7961494, PMID:28133611

[29] Ma C, Gao C, Hou S, Li S. Development of attention-based robust deep learning model for tongue diagnosis by smartphone. bioRxiv. 2023:2023.02.02.526804. https://doi.org/10.1101/2023.02.02.526804

[30] Redmon J, Farhadi A. YOLO9000_ Better, Faster, Stronger. 2017. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR):pp. 7263-71.

[31] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016. In: Proceedings of the IEEE conference on computer vision and pattern recognition:pp. 770-8.

[32] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. 2016. In: Proceedings of the IEEE conference on computer vision and pattern recognition:779-88.

[33] Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal visual object classes (VOC) challenge. Int J Comput Vis. 2010;88:303–38. https://doi.org/10.1007/s11263-009-0275-4

[34] Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull. 1971;76:378–82. https://doi.org/10.1037/h0031619

[35] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37–46. https://doi.org/10.1177/001316446002000104

[36] Cohen J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychol Bull. 1968;70:213–20. https://doi.org/10.1037/h0026256, PMID:19673146

[37] Bartko JJ. The intraclass correlation coefficient as a measure of reliability. Psychol Rep. 1966;19:3–11. https://doi.org/10.2466/pr0.1966.19.1.3, PMID:5942109

[38] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74. https://doi.org/10.2307/2529310, PMID:843571

[39] Bianco S, Cadene R, Celona L, Napoletano P. Benchmark analysis of representative deep neural network architectures. IEEE Access. 2018;6:64270–7. https://doi.org/10.1109/ACCESS.2018.2877890

[40] Jiang P, Ergu D, Liu F, Cai Y, Ma B. A review of Yolo algorithm developments. Procedia Comput Sci. 2022;199:1066–73. https://doi.org/10.1016/j.procs.2022.01.135

[41] Adachi M, Ishihara K, Abe S, Okuda K. Professional oral health care by dental hygienists reduced respiratory infections in elderly persons requiring nursing care. Int J Dent Hyg. 2007;5:69–74. https://doi.org/10.1111/j.1601-5037.2007.00233.x, PMID:17461957

[42] Danckert R, Ryan A, Plummer V, Williams C. Hospitalisation impacts on oral hygiene: an audit of oral hygiene in a metropolitan health service. Scand J Caring Sci. 2016;30:129–34. https://doi.org/10.1111/scs.12230, PMID:25962409

[43] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learing. J Big Data. 2019;6:60. https://doi.org/10.1186/s40537-019-0197-0