PRTMTM: A Priori Regularization Method for Tooth-Marked Tongue Classification

Jingqiao Lu
Health Testing Technique And Equipment Research Center
Xin-Huangpu Joint Innovation Institute of Chinese Medicine
Guangzhou, China
lu.jingqiao@139.com

Mingxuan Liu
Department of Biomedical Engineering
Tsinghua University
Beijing, China
liumx19@mails.tsinghua.edu.cn

Hong Chen*
School of Integrated Circuits
Tsinghua University
Beijing, China
hongchen@tsinghua.edu.cn

***Abstract*—Tooth-marks on tongues usually indicate the weakness of the spleen and stomach in traditional Chinese medicine (TCM). Therefore, tooth-marked tongue classification is important to health diagnosis in TCM clinic. Existing classification methods usually do not use the prior knowledge such as the location and width of tooth-marks, resulting in easily misclassification of unremarkable tongues. In this paper, we propose a prior regularization tooth-marked tongue method (PRTMTM), which makes full use of the prior knowledge of the position and width of tooth-marks. With PRTMTM, the original tongue image is first segmented to obtain the tongue edge position. Then, the prior mask map of tooth-marks is obtained by corroding the tongue from edge to interior according to the tooth-mark width. Finally, with a proposed regularization method, the tooth-marked tongues are classified accurately in the training process together with the prior mask map of tooth-marks. To verify our method comprehensively, we build twenty-five sub-training sets with different number of images and label distributions. Compared with state-of-the-art methods, the accuracy of our method is improved by 4.78% on average and 10.61% on maximum, the AUC by 0.04 on average and 0.07 on maximum, and the generated heatmap can highlight tooth-marked regions.**

Keywords—traditional Chinese medicine, tooth-marked tongue classification, convolutional neural network, priori regularization

I. INTRODUCTION

In traditional Chinese medicine (TCM), tongue diagnosis is an essential component of TCM inspection diagnosis [1,2,3]. Tooth-marked tongue, also known as dentate-marked tongue, is typically caused by the tongue being fat, which is usually due to dysfunction of the liver, spleen and kidneys [4,5,6]. Accurate classification of the tooth-marked tongue is challenging because of the diversity of color, shape and texture of tooth-marks, which make them difficult to identify.

Many efforts have been made to classify tooth-marked tongues. In the early stages, most methods use convex hull [7,8,9,10] to extract hand-crafted features, and then use a machine learning classification model to determine whether there are tooth-marks. Li et al. [7] used concave features to find candidate tooth-marked regions, followed by extracting features and building a multi-instance classifier for classification. Wang et al. [8] calculated the slope of the tongue edge and the length of the concave region, and classified the tooth-marked tongue according to the indicator threshold. Shao et al. [9] adopted concave features and brightness to classify the tooth-marked tongue. Li et al. [10] used the curvature and gradient of tongue contour points to predict the number and degree of tooth-marks.

In recent decades, convolutional neural networks (CNNs) have been extensively employed in tooth-marked tongue classification by automatically extracting image features [11,12,13,14]. Sun et al. [11] designed a 7-layer CNN and used Grad-CAM [12] to visualize key regions in the classification of tooth-marked tongue. Wang et al. [13] constructed a new tooth-marked tongue dataset with 1548 images and used the pre-trained Resnet34 based on the ImageNet dataset to fine-tune the tooth-marked tongue dataset to get the classification result. Tang et al. [14] proposed a two-stage tongue classification method. In the first stage, the cascaded CNN was used to detect the tongue region and key points of the tongue edge, which were combined in the second stage for classification. In addition, some researchers focused on how to locate the position of tooth-marks using diverse object detection frameworks [15,16,17].

However, these early methods suffer from poor generalization due to hand-crafted features. The CNN-based methods do not use the prior knowledge of tooth-marks, which easily leads to the misclassification of tooth-marked tongues. To address these problems, in this paper we propose a priori regularization tooth-marked tongue method (PRTMTM) using the prior knowledge that tooth-marks are located on the edge of the tongue and the width of tooth-marks has a certain range. In the training process, an additional loss for the attention map is added to guide the model to pay attention to the region where tooth-marks are located, which improves the classification accuracy and enhances the interpretability of the trained model.

II. PROPOSED METHOD

A. Problem Formulation

Given the dataset of tongue images $X = \{x_1, \ldots, x_N\} \in R^{H\times W\times 3}$ and $Y = \{y_1, \ldots, y_N\} \in \{0, 1\}$, x_i is the i-th tongue image, y_i is the corresponding label, where 0 and 1 indicate that x_i belongs to the non-tooth-marked tongue and the tooth-marked tongue respectively. A tooth-marked tongue classification model f is trained to predict the probability $f(x_i)$ that x_i is considered as the tooth-marked tongue. $Loss(f(x_i),y_i)$ represents the error between the predicted value and the true value, and the total loss $Loss(X,Y)$ is obtained by summing the loss of all tongue images:

$$Loss(X,Y) = \sum_{i=1}^{N} Loss(f(x_i), y_i) \quad (1)$$

This work is supported by the National Science and Technology Major Project from Minister of Science and Technology, China (Grant No. 2018AAA0103100), Guangzhou Foshan Science and Technology Innovation Project (No. 2020001005585), and National Natural Science Foundation of China (No. 92164110 and U19B2041), partly supported by Beijing Engineering Research Center (No. BG0149) and Tsinghua National Laboratory for Information Science and Technology (No. 042003266).

2023 IEEE International Symposium on Circuits and Systems (ISCAS) | 978-1-6654-5109-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISCAS46773.2023.10181870

Ideally, the model *f* has better classification performance when the *Loss*(*X,Y*) is small.

B. Prior Mask Map of Tooth-Marks

The PRTMTM is built first by creating the prior mask map of tooth-marks, which is then used to apply regularization to the attention map in the training process, limiting the attention region of the attention map and guiding the model to learn the relevant features of tooth-marks.

For a given original tongue image *x* (Fig. 1 left), the tongue segmentation mask image *M0* (Fig. 1 middle) is obtained by the tongue image segmentation algorithm [18]. *M0* is then converted into a single-channel grayscale image *M1*. To get the prior mask map of tooth-marks, we introduce an eroded grayscale image *M2* with the help of the corrosion. During the corrosion, the 3 × 3 kernel *K* is convolved across the width and height of the grayscale image *M1*, computing the maximal pixel value overlapped by *K* and replacing the image pixel in the anchor point position with that maximal value,

$$M2(x,y) = \min_{K(x',y')\neq 0} M1(x+x', y+y') \tag{2}$$

Finally, the prior mask map of tooth-marks *M* (Fig. 1 right) is obtained by the following formula,

$$M = M0 \times (1.0 - M2) \tag{3}$$

Fig. 1. Original tongue image (left), tongue segmentation mask image (middle), prior mask map of tooth-marks (right)

C. Schematic of PRTMTM and Training Process

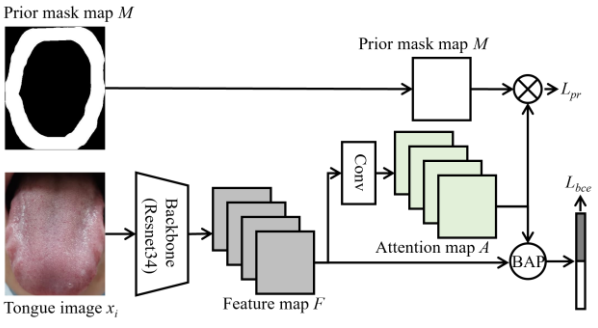


Fig. 2. Schematic of proposed PRTMTM

As we know, tooth-marks usually locate on the edge of the tongue and have a certain width. Therefore, the model performance can be improved by restricting the attention region on the attention map of the tongue image to focus on specific regions of the tongue edge. The schematic of proposed PRTMTM is shown in Fig. 2, from which we can see that a tongue image x_i and its prior mask map of tooth-marks *M* are first processed by a backbone network such as Resnet34 [19] in our work to get the feature map *F*:

$$F = CNN(x_i) = [F_1, ..., F_C] \in R^{C\times H\times W} \tag{4}$$

where *H*, *W* and *C* represents feature map's height, width and the number of channels respectively.

Next, the attention map *A* is obtained by convolution and up-sampling of the feature map *F*, which is then multiplied and summed by the corresponding position of the prior mask map *M* to obtain the prior regularization loss L_{pr} as follows:

$$A = Up(Conv(F)) = [A_1, ..., A_C] \in R^{C\times H\times W}) \tag{5}$$

$$L_{pr} = \frac{1}{C\times H\times W}\sum_{c=1}^{C}\sum_{h=1}^{H}\sum_{w=1}^{W}(1-M_{hw})\times A_{hw}^{c} \tag{6}$$

where 1-M_{hw} corresponds to the non-tooth-marked region. The smaller the loss L_{pr}, the less the model concerns the non-tooth-marked region. That is, the model focuses more on the tooth-marked region.

Then, we element-wise multiply feature map *F* by attention map *A* using Bilinear Attention Pooling (BAP, shown in Fig. 3) to obtain the final feature representation of x_i, which is used to get the binary cross-entropy loss of the predicted value and the true value L_{bce}:

$$L_{bce} = -\sum_{i=1}^{N} y_i \log p_i + (1-y_i)(1-\log p_i) \tag{7}$$

where y_i is the true value of the input x_i, 0 and 1 means x_i belongs to the non-tooth-marked and tooth-marked tongue respectively, and p_i is the probability of the tongue being predicted as the tooth-marked tongue, between 0 and 1.

With the previously defined loss L_{pr} and L_{bce}, the full loss *L* for our PRTMTM is summarized as:

$$L = \alpha \times L_{pr} + \beta \times L_{bce} \tag{8}$$

where α and β are loss weights, which are used to adjust the relative influence of two losses. We find that equal weights α=β=1.0 yield the best performance.

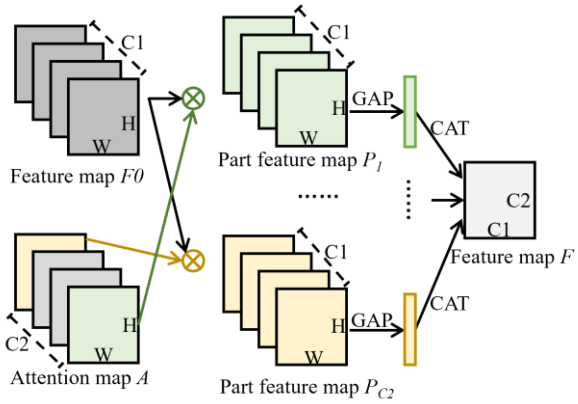


Fig. 3. Bilinear Attention Pooling (BAP)

D. Inference

The prior mask map of tooth-marks is only involved in the calculation of loss function, so there is no need to segment the tongue image, nor to generate the prior mask map of tooth-marks. The predicted label can be obtained by directly inputting the original tongue image into the trained model.

III. Experiments

A. Datasets and Experimental setup

In our work, 2362 tongue images are used, including 1215 tooth-marked tongue images and 1147 non-tooth-marked tongue images. Each tongue image is labeled by several senior TCM doctors, and the final tongue image label is determined by majority voting to exclude personalized labeling results by individual TCM doctors. The complete dataset is divided into an initial training set and a test set with the statistics shown in Table I.

TABLE I. Initial Training Set And Test Set

	No. of tooth-marked tongue images	No. of non-tooth-marked tongue images
Initial training set	917	917
Test set	298	230

TABLE II. Sub-training Sets With Different Number Of Images And Label Distributions

Group	Sub-training set	No. of tooth-marked tongues (*N0*)	No. of non-tooth-marked tongues (*N1*)	Label distribution (*N0:N1*)
1	1	183	917	1:5
1	2	458	917	1:2
1	**3**	**917**	**917**	**1:1**
1	4	917	458	2:1
1	5	917	183	5:1
2	6	128	641	1:5
2	7	320	641	1:2
2	**8**	**641**	**641**	**1:1**
2	9	641	320	2:1
2	10	641	128	5:1
3	11	91	458	1:5
3	12	229	458	1:2
3	**13**	**458**	**458**	**1:1**
3	14	458	229	2:1
3	15	458	91	5:1
4	16	55	275	1:5
4	17	137	275	1:2
4	**18**	**275**	**275**	**1:1**
4	19	275	137	2:1
4	20	275	55	5:1
5	21	18	91	1:5
5	22	45	91	1:2
5	**23**	**91**	**91**	**1:1**
5	24	91	45	2:1
5	25	91	18	5:1

To evaluate the effect of prior regularization more comprehensively, 25 sub-training sets with different number of images and label distributions are created based on the initial training set. As shown in Table II, these 25 sub-training sets are organized into 5 groups, and each group contains 5 sub-training sets. The number of images with the majority class varies between different groups. For example, the majority class in group 1 has the largest number of 917 tongue images and that in group 5 has the smallest number of 91 tongue images. Within a single group, the numbers of images with the majority class of 5 sub-training sets are the same, where the third sub-training set is a fully balanced dataset (marked in black bold) and the others are non-balanced datasets. It is important to note that the test set is unique and the label distribution of the test set is balanced.

B. Metrics for Evaluation

The precision, recall, F1-score, accuracy and area under curve of the receiver operating characteristic (AUC) [21] are used to evaluate the performance of each method on different sub-training sets.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{9}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{10}$$

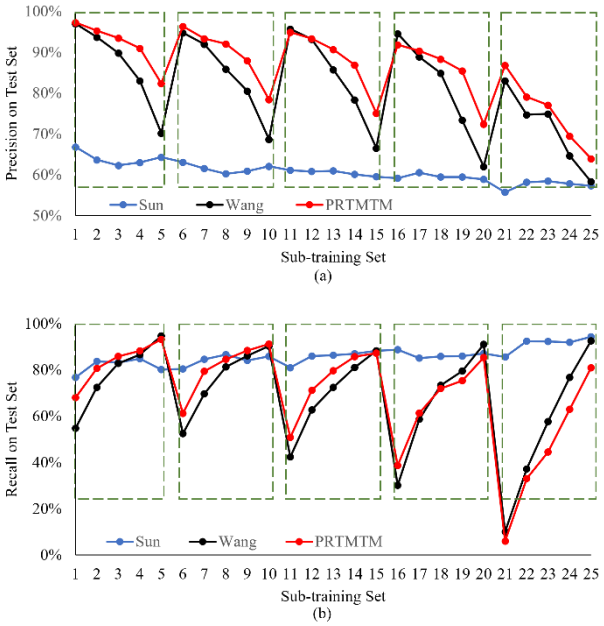
$$F1 - score = \frac{2 \times (\text{Precison} \times \text{Recall})}{\text{Precison} + \text{Recall}} \tag{11}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

where *TP* and *TN* represent the numbers of correctly predicted tooth-marked and non-tooth-marked tongues respectively, and *FP* and *FN* stand for the numbers of incorrectly predicted tooth-marked and non-tooth-marked tongues respectively. For these five metrics, larger metric values indicate better performance.

C. Experimental Results and Discussion

We train the model using Stochastic Gradient Descent (SGD) with the momentum of 0.9, epoch number of 200, weight decay of 1e-5, and a mini-batch size of 32. The initial learning rate is set to 0.001, with linear decay of 0.5 after every 50 epochs. We use Resnet34 [18] as the backbone to extract image features in PRTMTM. During the training phase, the input tongue image is randomly scaled to 230 × 230, then randomly cropped to 224 × 224, and augmented with data enhancement such as left-right flip, gaussian blur. In the testing phase, the input tongue image is scaled to 224 × 224 without any data enhancement.



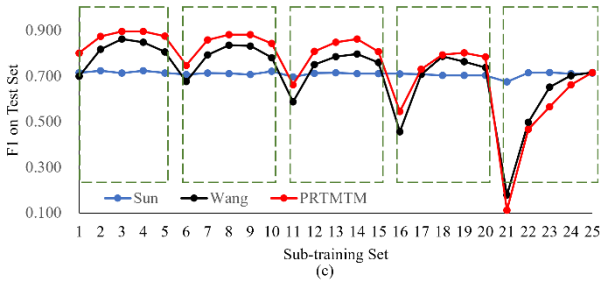


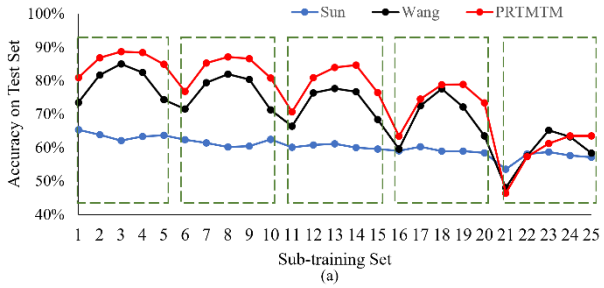
Fig. 4. Precision, recall and F1-score on test set. The green dashed box includes the metrics on different sub-training sets in the same group.

TABLE III. AVERAGE F1-SCORE ON ALL SUB-TRAINING SETS

	Sun	**Wang**	**PRTMTM**
Average F1-Score	0.709	0.713	**0.749**

To compare with state-of-the-art methods, we also conduct experiments on the datasets using Sun's method [11] and Wang's method [13] respectively, which have gained higher accuracy than previous methods. It should be emphasized that Wang's method also uses Resnet34 as the backbone, resulting in ablation studies with our PRTMTM.

First, the accuracy, recall and F1-score on the test set with the three models (Sun's, Wang's and our PRTMTM) trained on the 25 sub-training sets are shown in Fig. 4 (a), (b) and (c) respectively. The green dashed box includes the metrics on different sub-training sets in the same group. From Fig. 4 (a) and (b), we find that the precision of Sun's method is the lowest but its recall is very high, which indicates that with Sun's method the precision and recall can't be balanced simultaneously and hence the classification performance of the method is poor. With Wang's method and proposed PRTMTM, the precision in the same group decreases with the increase of recall. Therefore, only precision or recall does not represent overall performance. But F1-score is more comprehensive considering both precision and recall, which is illustrated in Fig. 4 (c) and Table III, from which we find that the F1-score of PRTMTM is 0.036 higher than that of Wang's method, indicating that the performance of the proposed PRTMTM is better than that of Wang's method.



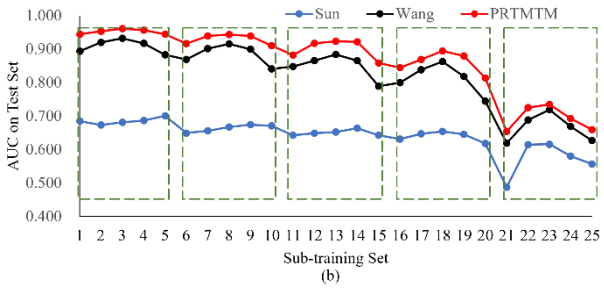


Fig. 5.Accuracy and AUC on test set

TABLE IV. EXPERIMENTAL NUMERICAL RESULTS OF ACCURACY AND AUC

Sub-training set	**Accuracy (%)**		**AUC**	
	Wang	PRTMTM	Wang	PRTMTM
1	81.74	86.87(+5.13)	0.92	0.95(+0.03)
2	73.48	80.92(+7.44)	0.89	0.94(+0.05)
3	85.00	88.67(+3.67)	0.93	0.96(+0.03)
4	82.49	88.44(+5.95)	0.92	0.96(+0.04)
5	74.33	84.94(+10.61)	0.88	0.94(+0.06)
6	79.42	85.25(+5.83)	0.90	0.94(+0.04)
7	71.57	76.80(+5.23)	0.87	0.92(+0.05)
8	81.95	87.16(+5.21)	0.92	0.94(+0.03)
9	80.38	86.61(+6.23)	0.90	0.94(+0.04)
10	71.33	80.84(+9.51)	0.84	0.91(+0.07)
11	76.39	80.91(+4.52)	0.87	0.92(+0.05)
12	66.44	70.73(+4.29)	0.85	0.88(+0.03)
13	77.66	83.95(+6.29)	0.88	0.92(+0.04)
14	76.70	84.66(+7.96)	0.87	0.92(+0.06)
15	68.38	76.44(+8.06)	0.79	0.86(+0.07)
16	72.50	74.50(+2.00)	0.84	0.87(+0.03)
17	59.59	63.48(+3.89)	0.80	0.84(+0.04)
18	77.56	78.85(+1.29)	0.86	0.89(+0.03)
19	72.20	78.90(+6.70)	0.82	0.88(+0.06)
20	63.54	73.44(+9.90)	0.74	0.81(+0.07)
21	57.46	57.29(-0.17)	0.69	0.73(+0.04)
22	48.04	46.40(-1.64)	0.62	0.65(+0.04)
23	65.20	61.25(-3.95)	0.72	0.74(+0.02)
24	63.23	63.54(+0.31)	0.67	0.69(+0.02)
25	58.40	63.56(+5.16)	0.63	0.66(+0.03)
		Avg+4.78		**Avg+0.04**
		Max+10.61		**Max+0.07**

Secondly, the accuracy and AUC are shown in Fig. 5 (a) and (b) respectively. It is observed that the accuracy and AUC of Wang's method and PRTMTM are similar, and decrease on the datasets with fewer images (such as the 25th sub-training set has the least images). Moreover, the metrics on the balanced datasets (such as the 3rd sub-training set in group 1) are better than that on the unbalanced datasets in the same group. Further observation shows that the accuracy and AUC of PRTMTM are higher than those of Wang's method. The detailed experimental results are illustrated in Table IV, from which we find that the accuracy of PRTMTM is 4.78% higher on average than that of Wang's method and the maximum is 10.61%, and the AUC of PRTMTM has an average improvement of 0.04 and a maximum improvement of 0.07 compared with that of Wang's method. It should be particularly emphasized that the poor results for each metric in the 21st sub-training set are because there are so few tongue images in that set, and the number of tooth-marked tongue images is even lower.

D. Visualization

To further verify PRTMTM, Gradient-weighted Class Activation Mapping (Grad-CAM) technology [12] is used to obtain the heatmap of the tongue image to prove the tooth-marked regions are focused by our method. The red region of the heatmap indicates that the model pays more attention to that region. As shown in Fig. 6, with Wang's method, some regions that are not related to tooth-marks are focused, such as the black dashed regions. In contrast, with PRTMTM, the most informative parts of tooth-marks (the more compact red regions) are localized as shown in Fig. 7, which indicates that the prior regularization allows the model to learn the tooth-mark pattern.

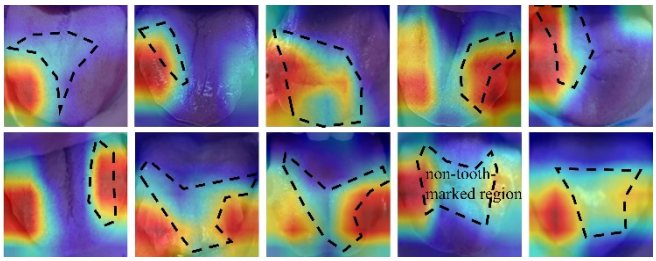


Fig. 6.Heatmap of Wang's method without prior regularization

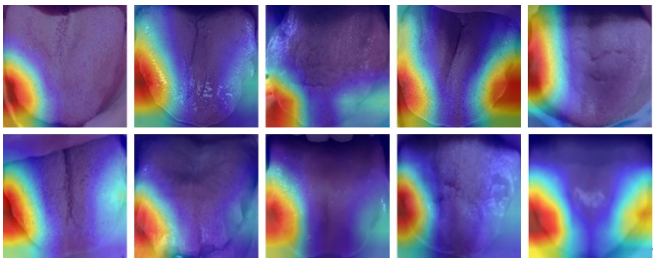


Fig. 7.Heatmap of PRTMTM with prior regularization

IV. Conclusion

In this paper, we put forward a tooth-marked tongue classification method PRTMTM by adding a prior regularization loss to the attention map. Experimental results show that our method improves the accuracy by 4.78% on average and 10.61% on maximum compared with state-of-the-art methods on various sub-training sets with different number of images and label distributions, and our method pays more attention to the related region of tooth-marks, which enhances the interpretability of the trained model.

References

[1] Zhang, David, Hongzhi Zhang, and Bob Zhang. Tongue image analysis. Singapore: Springer, 2017.

[2] Lozano, Francisco. "Basic theories of traditional Chinese medicine." Acupuncture for pain management (2014): 13-43.

[3] Pang, Bo, David Zhang, and Kuanquan Wang. "Tongue image analysis for appendicitis diagnosis." Information Sciences 175.3 (2005): 160-176.

[4] Laskaris, George. "Color atlas of oral diseases." Perative Denti Try (2003): 213.

[5] Li W, Luo J, Hu S, et al. "Towards the objectification of tongue diagnosis: the degree of tooth-marked," in 2008 IEEE international symposium on IT in medicine and education, 2008, pp. 592-595.

[6] Umadevi G, Malathy V, Anand M. "Diagnosis of Diabetes from Tongue Image Using Versatile Tooth-Marked Region Classification," TEST Engineering & Management, vol. 81, no. 19, pp. 5953-5965, 2019.

[7] Li X, Zhang Y, Cui Q, et al. "Tooth-marked tongue recognition using multiple instance learning and CNN features," IEEE transactions on cybernetics, vol. 49, no. 2, pp. 380-387, 2018.

[8] Wang H, Zhang X, Cai Y. "Research on teeth marks recognition in tongue image," in 2014 International Conference on Medical Biometrics. IEEE, 2014, pp. 80-84.

[9] Shao Q, Li X, Fu Z. "Recognition of teeth-marked tongue based on gradient of concave region," in 2014 International Conference on Audio, Language and Image Processing. IEEE, 2014, pp. 968-972.

[10] Li W, Luo J, Hu S, et al. "Towards the objectification of tongue diagnosis: the degree of tooth-marked," in 2008 IEEE international symposium on IT in medicine and education. IEEE, 2008, pp. 592-595.

[11] Sun Y, Dai S, Li J, et al. "Tooth-marked tongue recognition using gradient-weighted class activation maps," Future Internet, vol. 11, no. 2, pp. 45, 2019.

[12] Selvaraju R R, Cogswell M, Das A, et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization," in Proceedings of the IEEE international conference on computer vision. 2017, pp. 618-626.

[13] Wang X, Liu J, Wu C, et al. "Artificial intelligence in tongue diagnosis: Using deep convolutional neural network for recognizing unhealthy tongue with tooth-mark," Computational and structural biotechnology journal, 18, pp. 973-980, 2020.

[14] Tang W, Gao Y, Liu L, et al. "An automatic recognition of tooth-marked tongue based on tongue region detection and tongue landmark detection via deep learning," IEEE Access, vol. 8, pp. 153470-153478, 2020.

[15] Kong X, Rui Y, Dong X, Cai J, Liu Y. "Tooth-Marked Tongue Recognition Based on Mask Scoring R-CNN," in Journal of Physics: Conference Series. IOP Publishing, 2020, vol. 1651, no. 1, pp. 012185.

[16] Weng H, Li L, Lei H, et al. "A weakly supervised tooth‐mark and crack detection method in tongue image," Concurrency and Computation: Practice and Experience, vol. 33, no.16, pp. e6262, 2021.

[17] Zhou J, Li S, Wang X, Yang Z, Hou X, Lai W, Zhao S, Deng Q, Zhou W. "Weakly Supervised Deep Learning for Tooth-Marked Tongue Recognition." Frontiers in Physiology (2022): 567.

[18] Gu H, Yang Z, Chen H. "Automatic Tongue Image Segmentation Based on Thresholding and an Improved Level Set Model," in 2020 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS). IEEE, 2020, pp. 149-152.

[19] He K, Zhang X, Ren S, Sun J. "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016, pp. 770-778.

[20] Hu T, Qi H, Huang Q, Lu Y. "See better before looking closer: Weakly supervised data augmentation network for fine-grained visual classification," arXiv preprint arXiv:1901.09891 (2019).

[21] Omar L, Ivrissimtzis I. "Using theoretical ROC curves for analysing machine learning binary classifiers," Pattern Recognition Letters, vol. 128, pp. 447-451, 2019.