

A survey of artificial intelligence in tongue image for disease diagnosis and syndrome differentiation

DIGITAL HEALTH Volume 9: 1-25 © The Author(s) 2023 Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/20552076231191044 journals.sagepub.com/home/dhj



Qi Liu^{1,2,#} (D), Yan Li^{1,#}, Peng Yang¹, Quanquan Liu¹, Chunbao Wang¹, Keji Chen³ and Zhengzhi Wu¹

Abstract

The rapid development of artificial intelligence technology has gradually extended from the general field to all walks of life, and intelligent tongue diagnosis is the product of a miraculous connection between this new discipline and traditional disciplines. We reviewed the deep learning methods and machine learning applied in tongue image analysis that have been studied in the last 5 years, focusing on tongue image calibration, detection, segmentation, and classification of diseases, syndromes, and symptoms/ signs. Introducing technical evolutions or emerging technologies were applied in tongue image analysis; as we have noticed, attention mechanism, multiscale features, and prior knowledge were successfully applied in it, and we emphasized the value of combining deep learning with traditional methods. We also pointed out two major problems concerned with data set construction and the low reliability of performance evaluation that exist in this field based on the basic essence of tongue diagnosis in traditional Chinese medicine. Finally, a perspective on the future of intelligent tongue diagnosis was presented; we believe that the self-supervised method, multimodal information fusion, and the study of tongue pathology will have great research significance.

Keywords

Tongue diagnosis, deep learning, syndrome differentiation, artificial intelligence, tongue image

Submission date: 2 February 2023; Acceptance date: 13 July 2023

Introduction

Tongue diagnosis is an important means to obtain disease information in traditional medicine, especially in traditional Chinese medicine (TCM). TCM physicians rely heavily on it as they obtain the whole human health state through it, so they arrange the herbs to balance the body's health state. Because of the abundant superficial vascular tissue, the tongue can transmit many useful signals in real time,^{1–3} which may indicate important medical information. Nowadays, tongue diagnosis can even be used for disease prediction or disease stage estimation.^{4–6} In this point, as the only noninvasive method to explore organs in vivo, its diagnosis value has long been underestimated by modern medicine. Studying the classification or quantification of tongue images will not only help promote the modernization of TCM but also help to reveal the pathological basis behind the tongue organ.

Instead of being used to collect clinical information on the tongue organ for diagnosis of tongue diseases only, tongue diagnosis is often used to perceive the healthy state of the whole body for syndrome differentiation (SD) in TCM. As a

²Shenzhen Institute of Advanced Technology of the Chinese Academy of Science, Shenzhen, Guangdong, China

³Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing, China

[#]QL and YL contributed equally to this work.

Corresponding authors:

Keji Chen, Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing, Guangdong, China. Email: kjchenvip@163.com

Wu Zhengzhi, Shenzhen Institute of Translational Medicine, The First Affiliated Hospital of Shenzhen University, Shenzhen Second People's Hospital, Shenzhen 518035, Guangdong, China. Email: szwzz001@163.com

Creative Commons NonCommercial-NoDerivs CC BY-NC-ND: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License (https://creativecommons.org/licenses/by-nc-nd/4.0/) which permits non-commercial use, reproduction and distribution of the work as published without adaptation or alteration, without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).

¹Shenzhen Institute of Translational Medicine, The First Affiliated Hospital of Shenzhen University, Shenzhen Second People's Hospital, Shenzhen, Guangdong, China

traditional diagnosis method, tongue diagnosis refers to the physician collecting disease information by "watching" the tongue symptoms of coating color, tongue color, and tongue texture. However, the tongue symptoms are neither objective as they can be vulnerable to human eyes "watching" bias nor easy to quantify. Meanwhile, syndrome (or Zheng⁷) is not always a physical state that can be perceived clearly, and the relationship between tongue features and the syndrome is a kind of uncertain knowledge. To extract the objective features and fit the uncertain correlation, many studies have been launched using artificial intelligence (AI) technologies to extract tongue features automatically or model the uncertain relationships.

Up to now, the field of intelligent tongue diagnosis has gained new features and presented many traceable technological trends. Attention mechanism helps to improve the recognition and segmentation of tongue images and is receiving increasing attention and application.⁸⁻¹⁰ Due to the abundant health information on tongue images, extracting multiscale features is helpful for diagnosis or classification based on tongue images.^{11,12} Although deep learning methods have the ability to extract deeper or more abstract features, they have shown better results when combined with traditional methods. Traditional methods are increasingly being valued to compensate for some of the shortcoming of deep learning. For example, with the help of convolutional neural network (CNN) and latent Dirichlet allocation (LDA) model, Hu et al.¹³ have modeled a relatively robust correlation between Chinese herbal prescriptions and tongue images. Based on the tongue image feature extracted by ResNet50, the Extreme Gradient Boosting (XGBT) model optimized by the genetic algorithm (GA) method was able to predict prediabetes and diabetes.¹⁴ The multiple-instance support vector machine (SVM) method, the deep learning model VGG-16 and the SVM algorithm combined, had shown a great performance improved in classification or diagnosis of tongue images.^{15,16} The prior knowledge is also a useful aid when combined into the deep learning model, such as the morphological operation^{17–19} or knowledge from TCM experts.^{15,20}

This paper intended to report the progress of methods or technologies used to promote the automation and objectification of tongue diagnosis. Many related reviews have also been published,^{21–25} but none have provided a global perspective. Reference^{21,24} mainly focused on reporting the pathological meaning aspect,^{22,23} focused on feature engineering,²⁵ and depicted the history of tongue diagnosis in China. This paper focuses on the tongue image intelligence analysis methods used for calibration, segmentation, and classification of symptoms, syndrome types, and diseases and will review the related studies included in the Web of Science and IEEE mainly published in the last 5 years.

Tongue diagnosis in TCM (background)

Tongue diagnosis is a healthy status diagnosis method that the Chinese have followed for thousands of years and has made outstanding contributions to the history of fighting against diseases.²⁵ It is used to diagnose syndrome types or diseases by observing the color, texture, mobility, humidity, and some position information of the tongue, but mainly used for SD. In summary, it has the following 12 characteristics: tongue color, tongue shape, tongue motion, tongue coating color, tongue coating thickness, tongue coating quantity, tongue coating texture, tongue coating position, crack, tooth mark, the quantity of saliva and red dots, and the sublingual vein, as shown in Figure 1.

TCM theory believes that the whole is connected with the part, different internal organs' function states may show different features on the tongue, as shown in reference,^{2,26} and these different internal organs' function states can be understood as "syndrome." The empirical mapping between tongue and syndrome is established through the summary and induction of various manifestations of the above characteristics, which corresponds to the various syndrome types of the internal environment of the human body. The prescription of traditional Chinese herbs based on these empirically defined syndrome types can often obtain curative effects that modern science cannot explain.

In the past, because of the estrangement between tradition and modernity, few people cared about the evidence to support tongue diagnosis, so its diagnostic value was ignored by modern medicine for a long time. With the rapid development of society and the continuous collision between tradition and modern science, more and more tongue diagnostic researchers full of curiosity have carried out active exploration. Many studies have shown that the external manifestations of the tongue are significantly related to some diseases and health indicators in the body^{3,27–30} and even COVID-19.^{4,31–} ³³ Meanwhile, the movement of the tongue also relies on the function of the brain^{34,35} and can be used for identity verification.³⁶ Thus, tongue diagnosis has its inherent rationality, suggesting that TCM tongue diagnosis is expected to obtain evidence-based support. As the tongue is the only noninvasive visible internal organ in the human body, its capillaries are more superficial, which allows us to directly access internal information simply. All these indicate that the tongue has great diagnostic value, which needs to be paid more attention to and explored by modern medicine.

However, because the visual features used for diagnosis in tongue diagnosis cannot be objectively quantified, this poses a great challenge to its clinical research and pathological study. In addition, tongue diagnosis is mainly used for SD in TCM. However, the syndrome concept is vaguely defined and is an abstract description of the overall health or functional state of the body, and it is also impossible to quantify the TCM syndrome types, which increases the difficulty of tongue diagnosis research. With the development of AI technology, this kind of model technology that can simulate human decision-making with powerful feature extraction ability is favored. Many researchers expect to solve the problem of objectively quantifying tongue images and assisting decision-making through AI.



Figure 1. The tongue feature that TCM doctors care about. In TCM theory, these features are considered to be related to the body's health status. For example, the chubby tongue may indicate a "wet" body status, so as the lubricant tongue, the thin tongue may indicate a "dry, hot" status, the askew tongue may suggest an "insane" brain status and the flaccid tongue may imply a "neurasthenia" brain status, coating features in the tongue are closely related to the function of the stomach, a considerable number of red-dot imply a "fiery" status of the upper body, and a purple sublingual vein is related to a "blood stasis" body status.

Tongue image data preprocessing

Before analyzing the tongue image, there will be three important tasks: tongue image correction, tongue image recognition, and segmentation. Because tongue image data acquisition will be affected by different acquisition devices and external light sources, resulting in color distortion and nonstandard format of tongue image, such distortion harms the diagnosis from doctors. In addition, the collected tongue image data may be unqualified; for example, the camera does not capture the tongue image, or the patient does not fully spit out the tongue. Jiang et al.³⁷ had summarized some unqualified tongue images; more critically, they trained a deep model to distinguish the qualified tongue image using ResNet-152. Meanwhile,

a multitask deep learning model method was introduced³⁸ to classify the tongue image into high-quality and unqualified tongues to assess its quality. The tongue image recognition is to quickly identify whether the complete tongue image is captured in the auxiliary scene. Tongue image segmentation separates the tongue body from the background so that the image contains only the tongue body, eliminating redundant information.

Tongue image calibration

Aiming at the problem of color distortion in tongue image acquisition, some researchers use unified acquisition equipment in a closed acquisition environment to eliminate the interference of different equipment and light sources in clinic information transmitted by tongue image. Tongue image correction is mainly aimed at correcting the color and brightness of the tongue image acquired in an open environment, such as outdoor collection and collection by mobile phone equipment. The correction methods can be divided into three categories: deep learning algorithm,^{39,40} polynomial regression algorithm,^{41–44} and support vector regression (SVR) method.^{45,46}

The method based on the polynomial regression algorithm is the most commonly used because of its low computational complexity and short training time, which is crucial for online applications. Wang and Zhang⁴³ employed the classical polynomial regression algorithm accompanied with a new tongue colorchecker for tongue color space to improve calibration accuracy. The polynomial regression method always needs a colorchecker to be a reference for generating the color matrix,⁴² and a colorchecker specially established for tongue images would improve the accuracy. Furthermore, Sui et al.⁴¹ proposed a root polynomial regression method that has achieved a more stable and even better correction effect than the traditional polynomial regression under different illumination conditions. Zhuo et al.⁴⁰ have firstly introduced the deep learning method to the tongue image correction problem, and they built a simulated annealing (SA)-GA-backpropagation (BP) neural network based on a partial color gamut that was similar to those of the tongue body, tongue coating, and skin. They further⁴⁴ proposed a kernel partial least squares regression (K-PLSR)-based correction method which has obtained a superior color correction performance against classical polynomial-based and SVR-based methods under different lighting conditions. Lu et al.³⁹ proposed a two-phase deep color correction network (TDCCN) to establish the tongue color mapping model under a standard lighting condition, and they further proposed flexible color-adjusting options to conquer the differences between standard lighting conditions and the environments that doctors are familiar with. Zhang et al.46 used the SVR method to study the color rendition chart specially applied for tongue image calibration and proved that 24 was the optimal number of color patches.

Tongue image detection and segmentation

Tongue image detection and segmentation is an important and necessary step in the preprocessing steps of evaluating the health status based on tongue image. Tongue image segmentation excluding other redundant information is very important for subsequent downstream tasks. Since tongue image segmentation needs to detect the tongue image first, tongue image detection and tongue image segmentation have the same feature extraction process, often completed simultaneously in many studies, and they are discussed in this section.

Tongue image detection and segmentation methods are divided into two categories according to feature extraction methods: traditional and deep learning-based methods. Traditional methods include color threshold,^{47,48} edge detection,⁴⁹ active contour model (ACM),^{17,50} and region growing and merging.⁵¹ These methods are all based on manually set rules or prior knowledge to extract image features, and finally used for classification, among them, ACM is most used. For example, Guo et al.⁵⁰ firstly do the twostage K-means clustering method based on the extracted initial tone boundary and finally apply the ACM to segment the tongue image. Study⁴⁷ performed image thresholding in hue saturation intensity (HSI) color space and subsequent morphological operations to get an initial tongue region. Then, a gray projection technique is used to determine the upper bound of the tongue body root for refining the initial region. Wu and Zhang⁴⁹ fused the region base method with the edge-based, and they extracted region of interest (ROI) and subsequently merged adjacent regions utilizing the histogram-based color similarity criterion. Hence, the results are less sensitive to cracks and fissures on the tongue. Then, they adopted a fast marching method to get a close curve based on edge features. The contour obtained by the region-based approach was utilized as a mask during the fast marching process (edge-based) to make the ultimate contour more robust. Liu et al.¹⁷ proposed a path-driven segmentation method, and each patch in the testing image is sparsely represented by patches in the spatially varying dictionaries, constructed by the local patches of training images. The derived sparse coefficients are then employed to estimate the tongue probability. Finally, the hard segmentation is obtained by applying the maximum a posteriori (MAP) rule on the tongue probability map and further polished with morphological operations. However, the traditional methods are based on simple pixel values or low-level features of tongue images (such as color, edge, brightness, and texture) and cannot extract high-order features. Furthermore, these handcrafted features were extremely time consuming and tedious. Therefore, with the breakthrough of AI technology, most of the research has turned to the method of deep learning in recent years.

In the deep learning method, the distribution of the observed data is calculated by feedback propagation, and the parameter is updated based on this distribution to model the relationship between the label and the input data. In this process, the characteristics of the data can be automatically extracted. Full convolution network (FCN) model is the pioneering achievement in applying the deep learning method to the field of image segmentation and achieved extraordinary performance at that time.⁵² It applies the previous method based on global image classification to pixel-level classification, to achieve image segmentation and achieved better results than traditional methods. Huang et al.⁵⁴ improved the FCN by designing the receptive field block module (including a multibranch



Figure 2. The morphological layer for processing incorrect predictions. Using the morphological layer to reconstruct the incorrect morphological prediction or filter it out.

evolutionary block and a shortcut connection) and could extract higher-level or global features. Subsequently, Ronneberger et al.⁵⁵ developed the U-Net model based on the FCN model. Due to its excellent performance, it was widely used in computer vision. Li et al.⁵⁶ applied this model to the tongue crack segmentation, and they improved its encoder to extract relatively more abstract high-level semantic features. Similarly, Peng et al.⁵⁷ also improved the U-Net framework and designed a lightweight model P-Net with the letter "P" structure to be suitable for remote tongue image segmentation. Zhou et al.¹⁸ adding a morphological layer to U-Net aim at refining the obvious morphological errors in U-Net segmentation. Other important models are Faster R-CNN and Mask R-CNN. Reference⁵⁸ based on the tongue images segmented by Mask R-CNN, and Faster R-CNN is used to detect and classifv the tongue images of cracked, tooth marked, spotted or rotten, and so on. Besides, Yuan et al.⁵⁹ designed a cascaded CNN model for tongue image segmentation for mobile and embedded devices, and its prediction speed is significantly improved compared with other deep learning methods. Zhou et al.⁶⁰ referred to the idea of generate adversarial network (GAN), the generation module is used to generate a segmented image, and the identification module is used to determine whether the generated segmented image is true, to reduce the dependence on the annotated data set.

Combining some advantages of traditional methods with deep learning methods also helps to improve the model's performance. Zhou et al.¹⁸ designed a morphological processing layer based on morphological inductive bias, including some specifically designed filters to refine any morphologically incorrect coarse mask image, as shown in Figure 2. Similarly, a study¹⁹ also designed a tongue assessment filter to filter out some segmented tongue images predicted by U-Net with wrong contours. These handcraft features were based on morphology of the tongue, and Gao et al.⁶² used the geometric features of the tongue based on the level set method derived from ACM, combined with the CNN network, and proposed a level set model with symmetry and edge constraints. However, Yuan and Liao⁶¹ argued that tongue body and tongue coating segmentation can be clustered according to differences of the color blocks in Lab color space, and the *K*-means segmentation method based on Lab color space was proposed and was better than the deep learning methods of FCN, U-Net, and Deeplab-v3.

Researchers creatively combined or proposed many ideas for model building and training to make the model more robust and more targeted to segment the tongue image. Multitask learning always shares a feature extraction backbone to improve generalization performance and mitigate manual labeling consumption, and Xu et al.⁶⁶ used the U-Net framework as a common feature extraction module to segment tongue and classify tongue coating in a multitask learning way. Zhou et al.⁶³ designed two different loss functions based on the multitask learning method to anchor the tongue image localization and segmentation tasks, and decoupled the tongue image segmentation and localization task. Furthermore, Tang et al.¹¹ proposed cascaded CNNs with multitask learning to predict tongue region, tongue landmarks, and tooth-marked tongue. The multitask learning method often requires a special loss function. Cai et al.⁶⁷ changed the loss function for tongue segmentation and proposed a function that would decrease the intraclass distance and increase the interclass distance. To change the model to extract specific concerns, Li et al.⁵⁶ introduced a global convolution network module to extract relatively abstract high-level semantic features, while Huang et al.⁵⁴ constructed a receptive field block based on the receptive field theory that the region closer to the center of retinotopic maps is more important than others in distinguishing objects, making the model deal more with the blurred edge of the tongue body. Similarly, Peng et al.⁵⁷ applied an attention module to intensify the attention to the boundary and suppress useless information.

In addition, some studies report special solutions for other tongue segmentation scenarios. Tang et al.¹¹ reported a solution of tooth-marked recognition based on the segmented tongue, and Li et al.⁵⁶ refined the U-Net model to especially segment the cracks on the tongue, the same as the reference.⁵⁷ The feature of sublingual vein had also been noticed and extracted,^{10,68} Qiu et al.'s¹⁰ study added the cross-channel attention module to assign more weight to the target area, which shows an improved performance in their lightweight tongue diagnosis system on mobile devices, and another paper⁶⁸ proposed a two-stage segmentation method: a fully CNN network without downsampling to reduce the loss of spatial feature information effectively and another fully CNN network with proper dilated convolution to avoid the gridding issue. For real-time diagnosis, the size of the popular models is generally too large; Li et al.⁶⁴ proposed a lightweight architecture for tongue image segmentation in real-time, and they further proposed the P-Net model⁵⁷ refined from U-Net and used for realtime tongue crack extraction. The data set to supervise the tongue segmentation model is relatively small, which may cause generalization issues; Li et al.¹⁹ applied an iterative learning method to train the network on the good samples

repeatedly judging by some specific filters. Others^{54,56} used pretrained methods to alleviate the generalization problem. Some studies believe that tongue features must be extracted from the global image level to make segmentation, so the dilated convolution block was introduced. Zhou et al.⁶³ used a context-aware dilated residual block to ensure the efficiency of information without increasing extra parameters and computation; Peng et al.⁵⁷ used dilated convolution for dense feature extraction and field-of-view enlargement; and Tang et al.¹² proposed a hybrid cascade dilated convolution to extract multiscale features. The division of the tongue is also important for assessing health status, and Wu et al.⁶⁹ reported regional alignment of tongue images to improve the accuracy of disease diagnosis. Using medical ultrasound image to visualize and characterize human tongue shape and motion in real-time would do a great favor to study healthy or impaired speech production; thus, some tongue contour segmentation methods^{70–72} of ultrasound image in real time have been proposed by the team of Hamed Mozaffari; furthermore, they⁷³ have used the augmented reality of ultrasound data from the extracted tongue movement to provide a real-time visual feedback to improve the training trend of language learners. There are also some researchers⁷⁴ who used magnetic resonance signals of tongue images to perform tongue segmentation.

To sum up, different from the general field, most of the tongue image recognition and segmentation scenes used in tongue diagnosis only need to recognize and segment one tongue instance and the tongue body often occupies more pixels in the tongue image; as we can see, ^{12,63} dilated convolution block or extract from a coarser or more global feature level can be useful. In addition, because the tongue image has a relatively stable shape, that is, a relative ellipse of convex shape, the morphological filter method is also suitable for tongue image segmentation scenes, as shown in Figure 2. Since the feature of tongue body edge is the key to tongue image segmentation, it would be a favor to focus the main attention of algorithm on the edge area.⁵⁷ The cross-channel attention module promotes the information exchange between feature maps and assigns weight among them, which may explain why it can alleviate the problem of difficult tongue image samples.¹⁰ Most studies draw on the modules, frameworks, or methods successfully used in the general field. Although these can improve the accuracy when applied to the tongue analysis field, tongue image detection and segmentation remains relatively difficult task, as the reference states⁶³ and shown in Figure 3. The tongue in images has a large variation in shape and color. Its edge pixels are always difficult to distinguish from the lips or skin. Meanwhile, few public data sets can be used for unified evaluation, and most of the proposed model codes are not published, so the declared effect cannot be verified; as shown in Table 1, some of the studies did not report the data set construction and its



Figure 3. The variation in tongue appearance. The variation of tongue appearance in tongue image increases the difficulty of tongue segmentation, and the variation in shape and tongue extension may reduce extracted features and weaken the performance of downstream tasks.

labeling quality. And the open scene and fixed scene often have a more complex variation, so data set quality should be paid more attention.

Tongue image for disease diagnosis

The tongue appearance is believed to be sensitive to some diseases,^{21,23} especially diabetes, whose association is most studied^{11,75} in recent years. Zhang et al.⁷⁶ used the SVM method to diagnostic diabetes based on the tongue image features of tongue color values and tongue texture, and these features were extracted by the division-merging method and chrominance-threshold method. Selvarani and Suresh⁷⁷ further proposed the SVM classifier with multiple kernels, named kernel ensemble classification method, to classify diabetes from the healthy person based on the tongue color distribution and texture. The study from Fan et al.⁷⁸ showed a better diabetes diagnosis performance of random forest (RF) than SVM. They combined texture features and four TCM tongue features of constitution color, coating color, cracks, plumpness, and slenderness as input. Deepa and Banerjee⁷⁹ also applied the SVM method as a classifier and used particle swarm optimization (PSO) technique to tune the parameters and enhance its performance; this time, deep features of the tongue (such as color, texture, coating, tooth-marked, and red spots) were utilized from CNN DenseNet framework. Mathew and Sathyalakshmi⁸⁰ also developed an optimization-driven hybrid deep learning method for diabetes detection based on tongue images. A proposed ExpACVO optimization algorithm was used for Deep Q-Network classifier training. The proposed ExpACVO algorithm combined anticorona virus optimization with exponential weighted moving average and has achieved improved

performance. Li et al.14 fused prior knowledge of tongue images and the deep features of tongue images from ResNet50 to diagnose diabetes and prediabetes based on the XGBT algorithm, and they optimized the parameters in the XGBT model with GA which has shown a further performance improved of XGBT. Zhang et al.⁸¹ also proposed a fusion method of color features from tongue images for diabetes diagnosis (Table 2). They used a novel clustering-based color descriptor to represent and fuse the RGB, HSV, and Lab color space of tongue images, and K-nearest neighbor (KNN), SVM, minimum squared error (MSE), lasso, and ridge regression were used as classifier and evaluation method. Another deep feature of tongue image used for diabetes diagnosis was extracted from ResNet50,⁸² a popular CNN base framework. On the other side, Vijayalakshmi et al.⁸³ used a CNN block as a classifier and was proved to have a better performance against SVM when performing classification based on the three tongue quantitative features of geometry, color, and texture, which are measured by MATLAB. They argued that a person with diabetes would have a gray color coating at the center of the tongue. Srividhya and Muthukumaravel⁸⁴ applied a self-organizing map (SOM) Kohonen classifier to classify diabetes or nondiabetes based on quantified features of tongue color and gist.

The second most studied is stomach disease; Gholami et al.⁸⁵ compared the accuracy of different CNN frameworks in diagnosing stomach cancer based on tongue color and its lint features, and the best model eventually comes to the DenseNet. Wu et al.⁵ studied the tongue diagnosis indices for gastroesophageal reflux disease (GERD), and they used an automatic tongue diagnosis system (ATDS) to extract the tongue indices and found that the saliva amount (p = .009) and thickness of the tongue's fur (p = .036), especially that in the spleen–stomach area (%)

Table 1. Summary of machine learning modeling approaches for tongue detection and segmentation. The open scene is more complex than the fixed scene and requires more in terms of data quality and model performance.

							e
Scene	Fixed scene	No report	Fixed scene	No report	No report	Fixed scene	Fixed scene (tong images were captured by a camera with
Annotation	2 experts using the public software ImageJ to annotation	Detection: a visible tongue in image is positive, no tongue or incomplete tongue is negative	Manual segmentations	Manually marked by experts	Manual segmentation	No report	30 landmarks for tongue segmentation; the tooth-marked label was cross-validated by 2 primary physicians, and a resident physician conducts final validation
Data Set	16 images randomly selected from a tongue database of a medical school by the same device	864 images for tongue detection; 300 images from BioHit ^a for tongue segmentation	290 tongue images by a tongue imaging system	550 tongue images, 300 of them from BioHit	150 tongue images from the same device	300 images from BioHit, 331 images from the hospital, 290 images from the same device	1858 images were captured from the hospital since October 2018
Highlights	Adaptive active contour model	HOG features; Lab color space	Spatially varying dictionaries	Symmetry detection constraint; novel level set initialization method	Decorrelation and stretch algorithm; fusion algorithm of two approaches	End-to-end model/ multitask learning	Multitask learning/ tongue segmentation and tooth-marked classification
Feature Backbone/ Model Method	Otsu's threshold on HSV color, K-means	SVM, <i>K</i> -means	Patch-based sparse representation	Level set model; CNN	Maximal similarity-based region merging; edge-based	Context-aware residual blocks	Cascaded CNNs
Study	Adaptive active contour model-based automatic tongue image segmentation ⁵⁰	Design and Implementation of the Traditional Chinese Medicine Constitution System Based on the Diagnosis of Tongue and Consultation ⁶¹	Patch-Driven Tongue Image Segmentation Using Sparse Representation ¹⁷	LSM-SEC: Tongue Segmentation by the Level Set Model with Symmetry and Edge Constraints ⁶²	Robust tongue segmentation by fusing region-based and edge-based approaches ⁴⁹	TongueNet: Accurate Localization and Segmentation for Tongue Images Using Deep Neural Networks ⁶³	An Automatic Recognition of Tooth Marked Tongue Based on Tongue Region Detection and Tongue Landmark Detection via Deep Learning ¹¹

Table 1. Continued.					
Study	Feature Backbone/ Model Method	Highlights	Data Set	Annotation	Scene
					standard illumination)
Application of U-Net with Global Convolution Network Module in Computer-Aided Tongue Diagnosis ⁵⁶	U-Net	Transfer learning/ boundary refinement block	316 tongue images, <i>no</i> report where it comes from	Labelme software annotates the crack area; check and confirm the annotation	Fixed scene
Automatic Tongue Image Segmentation For Real-Time Remote Diagnosis ⁶⁴	Cascaded dilated depthwise convolutions	Lightweight model; PReLU/nonlinear active layer/group normalization	5600 tongue images, <i>no</i> report where it comes from	Photoshop quick selection was used for labeling, <i>no report on how to</i> <i>ensure the labeling quality</i>	Open scene
TISNet-Enhanced Fully Convolutional Network with Encoder-Decoder Structure for Tongue Image Segmentation in Traditional Chinese Medicine ⁵⁴	ResNet101	Receptive field block/ transfer learning/ data augmentation	300 images from BioHit ^a ; 700 images from the hospital	No report	Fixed scene
An iterative transfer learning framework for cross-domain tongue segmentation ¹⁹	U-Net	lterative transfer learning/the contour filter	756 images and 1572 images from the hospital by 2 different tongue diagnosis devices	No report	Fixed scene
Automatic tongue image matting for remote medical diagnosis ⁶⁵	FCN	End-to-end iterative network/error correcting mechanism	1578 tongue images, <i>no</i> report where it comes from	Photoshop quick selection was used for labeling, <i>no report on how to</i> <i>ensure the labeling quality</i>	Open scene
Automatic Tongue Crack Extraction For Real-Time Diagnosis ⁵⁷	U-Net	Dual attention gates/ lightweight model	281 tongue images, no report where it comes from	Photoshop was used for labeling, professional TCM doctors check the accuracy	Fixed scene
Multi-task Joint Learning Model for Segmenting and Classifying	U-Net	Multitask joint Iearning/	1858 images from the hospital since October 2018	Labels were cross-validated by 2 primary physicians, and a resident physician conducts final validation	Fixed scene (tongue images were captured by a
					(continued)

Table 1. Continued.

Scene	camera with standard illumination)	Fixed scene (tongue images were captured by TFDA-1 tongue diagnosis instrument)	
Annotation		Labels were marked manually using Labellmg software	
Data Set		1778 images from the hospital under fasting conditions. The collection time was 8:00-10:00 a.m.	
Highlights	discriminative filter learning	Split and merge algorithm; color threshold method	
Feature Backbone/ Model Method		Mask R-CNN; Faster R-CNN	
Study	Tongue Images Using a Deep Neural Network ⁶⁶	Application of computer tongue image analysis technology in the diagnosis of NAFLD ⁵⁸	

^ahttps://github.com/BioHit/TongelmageDataset. SVM: support vector machine; HOG: histogram of oriented gradients; HSV: hue-saturation-value; CNN: convolutional neural network; FCN: full convolution network.

Study	Diagnostic Task	Feature Backbone/Model Framework	Highlights	Labels	Methods to Ensure Annotation Quality
Automated Screening of COVID-19-Based Tongue Image on Chinese Medicine ⁴	Diagnosis COVID-19	AlexNet		COVID-19 patients/ non-COVID-19 patients	The official diagnostic criteria
Tooth-Marked Tongue Recognition Using Multiple Instance Learning and CNN Features ¹⁵	Tooth-marked tongue classification	VGG-16	Multiple-instance SVM classifier, transfer learning; handcraft features	Tooth-marked for tongue image; bounding boxes of tooth-marked regions	The labels were voted by multiple TCM practitioners
Artificial intelligence in tongue diagnosis: Using deep convolutional neural network for recognizing unhealthy tongue with tooth-mark ¹⁰³	Tooth-marked tongue classification	ResNet34	Grad-CAM for network visualization	Tooth-marked and nontooth-marked tongue	First, professionals discuss the diagnostic criteria. Second, one label, two professionals review, respectively. For instances of disagreement, three professionals will discuss and make the final decisions.
A novel tongue feature extraction method on mobile devices ¹⁰	Coating classification; sublingual vein classification	MobileNet V2	Cross-channel attention mechanism	Thin and thick coating; thin and thick sublingual vein	Two experienced physicians perform label, another two physicians discuss and confirm the different annotation
A multi-step approach for tongue image classification in patients with diabetes ¹⁰²	Syndrome clustering	VQ-VAE	Self-supervised features; ViT verifies the clustering result: without prior knowledge	No need	No need
Panoramic tongue imaging and deep convolutional machine learning model for diabetes diagnosis in humans ⁸²	Diabetes diagnosis	ResNet50	Deep radial basis function neural network (RBFNN) algorithm as a classifier	Diabetes patients and nondiabetes patients	The official diagnostic criteria
Human-computer interaction based health diagnostics using ResNet34 for tongue image classification ¹⁰⁴	Coating classification	ResNet34		Nongreasy fur, greasy fur, thick greasy fur	Three experts label, the eyes and the display screen is 45 cm; each tongue image is 20 s.

Table 2. Summary of machine learning modeling approaches for tongue image classification.

(continued)

Methods to Ensure Annotation Quality	First, professionals discuss the diagnostic criteria. Second, one label, two professionals review, respectively. For instances of disagreement, three professionals will discuss and make the final decisions.	Professionals propose the diagnostic criteria, two TCM practitioners label, the third TCM professional joins and makes the final decision for any dispute	The labels were cross-validated by two primary physicians, and a resident physician conducts final validation	Labeled manually by outpatient doctors, and double-checked by more than three chief physicians	Labeled by TCM	Labeled by TCM	The official diagnostic criteria
Labels	Tooth-marked and nontooth-marked tongue	Nongreasy, greasy, and thick greasy for tongue image	Six tongue coating types for tongue image	12 categories for disease locations for tongue image	Nine body constitution types for tongue image	Nine body constitution types for tongue image	NAFLD patients, people with normal liver function for tongue image
Highlights	Using image-level labels to locate the tooth-marked area	Quantify the greasy coating level; transfer learning	Multitask joint learning; discriminative filter learning	Stochastic regional area; inner-imaging channel-wise attention network	Difficult model/easy model; difficult samples/easy samples	Domain knowledge; zero-shot learning	TCM TDAS V1.0
Feature Backbone/Model Framework	ResNet34	ResNet18, ResNet34, ResNet50	U-Net	CNN	ResNet50, VGG-16, Inception-V3	ResNet18	Faster R-CNN
Diagnostic Task	Tooth-marked tongue classification	Three-level greasy coating	Coating color classification	Tongue area of disease location classification	Body constitution classification	Body constitution classification	Nonalcoholic fatty liver disease classification
Study	Weakly Supervised Deep Learning for Tooth-Marked Tongue Recognition ¹⁰⁵	Constructing tongue coating recognition model using deep transfer learning to assist syndrome diagnosis and its potential in noninvasive ethnopharmacological evaluation ³¹	Multi-task Joint Learning Model for Segmenting and Classifying Tongue Images Using a Deep Neural Network ⁶⁶	Fully-channel regional attention network for disease-location recognition with tongue images ⁸	Complexity perception classification method for tongue constitution recognition ¹⁰⁰	Grouping attributes zero-shot learning for tongue constitution recognition ²⁰	Application of computer tongue image analysis technology in the diagnosis of NAFLD ⁵⁸

(continued)

Table 2. Continued.

ole 2. Continued.					
Jy	Diagnostic Task	Feature Backbone/Model Framework	Highlights	Labels	Methods to Ensure Annotation Quality
lew Method for Syndrome Classification of NonSmall-Cell Lung Cancer Based on Data of Tongue and Pulse with Machine Learning ¹⁰⁶	Diagnosis Qi deficiency syndrome and Yin deficiency syndrome	Tongue Image Diagnostic Analysis System (TDAS) V2.0	TDAS v2.0; body symptom and pulse features	Qi deficiency syndrome, Yin deficiency syndrome for tongue image	TCM syndrome differentiation standard and was determined by at least three senior physicians
gnostic Method of Diabetes Based on Support Vector Machine and Tongue Images ⁷⁶	Diabetes disease classification	TDAS V1.0	RGB color feature; LAB and HIS feature	Diabetes outpatient and nondiabetes outpatient for tongue image	The official diagnostic criteria
search on Multiple-Instance Learning for Tongue Coating Classification ¹⁶	Rotten-greasy coating classification	AlexNet; VGG-16; ResNet	Multiple-instance SVM; suspected rotten-greasy coating patches	Rotten-greasy coating and normal coating for tongue image	The label was voted by five TCM practitioners and only tongue images with a judgment rate of 80% or above were accepted
rgue diagnosis indices for gastroesophageal reflux disease: A cross-sectional, case-controlled observational study ⁵	Gastroesophageal reflux disease classification	An automatic tongue diagnosis system (ATDS)	Analysis of variance method (ANOVA); the saliva amount; thickness of the tongue's fur	Gastroesophageal reflux disease patient	Diagnosis through endoscopy
Itiple color representation and fusion for diabetes mellitus diagnosis based on back tongue images	Diabetes disease classification	KNN, SVM, MSE, lasso, ridge regression	Including sublingual vein; a color descriptor on RGB, HSV, and Lab feature, a color fusion method of the three features	Diabetes mellitus sample and healthy sample	The official diagnostic criteria
ACVO-Hybrid Deep learning: Exponential Anti Corona Virus Optimization enabled Hybrid Deep learning for tongue image segmentation towards diabetes mellitus detection	Diabetes disease classification	Deep Q-Network (CNN)	ExpACVO: (optimization algorithm) anticorona virus optimization (ACVO) with exponential weighted moving average (EWMA)	No report	No report
chine learning algorithms in classifying TCM tongue features in diabetes mellitus and symptoms of gastric disease	Diabetes disease and gastric disease classification	SVM, RF	Feature engineering. Features (mean, variance, and skewness) from RGB, HSV, Lab. Texture features	Diabetes mellitus, healthy, and gastric patients (liver-stomach	The official diagnostic criteria for the disease, the two symptoms of spleen-stomach deficiency

(continued)

Study	Diagnostic Task	Feature Backbone/Model Framework	Highlights	Labels	Methods to Ensure Annotation Quality
			and four TCM tongue features	disharmony, spleen- stomach deficiency)	and liver-stomach disharmony were diagnosed by two doctors
BESS: Balanced evolutionary semi-stacking for disease detection using partially labeled imbalanced data	Disease classification	Different combinations of LGBM, Ada, RF, KNN, SVM	Color alignment features, color gamut features, texture Gabor features, geometry FFT features	Diabetes mellitus, chronic kidney disease, breast cancer, chronic gastritis, and healthy	No report
A tongue features fusion approach to predicting prediabetes and diabetes with machine learning	Diabetes disease classification	XGBT optimized by GA	Prior features from TDAS, and deep features extracted by ResNet50.	Diabetic subjects, prediabetic subjects, and normal subjects	The official diagnostic criteria. Prediabetics: 11.1 mmol/L \geq 2-h postprandial blood glucose \geq 7.8 mmol/L, or 7.0 mmol/L \geq fasting blood glucose \geq 6.1 mmol/L, or \geq 5.7% glycosylated hemoglobin \geq 6.5%
A Framework to Predict Gastric Cancer Based on Tongue Features and Deep Learning	Gastric cancer classification	EfficientNet		Gastric cancer patients and nongastric cancer subjects	Gastric cancer was diagnosed at the hospital
Construction of Tongue Image-Based Machine Learning Model for Screening Patients with Gastric Precancerous Lesions	Gastric precancerous lesion diagnosis	ResNet50; logistic regression model	Combined deep learning features from tongue image and canonical risk factors	High risk and low risk	Using video endoscopes (Olympus Corp), upper gastroscopic examinations. Tissue samples for biopsy were reviewed blindly by the two pathologists according to the criteria
Reliability of noncontact tongue diagnosis for Sjogren's syndrome using machine learning method	Sjogren's syndrome classification	SVM, RF, LR, bagging, and stacking model	Bagging of three SVM, and stacking of SVM, RF, and LR	Sjogren's syndrome and non-Sjogren's syndrome normal healthy individuals	Saxon test by the Japan Ministry of Health, Labor, and Welfare
Missing-view completion for fatty liver disease detection	Fatty liver disease classification	Low-rank learning method	Multifeature view of facial,	Fatty liver disease	All patients were diagnosed by medical experts
					(continued)

Table 2. Continued.

Table 2. Continued.					
Study	Diagnostic Task	Feature Backbone/Model Framework	Highlights	Labels	Methods to Ensure Annotation Quality
			sublingual vein, and tongue images	individual and healthy samples	
Exploring the pivotal variables of tongue diagnosis between patients with acute ischemic stroke and health participants	Acute ischemic stroke diagnosis	The automatic tongue diagnosis system (ATDS)	Multiple logistic regression analysis	Acute ischemic stroke patient and healthy participants	Acute ischemic stroke patient: diagnosed as ischemic stroke by neurologist and were approved by head CT or MRI examination
/GG: visual geometry group; ResNet: residual ninimum squared error; ACVO: anti corona vir	l net; ViT: vision transformer; us optimization; EWMA: expo	VQ-VAE: vector quantiz nential weighted moving	ed variational autoencoder; TDAS: T. average; TCM: traditional chinese m	ongue Image Diagnostic Analysis nedicine; FFT: fast fourier transforı	System; KNN: k-nearestneighbor; MSE: m; LGBM: LightGBM, Ada: Adaboost; LGBM

light gradient boosting machine; XGBT: extreme gradient boosting; GA: genetic algorithm; RF: random forest; LR: logistic regression; CT: computed tomography; MRI: magnetic resonance imaging.

(p = .029), were significantly greater in patients with GERD. Handcraft features of tongue images from physicians were weighted and filtered by the XGBT algorithm in research,⁸⁶ and the EfficientNet network was used to classify gastric cancer based on the selected features. Meng et al.87 used tongue image high-level features extracted from a CNN framework to diagnose gastritis and found that the deep learning framework extracted more suitable features than histogram of oriented gradients (HOG), local binary pattern (LBP), and scale-invariant feature transform (SIFT), which extract the handcrafted low-level features. And they also argued that LIBLINEAR SVM classifier could handle the imbalanced data well. They⁸⁸ further introduced a high dispersal and local response normalization operation to the CNN framework to reduce redundancy and a multiscale features analysis to avoid its sensitivity to tongue deformation. Ma et al.⁸⁹ used the logistic regression (LR) model to integrate deep learning features of tongue images and canonical risk factors to screen patients with gastric precancerous lesions (PLGC), and the result showed 10.3% higher than that of the model only including canonical risk factors, which has demonstrated the value of tongue image characteristics in PLGC screening and risk prediction.

It is worth noting that the tongue was reported to have a strong association with COVID-19. The study said that⁴ the tongue image has excellent discriminative ability for screening COVID-19 cases when using deep learning framework. And patients with mild and moderate COVID-19 commonly would hold a light red tongue and white coating. In contrast, more severe patients had a purple tongue and yellow coating,³² highlighting that the fatty coating is a significant feature of COVID-19. Liang et al.³³ reported a cured case report of COVID-19 on its tongue diagnosis index and its Chinese medicine formula treatment which was also tweaked by the tongue features. They argued that the tongue color, fur thickness, and fur color were closely related to the progression of COVID-19. Wang et al.³¹ had demonstrated that a convolutional network with a transfer learning method could construct a robust classifier of COVID-19 based on tongue features.

Besides, there is some other creative or meaningful research on tongue diagnosis. Noguchi et al.⁹⁰ studied using the principal component scores of tongue color, gender, and age to diagnose Sjogren's syndrome through machine learning methods of SVM, RF, LR, bagging model of three SVMs, and stacking model of them. They found that SVM trained using principal component scores of tongue color, sex, and age showed the best accuracy, achieved significant values than other classifiers and different feature combinations, and reached a level comparable to machine learning models trained using the Saxon test. Another multifeature combination research from Zhang et al.⁹¹ used a low-rank representation model to form a multiview completion method to complete the missing view information of facial, sublingual vein, and tongue images for diagnosing fatty liver disease. And they used KNN, LDA, RF, least squares regression (LSR), or sparse representation classifier (SRC) as the classifier, which all showed better diagnostic results with the proposed approach. Jiang et al.⁵⁸ applied a Faster R-CNN model to recognize the region of cracked, tooth-marked, stasis spotted, greasy coating, peeled coating, and rotten coating of the tongue and then used the split and merge algorithm and color threshold method on this region to extract color feature and area value. Coupled with some classifiers, they found that these features would improve the accuracy in diagnosing nonalcoholic fatty liver disease (NAFLD). Huang et al.⁶ have studied different variables on tongue images between patients with acute ischemic stroke and health participants, and they have found that pale tongue color, bluish tongue color, ecchymoses, and tongue deviation angle were associated with significantly increased odds ratios for acute ischemic stroke through multiple LR analysis. Ning et al.⁹² employed a specially designed evaluation algorithm, balanced evolutionary semi-stacking (BESS) to simultaneously enhance balanced bagging and cotraining procedures when detecting diabetes mellitus, chronic kidney disease (CKD), breast cancer, and chronic gastritis and studied the tongue image classification performance of diseases of different machine learning methods. Therefore, the data and classifier diversity generated from it were fully considered to create multiple metafeatures for the stacking ensemble of BESS. and their result showed "SVM+LGBM (LightGBM)" achieved the best performance. Devi and Anita⁹³ used tongue images to diagnose thyroid and ulcers based on a semisupervised algorithm. Mansour et al.⁹⁴ reported using the Internet of Things technology and ResNet50 backbone in remote diagnosis of 12 diseases such as CKD, nephritis, verrucous gastritis, nephritis syndrome, chronic cerebral circulation insufficiency, and coronary heart disease. Furthermore, Thanikachalam et al.95 used a SqueezeNet model to extract tongue image features for classifying the same 12 diseases.

Coupled with other reviews, we have found that diagnostic models of diabetes and stomach diseases have been studied most in tongue diagnosis research. Among them, tongue images of diabetes patients have macroscopic changes, as we could easily perceive internal changes on the tongue which has no skin covering. And both the stomach and tongue belong to the esophagus, which may lead to their frequent association.96 Traditional feature engineering methods (handcraft features or manual features) and deep learning methods are all helpful in extracting disease-related features from tongue images, and combining the two can perform better.⁹⁷ Although the performance of tongue image is highly related to some diseases, it still cannot replace the existing diagnosis methods in terms of accuracy and timeliness, and it does not show the advantages of being able to replace the existing examination methods. As the default in TCM, the tongue is more used for SD than disease differentiation. In TCM, features from the tongue will always need to be combined with other symptoms in diagnosis, and it is not compliant with the holistic concept of TCM theory to diagnose diseases by the tongue alone. Therefore, the model of disease diagnosis based on the tongue may not only fail to achieve the expected effect but also face the situation of less clinical application scenarios.

Tongue image for syndrome differentiation

Syndrome (or Zheng⁷) differentiation plays a crucial role in TCM, and the prescription of compound Chinese herbs, acupuncture, massage, and so on all mainly depend on it. But the syndrome is not easy to distinguish, which poses a great challenge to physicians. Compared with other diagnosis methods of the syndrome, the healthy signals expressed from the tongue are more concrete and easier to perceive, and tongue diagnosis plays a more important role in SD.^{24,98,99}

The team of Guihua Wen has done a lot of work in this field.^{8,13,20,100} Body constitution type is another syndrome that only occurs in the healthy or subhealthy population, because of its inconspicuous symptoms; using the tongue to classify would be a great challenge. Even so, this team used a so-called complexity perception classification method¹⁰⁰ to classify constitution types based on tongue on the hard and easy samples, respectively, to solve the variation of environmental conditions and the uneven distribution of tongue images. A data set of 22,482 tongue images was collected from a local hospital and labeled body constitution types by TCM doctors. Faster R-CNN was used to detect a modified VGG-16 as a feature extractor. They applied a LR to judge whether the sample was hard or easy. Thus, if a sample is easily classified, it would be decided as an easy sample, otherwise a hard one. Then, the hard samples were used to train a difficult body constitution type classification model; on the contrary, the easy samples trained an easy model and found that the hard samples have a more dense distribution than the easy one; their method does improve the classification accuracy though its highest record is only 61%. Furthermore, Wen et al.²⁰ proposed grouping attributes zero-shot learning methods based on prior knowledge of TCM to solve the imbalanced constitution class problem in the next year. This time, they expanded their tongue image data set to 46,753, and 15 attributes of the tongue were proposed as a 15D semantic vector to represent the tongue image based on the knowledge of tongue diagnosis, as shown in Table 3. Combined with this prior knowledge, their method alleviated the uneven distribution problem and could classify the constitution type that has never been seen before. In addition, they⁸ studied the recognition of disease location of tongue image, a stochastic region pooling method was proposed to focus more on detailed regional features, and an inner-imaging channel-wise attention mechanism also was proposed to enhance the

sent tongue image	
vere used to repre	
15 attributes	
(TCM) experts. ²⁰ T	
Chinese medicine	
ing to traditional	
tution type accord	
agnosis for consti	
edge of tongue di	nantic vector.
ble 3. Prior knowl	mples in a 15D sei

as uct a ill saidillas															
	Tongue	Color		Tongue E	3ody				Tongue Na	ature					
Constitution Types	Pale Red	Reddish	Purple Black	Big Fat	Fat and Tender	Cracked	Tooth-Print	Varicose Veins	Thin White	Thick White Greasy	Yellow Greasy	Few Moss	Moist [Dry Ot	thers
Qi-deficiency	>			>			\mathbf{i}								
Yin-deficiency		\mathbf{i}				>						>		>	
Yang-deficiency					\geq		>						>		
Phlegm-wetness				>						\mathbf{i}					
Wetness-heat		\geq									>				
Blood-stasis			>					\geq							
Qi-depression	>								>						
Special-diathesis															
Gentleness	>								>				>		>

robustness of modeling relationships between CNN components. These methods have shown great adaptability in automatic tongue disease location prediction. Moreover, they studied^{13,101} an automatic prescription system of Chinese herbal based on CNN feature extractor and an auxiliary therapy topic loss mechanism using tongue image, which generated a prescription that relatively matched the real prescription. Li et al.¹⁰² proposed a complete diabetic tongue image classification method based on selfsupervised features from vector quantized variational autoencoder (VQ-VAE) and *K*-means clustering classifier; the classification result showed that the self-supervised method could also extract features of TCM health signals from diabetic tongue image to nicely separate different syndromes without human annotation.

Sometimes combining with other healthy signals would improve the classification accuracy of syndromes. Yuan and Liao⁶¹ combined tongue Lab color value and consultation based on a questionnaire to diagnosis constitution. They used a relationships table between tongue Lab color value and constitutions, and sentence similarity as the features to implement classification. Huang et al.¹⁰⁷ also integrated tongue image features with the indices of acoustic sound, and pulse signal, to imitate the "integrating four diagnoses into one reference" in the TCM clinical practice, which means the cross-reference between four diagnosis methods. They built an equation based on linear regression to model the association between these features and body constitution. The features of RGB and HSB color value of the tongue body were measured by Photoshop software, tongue coating features were represented by the modified Winkel tongue coating indices, and length and width of the sublingual vein were also included. The blood pressure feature in another study¹⁰⁸ was reported to be fused with the tongue feature and performed better than the tongue feature alone. Shi et al.¹⁰⁶ compared the effects of different feature combinations (tongue image, radial pulse wave, and body symptom) on classification accuracy, and the result was as follows: tongue and pulse < symptom < symptom and tongue and pulse. They also compared the syndrome classification performance of different classification methods classifying Qi deficiency and Yin deficiency in nonsmall cell lung cancer patients, which turned out the neural network performs better than RF, SVM, and LR. A similar result was received by reference¹⁰⁹ that the multilayer neural network was more adequate for modeling complex relationships between tongue color features and Zheng (syndrome)/ coating classes than SVM and AdaBoost. They used a feature vector that combined different pixel values of tongue image from different color spaces to represent tongue features and found that the tongue color features were more suitable to discriminate Zheng classes rather than the western groups (superficial vs. atrophic, Helicobacter pylori positive vs. negative).

As we all know, the syndrome in TCM is relatively abstract and vague; therefore, it is crucial to ensure the high quality of the training data when making the DIGITAL HEALTH

machine automatically differentiate syndromes. However, only part of the researchers have reported the details of how to label the syndrome type when building the training data set; as shown in Table 2, this decreases the reliability of their research. What is more, fewer studies discussed the consensus on syndrome labels, restricting their machine learning model's generalization. Nevertheless, the self-supervised method may have the potential quantitative ability to TCM symptoms or syndromes.¹⁰² There are huge benefits in accuracy when combining the features of tongue images and other symptoms in SD,^{106,107} which means it is strongly recommended to introduce other symptoms when using tongue images to classify syndrome types.

Tongue image for the symptom (sign) differentiation

Symptom/sign is essential for SD and sometimes for disease diagnosis. Classifying the symptoms/signs of tongues helps to explore AI's potential to quantify the changes in the tongue and the mechanism behind its symptoms/signs. However, the symptoms on the tongue are very difficult to automatically identify or quantify, which has become a core challenge in SD. In recent years, many studies in this area have paid much attention to using deep learning methods to automatically classify or identify tongue symptoms, like tooth-marked tongue,^{11,15,103–105,110,111} tongue coating,^{10,16,31,42,112,113} coating color,^{66,113,114} tongue color,^{22,113,115} cracked tongue,^{9,110,111,116,117} sublingual vein,¹⁰ and fungiform papillae on tongue.¹¹⁸

Using the tongue image to discriminate symptoms slightly differs from using it for SD. Because classifying a syndrome is an entire image-level classification task, the symptom classification is based on local features, which always show a low-level differentiation compared to the surrounding area on the tongue. Especially the toothmarked area on the tongue, to solve this problem, Li et al.¹⁵ proposed a multiple-instance SVM method. They first annotated the tooth-marked region and based on its convex hull features generated all suspected regions of tooth-marked through a color threshold method and classified the tooth-marked tongue based on this suspected region and tooth-marked region using multiple-instance SVM. In this way, they not only could extract features more specifically but also combined the advantages of both traditional and deep learning methods effectively. Similarly, they utilized this method to classify rotten greasy coating of the tongue¹⁶ and cracked tongue¹¹⁶ based on the deep features, which outperforms other state-of-the-art methods. Another tooth-marked classification model based on image-level annotation was proposed by Zhou et al.,¹⁰⁵ and it could also locate the tooth-marked area with the weakly supervised method. When classifying coating features, Wang et al.³¹ proposed a GreasyCoatNet framework that could classify three-level greasy coating robustly, indicating the

potential ability to quantify tongue greasy coating. Similarly, Zhuang et al.¹⁰⁴ deployed an intelligent detector using refined ResNet34 to discriminate the three-level thickness of the coating, which showed a better performance than VGG-16. The cross-channel attention module combined with the MobileNet V2 network¹⁰ got a competitive accuracy when classifying coating and sublingual vein thickness in a lightweight model. Ni et al.¹¹⁵ combined capsule network and residual block to launch a lightweight CapsNet model, which achieved a competitive performance when classified tongue color.

Labeled symptom data with high quality is hard to obtain; thus, we always face a lack of effective training samples. The transfer learning method is always applied to transfer general domain knowledge to alleviate small training data. Song et al.¹¹¹ utilized a pretrained model training from ImageNet to classify tooth-marked, cracked, and thick coating, which obtained a compatible result when compared with ResNet50 and Inception v3. Zhang et al.¹¹³ reported that the transfer learning method was used to differentiate tongue body color, coating color, and coating thickness for remote tongue diagnosis. Multitask learning could solve the classification and location tasks together; Weng et al.¹¹⁰ proposed a weakly supervised method to perform a coarse classification of tooth-marked and cracked tongues first and then a detection branch to locate the position of the features. For the problem of imbalanced data, Cao et al.¹¹⁹ create new samples with a linear interpolation method, which not only retained the characteristics of the original data but also avoided overfitting. The attention mechanism maximized the use of data in improving classification accuracy on cracked tongues.9

The handcrafted features are also useful to differentiate the tongue symptoms; Zhang et al.¹¹² used fractal spectra as a differentiating feature with high accuracy in the detection and classification of greasy and thin/thick tongue coatings based on fractal theory. Statistic features of Lab color value¹¹⁹ could be used to classify the unhealthy tongue images and prove that the XGBoost classifier has better accuracy than KNN, SVM, and RF. CIE L* a* b* value was applied to quantify the tongue color and classify its color type.¹²⁰ Another statistical feature of wide line on the tongue image would do a great favor in classifying cracked tongue.¹¹⁷ Features of too many irrelevant areas may negatively impact discrimination tasks. Wang et al.¹⁰³ utilized the segmented tongue image would have a 0.97% higher classification accuracy than the raw tongue image (no segmentation and having irrelevant facial portions and background surrounding the tongue) when classifying tooth-marked tongue.

The severity of the symptom/sign is closely related to the accuracy of SD and determines the dosage of the herbs. Using deep learning to classify the tongue's appearance has demonstrated the possibility of automatically classifying the body symptom/sign severity. Even in the face of low-feature level or dense prediction on small targets, they can be commendably utilized by the multiple-instance SVM method^{15,16,116} when compared with tongue color or coating color classification tasks based on image-level features. The irrelevant features or background pixels may reduce the discrimination accuracy.¹⁰³ Quantifying the severity of symptoms/signs through the machine is another challenge that may also be resolved through deep learning.³¹

Discussion and future directions

In this review, we have reported studies using AI methods in the tongue diagnosis area in recent years. As a traditional medical issue, the diagnosis from tongue images is different from other medical images that only focus on the size of space-occupying lesions, and tongue diagnosis needs to summarize 12 feature types, as depicted in Figure 1. So the automatic process of the tongue image has many differences from other medical images, and the machine needs to utilize and handle 12 different characteristics or volatile tongue appearance, as shown in Figure 3, which pose a greater challenge than other medical image and result in a later modernization.²² Just like tongue image segmentation, the performance has not been ideal for a long time, let alone the automation diagnosis based on tongue image. Until a wide range using the deep learning method and different kinds of CNN framework, we have witnessed the challenge being alleviated rapidly through this review.

Nevertheless, we have observed that most of the research was trapped in two major problems in research methods: data set construction and low reliability of performance evaluation. Since there is no recognized authoritative data set, on the one hand, most researchers can only build their data sets; on the other hand, their claimed model performance was based on their validation data sets, and no public channel is provided to confirm its performance. First of all, few studies reported the details of data collection, annotation, and annotation consistency when creating data sets, so the quality of data used for model training is questionable. As shown in Tables 1 and 2, only Shi et al.¹⁰⁶ report the collection details, such as the collection time and the body status at the time of collection; Zhuang et al.¹⁰⁴ even report that the distance between the eves and the display screen is 45 cm and each tongue image retention period is 20 s. Less literature reports the details of how to ensure the quality of labeling. Because the color of the tongue is easily affected by diet, it is possible to collect tongue images with deviation if the collection environment is not standardized. Thus, when labeling tongue images, if it cannot be confirmed that the labeling results are based on a wide range of consensus, the generalization performance of the model will be affected. Moreover, the definition of symptoms and syndromes is not clear enough, and its classification lacks recognized standards, which further reduces the generalization performance. So as the label setting, many binary classification tasks set their negative label as non or no tags, such as COVID-19 patients vs. non-COVID-19 patients,⁴ toothmarked vs. nontooth-marked tongue,¹⁰³ and diabetes tongue vs. nondiabetes tongue,⁷⁶ and these negative labels having a large value range are hard to cover in reality. It is hard to believe that the model discriminates against the right class when having partial negative labels. Secondly, the model performance claimed by the researches was mostly evaluated by their own data sets, and only a small part of tongue segmentation researches^{63,64} used BioHit public data set for verification. In addition, no reference disclosed its code and model weight files and provided relevant channels for the public to verify its performance. This not only makes it impossible to compare the studies horizontally but also harms their credibility.

Besides, we also observed some problems in research design, which can be roughly classified into two major categories: misunderstanding the application scenario of tongue diagnosis in TCM and SD based on a single feature or single diagnosis method. First, although tongue image has great advantages in the diagnosis and classification of diabetes and gastric diseases and may reflect the progress of COVID-19, its characteristics are more suitable for the classification of syndromes than diseases.¹⁰⁹ Because disease diagnosis requires more strict classification boundaries while syndromes are relatively unclear, at present, the optical signal of the tongue cannot be quantified to provide disease diagnostic indicators or specific substances with clear boundaries. Using tongue images to diagnose disease may have a high recall but a low accuracy. Second, there are many tongue image feature types, as shown in Figure 1, and they are more likely to be comprehensively judged as the probability of a certain type of health state.^{10,121} The health state abstracted from the whole body system is related to the concept of "syndrome." TCM theory believes that the human body is an interrelated system, the running state and abnormal symptoms of different body parts are interrelated and mutually affected, 1,122 showing a certain distribution law, and syndrome types could be treated as the clusters summarized by this distribution. Therefore, not only do the features of the tongue image need to be integrated but also the features from other diagnostic methods, to better distinguish syndromes.^{106–108} This is also the opinion of "integrating four diagnoses into one reference" emphasized by TCM, that is, to perceive the overall health state (or syndrome type) from multiple local abnormalities. Treatment should be based on the overall state, rather than focusing on a specific or specific part, such as an organ, signal pathway, or protein target.

Even though intelligent tongue diagnosis has many difficulties, there are still many solutions with potential utilization value. The self-supervised model¹⁰² obtains the feature embedding of tongue images through a self-encoder. It can preliminarily distinguish syndrome types without artificial subjective labels, which shows the potential ability of selfsupervised learning in the quantitative representation of TCM symptoms/signs. It is expected to bypass the deviation caused by empirical labels. It has also been proved that some traditional feature engineering methods can help improve deep learning performance, for example, repair or filtering of morphological layer,^{18,19} reasonably using the prior knowledge of experts,^{15,20} multiple-instance SVM method^{15,16,21,116} for small targets like the cracked feature or tooth-marked feature, and channel attention mechanism⁸⁻¹⁰ to better locate the tongue region while restraining useless parts. Some methods are particularly suitable for intelligent tongue diagnosis. Features such as tooth marks, cracks, fur color, and tongue color are often required to be comprehensively used when making clinical diagnoses, and the attention mechanism^{8,57} and multiscale feature extraction^{12,88} are conducive to integrating different features on tongue images. In addition, multitask learning, $^{66,110}_{64,110}$ data augmentation, $^{54,119}_{54,110}$ and transfer learning, 54,56,111 are effective ways to alleviate the problems of low data set quality and small data volume.

Based on the above discussion, there may be greater research value in the following aspects in the future. First is using the self-supervised learning method to embed tongue image features as input; as shown in the literature, 102 the prescription drugs and their doses are used as an output to bypass the complex and fuzzy intermediate process of SD and build an end-to-end clinical decision support system (CDSS) model, such as researches.^{13,101} Taking the process of SD as a black box model in the middle, directly modeling the relationship between symptoms/signs and doctor's prescriptions can not only effectively play the advantages of deep learning but also avoid error transmission in SD. Second is the correlation between the optical features of the tongue and other diagnostic features. In the long-term clinical practice of TCM, abnormal manifestations of various body parts often occur in association, showing a more robust joint distribution law with different body statues or syndromes than the diseases. The performance of the tongue body is also related to the overall health status. The effective integration or multimodal information fusion method of various parts or health features to obtain the embedding of overall health status is expected to deepen the understanding of the syndrome and reveal its essence. Third is the pathological basis behind the changes in tongue appearance, based on quantifying the symptoms of tongue images and exploring the relationship between them and genomics, proteomics, microbial populations, etc. to deliberate the essence behind various changes in tongue image.

Conclusion

This article provided a comprehensive overview of relevant works in intelligent tongue diagnosis over the past 5 years, including the progress of work contents and algorithm models, covering a wide range of tongue image calibration, detection, segmentation and classification of diseases, syndromes, and symptoms, as well as existing algorithm model approaches of manual feature methods, traditional feature engineering methods, and deep learning methods. In particular, we outlined future potential and the remaining limitations of these approaches toward intelligent tongue diagnosis that may hinder widespread clinical deployment. In the past, there was insufficient understanding of the value of traditional medicine and insufficient promotion of its "renaissance." This review shows that due to the characteristics of the noninvasive, rich vascular network, and rich microecology in tongue diagnosis, intelligent tongue diagnosis has enormous value for the future diagnosis and treatment of diseases, and this value needs to be further explored. We hope that this review may provide an intuitive understanding of this and also increase the awareness of common challenges in this field that call for future contributions.

Contributorship: All authors have made a substantial contribution to the development, drafting, and revising of this manuscript.

Declaration of Conflicting Interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Central Finance Improvement Project of the State Key Laboratory of Traditional Chinese Medicine (Central Finance. CS (2021) No. 151); National Natural Science Foundation of China, No. 81574038; China Postdoctoral Science Foundation, No. 2022M722210, and Shenzhen Basic Discipline Layout Project (JCYLL2022081 8101806014).

Informed consent: The images used in this paper were collected with informed consent, and as a review, the remaining content does not require informed consent.

Guarantor: QL

ORCID iD: Qi Liu (D) https://orcid.org/0000-0002-6113-856X

References

- He CH, Liao QY, Fu P, et al. Microbiological characteristics of different tongue coatings in adults[J]. *BMC Microbiol* 2022; 22: 214.
- Li YW, Cui J, Liu YF, et al. Oral, tongue-coating microbiota, and metabolic disorders: a novel area of interactive research[J]. *Front Cardiovasc Med* 2021; 8: 922.
- Cui JX, Cui HF, Yang MR, et al. Tongue coating microbiome as a potential biomarker for gastritis including precancerous cascade[J]. *Protein Cell* 2019; 10: 496–509.

- Zhang G, He XY, Li DL, et al. Automated screening of COVID-19-based tongue image on Chinese medicine[J]. *BioMed Res Int* 2022; 2022: 6825576.
- Wu TC, Lu CN, Hu WL, et al. Tongue diagnosis indices for gastroesophageal reflux disease: a cross-sectional, casecontrolled observational study[J]. *Medicine (Baltimore)* 2020; 99: 29.
- Huang YS, Wu HK, Chang HH, et al. Exploring the pivotal variables of tongue diagnosis between patients with acute ischemic stroke and health participants[J]. J Tradit Complement Med 2022; 12: 505–510.
- Wang YH and Xu AL. Zheng: a systems biology approach to diagnosis and treatments[J]. Science 2014; 346: S13–S15.
- Hu Y, Wen GH, Luo MN, et al. Fully-channel regional attention network for disease-location recognition with tongue images[J]. *Artif Intell Med* 2021; 118: 102110.
- Chen P, Men S, Lin H, et al. Detection of local lesions in tongue recognition based on improved Faster R-CNN[C].
 2021 6th International Conference on Computational Intelligence and Applications (ICCIA); 2021 11–13 June 2021; 2021. p. 165–168.
- Qiu DH, Zhang XY, Wan XH, et al. A novel tongue feature extraction method on mobile devices[J]. *Biomed Signal Process Control* 2023; 80: 104271.
- Tang WJ, 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[J]. *IEEE Access* 2020; 8: 153470–153478.
- Tang H, Wang B, Zhou J, et al. Ieee Comp SOC. DE-Net: dilated encoder network for automated tongue segmentation[C]. 2020 25TH INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION (ICPR); 2021; 2021. p. 2575–2581.
- Hu Y, Wen GH, Liao HQ, et al. Automatic construction of Chinese herbal prescriptions from tongue images using CNNs and auxiliary latent therapy topics[J]. *IEEE Transactions on Cybernetics* 2021; 51: 708–721.
- Li J, Yuan P, Hu XJ, et al. A tongue features fusion approach to predicting prediabetes and diabetes with machine learning[J]. J Biomed Inform 2021; 115.
- Li XQ, Zhang Y, Cui Q, et al. Tooth-marked tongue recognition using multiple instance learning and CNN features[J]. *IEEE Transactions on Cybernetics* 2019; 49: 380–387.
- Tang YH, Sun Y, Chiang JY, et al. Research on multiple-instance learning for tongue coating classification[J]. *IEEE Access* 2021; 9: 66361–66370.
- Liu WX, Zhou CG, Li ZY, et al. Patch-driven tongue image segmentation using sparse representation[J]. *IEEE Access* 2020; 8: 41372–41383.
- Zhou JH, Zhang Q, Zhang B, et al. TongueNet: a precise and fast tongue segmentation system using U-Net with a morphological processing layer[J]. *Applied Sciences-Basel* 2019; 9: 3128.
- Li L, Luo ZM, Zhang MT, et al. An iterative transfer learning framework for cross-domain tongue segmentation[J]. *Concurrency and Computation-Practice & Experience* 2020; 32: e5714.
- Wen GH, Ma JJ, Hu Y, et al. Grouping attributes zero-shot learning for tongue constitution recognition[J]. *Artif Intell Med* 2020; 109: 101951.

- Vocaturo E and Zumpano E. Machine learning opportunities for automatic tongue diagnosis systems[C]. 2020 IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE; 2020; 2020. p. 1498–1502.
- Peng R, Li Y, Rui Y, et al. Research progress of tongue image segmentation through artificial intelligence and deep learning[C]. 2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA); 2021 27–28 Aug. 2021; 2021. p. 677–683.
- Tania MH, Lwin K and Hossain MA. Advances in automated tongue diagnosis techniques[J]. *Integrative Medicine Research* 2019; 8: 42–56.
- Sun SP, Wei HM, Zhu RL, et al. Biology of the tongue coating and its value in disease diagnosis[J]. Complementary Medicine Research 2018; 25: 191–197.
- 25. Solos I and 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]. *Journal of Integrative Medicine-JIM* 2018; 16: 141–146.
- Cui JX, Hou SY, Liu B, et al. Species composition and overall diversity are significantly correlated between the tongue coating and gastric fluid microbiomes in gastritis patients[J]. *BMC Med Genet* 2022; 15: 60.
- Park SH, Shin NR, Yang M, et al. A clinical study on the relationship among insomnia, tongue diagnosis, and oral microbiome[J]. *Am J Chin Med* 2022; 50: 773–797.
- Kang XY, Lu B, Xiao P, et al. Microbial characteristics of common tongue coatings in patients with precancerous lesions of the upper gastrointestinal tract[J]. *J Healthc Eng* 2022; 2022: 7598427.
- 29. Lu HF, Ren ZG, Li A, et al. Deep sequencing reveals microbiota dysbiosis of tongue coat in patients with liver carcinoma[J]. *Sci Rep* 2016; 6: 33142.
- Han SW, Chen Y, Hu J, et al. Tongue images and tongue coating microbiome in patients with colorectal cancer[J]. *Microb Pathog* 2014; 77: 1–6.
- Wang X, Wang XR, Lou YN, et al. Constructing tongue coating recognition model using deep transfer learning to assist syndrome diagnosis and its potential in noninvasive ethnopharmacological evaluation[J]. *J Ethnopharmacol* 2022; 285: 114905.
- Vocaturo E and Zumpano E. On the development of a tool for tongue images analysis[C]. 2020 IEEE International Conference on Bioinformatics and Biomedicine 2020; 2020: 2318–2319.
- Liang KJ, Huang XD, Chen H, et al. Tongue diagnosis and treatment in traditional Chinese medicine for severe COVID-19: a case report[J]. Ann Palliat Med 2020; 9: 2400–2407.
- Kaeseler RL, Johansson TW, Struijk L, et al. Feature and classification analysis for detection and classification of tongue movements from single-trial pre-movement EEG[J]. *IEEE Trans Neural Syst Rehabil Eng* 2022; 30: 678–687.
- Kaeseler RL, Struijk L and Jochumsen M, Ieee. Detection and classification of tongue movements from single-trial EEG[C].
 2020 IEEE 20TH INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOENGINEERING (BIBE 2020); 2020; 2020. p. 376–379.
- 36. Xin Y, Cao YK, Liu Z, et al. Automatic tongue verification based on appearance manifold learning in image sequences

for the internet of medical things platform[J]. *IEEE* ACCESS 2018; 6: 43885–43891.

- Jiang T, Hu XJ, Yao XH, et al. Tongue image quality assessment based on a deep convolutional neural network[J]. BMC Med Inform Decis Mak 2021; 21: 147.
- Xian HM, Xie YY, Yang ZZ, et al. Automatic tongue image quality assessment using a multi-task deep learning model[J]. *Front Physiol* 2022; 13: 966214.
- Lu YX, Li XG, Gong ZP, et al. TDCCN: a two-phase deep color correction network for traditional Chinese medicine tongue images[J]. *Applied Sciences-Basel* 2020; 10: 1784.
- Zhuo L, Zhang J, Dong P, et al. An SA-GA-BP neural network-based color correction algorithm for TCM tongue images[J]. *Neurocomputing* 2014; 134: 111–116.
- Sui JY, Xia CM, Yang K, et al. Tongue image color correction method based on root polynomial regression[C]. PROCEEDINGS OF 2019 IEEE 8TH JOINT INTERNATIONAL INFORMATION TECHNOLOGY AND ARTIFICIAL INTELLIGENCE CONFERENCE (ITAIC 2019); 2019; 2019. p. 1337–1342.
- Hu MC, Lan KC, Fang WC, et al. Automated tongue diagnosis on the smartphone and its applications[J]. *Comput Methods Programs Biomed* 2019; 174: 51–64.
- 43. Wang XZ and Zhang D. A new tongue colorchecker design by space representation for precise correction[J]. *IEEE J Biomed Health Inform* 2013; 17: 381–391.
- Zhuo L, Zhang P, Qu PL, et al. A K-PLSR-based color correction method for TCM tongue images under different illumination conditions[J]. *Neurocomputing* 2016; 174: 815–821.
- Zhang HZ, Wang KQ, Jin XS, et al. Ieee. SVR based color calibration for tongue image[C]. Proceedings of 2005 International Conference on Machine Learning and Cybernetics, Vols 1–9; 2005; 2005. p. 5065–5070.
- Zhang B, Nie W and Zhao SP. A novel color rendition chart for digital tongue image calibration[J]. *Color Res Appl* 2018; 43: 749–759.
- Liu WX, Hu JM, Li ZY, et al. Tongue image segmentation via thresholding and gray projection[J]. KSII Transactions on Internet and Information Systems 2019; 13: 945–961.
- Liu Y, Chen D, Yu F, et al. A tongue segmentation algorithm based on LBP feature and cascade classifier[C]. 2020 2nd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM); 2020 15–17 Oct. 2020; 2020. p. 109–112.
- Wu KB and Zhang D. Robust tongue segmentation by fusing region-based and edge-based approaches[J]. *Expert Syst Appl* 2015; 42: 8027–8038.
- Guo JW, Yang YK, Wu QW, et al. Adaptive active contour model based automatic tongue image segmentation[C]. 2016 9TH INTERNATIONAL CONGRESS ON IMAGE AND SIGNAL PROCESSING, BIOMEDICAL ENGINEERING AND INFORMATICS (CISP-BMEI 2016); 2016; 2016. p. 1386–1390.
- Ning JF, Zhang D, Wu CK, et al. Automatic tongue image segmentation based on gradient vector flow and region merging[J]. *Neural Comput Appl* 2012; 21: 1819–1826.
- Long J, Shelhamer E and Darrell T, Ieee. Fully convolutional networks for semantic segmentation[C]. 2015 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR); 2015; 2015. p. 3431–3440.

- Wang L, He X, Tang Y, et al. Tongue semantic segmentation based on fully convolutional neural network[C]. 2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS); 2019 6–8 Dec. 2019; 2019. p. 298–301.
- Huang XD, Zhang H, Zhuo L, et al. TISNet-enhanced fully convolutional network with encoder-decoder structure for tongue image segmentation in traditional Chinese medicine[J]. *Comput Math Methods Med* 2020; 2020: 6029258.
- Ronneberger O, Fischer P and Brox T. U-Net: convolutional networks for biomedical image segmentation[C]. MEDICAL IMAGE COMPUTING AND COMPUTER-ASSISTED INTERVENTION, PT III; 2015; 2015. p. 234–241.
- Li MY, Zhu DJ, Xu W, et al. Application of U-Net with global convolution network module in computer-aided tongue diagnosis[J]. *J Healthc Eng* 2021; 2021: 5853128.
- Peng JQ, Li XL, Yang DW, et al. Automatic tongue crack extraction for real-time diagnosis[C]. 2020 IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE; 2020; 2020. p. 694–699.
- Jiang T, Guo XJ, Tu LP, et al. Application of computer tongue image analysis technology in the diagnosis of NAFLD[J]. *Comput Biol Med* 2021; 135: 104622.
- Yuan W and Liu CS, Ieee. Cascaded CNN for real-time tongue segmentation based on key points localization[C]. 2019 4TH IEEE INTERNATIONAL CONFERENCE ON BIG DATA ANALYTICS (ICBDA 2019); 2019; 2019. p. 303–307.
- 60. Zhou CG, Fan HY, Zhao W, et al. Reconstruction enhanced probabilistic model for semisupervised tongue image segmentation[J]. *Concurrency and Computation-Practice & Experience* 2020; 32: e5844.
- Yuan Y and Liao W. Design and implementation of the traditional Chinese medicine constitution system based on the diagnosis of tongue and consultation[J]. *IEEE access* 2021; 9: 4266–4278.
- Gao SS, Guo NN and Mao DQ. LSM-SEC: tongue segmentation by the level set model with symmetry and edge constraints[J]. *Comput Intell Neurosci* 2021; 2021: 6370526.
- Chou CG, Fan HY and Li ZY. TongueNet: accurate localization and segmentation for tongue images using deep neural networks[J]. *IEEE access* 2019; 7: 148779–148789.
- 64. Li XL, Yang DW, Wang Y, et al. Automatic tongue image segmentation for real-time remote diagnosis[C]. 2019 IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE (BIBM); 2019; 2019. p. 409–414.
- Li XL, Yang T, Hu YY, et al. Automatic tongue image matting for remote medical diagnosis[C]. 2017 IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE (BIBM); 2017; 2017. p. 561–564.
- 66. Xu Q, Zeng Y, Tang WJ, et al. Multi-task joint learning model for segmenting and classifying tongue images using a deep neural network[J]. *IEEE J Biomed Health Inform* 2020; 24: 2481–2489.
- 67. Cai YZ, Wang T, Liu W, et al. A robust interclass and intraclass loss function for deep learning based tongue segmentation[J]. *Concurrency and Computation-Practice* & *Experience* 2020; 32: e5849.
- 68. Xu H, Chen XF, Qian P, et al. A two-stage segmentation of sublingual veins based on compact fully convolutional

networks for traditional Chinese medicine images[J]. *Health Inf Sci Syst* 2023; 11: 19.

- Wu J, Zhang B, Xu Y, et al. Tongue image alignment via conformal mapping for disease detection[J]. *IEEE Access* 2020; 8: 9796–9808.
- Mozaffari MH and Lee WS. Encoder-decoder CNN models for automatic tracking of tongue contours in real-time ultrasound data[J]. *Methods* 2020; 179: 26–36.
- Mozaffari MH, Yamane N and Lee WS. Deep learning for automatic tracking of tongue surface in real-time ultrasound videos, landmarks instead of contours[C]. 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2020 16–19 Dec. 2020; 2020. p. 2785–2792.
- Mozaffari MH and Lee WS. Dilated convolutional neural network for tongue segmentation in real-time ultrasound video data[C]. 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2021 9-12 Dec. 2021; 2021. p. 1765–1772.
- Mozaffari MH and Lee WS. Second language pronunciation training by ultrasound-enhanced visual augmented reality[C].
 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2021 9-12 Dec. 2021; 2021. p. 3043–3050.
- 74. Delmoral JC, Ventura SMR and Tavares J. Segmentation of tongue shapes during vowel production in magnetic resonance images based on statistical modelling[J]. *Proceedings* of the Institution of Mechanical Engineers Part H-Journal of Engineering in Medicine 2018; 232: 271–281.
- Zhang B and Zhang H. Significant geometry features in tongue image analysis[J]. *Evidence-Based Complementary Altern Med* 2015; 2015: 897580.
- Zhang JF, Xu JT, Hu XJ, et al. Diagnostic method of diabetes based on support vector machine and tongue images[J]. *BioMed Res Int* 2017; 2017: 7961494.
- Selvarani A and Suresh GR. Decision support system for diabetes using tongue images[C]. 2020 International Conference on Communication and Signal Processing (ICCSP); 2020 28–30 July 2020; 2020. p. 0012–0016.
- Fan SY, Chen B, Zhang XR, et al. Machine learning algorithms in classifying TCM tongue features in diabetes mellitus and symptoms of gastric disease[J]. *Eur J Integr Med* 2021; 43: 101288.
- Deepa SN and Banerjee A. Intelligent decision support model using tongue image features for healthcare monitoring of diabetes diagnosis and classification[J]. *Netw Model Anal Health Inform Bioinform* 2021; 10: 41.
- Mathew JK and Sathyalakshmi S. ExpACVO-hybrid deep learning: exponential anti corona virus optimization enabled hybrid deep learning for tongue image segmentation towards diabetes mellitus detection[J]. *Biomed Signal Process Control* 2023; 83: 104635.
- Zhang NN, Jiang ZX, Li JX, et al. Multiple color representation and fusion for diabetes mellitus diagnosis based on back tongue images[J]. *Comput Biol Med* 2023; 155: 106652.
- Balasubramaniyan S, Jeyakumar V and Nachimuthu DS. Panoramic tongue imaging and deep convolutional machine learning model for diabetes diagnosis in humans[J]. *Sci Rep* 2022; 12: 186.
- 83. Vijayalakshmi A, Shahaana M, Nivetha NCD, et al. Ieee. Development of prognosis tool for type-II diabetics using

tongue image analysis[C]. 2020 6TH INTERNATIONAL CONFERENCE ON ADVANCED COMPUTING AND COMMUNICATION SYSTEMS (ICACCS); 2020; 2020. p. 617–619.

- 84. Srividhya E and Muthukumaravel A. Diagnosis of diabetes by tongue analysis[C]. PROCEEDINGS OF 2019 INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE AND KNOWLEDGE ECONOMY (ICCIKE' 2019); 2019; 2019. p. 256–259.
- Gholami E, Tabbakh SRK and Kheirabadi M. Increasing the accuracy in the diagnosis of stomach cancer based on color and lint features of tongue[J]. *Biomed Signal Process Control* 2021; 69: 102782.
- Zhu XL, Ma YH, Guo D, et al. A framework to predict gastric cancer based on tongue features and deep learning[J]. *Micromachines* 2023; 14: 53.
- Meng D, Cao GT, Duan Y, et al. A deep tongue image features analysis model for medical application[C]. 2016 IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE (BIBM); 2016; 2016. p. 1918–1922.
- Meng D, Cao GT, Duan Y, et al. Tongue images classification based on constrained high dispersal network[J]. *Evidence-Based Complementary Altern Med* 2017; 2017: 7452427.
- Ma CZ, Zhang P, Du SY, et al. Construction of tongue image-based machine learning model for screening patients with gastric precancerous lesions[J]. *J Pers Med* 2023; 13: 271.
- Noguchi K, Saito I, Namiki T, et al. Reliability of noncontact tongue diagnosis for Sjogren's syndrome using machine learning method[J]. *Sci Rep* 2023; 13: 1334.
- Zhang Q, Wen J, Zhou JH, et al. Missing-view completion for fatty liver disease detection[J]. *Comput Biol Med* 2022; 150: 106097.
- Ning ZH, Ye ZQ, Jiang ZX, et al. BESS: balanced evolutionary semi-stacking for disease detection using partially labeled imbalanced data[J]. *Inf Sci (Ny)* 2022; 594: 233–248.
- Devi GU and Anita EAM. A novel semi supervised learning algorithm for thyroid and ulcer classification in tongue image[J]. Cluster Computing-the Journal of networks Software Tools and Applications 2019; 22: 11537–11549.
- Mansour RF, Althobaiti MM and Ashour AA. Internet of things and synergic deep learning based biomedical tongue color image analysis for disease diagnosis and classification[J]. *IEEE Access* 2021; 9: 94769–94779.
- Thanikachalam V, Shanthi S, Kalirajan K, et al. Intelligent deep learning based disease diagnosis using biomedical tongue images[J]. CMC-Computers Materials & continua 2022; 70: 5667–5681.
- Shang ZM, Du ZG, Guan B, et al. Correlation analysis between characteristics under gastroscope and image information of tongue in patients with chronic gastritis[J]. J Tradit Chin Med 2022; 42: 102–107.
- Feng L, Huang ZH, Zhong YM, et al. Research and application of tongue and face diagnosis based on deep learning[J]. *Digital Health* 2022; 8: 20552076221124436.
- Chengdong P, Li W, Dongmei J, et al. Establishing and validating a spotted tongue recognition and extraction model based on multiscale convolutional neural network[J]. *Digital Chinese Medicine* 2022; 5: 49–58.

- 99. Ding L, Zhang XY, Wu DY, et al. Application of an extreme learning machine network with particle swarm optimization in syndrome classification of primary liver cancer[J]. *Journal of Integrative Medicine-JIM* 2021; 19: 395–407.
- Ma JJ, Wen GH, Wang CJ, et al. Complexity perception classification method for tongue constitution recognition[J]. *Artif Intell Med* 2019; 96: 123–133.
- Wen GH, Wang KW, Li HH, et al. Recommending prescription via tongue image to assist clinician[J]. *Multimed Tools Appl* 2021; 80: 14283–14304.
- 102. Li J, Huang JB, Jiang T, et al. A multi-step approach for tongue image classification in patients with diabetes[J]. *Comput Biol Med* 2022; 149: 105935.
- 103. Wang X, Liu JW, Wu CY, et al. Artificial intelligence in tongue diagnosis: using deep convolutional neural network for recognizing unhealthy tongue with tooth-mark[J]. *Comput Struct Biotechnol J* 2020; 18: 973–980.
- Zhuang QB, Gan SZ and Zhang LY. Human-computer interaction based health diagnostics using ResNet34 for tongue image classification[J]. *Comput Methods Programs Biomed* 2022; 226: 107096.
- 105. Zhou JG, Li SX, Wang XS, et al. Weakly supervised deep learning for tooth-marked tongue recognition[J]. Front Physiol 2022; 13: 847267.
- 106. Shi YL, Liu JY, Hu XJ, et al. A new method for syndrome classification of non-small-cell lung cancer based on data of tongue and pulse with machine learning[J]. *BioMed Res Int* 2021; 2021: 1337558.
- 107. Huang CJ, Lin HJ, Liao WL, et al. Diagnosis of traditional Chinese medicine constitution by integrating indices of tongue, acoustic sound, and pulse[J]. *Eur J Integr Med* 2019; 27: 114–120.
- 108. Li ZR, Ren XJ, Xiao L, et al. Research on data analysis network of TCM tongue diagnosis based on deep learning technology[J]. *J Healthc Eng* 2022; 2022: 9372807.
- 109. Kanawong R, Obafemi-Ajayi T, Liu DH, et al. Tongue image analysis and its mobile app development for health diagnosis[M]. In: Shen B (ed.) *Translational informatics in smart healthcare*. Singapore: Springer, 2017, vol. 1005, pp. 99–121.
- 110. Weng H, Li L, Lei HW, et al. A weakly supervised tooth-mark and crack detection method in tongue image[J]. *Concurrency and computation-Practice & experience* 2021; 33: e6262.
- 111. Song C, Wang B and Xu JT, Ieee. Classifying tongue images using deep transfer learning[C]. 2020 5TH INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE AND APPLICATIONS (ICCIA 2020); 2020; 2020. p. 103–107.
- 112. Zhang J, Qian J, Yang T, et al. Analysis and recognition of characteristics of digitized tongue pictures and tongue coating texture based on fractal theory in traditional Chinese medicine[J]. *Computer Assisted Surgery* 2019; 24: 62–71.
- Zhang X, Huang W, Gao J, et al. Deep sparse transfer learning for remote smart tongue diagnosis[J]. *Math Biosci Eng* 2021; 18: 1169–1186.
- 114. Wang ZC, Zhang SP, Yuen PC, et al. Intra-rater and interrater reliability of tongue coating diagnosis in traditional Chinese medicine using smartphones: quasi-Delphi study[J]. *JMIR Mhealth Uhealth* 2020; 8(7): e16018.

- 115. Ni JH, Yan ZZ and Jiang JH. TongueCaps: an improved capsule network model for multi-classification of tongue color[J]. *Diagnostics* 2022; 12(3): 653.
- 116. Xue YS, Li XQ, Cui Q, et al. Cracked tongue recognition based on deep features and multiple-instance SVM[C]. ADVANCES IN MULTIMEDIA INFORMATION PROCESSING - PCM 2018, PT II; 2018; 2018. p. 642–652.
- 117. Li XQ, Shao Q and Yao Q. Cracked tongue recognition using statistic feature[C]. 2014 IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE (BIBM); 2014; 2014.
- 118. Cattaneo C, Liu J, Wang CH, et al. Comparison of manual and machine learning image processing approaches to determine fungiform papillae on the tongue[J]. *Sci Rep* 2020; 10: 18694.

- Cao GT, Ding L, Duan Y, et al. Classification of tongue images based on doublet and color space dictionary[C].
 2016 IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE (BIBM); 2016; 2016. p. 1170–1175.
- Chen HS, Chen SM, Jiang CY, et al. Computational tongue color simulation in tongue diagnosis[J]. *Color Res Appl* 2022; 47: 121–134.
- 121. Wang J, Wang HJ, Xiao Y, et al. Microecology-turbidity toxin theory: correlation between *Helicobacter pylori* infection and manifestation of tongue and gastroscopy[J]. *J Tradit Chin Med* 2022; 42: 458–462.
- 122. Jiang T, Lu Z, Hu XJ, et al. Deep learning multi-label tongue image analysis and its application in a population undergoing routine medical checkup[J]. *Evidence-Based Complementary Altern Med* 2022; 2022: 3384209.