

## Review

Jiacheng Xie, Congcong Jing, Ziyang Zhang, Jiatio Xu, Ye Duan and Dong Xu\*

# Digital tongue image analyses for health assessment

<https://doi.org/10.1515/mr-2021-0018>

Received July 14, 2021; accepted December 13, 2021;

published online February 14, 2022

**Abstract:** Traditional Chinese Medicine (TCM), as an effective alternative medicine, utilizes tongue diagnosis as a major method to assess the patient's health status by examining the tongue's color, shape, and texture. Tongue images can also give the pre-disease indications without any significant disease symptoms, which provides a basis for preventive medicine and lifestyle adjustment. However, traditional tongue diagnosis has limitations, as the process may be subjective and inconsistent. Hence, computer-aided tongue diagnoses have a great potential to provide more consistent and objective health assessments. This paper reviewed the current trends in TCM tongue diagnosis, including tongue image acquisition hardware, tongue segmentation, feature extraction, color correction, tongue classification, and tongue diagnosis system. We also present a case of TCM constitution classification based on tongue images.

**Keywords:** computerized diagnosis; image analysis; machine learning; mobile app; tongue diagnosis; traditional Chinese medicine.

## Introduction

Traditional Chinese Medicine (TCM) [1] has a long history in health care. TCM diagnosis is generally based on the

information obtained from four diagnostic processes: inspection, auscultation and olfaction, inquiry, and palpation. Inspection tops the four ways of TCM diagnosis, and the tongue is the primary subject for TCM inspection [2, 3]. As a simple, non-invasive, and valuable diagnostic procedure, tongue diagnosis [4] has been successfully utilized by TCM practitioners for at least 2,000 years [5]. In the TCM theory, the tongue has connections to the conditions of organs and body fluids, as well as the degree and progression of the disease [6, 7]. The main features used in tongue diagnosis are color and coating. For example, the normal tongue is red with a thin white coating [8]. Some characteristic changes like red body, yellow coating, or thick coating occur in the tongue in some diseases. For example, tongues of patients with diabetes mellitus are often yellow with thick moss [9]; in cancer patients, the tongue color is mainly purple and the tongues often do not have coating, but with thick greasy moss and slippery moss [10]; in patients with acute ischemic stroke, the tongues are often red and crooked, with greasy white moss [11]; red tongues with greasy white moss are also observed in patients with refractory *Helicobacter pylori* infection [12]. Tongue shape assessment is also an important component of tongue diagnosis. Such geometrical shape information includes thickness, size, cracks, and teeth-marks. In patients with primary Sjögren's syndrome, the tongues are typically thin, reddish-red, non-mossy, or cracked [13]. In patients with primary insomnia, tongues are predominantly red and fat, with yellow and white greasy moss, and tooth marks [14]; HIV patients show red and fat tongues as well, but with white thick moss and tooth marks [15]. Tongue coating, which is covered on a tongue like moss, is an important factor with many features, including color, degree of wetness, thickness, form, and distributed range, reflecting a patient's disease and body condition.

The tongue's body, tip, and root have changes, which may indicate particular pathologies [16]. The normal tongue is an ellipse tongue, but there are also six other classes of tongue shapes: square, rectangular, round, acute triangular, obtusely triangular, and hammer [17]. Numerous clinical reports [18–20] have associated tongue

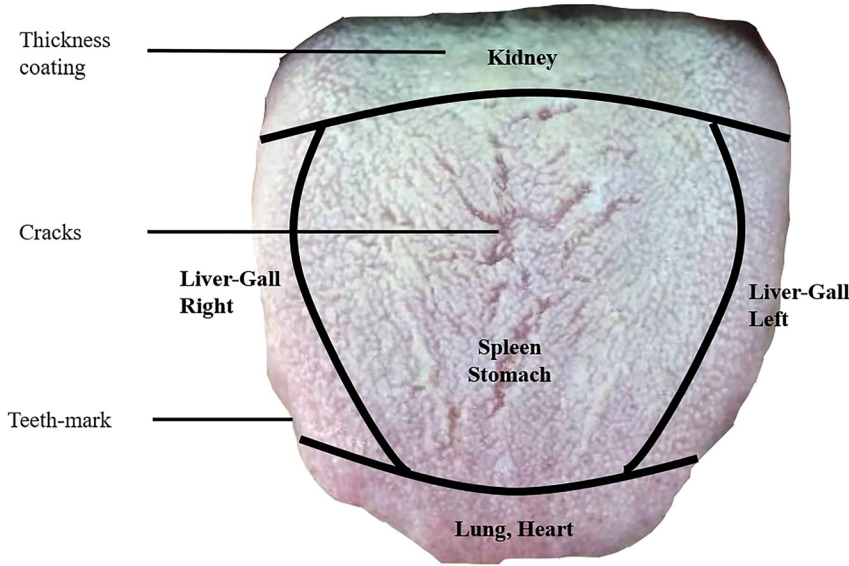
---

\*Corresponding author: Dong Xu, Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211, USA; and Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO 65211, USA, E-mail: xudong@missouri.edu. <https://orcid.org/0000-0002-4809-0514>

Jiacheng Xie and Ziyang Zhang, Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA; and Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO, USA

Congcong Jing and Jiatio Xu, School of Basic Medical Sciences, Shanghai University of Traditional Chinese Medicine, Shanghai, China

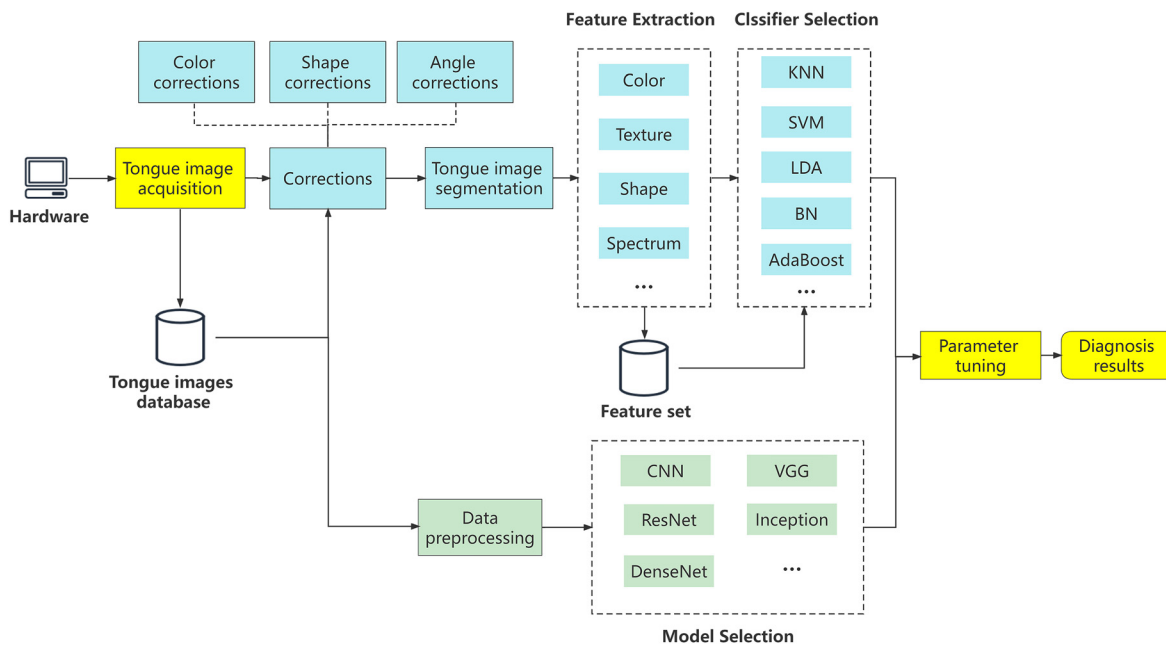
Ye Duan, Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, MO, USA



**Figure 1:** Organ correspondence of tongue regions.

shapes with diseases. For instance, a round tongue is associated with gastritis, an obtuse triangular tongue with hyperthyroidism, and a square tongue with coronary heart disease or portal sphygmo-hypertension. Organ conditions, properties, and variations of pathogens can also be found through observation of the tongue. For example, variations in tongue fur represent exogenous pathogenic factors and the flow of the stomach [21]. TCM usually divides the tongue into five areas, as shown in Figure 1. The left and right areas, the tip, the middle, and the tongue base reflect the conditions of the liver and gallbladder, heart and lungs, spleen and stomach, and kidneys, respectively [22–25].

The advantage of tongue diagnosis is that it is a simple and non-invasive technique. However, it is difficult to achieve an objective and standardized examination. Changes in inspection circumstances, such as light sources, affect results significantly. Moreover, because the diagnosis relies on the doctor’s experience and knowledge, it is hard to obtain a standardized result. Recently, various researches are being carried out to solve these problems [26]. In this review, we summarize the development of tongue diagnosis and current technologies. As shown in Figure 2, the general computerized tongue diagnosis process can be divided into two approaches. One is the traditional machine learning method, as shown in the blue part



**Figure 2:** General process of computerized tongue diagnosis. The blue parts represent the tongue diagnosis process using traditional machine learning methods, and the green parts represent the deep learning process.

of the figure. This method generally segments the raw tongue image, and then extracts features such as color, texture, shape, and spectrum from the segmented tongue image, and then selects the classifier to finally achieve tasks such as classification and recognition. The other is the deep learning method, which usually uses raw data for training and feature extraction by convolution operations, as shown in the green part of the figure. In this paper, we will conduct a review for both approaches.

The remainder of this paper be organized as follows. In “Hardware for tongue images collection” section, we reviewed the development of existing hardware for tongue image collection. Then, in “Tongue image analysis” and “Tongue diagnosis system (TDS)” sections, we reviewed studies on tongue diagnosis, including tongue image segmentation, feature extraction, and color correction, as well as tongue classification and diagnosis system. In “An intelligent TCM constitution classification system based on tongue image” section, we demonstrated our study as an application example of TCM constitution classification. Finally, we discussed the current tongue diagnosis work and gave possible future research directions in “Discussions” section.

## Hardware for tongue images collection

Tongue image acquisition is the first step in computerized tongue diagnosis. The quality of tongue image acquisition has an important impact on the subsequent sample labeling and analysis. In particular, clear and complete tongue images benefit tongue segmentation and feature extraction in the training phase of the machine-learning model development. However, the quality of tongue images is highly dependent on the hardware of the tongue image collection. The mainstream tongue image acquisition devices include commercial digital cameras and smartphones. Different manufacturers and brands of cameras/phones use different internal sensors. Devices with high-performance optical sensors are more likely to capture clear, high-resolution tongue images. Table 1 summarizes major hardware for tongue image collections.

Most recently developed hardware consists of three major components: an image sensing module, an illumination module, and a computing and control module [27] to obtain high-quality and reproducible tongue images under

**Table 1:** Tongue image collection hardware.

Author	Year	Imaging camera (pixels)	Illumination	Color temperature	Color correction
Chiu [25]	2000	440 × 400 in a 2/3 CCD camera	Two fluorescent lamps installed in office environment	5,400 K	Printed color card
Cai [28]	2002	Commercial digital still camera, 640 × 480	Office illumination		Munsell ColorChecker
Jang [42]	2002	Watec WAT-202D CCD camera, 768 × 494	Optical fiber source (250 W halogen lamp)	4,000 K	None
Wei [60]	2002	Kodak DC 260, 1,536 × 1,024	OSRAM L18/72–965 fluorescent lamp	6,500 K	Printed color card
Wang [29]	2004	Canon G5, 1,024 × 768	Four standard light sources installed in dark chest	5,000 K	Printed color card
Zhang [43]	2005	Sony 900E video camera, 720 × 576	Two 70 W cold-light type halogen lamps	4,800 K	Printed color card
He [61]	2007	DH-HV3103 CMOS camera, 2,048 × 1,536	PHILIPS YH22 circular fluorescent lamps	7,200 K	
Liu [61]	2007	Hitachi KP-F120 CCD camera	KOHLER illumination light source		
Zhi [32]	2007	Hyperspectral camera	KOHLER illumination light source		
Lo [37]	2013	CCD camera	Circular LED lighting		Datacolor Spyder 3 ELITE
Lu [38]	2018	EOS 1200D	Simulated D65 illuminant environment	6,500 K	Color Checker Digital SG
Zhuo [39]	2016	Logitech HD Pro C920 camera	Simulated D65 illuminant environment	6,500 K	Munsell color checker
Yamamoto [36]	2010	Hyperspectral camera, 480 × 640 pixels	Artificial sunlight lamp		
Kim [41]	2008	Digital camera, 1,280 × 960	Standardized light sources	5,500 K	Local minimum correction
Qi [40]	2016	Eolane digital camera, 2,048 × 1,536	LED illuminator	6,447 K	Color Checker Digital SG

varying conditions. In the early days, Cai et al. [28] applied a modified handheld color scanner with a microscopy slide on top of the tongue. The method can remove artifacts and avoid major color calibration. However, the scanner requires contact with the tongue and it is undesirable in a clinical setting. Hence, they later used a commercial digital camera ( $640 \times 480$  pixels) and a ColorChecker embedded inside the image for non-contact tongue image acquisition, a reference for subsequent tongue image acquisition devices. After this, various research groups developed and implemented their digital acquisition devices for tongue data collection. For instance, Wang et al. [29] employed a CCD digital camera with a resolution of  $1,024 \times 768$  pixels, and they mounted this camera on a face-supporting device. Jiang et al. [30] developed a tongue diagnosis system using a high-quality digital camera with a 7.2-megapixel resolution. To minimize color errors, a Munsell color checker was embedded inside their hardware for color calibration. Cijin et al. [31] collected tongue images through an in-house tongue capture device consisting of a three-chip CCD camera with eight-bit resolution. Two D65 fluorescent tubes were placed symmetrically around the camera to produce a uniform illumination, as shown in Figure 3A. The images captured in the JPEG format range from  $257 \times 189$  pixels to  $443 \times 355$  pixels.

CCD camera has the advantages of small size, high reliability, and simple operation, but there are some limitations for images captured by traditional CCD cameras. In particular, it is difficult to distinguish the RGB color space between the tongue and neighboring tissues, as well as between the tongue coating and the tongue body [32]. Some methods [33, 34] perform well only on tongue images acquired under some particular conditions but often fail when the quality of the image is less than ideal. The discriminative capacities of multispectral sensors can be improved in spectral resolution by using hyperspectral sensors with hundreds of observation channels. Spectroscopy is a valuable tool for many applications. For example, in remote sensing applications, researchers have shown that hyperspectral data are adequate for material identification in scenes where other sensing modalities are ineffective [35]. In addition, spectral measurements from human tissues have been used for biomedical applications. Zhi et al. [32] obtained the hyperspectral images by a particular capture device, and they used the hyperspectral properties of the tongue coating to build a Support Vector Machine (SVM) classifier. Yamamoto et al. [36] also used a hyperspectral camera to acquire hyperspectral tongue surface images. Their hyperspectral camera contains an array of transmissive grating sensors and an eight-bit monochrome CCD camera with  $480 \times 640$  pixels. Moreover,

this camera is capable of taking a full-sized hyperspectral image every 16 s. They also used multiple scattering reflection technologies for the lamp to reduce reflection and a semi-closed box to avoid unexpected light. However, they did not have a color correction in their hardware system. Overall, hyperspectral sensors in high spectral resolution can achieve better tongue classification performance than optical images.

In addition to the acquisition devices mentioned above, more sophisticated hardware was also developed. Lo et al. [37] developed an automatic tongue diagnosis system (ATDS) to assist TCM practitioners. This system consists of a camera with circular LED lighting, camera support, and a color card, with a sliding trail for vertical adjustment, as shown in Figure 3B. The circular LED lighting can provide a consistent and stable light source, and compensate for the variations in intensity and color temperature by calibrating brightness and color. The ATDS can automatically correct lighting and color deviation caused by the change of background lighting by a color bar placed beside the subject. The color calibration bar can make sure the image quality is consistent even when taken in different circumstances.

Furthermore, Lu et al. [38] utilized a device equipped with a manual camera (Canon EOS 1200D) with a simulated D65 illuminant environment as shown in Figure 3C, and it is intended to render the average daylight with a correlated color temperature of approximately 6,500 K. Although such commercial cameras can obtain higher-resolution images, it makes the whole tongue image acquisition system very bulky. To improve the device's portability, Zhuo et al. [39] developed a tongue image capturing device by adopting the Logitech Pro C920 camera and the D65 illuminant, whose brightness can be adjusted according to the actual requirements. Its appearance and internal structure are shown in Figure 3D. In addition, Qi et al. [40] introduced a handheld tongue imaging device called TDA-1, which consists of an image acquisition system, LED illuminator, and removable collecting ring. As shown in Figure 3E, it is equipped with a CCD camera as its photosensitive components, including a small-sized Eolane digital camera (Altek A12, China). Their hardware system can capture color images with a resolution of  $2,048 \times 1,536$  pixels, and operators can adjust white balance, exposure time, exposure compensation, ISO speed, metering modes, flash mode, etc. Thus, the whole hardware system can create a stable light source environment for tongue image acquisition.

In other cases, many tongue acquisition devices use other standards. Kim et al. [41] used standardized light sources, a digital camera, and color correction to acquire tongue images. As shown in Figure 3F, they designed the camera according to the facial contour, and a strobe light

with 5,500 K color temperature was used to simulate the standard light source. Then an image with the resolution of  $1,280 \times 960$  pixels with RGB 24 bit-BMP format was obtained with a cross mark on the center of the image itself. The advantage of this tongue acquisition system is a graphic user interface (GUI) on the monitor showing the tongue image in real-time. The operator can locate the center of the tongue and draw a cross on an image, which significantly improves the quality of the tongue image acquisition. Jang et al. [42] chose the Watec WAT-202D CCD camera for the collection of tongue images. This type of camera has less distortion and lower price than the CMOS-type array sensor. It provided a maximum of  $768 \times 494$  pixels with push-lock white balance, 50 dB S/N ratio, and NTSC signal. In their hardware system, the light source was a 250 W halogen with 4,000 K color temperature. Unlike the previously designed acquisition hardware, their hardware system takes more into account the durability, power consumption, and cost of the device. Zhang et al. [43] chose a Sony 900E video camera as the image capture device, with a new type 3 CCD kernel with relatively less distortion. This camera provides images with a maximum of  $720 \times 576$  pixels, 50 dB S/N ratio, as well as both PAL and NTSC signals. The light source of the acquisition device is two 70 W cold-light type halogen lamps with a 4,800 K color temperature. To compensate for the high heat emission, they used optical fiber as a waveguide and installed the light source separated from the image acquisition part. Their tongue acquisition hardware was approved by six TCM experts and received a Chinese Invention patent (NO. 021324581).

In general, most acquisition devices can provide a stable light source environment for tongue diagnosis. However, there is still no uniform standard for tongue acquisition hardware parameters, such as the type of camera and light source conditions.

## Tongue image analysis

TCM diagnosis based on tongue images often includes image segmentation, color correction, and tongue type classification, although the former two may be skipped when using deep-learning methods. In this section, we will explain each of them in detail.

### Tongue segmentation and effective part extraction

In tongue image diagnosis, color, shape, movement, and coating are the main factors for consideration. The

thickness of the tongue, the size and cracks of the tongue surface, and the tooth marks on the edge of the tongue are also important for tongue diagnosis. Therefore, the segmentation of the tongue image not only can effectively filter out the interference of background information but also has great significance for the subsequent classifier training. However, there are many challenges in tongue image segmentation. The main challenge is that the tongue color is similar to the color of the lips and face, making it difficult to segment the tongue from face and lips.

The existing approaches for tongue image segmentation can be divided into two types: one based on traditional image processing technology and the other based on deep learning. In the early research, manual segmentation and automatic segmentation algorithms were used to segment the tongue area. Some methods employed the techniques of the adaptive threshold [44], region growing and merging [45], and edge detection [46, 47] to segment the tongue body from tongue images. These algorithms require the user to draw an initial boundary but often fail to completely extract the tongue body from the surroundings. To make segmentation more automatic, various advanced image processing techniques were proposed. The active contour model (ACM) or Snakes [48] was developed for contour extraction and image interpretation. Snakes incorporates a global view of edge detection by assessing continuity and curvature combined with the local strength of an edge. A major advantage of this approach over other techniques is the integration of image data, an initial estimate of the contour, the desired contour properties, and knowledge-based constraints in a single extraction process [49].

Many variants of the ACM-based algorithms are developed following the initialization methods and the strategy of curve evolution. Regarding the initialization, Pang et al. [49, 50] combined the deformable template techniques with ACM to build an original model called Bi-Elliptical Deformable Contour (BEDC) by introducing a new term, i.e., addressed template force, to maintain the global shape while locally deforming the details. In addition, the template force can prevent undesirable deformation effects of traditional Snakes, like shrinking and clustering. Their method exhibited better performance than traditional Snakes in dealing with noises. However, the BEDC fails to find the correct tongue contour when the tongue edges become very vague or even totally missing. Zuo et al. [33] developed a tongue segmentation method by combining the polar edge detector, edge filtering, edge binarization, active contour model, and a method to filter out useless edges for tongue segmentation. Furthermore, Wu et al. [51] introduced the watershed transform for acquiring initial contour, which was later used as a Snake to converge to the

exact edge. Regarding the curve evolution, Yu et al. [52] extracted the tongue body by adding a color gradient to the gradient vector flow (GVF) Snake. Shi et al. [53] applied double geodesic flow to extract tongue body based on the prior information of tongue shape and location. Later, Shi et al. [54] continued the work with the color control-geometric and gradient flow Snake algorithm-enhanced curve velocity. Ling et al. [55] presented a segmentation method based on the combination of gray histogram projection and automatic threshold selection. They determined the area of the tongue by performing horizontal and vertical gray-scale projection of the gray-scale image, and used the Otsu method to select the threshold to segment. Although these mentioned algorithms can segment the tongue part from the original images, active geodesic contour costs much more time than the parameterized mode. Based on the statistical distribution characteristics of tongue color, Wang et al. [56] introduced a mathematically described tongue color space for diagnostic feature extraction. The characteristics of tongue color space includes a tongue color gamut that defines the range of colors, color centers of 12 tongue color categories, and color distribution of typical image features in the tongue color gamut. They built the tongue color gamut in the CIE chromaticity diagram by color gamut boundary descriptor using one-class SVM. Then, they defined centers of 12 tongue color categories and built a relationship between the tongue color space and color distributions of various tongue features. In the end, the descriptor can cover 99.98% of tongue colors, and 98.11% of them are densely distributed in 98% tongue color gamut. These tongue feature extraction methods provide a lot of valuable feature information for computerized tongue diagnosis.

In some other studies, spectral image data have also been applied to tongue segmentation. Li et al. [57] presented a segmentation method based on the hyperspectral tongue image data. They constructed a transformed data cube by finding the spectral angle (SA) between each pixel and every other pixel in the original data cube, and then used a one-dimensional edge detector to analyze each spectrum in the transformed SA cube. The contour is extracted from the hyperspectral tongue image according to the detected edges. This study provides a new perspective on the method of segmentation of the tongue image. In addition to the methods mentioned above, some studies further improved the Snake method. For example, Liang et al. [58] proposed a new efficient tongue segmentation approach based on the combination of a feature of tongue shape and the Snakes correction model. They used the tongue image in the HIS color model to obtain the tongue's

contour, then used the shape of the tongue to correct a preliminary tongue contour and applied it to the Snakes model to get the final result. Some researchers have also done theoretical studies of the method, such as Wei et al. [59] compared the Canny, Snake, and threshold (Otsu's thresholding algorithm) methods for edge segmentation. They found that the Canny algorithm is not suitable for tongue segmentation because it may produce many false edges after cutting; The Snakes segmentation in the tongue requires a larger convergence number, which is time-consuming computationally. The threshold method using Otsu's thresholding algorithm and filtering process can achieve an easy, fast and effective segmentation result in tongue diagnosis. This study has some reference value for the selection of tongue segmentation algorithms.

Some methods utilized multiple steps to achieve optimal tongue image segmentation and feature extraction. Kim et al. [41] used several steps to segment the tongue images, including preprocessing, over-segmentation, region merging, local minimum detection, local minimum correction, color edge detection, and curve fitting, and edge selection and edge smoothing. During the preprocessing period, tongue images are changed to the resolution of  $533 \times 400$  pixels before histogram equalization and edge enhancement. Then a graph-based segmentation method is used based on edge selection. However, the tongue segmentation in this study was performed using a natural number scoring from 1 to 5, which lacks a precise assessment method. Zhang et al. [62] used the HSV color space for color feature extraction. After getting tongue images, they changed them from RGB to HSV and then extracted H and S components as color features. Texture features were extracted based on a statistical analysis of the gray level co-occurrence matrix. Specifically, four feature vectors, including contrast (CON), angular second moment (ASM), entropy (ENT), and correlation (COR), were used to determine the features of greasiness in the tongue coating, as well as smoothness and wetness in the tongue body. For teeth mark extraction, the Graham convex hull algorithm was used to construct the convex hull of the tongue body. This study used HSV color space for tongue segmentation and texture and color feature extraction, but the number of samples during the experiment was a bit small. Li et al. [63] established a method to detect the amount of tooth-mark. It applies the characters of Gradient Vector Flow Snakes (GVF) Snakes, the features of the curvatures, and gradients in all the points of tongue contour. A tooth-mark is defined by the curvatures and gradients of the points in the boundary contour of the image. In their experiments, the accuracy achieved 98% compared with the results determined by

doctors. These traditional methods can produce satisfactory segmentation results to some extent. However, they have some disadvantages, mainly in three aspects: (1) these methods are sensitive to illumination changes and cluttered backgrounds; (2) they cannot segment the tongue from the lips accurately due to their similar colors, especially in the Snake-based methods; (3) most of these methods require preprocessing, such as tongue-body detection or require the initial region to be specified before segmentation begins.

In addition to traditional machine learning methods, some studies applied deep learning for tongue image segmentation. Li et al. [64] proposed a real-time, automatic tongue images segmentation method using a lightweight architecture based on the encoder-decoder structure. They also constructed a tongue image dataset for model training and testing, which contained 5,600 tongue images and corresponding high-quality segmentation labels. They demonstrated the model's effectiveness on BioHit, PolyU/HIT, and their datasets, and achieved the performance of 99.15%, 95.69%, and 99.03% intersection over union (IoU) accuracy, respectively. Li et al. [65] proposed a three-step iterative fully convolutional network (TFCN) to extract the tongue body from the original tongue image. Their method can directly learn the alpha matte from the input image by correcting misunderstanding in intermediates steps without user interaction or initialization. Compared with GrabCut [66], the Closed-Form matting [67], and KNN matting [68], their approach can achieve 97.94% IoU accuracy, far better than other methods. Qu et al. applied an encoder-decoder model called SegNet to segment the tongue image automatically. However, these mentioned solutions have some drawbacks. They need some additional preprocessing operations, such as brightness discrimination and image enhancement which complicate the whole segmentation process.

The above methods lose some information on image detail after the successive pooling layers. To solve these drawbacks, Lin et al. [69] presented DeepTongue, an end-to-end trainable segmentation method using a deep convolutional neural network (CNN) based on ResNet. Without preprocessing, tongue images can be segmented using a forward network of 50 layers with fast speed and high accuracy. Huang et al. [70] propose an end-to-end network called Tongue U-Net (TU-Net), which combines the classical U-Net structure with squeeze-and-Excitation (SE) block, Dense Atrous Convolution (DAC) block, and Residual Multi-kernel Pooling (RMP) block. They applied their method on a tongue dataset with 300 images, and it performed better than other segmentation methods such as U-Net, Attention U-Net. Zhou et al. [71] presented a tongue segmentation method using a multi-task, end-to-end

learning model named TongueNet, for supervised deep CNN training. They used a feature pyramid network based on the designed context-aware residual blocks to extract multi-scale tongue features. The region of interests (ROIs) from feature maps was also used for finer localization and segmentation. In a small-scale tongue dataset, they also applied U-Net for fast tongue segmentation, and achieved the highest accuracy of 98.45% and consumed 0.267 s per picture on average [72]. Also, using the U-Net structure, Zhu et al. [73] explored to extract the ultrasound tongue contour using the U-Net, and superior performance has been obtained. Their experiment shows that the U-Net model can extract frame-specific contours and be robust to misleading features in the ultrasound tongue image. Tang et al. [74] proposed a Dilated Encode Network (De-Net) for automated segmentation of tongue image acquired from a mobile device in an opening environment. Unlike the previous deep learning methods, which use continuous pooling operations to increase the perceptive field and capture more abstract features, their model designed an HCDC block to expand the receptive field without losing resolution by using dilated convolutions, to ensure more high-level features and high-resolution output. They also compare with some competitive methods, including SegNet, U-Net, and DeepTongue. The result shows their method obtains more accurate segmentation results on their Tongue database, and it is effective for tongue image segmentation tasks with mobile devices.

In another study, Huang et al. [75] presented an automated tongue image segmentation method using an enhanced fully convolutional network with an encoder-decoder structure. In the quantitative evaluation of the segmentation results of 300 tongue images from their tongue image dataset, the average precision was 95.66%. Xu et al. [76] proposed a Multi-Task Joint learning (MTL) method for segmenting and classifying tongue images. The method shares the underlying parameters and adds two different task loss functions. Moreover, two deep neural network variants (U-Net and Discriminative Filter learning) are fused into the MTL. The experimental results show that the joint method outperforms the existing tongue characterization methods. Yuan et al. [77] proposed a framework to integrate tongue detection and segmentation using cascaded CNNs by multitask learning. The advantage of this method is specifically designed for mobile and embedded devices. The size of their model is hundreds of times smaller than other deep learning models such as SegNet, Iterative TFCN, but the accuracy is comparable to other methods. Tongue crack segmentation is also an essential component of computer-aided diagnosis applied. Tongue cracks refer to fissures with different depths and

shapes on the tongue's surface and muscle layer. In most cases, it can be viewed as a changeful curve structure on the tongue's surface, and its depth is determined by the severity of atrophy and lesions of tongue mucosa. Also, the quantitative value of fissured tongue reflects the health condition of internal organs [78].

Another study by Chen et al. [79] proposed a tongue crack extraction method based on Bot-hat transform and Otsu adaptive threshold, which achieved an extraction accuracy of over 90%. Xue et al. [80] utilize Alexnet to extract the deep features of the crack region and train a multi-instance support vector machine (SVM) to make the final decision to obtain the object detection results. Chang et al. [81, 82] take advantage of ResNet50 as the model's backbone to recognize and localize tongue crack regions and visualize the fissure regions with Gradient-weighted Class Activation Mapping (Grad-cam). However, these methods have not achieved pixel-level precise extraction due to the vague tongue crack boundary. Peng et al. [83] propose a P-shaped neural network architecture based on a lightweight encoder-decoder structure to extract tongue cracks. They improved the U-Net framework structure and applied dual attention gates for better information fusion. For the class imbalance issue, they applied oversampling pre-training strategy to solve this problem. Their methods achieved better results and less time consumption on the extraction of tongue cracks. In addition, applying an attention mechanism to discover new tongue features (e.g., a particular color in a region that associates with a disease) may help TCM professionals to define robust diagnosis protocols.

In general, most segmentation methods can split out the tongue region, but for some images with low resolution or blurred tongue edges, the accuracy and effectiveness of tongue segmentation are still significant challenges.

## Tongue color correction

In the TCM tongue analysis, color and color differences convey important diagnostic information. However, there are often color deviations in the tongue image acquisition process, as different types and brands of digital cameras may use different color spaces to acquire images. As a result, it will be difficult to exchange or compare images reliably and meaningfully. The developed methods and obtained results on these device-dependent images may suffer from limited applicability. The other problem is that the color and intensity of the external light source have an impact on tongue pictures. Therefore, tongue image color

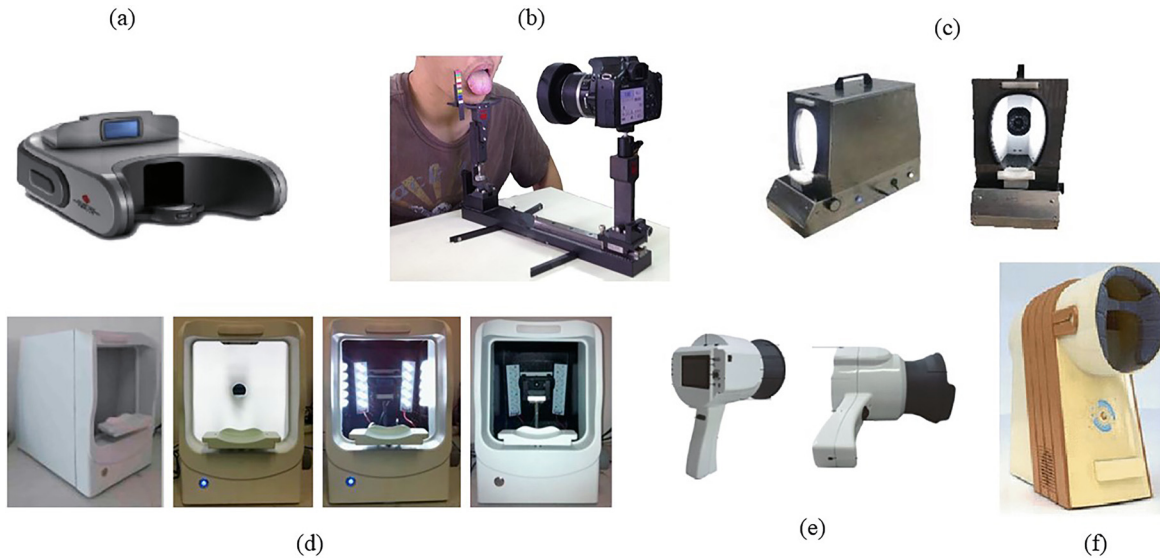
correction is an essential issue in the field of automatic tongue diagnosis.

To reduce the interference of the external light sources on tongue diagnosis, several color correction methods have been proposed [23, 43, 84–86]. Existing color correction methods can be classified into four categories, i.e., methods based on simple image statistics, color temperature curve calibration, double exposures, and supervised learning [87]. Among the supervised learning methods, polynomial-based correction and network mapping are the most widely used. However, most related research has focused on color correction for general imaging devices, such as digital cameras, cathode ray tube/liquid crystal display (CRT/LCD) monitors, and printers, where the color gamut covers almost the whole visible color area. Since the color gamut of the human tongue images is much narrower, these color correction algorithms need to be adapted and optimized.

In the early research, Jang et al. [42] proposed a Trigonal Pyramid (TP) color reproduction method without any color checker for reference. The disadvantage of their approach is which lacks enough objectivity. To compensate for the errors in lighting, camera angle, and color representation in the tongue image acquisition process, the color calibration method uses an embedded color checkboard. The color correction process typically involves deriving a transformation between the device-dependent camera RGB values and device-independent chromatic attributes with the aid of several reference colors, which are often printed and arranged in a checkboard chart named ColorChecker [88–91]. The ColorChecker is the reference target for training a correction model, and it plays a crucial role in tongue color correction. Cai [28] developed a semi-automatic color calibration tool, the Munsell ColorChecker [92], which contains 24 scientifically selected color patches, including additive primaries and colors of ordinary natural objects. The Munsell Colorchecker was designed in 1967 and is widely used in the color reproduction process of photography, television, and printing. Cai's software can find the points in each square of the color checker and apply a linear color calibration model to recover the original color of the tongue under various lighting conditions. However, its linear model cannot obtain satisfactory calibration results.

Among the supervised learning methods, the polynomial regression-based correction method is most widely used for its low computational complexity. Based on Support Vector Regression (SVR), Wang and Zhang [84] used a polynomial regression-based correction method for tongue image analysis. They proposed an optimized tongue color correction scheme to achieve accurate tongue



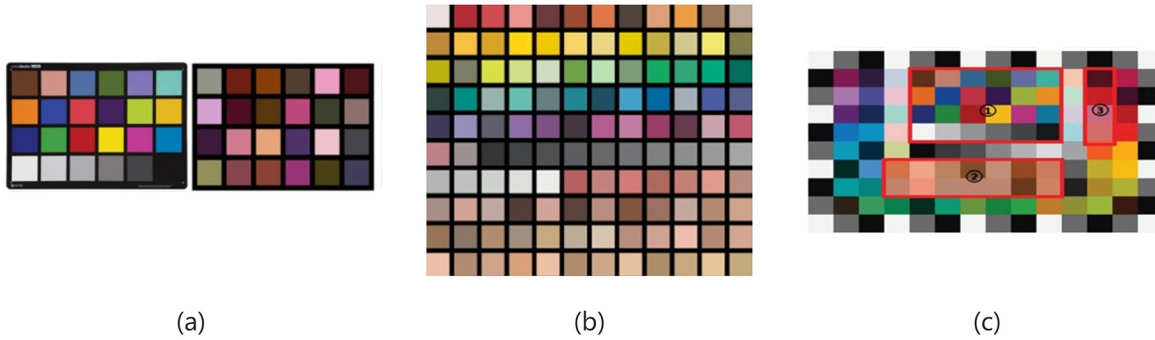


**Figure 3:** Tongue image collection hardware. (A) A tongue capture device consisting of a three-chip CCD camera [31]. (B) An automatic tongue diagnosis system using a camera with circular LED lighting, a color card, camera support, and a sliding rail for vertical adjustment [37]. (C) A device equipped with Canon EOS 1200D and a simulated D65 illuminant environment [38]. (D) A tongue image capturing device with Logitech Pro C920 camera and the D65 illuminant [38]. (E) The handheld TDA-1 tongue imaging capturing device [40]. (F) An integrated system with standardized light sources, a digital camera, and color correction [41].

image rendering. They chose the sRGB color space as selection criteria and then optimized the color correction algorithms accordingly. The method can reduce the color difference between images captured using different cameras or under different lighting conditions. The distances between the color centers of tongue images by more than 95%. Their further research [85] conducted a thorough study on the design of ColorChecker for more precise tongue color correction. Unlike the Munsell ColorChecker, they proposed a space-based color checker, as shown in Figure 4A. The Munsell ColorChecker was designed to process natural color but not especially for tongue colors, which means most colors like green and yellowish-green are unlikely to appear in tongue images. Therefore, this method is more accurate than the Munsell ColorChecker for tongue color correction.

To improve the correction accuracy of tongue colors, more studies have been reported. Hu et al. [93] applied the polynomial regression-based method for tongue image color correction and investigated the design of model parameters. Unlike the previous studies of tongue image analysis, the tongues were captured using a device with a fixed camera position, camera setting, and lighting condition to simplify the color correction procedure. They propose a lighting estimation method based on analyzing the CIE XYZ color information of photos taken with/without flash. Hence, their system works on smartphones and smartphone users do not have to take a color checker

with them. Hu et al. [94] also applied SVM to predict the lighting condition and the corresponding color correction matrix according to the color difference of images taken with and without flash. They also add a denoising step and the hue information of tongue images into the fur and fissure detection part for color correction. Zhang et al. [86] proposed a novel method to calibrate color distortion of tongue images captured under different lighting conditions. They adopted Li's regularized color clustering algorithm to produce the color codes of the new color checker [95]. To compare the difference of calibrations using different color codes, they combined standard 24 color values provided by Microsoft Windows, the cluster values from gray values into the color checker. Their experiments show that the SVR-based calibration method, cooperating with the proposed colorchecker, provides better overall performance than polynomial-based methods. In order to guarantee better color representation of the tongue body, researchers in Ref. [96] introduced a new tongue Color Rendition Chart for color calibration algorithms. They built a statistical tongue color gamut based on tongue image database, and determined different quantities of colors in the chart with experimentation. With the color difference calculation formula of CIELAB and CIEDE2000, they compared the results between tongue Color Rendition Chart and X-Rite's ColorChecker Color Rendition Chart, and got a smaller error rate with 24 colors (Figure 4).



**Figure 4:** Three common color checkers. (A) A checker with 24 colors in an improved version [85]. (B) A checker with 120 colors [97]. (C) Color checker SG highlights the interesting regions [38].

To overcome the problem of too few samples, Wei et al. [98] adopted Partial Least Squares Regression (PLSR) to correct tongue images in RGB color space. PLSR is a robust statistical analysis method for ordinary multiple regression, using relatively few samples, as the multiple corrections between variables. However, RGB color space is a device-dependent color space, making it difficult to calculate the difference in the device-dependent color space. Rosipalet et al. [99] improved the precision of fitting and prediction of the PLSR method by introducing a nonlinear kernel (named K-PLSR) to map the independent variable space to a high dimensional feature space. Zhuo et al. [39] proposed a K-PLSR-based color correction method for tongue images. The device-independent CIE LAB color space is adopted to train a K-PLASE model. Their method can correct tongue images under different illumination conditions to a consistent rendering result, suitable for standardized storage and automatic analysis in TCM.

Several studies applied neural networks for tongue image color correction. Zhuo et al. [97] proposed a simulated annealing (SA)–genetic algorithm (GA)–back-propagation (BP) neural network-based color correction algorithm for tongue images. They also only used several colors similar to the tongue body, tongue coating, and skin. The training samples consist of 120 color checkers, as shown in Figure 4B. They also established the color mapping model to improve the correction accuracy, with the captured samples of the color checkers under the capturing environment taken as the input data and the standard color data as the output. Their experimental results demonstrate that their color correction algorithm could improve the correction accuracy with a much lower computational complexity. Zhang et al. [100] developed a neural network-based color correction algorithm incorporating an evolutionary computation method. Their experimental results demonstrate that their method achieves better correction

performance than the polynomial regression model, the conventional back-propagation neural network, or the genetic algorithm-back-propagation neural network.

Lu et al. [38] proposed a Deep Color Correction Neural Network (DCCN) to model the relationship among the captured tongue images under different lighting conditions to the target images. The proposed DCCN learns the color mapping model with the operations in hidden layers. In their convolutional neural network architecture, the number of layers is determined via patches from the color checker, consisting of 140 patches, as shown in Figure 4C. Sui et al. [101] used the Root-Polynomial Color Correction (RPCC) algorithm to correct the color of the tongue. They collected the images of tongue and x-rite Color Checker Classic with a mobile phone simultaneously and used the CIR1976  $L \times a \times b$  color difference equation to evaluate the effect of their algorithm. The experimental results demonstrate that RPCC could improve color correction quality compared with the traditional polynomial regression algorithm and back-propagation neural network. Zhu et al. [102] used a new image retrieval method based on multi-feature of tongue images. For feature extraction, HIS color space is used; and statistics of gray difference presents texture. Then color and texture are uninformed from the 12-D eigenvector of colors and 8-D eigenvector of texture so that the matching degree of images could be judged by calculating the weighted distance of the uniform results.

## Tongue classification

In the TCM theory, a tongue is composed of a body, tip, and root [20]. These parts give the tongue its shape, which may change over time, and these changes may indicate pathologies. In oriental medicine, the classification of tongue images has important clinical diagnostic significance.

Therefore, researchers have conducted many studies on the classification of tongue images. For example, Huang et al. [103] presented a classification approach for automatically recognizing and analyzing tongue shapes based on geometric features. The approach corrects the tongue deflection by applying three geometric criteria and then classifies tongue shapes according to seven geometric features defined by various measurements of length, area, and angle of the tongue. The results show that the proposed shape correction method reduces the deflection of tongue shapes in a test on a total of 362 tongue samples. It achieved an accuracy of 90.3%, making it more accurate than either the K-nearest neighbors (KNN) or Linear Discriminant Analysis (LDA) method.

Zhang et al. [104] proposed a tongue images classification method via geometry features with Sparse Representation Classifier (SRC). Based on areas, measurements, distances, and ratios of foreground pixels, they extracted 13 geometry features in total in terms of width, length, length-width ratio, smaller-half-distance, center distance, center distance ratio, area, circle area, circle area ratio, square area, square area ratio, triangle area, and triangle area ratio. After that, they used two sub-tongue geometry feature dictionaries Healthy and Disease to classify SRC. In the end, they got an average accuracy of 79.23%, a sensitivity of 86.15%, and a specificity of 72.31% with 130 Healthy and 130 Disease samples. Zhang et al. [105] used geometric features for quantitative analysis through computerized methods. They used a decision tree model to classify the five tongue shapes (rectangle, acute and obtuse triangles, square, and circle) defined based on TCM. The experimental dataset included 130 healthy and 542 diseased Western medical calibration samples, and the average accuracy of the experiment was 76.24% for all shapes. This work provides a foundation for the objectification of tongue diagnosis. However, when extracting the shape and geometric features of the tongue, tongue image alignment faces significant challenges. One is that the tongue shape is highly unpredictable; the other is the consistency between the tongues of different people [106]. Wu et al. [107] present a conformal mapping method for tongue image alignment challenges. This method first establishes the mapping on the boundary by Fourier descriptors. It then extends the mapping to the interior region by means of the Coasean integration and finite difference methods. Their proposed method can be robust against tongue deformation and is faster and more accurate than the baseline method. Their work provides an effective and accurate tool for deformable medical image alignment and disease dispute.

In addition to classifying tongues based on their geometric features, some researchers have classified tongues

based on their textural features. For instance, Pang et al. [103] proposed a computerized tongue diagnosis method using a Bayesian network based on quantitative features, namely chromatic and textural measurements, as the decision models for diagnosis. Experiments were carried out on 455 inpatients affected by 13 common internal diseases and 70 healthy volunteers, with a prediction accuracy of up to 75.8% for diagnosing four groups, i.e., healthy, pancreatitis, hypertension, and cerebral infarction. Gao et al. [108, 109] presented a computerized tongue inspection method based on SVM. They used chromatic and textural measures extracted from tongue images as two quantitative features and then built the mapping relations between features and diseases using SVM and Bayesian networks. Experiments were carried out on 665 inpatients affected by six common internal diseases and 103 healthy volunteers. The estimated prediction accuracy of the multi-class SVM classification is 86.6%, which outperforms the Joint Bayesian Network (BN) Classification.

In our study, Ratchadaporn et al. [110] proposed a color-space-based feature set, which can be extracted from tongue images of clinical patients to build an automated ZHENG classification system. ZHENG (TCM syndrome) is an integral and essential part of TCM theory [111]. It is a characteristic profile of all clinical manifestations that can be identified by TCM practitioners, representing all the symptoms and signs (tongue appearance and pulse feeling included). Our experiment used Multilayer Perception (MLP) and SVM to establish the relationship between the tongue images features and ZHENG. The experimental results obtained over 263 gastritis patients, most of whom suffered Cold Zheng or Hot ZHENG, and a control group of 48 healthy volunteers demonstrate excellent performance.

Zhang et al. [112] introduced a new tongue color analysis system. They collected 1,045 images from 143 people as the control group (Healthy) and 902 patients with disease (Disease), and composed 13 ailment groups and one miscellaneous group. From CIE-xy chromaticity diagram, they marked 12 colors to represent the tongue color gamut and converted RGB value to CIELAB after images segmentation. With extracted features, they found Disease tongues have a higher ratio in red, deep red, black, gray, and yellow. They classified Healthy and Disease using KNN and SVM with quadratic kernel and selected features from sequential forward search. Then, 13 illnesses were grouped into three clusters by FCM to perform classification separately. It has an average accuracy of 91.99% to classify Healthy and Disease and more than 70% accuracy to distinguish between different illnesses. Lo et al. [21] investigated discriminating tongue features to differentiate between Breast cancer (BC) patients and the normal control

group, and established a differentiating index to facilitate the non-invasive detection of BC. They build an automatic tongue diagnosis system to extract the tongue features for 60 BC patients and 70 normal persons. The accuracy reached 80% by applying the seven tongue features. Some studies combine chromatic and textural features of the tongue for tongue classification. Researchers in Ref. [113] presented a tongue-computing model (TCoM) to diagnose appendicitis. Compared with other existing models and approaches, the underlying validity of the model is based on diagnostic results using Western medicine. The measurements of a tongue's chromatic and textural properties obtained via image processing techniques are compared with the corresponding diagnostic results from Western medicine instead of the judgment of a TCM doctor. This forms an evidence-based model and such an approach may avoid some subjective issues in judging Zheng. Their experiments selected 912 samples from a tongue-image database with more than 12,000 tongue images, including 114 images from appendicitis and 798 samples from 13 other familiar diseases. They evaluated the performance of the color metrics in each color space (RGB, CIEYxy, CIE-LUV, CIELAB) and textural metrics in different partitions of the tongue. The accuracy of the diagnosis of appendicitis is 92.28% after the identification of filiform papillae. In their experiments, integrating TCM tongue images with Western diagnostic results may provide a viable method for objectifying tongue diagnosis. Also based on a combination of features such as tongue color and texture, Zhang et al. [114] detect diabetes mellitus (DM) and nonproliferative diabetic retinopathy (NPDR) using tongue color, texture, and geometry features. They built a tongue color gamut with 12 colors, texture values with nine blocks, and geometry features from measurements, distances, areas, and ratios of the images. Then, these features were used to classify two groups healthy/DM and NPDR/DM without NPDR on a dataset of 130 Healthy and 296 DM samples, where 29 of those in DM are NPDR. In the end, the average accuracies of 80.52% for Healthy/DM group and 80.33% for NPDR/DM without NPDR group can be reached.

Spectral information is also an important feature used in the classification and diagnosis of tongue images. Hyperspectral images are carefully segmented in the spectral dimension. Unlike traditional color spaces such as RGB, hyperspectral images can obtain the spectral data of each point and the image information of either spectral band. Zhi et al. [32] presented a classification method for chronic cholecystitis based on hyperspectral medical tongue images. They collected 375 tongue images from 300 patients and 75 healthy volunteers by using a hyperspectral medical sensor. It shows that hyperspectral

tongue images can get better classification performance for patients than optical tongue images. This also indicates that hyperspectral cameras can capture more feature information compared to color digital cameras. Wan et al. [115] described the characteristic of tongue images of patients with lung cancer of different TCM syndromes and revealed the basic rule on the changes of the tongue images. A total of 207 patients with lung cancer were divided into four syndrome groups. The correct identification rate of the discriminant function on the raw data was 65.7%. Ding et al. [116] investigated an approach for the classification based on the doublet SVM. They acquired the pathological characteristics by using a robust approach with full use of the local information of tongue images. They extracted Histogram of Oriented Gradient (HOG) [117] features based on local object appearance and shape, and then calculated the distance metric by the SVM classifier and doublets with samples built by tongue images with different labels. The prediction accuracy of their method is 89.1%, with the specificity being 61.3% and the sensitivity being 95.8%.

Tooth marks are also an important feature used for tongue classification. In the actual diagnosis of TCM, a tooth-marked tongue or crenated tongue can provide valuable diagnostic information for TCM doctors. Li et al. [118] presented a multiple-instance method for the classification of teeth. The teeth marks are along the lateral borders. They generated the suspected regions, and then used a deep ConvNet to extract the feature. A bag of feature vectors represents a tongue and a multiple-instance SVM is used to make the final classification. Their experiments show that the proposed method dramatically improves accuracy and effectiveness. Zhang et al. [119] used transductive support vector machine (TSVM) for classification to improve the accuracy and reduce human labor from previous methods of others. They organized tongue images into 13 binary-category classification problems, and the classification rate is about 85%. However, the use of TSVM has high requirements for the quality of unlabeled tongue images. The classification rate would be lower, so the selection of images greatly influences the results of the classification. Then after TSVM, they also used Universum SVM to add labeled data and some selected irrelevant data (universum data) into the learning process to improve the classifier's performance [120]. Because not all universum data are useful for classification, they imported a selection algorithm to select the in-between universum samples (IBU). From their results, Universum SVM can get better performance than traditional SVM, and the parameters in kernel functions also play an important role.

Another common feature of classification is tongue coating. Li et al. [121] presented a method to classify rotten and greasy coating. They used random oversampling, Gabor feature, and distribution feature of tongue coating to solve the unbalanced classification and texture recognition problem. Gabor feature, as a rich and typical texture, is optimal in the sense of minimizing the joint two-dimensional uncertainty in space and frequency [122], and its micro-features are often used to characterize the underlying texture information [123]. Their experiment shows the method achieved high accuracy in the unbalanced data set. Huang et al. [124] employed a naïve Bayesian classifier to differentiate the three categories of fur coating, greasy fur coating, sub-greasy coating, and normal coating. Their experiment shows that the proposed method has a good performance in tongue coating classification. Xu et al. [125] established a tongue diagnosis method based on tongue images for tongue diagnosis. They used CCD devices to acquire images of the front and lateral of the tongue. After that, the tongue's length, width, and height are measured, and an optimum formula between the body surface area and the sum of the width of the tongue and height is established from the data. As a result of clinical studies, the accuracy rate for fat and thin tongues is 93.40% and 88.57%, respectively. This method is also helpful in the diagnosis of diabetes mellitus, hypertension, chronic gastritis, and hyperthyroidism.

In addition to the classification of tongue images using traditional machine learning methods, some studies have also used some deep learning methods to classify tongue images. For example, Tang et al. [126] have a more novel paradigm to classify tongue images by Multiple-Instance Learning (MIL) and in-depth features. They first selected suspected rotten-greasy tongue coating patches and then used a deep CNN to extract features of each patch. In the end, tongue coating is represented by a bag consisting of multiple feature vectors, and Multiple-Instance Support Vector Machine (MI-SVM) are used to perform the final classification. They achieved an accuracy of 85.0% and a recall rate (TPR) of 89.8%. Hou et al. [127] performed tongue color classification by modifying CaffeNet, and they constructed a tongue images dataset containing about 1,500 tongue images. They modified the traditional network parameters for tongue color classification and then fine-tuned the neural network model. Their experimental results show the method is practical and accurate in color classification. Huo et al. [128] presented a CNN method to classify three different tongue shapes, i.e., tooth-marked tongue, dot-sting tongue, and fissured tongue. They used the Gabor filtering algorithm and edge extraction approach for preprocessing and then optimized CNN for the tongue

image training. The experimental results indicate that the preprocessing methods increase the accuracy and decrease the time of the training process of tongue shape classification. Ful et al. [129] propose to computerize tongue coating nature using deep neural networks. Their method combines the characteristics of basic image processing and deep learning. Nonlinear activation function ReLU and dropout technique are used in their neural networks. They carried out their four-classification experiment and three-classification experiment, and the accuracy reached 0.87 and 0.95, respectively. Hou et al. [127] proposed a method combining deep learning with tongue color for classification. They created their tongue images database by preprocessing and enhancing tongue images. And then, they modified the parameters of the traditional networks. Their result shows that their method is more practical and accurate than the traditional methods. Meng et al. [130] propose a feature extraction framework called constrained high dispersal neural network (CHDNet) to extract unbiased features and reduce human labor for tongue diagnosis. High dispersal and local response normalization operation are introduced to address the issue of redundancy. They tested the proposed method on 267 gastritis patients and a control group of 48 healthy volunteers. The experiment results show that CHDNet is a promising method in tongue image classification for the TCM study. Song et al. [131] attempt to use deep transfer learning for tongue image classification. They extract the tongue features through the pre-trained networks (ResNet and Inception\_V3) and then rewrite the output layer of the original network with global average pooling and full-connected layer. A dataset of 2,245 tongue images collected from specialized TCM medical institutions is used for tongue classification performance evaluation. The proposed method performs well and achieves an average classification accuracy of 95.92%. Ma et al. [132] designed a system framework to take the constitution recognition. The system contains tongue images acquisition, image pre-processing, features extraction, and constitution recognition in total. At first, they take images by a camera from the wild environment directly. Then for the pro-processing part, tongue detection and images segmentation are included. They used a faster R-CNN method to detect tongue from the images, and another method, VGG, is also used to calibrate the detected tongue region. After features extraction, constitution recognition is performed with the complexity perception method. These studies of tongue classification provide methodological support for tongue diagnosis. We summarize the classification methods and the diseases classified in Table 2.

**Table 2:** Tongue classification and diseases.

Authors	Year	Classification type	Methods	Features	Sample size	Devices	Best accuracy
Pang [113]	2005	Appendicitis	quantitative measurements	Color and textural	912	3-CCD digital camera, standard D65 lights	92.98%
Zhi [32]	2007	Chronic Cholecystitis	SVM, RBFNN, KNN	Spectrum property	375	Hyperspectral medical sensor	93.11%
Gao [108, 109]	2007	Six common internal diseases	Bayesian Network, SVM	Color and textural metrics	768		86.60%
Huang [103]	2009	Tongue shapes	KNN, LDA	Geometric features	362		90.30%
Kanawong [110]	2012	ZHENG	AdaBoost, SVM, MLP	Color	263		
Lo [21, 133]	2016	Breast cancer	Logistic Regression	Color, thickness, shape, etc.	130	CCD camera with circular LED lighting	80.00%
Ding [116]	2016	Gastritis	KNN, RF, SVM	Textural	326	SONY 3-CDD camera	89.10%
Tang [126]	2021	Tongue coating	CNN, SVM	Coating	274	DS01-B Information Collection System of Tongue and Face Diagnosis	85.00%
Hou [127]	2017	Tongue color	CNN	Color	1,500		93.00%
Li [118]	2019	Teeth marks	CNN, SVM	Teeth marks	641		76.20%
Fu [129]	2017	Tongue coating	CNN	Textural	120		91.7%
Huo [128]	2017	Tongue shapes	CNN	Textural, tongue edge	718		81.2%
Meng [130]	2017	Gastritis	CNN, SVM	Color and textural metrics	315		91.14%
Song [131]	2020	Tooth-marks, cracks, etc.	ResNet, Inception	Tooth-marks, cracks, thickness	2,245	Tongue diagnostic instrument	95.92%

## Tongue diagnosis system (TDS)

A comprehensive tongue diagnosis system often has all the components, including hardware, image segmentation and extraction, color correction, and image classification. For example, Zhang et al. [43] presented a computer-aided tongue diagnosis system (CATDS), constituted by five components: user interface module, acquisition module, tongue image database, image pre-processing module, and diagnosis engine. This system aims to establish the relationship between quantitative features and disease via Bayesian networks. It is carried out on 544 patients affected by nine common diseases and 56 healthy volunteers. The results show that the system can adequately identify six groups: healthy, pulmonary heart disease, appendicitis, gastritis, pancreatitis, and bronchitis, with accuracy higher than 75%. The whole diagnosis process is less than 5 s. Jang et al. [42] developed a digital tongue inspection system, including hardware parts for tongue image acquisition, image processing for color interpolation, edge detection for tongue area separation, tongue color detection, and a database and user interface system for archiving and managing the acquired tongue images.

Kim et al. [134] presented a tongue diagnosis system to assess tongue coating thickness with functional dyspepsia. They obtain tongue images twice with a 30-min interval and then classify the type of tongues into three categories: no coating, thin coating, and thick coating. The system consists of an image acquisition system, LED illuminator, case, and analysis software. It is equipped with a vision camera (HVR-2130CPA, Hyvision System, Korea) and an H2Z0414C-MP lens (Hyvision System, Korea). The camera can take color images of SXGA resolution ( $1,280 \times 1,024$  pixels) and automatically adjust white balance, exposure time, and gain settings. To evenly illuminate the dorsal surface of the tongue, 12 white LED lamps are arranged around the camera, and thin double diffusion plates are placed in front of the lamps. To extract the tongue area, 17 nodes are generated to determine the boundary line of the tongue. Then, the image with RGB color space is converted into CIE-Lab color space to extract the coating area. For measurement of the system, they used a cutoff point of 29.06% to differentiate between no coating and thin coating, and a cutoff of 63.51% to differentiate between the thin coating and thick coating.

Lo et al. [133] presented an automatic tongue diagnosis system to use discriminating tongue features to distinguish

early-stage breast cancer (BC) patients. Nine tongue features, including tongue color, tongue quality, tongue fissure, tongue fur, red dot, ecchymosis, tooth mark, saliva, and tongue shape, are subdivided according to the areas located such as spleen–stomach, liver–gall-left, liver–gall-right, kidney, and heart–lung areas. The system’s main functions are image capturing and color calibration, tongues area segmentation, and tongue feature extraction. For image analysis, they isolated the tongue region within an image to eliminate irrelevant lower facial portions and background surrounding the tongue at first. Then, tongue features are extracted according to the aspect ratio, color composition, location, shape, and color distribution of the tongue. After that, the Mann–Whitney test, a non-parametric test, was used to compare two independent groups of sampled data with no assumption of normal distributions to identify features with significant differences. As for the result, the model employing five tongue features induced by logistic regression with independently significant meaning achieved 90% accuracy for non-breast cancer individuals and 50% accuracy for early-stage BC patients. The model employing six tongue features induced by logistic regression with independently significant meaning achieved 80% accuracy for non-breast cancer individuals and 60% accuracy for early-stage BC patients. Finally, the model employing seven tongue features achieved an accuracy of 80% on non-breast cancer individuals and 60% on early-stage BC patients.

Wu et al. [135] presented an automatic tongue diagnosis system (ATDS) to show the association between gastroesophageal reflux disease (GERD) and tongue manifestation and try to apply it to the process of noninvasive diagnosis of GERD. They used the system to acquire tongue images before the endoscope examination for some participants. In the process, nine tongue features were extracted, and a receiver operating characteristic (ROC) curve, analysis of variance, and logistic regression were used. The system contains a camera, light-emitting diode light, chin support, color bar, and adjustment. Three primary functions in terms of image capturing and color calibration, tongue area segmentation, and tongue feature extraction are processed successively. After that, tongue images are sub-divided into five segments: the spleen–stomach, liver-gall-left, liver-gall-right, kidney, and heart–lung areas; tongue features are also extracted, including tongue shape, tongue color, tooth marks, tongue fissure, fur color, fur thickness, saliva, ecchymosis, and red dots. With the records of endoscopic findings, GERD lesions are graded manually from A to D. Then, categorical test data and continuous data are tested by Chi-square tests and analysis of variance (ANOVA) tests, respectively. The odds

ratio and probability of a binary response are also estimated by logistic regression analysis. In the end, they got an AUC of  $0.606 \pm 0.049$  for the amount of saliva and  $0.615 \pm 0.050$  for tongue fur in the spleen–stomach area. Both the amount of saliva and percentage of tongue fur in the spleen–stomach area might predict the risk and severity of GERD, which might serve as noninvasive indicators of GERD. These systems have improved the efficiency of tongue diagnosis and helped TCM practitioners to manage patient health data more efficiently. We have organized the features and components of these tongue diagnosis systems in Table 3.

In addition to these desktop-based systems, there are some systems developed for mobile devices. For example, Ryu et al. [136] developed a tongue diagnosis system named TongueDx, including color correction and data history. Users can track their health condition by recording the color of tongue coating and body on smartphones. They also developed a line graph of tongue coating and body-color function, and users can know health conditions timely. Min-Chun et al. [25] presented an automatic tongue diagnosis framework to analyze tongue images taken by the smartphone. Since the same tongue’s images may look quite different in color due to various lighting conditions, they proposed a method to detect tongue features under different lighting conditions. They trained the lighting condition classifier based on the color distance of the tongue images pairs, which are captured with/without a flashlight. They used a color-checker based correction method to train the tongue image color correction matrices. They also trained a tongue feature detector for images under the standard lighting condition based on color features and SVM. They found that some tongue features have a strong correlation with the aspartate aminotransferase (AST) or alanine aminotransferase (ALT), which suggests the possible use of these tongue features captured on a smartphone for an early warning of liver diseases.

Hu et al. [137] proposed an automatic tongue diagnosis system with four main components, i.e., tongue photo-taking guide, tongue image color correction, tongue region segmentation, and tongue image diagnosis. In addition, a lighting condition estimation method based on the SVM classifier was used for unknown light sources and low-resolution images uploaded by users. Thus, tongue diagnosis systems can provide the general public and medical practitioners with convenient, prompt, and reliable diagnostic results for assessing health conditions. However, there is not yet a uniform standard for tongue diagnostic systems, leading to inconsistent analysis results in clinical diagnosis.

Table 3: Tongue diagnosis system.

Authors	Year	System name	System components	System features and functions
Min-Chu [25]	2000	Automatic tongue diagnosis framework		<ul style="list-style-type: none"> <li>- Tongue images taken by the smartphone</li> <li>- Lighting condition classifier is trained to eliminate influence of different lighting conditions</li> <li>- Color correction matrices</li> </ul>
Jang et al. [42]	2002	Digital tongue inspection system	<ul style="list-style-type: none"> <li>- Image acquisition module</li> <li>- Image processing module</li> <li>- Color interpolation</li> <li>- Edge detection algorithm</li> <li>- Tongue color detection algorithm</li> <li>- Database</li> <li>- User interface system</li> <li>- User interface module</li> <li>- Acquisition module</li> <li>- Tongue image database</li> <li>- Image preprocessing module</li> <li>- Diagnosis engine</li> </ul>	
Zhang [43]	2007	Computer-aided tongue diagnosis system	<ul style="list-style-type: none"> <li>- User interface system</li> <li>- Acquisition module</li> <li>- Tongue image database</li> <li>- Image preprocessing module</li> <li>- Diagnosis engine</li> </ul>	<ul style="list-style-type: none"> <li>- Bayesian networks are used to establish the relationship between quantitative features and disease</li> </ul>
Kim et al. [134]	2013	Tongue diagnosis system for tongue coating thickness assessment with functional dyspepsia	<ul style="list-style-type: none"> <li>- Image acquisition system</li> <li>- LED illuminator case</li> <li>- Analysis software</li> </ul>	<ul style="list-style-type: none"> <li>- White balance and exposure time adjusted automatically</li> <li>- White LED lamps are arranged around the camera</li> <li>- Thin double diffusion plates placed in front of the lamps</li> <li>- RGB color space is converted into CIE-Lab color space for tongue images</li> <li>- Isolated the tongue region to eliminate irrelevant lower facial portions and background</li> <li>- Using Mann-Whitney test to identify features with significant differences</li> </ul>
Lo et al. [133]	2015	Automatic diagnosis system for tongue diagnosis of patients with early breast cancer	<ul style="list-style-type: none"> <li>- Image capturing</li> <li>- Color calibration</li> <li>- Tongues area segmentation</li> <li>- Tongue feature extraction</li> </ul>	
Wu et al. [135]	2020	Automatic tongue diagnosis system for gastroesophageal reflux disease	<ul style="list-style-type: none"> <li>- Camera</li> <li>- Light-emitting diode light</li> <li>- Chin support</li> <li>- Color bar</li> <li>- Adjustment</li> </ul>	<ul style="list-style-type: none"> <li>- Receiver operating characteristic (ROC) curve</li> <li>- Analysis of variance</li> <li>- Logistic regression</li> <li>- Color calibration</li> <li>- Tongue area segmentation and feature extraction</li> <li>- Chi-square tests</li> <li>- Analysis of variance tests</li> </ul>

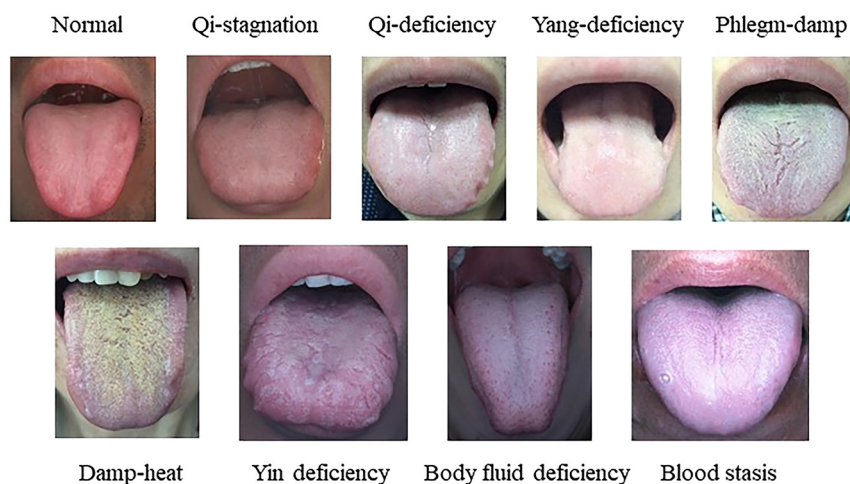


## An intelligent TCM constitution classification system based on tongue image

TCM focuses on preventive and personalized medicine by comprehensively assessing the physiological and pathological conditions (called constitution in TCM). In TCM, body constitution is regarded as a form of innate and acquired talent in the process of human life. It is a comprehensive expression of physiological function and psychological state [138]. The type of body constitution is highly related to some diseases and even determines the tendency of the disease [139, 140]. The classification of the TCM constitution is empirically based on many practitioners and a long TCM history. Since 2005, Wang Qi's classification of TCM constitutions has been regarded as the standard of TCM physique classification [141]. According to the national standard classification and determination of constitution in TCM published by China Association of Chinese Medicine in 2009, the constitution is categorized as into nine types: normal constitution, Qi-stagnation constitution, Qi-deficiency constitution, Yang-deficiency constitution, Phlegm-damp constitution, Damp-heat constitution, Yin deficiency constitution, Body fluid deficiency constitution and Blood stasis constitution [142]. The example samples are shown in Figure 5. The traditional constitution classification is generally determined by answers to a constitutional questionnaire that has been designed by TCM experts [143]. However, this method can be easily influenced by the individual subjective intention, and it will take a long time to finish the whole testing [132]. Zhang et al. [144] explored the correlation between TCM constitution and facial features. All face images were taken using a

facial biometric image acquisition system. Then facial features were extracted using a multi-channel detection method with color space and gradient and directional gradient histograms. They processed images by converting each pixel into a one-dimensional vector. After local binary pattern (LBP) feature maps were obtained, local histograms were used to analyze the regional distribution of texture. In the end, they got the result of  $36.72\% \pm 1.73\%$  classification accuracy rate based on LBP texture features for the 10 randomizations, and the experiment showed higher accuracy for LBP texture features compared with RGB pixel features. Lin et al. [145] analyzed clinical data with a topic model. The weighted mechanism was adopted for each feature word to improve the distinguishing ability and interpretability between the topics based on the LDA model. They adopted TF-IDF and Gauss function weighting, compared the KL distance, SVM classification accuracy, model complexity, and topic similarity, resulting improvement of the weighted LDA's performance. Then, a symptom-herb-therapy-diagnosis topic model was obtained and the Multi-Relationship LDA Model topic model combining symptom, herb, therapy, and diagnosis was proposed under TCM clinical background. As a result, they found weighted topic model can improve the performances of topic models at different levels. Their studies utilize biomedical and computer technologies to obtain constitution types more accurately and quickly. Tongue diagnosis offers a custom, immediate, inexpensive, and non-invasive solution to recognize the constitution [146, 147].

In this section, we implemented two types of deep learning methods: VGG and ResNet models, as well as two types of machine learning methods: Random Forest (RF) and SVM. The experiments results show that the deep learning method is feasible in TCM classification.



**Figure 5:** Tongue image of nine constitutions in traditional Chinese medicine.

## Data acquisition and preprocessing

All the data in our experiment comes from two parts, and the one part is from the Shanghai University of Traditional Chinese Medicine. There are 2,215 tongue images collected by professional personnel of TCM through the TFDA-1 digital tongue diagnosis instrument [148]. The other part is from the iTongue app, an intelligent system for personal health monitoring based on tongue image. It is also available in Apple App Store and other Android platforms. There are 2,572 tongue images from our app server database. The characteristic of this part of data is that the light source conditions of the tongue are complex and inconsistent, but it can well contain all kinds of complex light sources when users take photos.

As deep learning methods are applied to perform the constitution classification, it generally needs a large number of tongue images as training databases. It is common knowledge that the more data an ML algorithm has access to, the more effective it can be. Even when the data is of lower quality, algorithms can perform better, as long as useful information can be extracted by the model from the original data set [149]. Therefore, data augmentation is a common method to increase the size of the training database.

In the experiment, we performed data augmentation using Albumentations, a python library for fast and flexible image augmentations [150]. As shown in Figure 6, we performed some basic computer graphics transformations on the images in the original training database, such as blurring, random brightness, ISO noise, Gauss noise, and coarse dropout. In this way, we expanded the training dataset to 20 times the original. Some transformed parameters are shown in Table 3.

**Table 3:** Key parameters in images transformation.

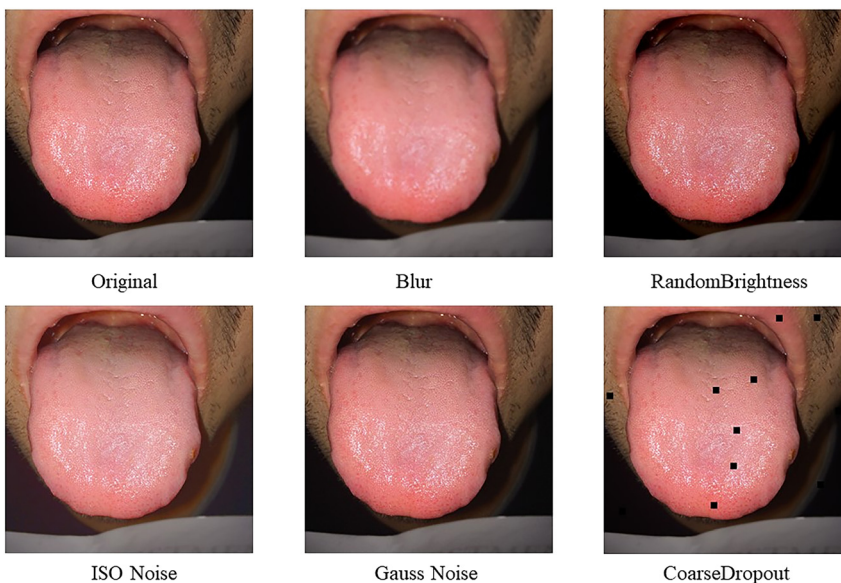
Methods	p	Range
Blur	0.3	blur_limit(3, 7)
RandomBrightness	0.5	limit(-0.2, 0.2)
ISO Noise	0.3	intensity(0.1, 0.5), color_shift(0.01,0.05)
Gauss Noise	0.2	var_limit(10.0, 50.0)
CoarseDropout	0.5	max_holes=8, max_height=8, max_width=8, min_holes=8, min_height=8, min_width=8

p represents the probability of applying the transform.

We used 70% of the mixed data set as the training set and the remaining data as the test set to preprocess the training set. In total, 11,200 tongue images were used as the training set. Then we train the model based on PyTorch, an open-source machine learning framework, and use ten-folds cross-validation to select the optimal model.

## Methods and experiments

In the experiment, we first used the RF and SVM methods to train the constitution classification model. In traditional digital image processing methods, the important step is to extract the features of tongue images. The changes in texture and color of the tongue image tend to reflect changes in health status. According to the TCM theory, a tongue image can be divided into five parts, each of which reflects different parts of the heart, liver, and so on. We divided the tongue into five parts and extracted the seven color spaces from the five parts, including RGB, HIS, YCrCb, CIE Lab, XYZ, LUV, and YUV. These seven color spaces have also been introduced in our previous research.



**Figure 6:** Example of tongue image data augmentation.

And then, we used these features to train our RF and SVM classifiers. These two classifiers are implemented by Scikit-learn [151], which is an efficient machine learning library in Python.

In addition to the traditional feature extraction methods, deep learning methods are recently proposed as new ways to perform feature extraction in segmentation, detection, and classification in computer vision. These methods have done well in ILSVRC (ImageNet Large Scale Visual Recognition Competition) that evaluates algorithms for images classification and object detection, such as AlexNet [152], VGG [152], GoogLeNet [153], ResNet [154], and SENet [155]. In the second experiment, we used VGG and ResNet to train the constitution classification model. These two classifiers were from the pre-trained models in Pytorch [156], and we fine-tuned the parameters in the models. Pytorch is a deep learning framework with a high-performance deep learning library. We set an initial learning rate of 0.001 for selecting parameters, and then we used the cosine learning rate decay method [157] to adjust the learning rate dynamically. We set the initial batch size to 32 for the batch size and conducted experiments with batch sizes of 16 and 64 based on our training results. Ten cross-validation method was used to evaluate our experimental results. We used GPU (GTX TITAN with 32 GB memory) to speed up the training process. After this experiment, we also compared our method with previous machine learning methods for constitution classification. To reduce the interference of non-tongue pictures on the physique classification results, we also trained a tongue image judgment model using the same development environment. We used the previous tongue images as the positive sample, and we downloaded 5,000 non-tongue pictures of 1,000 categories from the ImageNet database [152] as the negative samples. Ten-fold cross-validations are used to evaluate classification results. The recognition accuracy of tongue images is over 99%, and this tongue image recognition model will be applied to the system in “Application of the model” section.

## Results

We perform the classification experiment of nine constitution types on eight different classifiers to compare the effects of various methods. The accuracy of each method is shown in Table 4. The results of SVM and RF are close to 57.14% and 56.67%. The accuracy of VGG network deepening at any time has a slight improvement. The effect of the residual network is the best in these models, and the best result is close to 64.52%. In addition, we also tested the

**Table 4:** Comparison of results (the bold values show the best performance in the categories).

Method	Accuracy, %	MCC	F1
SVM	57.14	0.5373	0.5827
RF	56.67	0.5294	0.5696
VGG11	59.37	0.5535	0.6007
VGG13	60.61	0.5683	0.6107
VGG19	62.86	0.5922	0.6242
ResNet18	58.06	0.5513	0.6002
ResNet34	61.29	0.5876	0.6206
ResNet50	<b>64.52</b>	<b>0.6253</b>	<b>0.6486</b>

accuracy of each classification, randomly selected 100 samples from each classification for testing, the accuracy of normal constitution (0), Qi-Stagnation constitution (1), Qi-Deficiency constitution (2), Yang-Deficiency constitution (3), Phlegm-damp constitution (4), Damp-Heat constitution (5), Yin deficiency constitution (6), Body fluid deficiency constitution (7) and Blood stasis constitution (8) were 65%, 55%, 57%, 68%,63%,62%, 66%, 60%, 66%, respectively.

## Application of the model

We applied the trained model to WeChat public platform and iOS app store, and it also can get more information from our website [158]. Figure 7 shows the whole workflow of our app. Users can upload their tongue images to obtain their own TCM constitution classification information.

When a user uploads a non-tongue picture and gets some constitution classification information, the classification result is meaningless. Therefore, we added a tongue image judgment mechanism to determine whether the tongue pictures uploaded by the users are tongue images. If the user uploads a picture that is not a tongue, the system will display a prompt message to remind them to upload tongue images again. After judging that the picture uploaded contains the tongue part, the images will be transmitted to the model on the server for constitution classification. And then, the system will show the corresponding constitution characteristic information, disease tendency, and the way of recuperation according to the type of the user’s constitution. In addition, the function of history information management was added to the system, and users can view their previously uploaded tongue information and results, which helps users analyze their body changes. To enrich the system’s functions, we also added the traditional constitution examination questionnaire in the system.

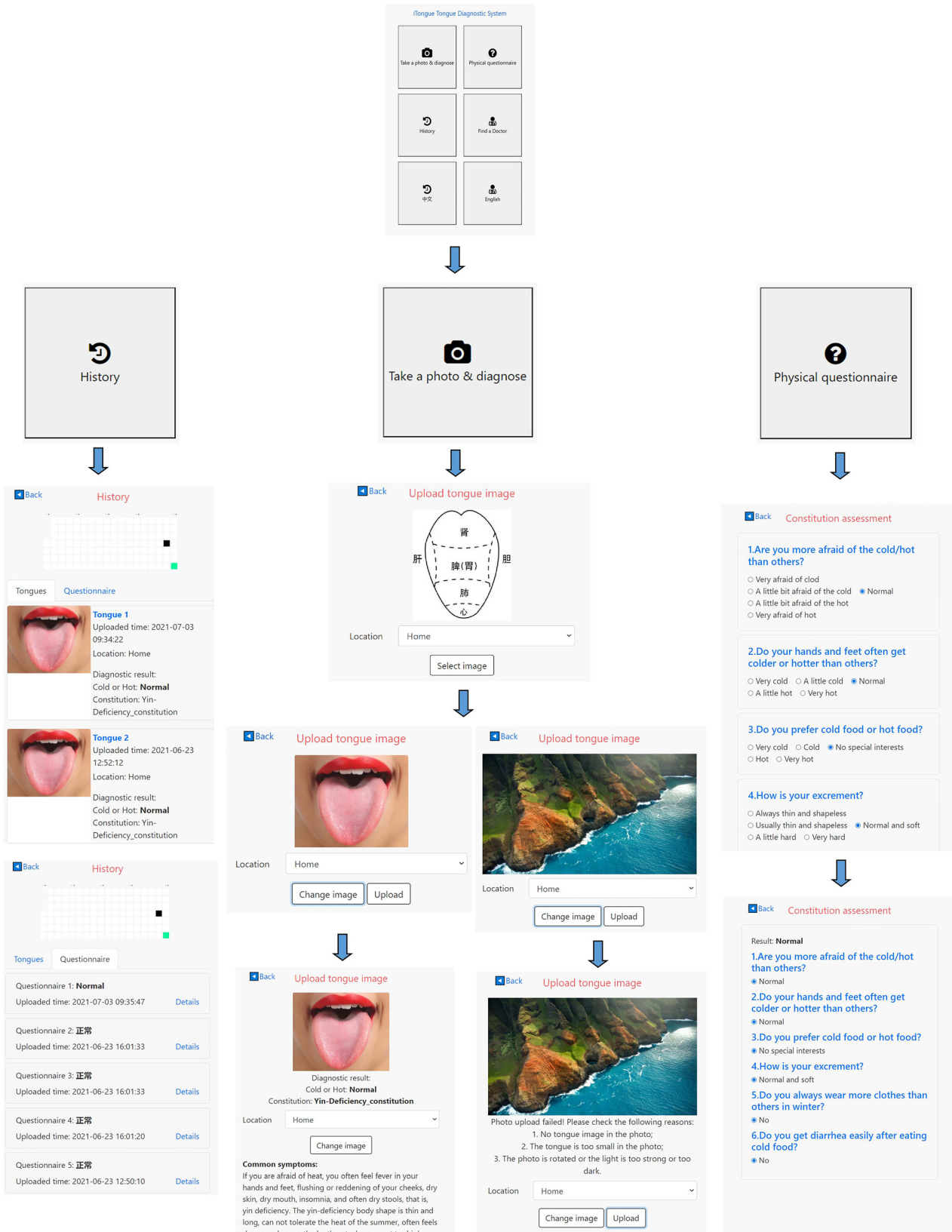


Figure 7: Operation flowchart of iTongue app.

## Discussions

In this section, we summarize the process of computerized tongue imaging, and provide discussions and outlook. With the development of digital imaging technology, tongue image acquisition hardware has been studied for several years. The framework of a general tongue image acquisition system generally includes the following four parts: imaging camera, light source, light path, and color correction tool. In this paper, we reviewed more than 10 tongue image acquisition systems. Although these systems can collect tongue images, there are still two problems that remain to be solved. First, the design of these tongue image acquisition systems does not have any general standards, which leads to significant differences in the quality of the tongue images obtained. Secondly, several systems [42, 61, 159] did not involve a color correction procedure aiming to correct color variations caused by system components. Thus their produced images may be unstable and unreliable. To obtain quantitative and objective diagnostic results, the tongues images must be reproducible under varying light environmental conditions. In terms of light source selection, some studies [42, 43, 159] have tried to use halogen tungsten lamps to supplement external light sources, but their color temperature is too low to render colors with high fidelity, which will make the acquired images reddish, and the acquired tongue photos and the actual application situation is different. Some studies [25, 28, 160] use image acquisition in a public office environment, which will cause the collected images to be affected by ambient light. In terms of camera selection, most systems [29, 38, 43] choose digital cameras. The advantage is that the pictures acquired by these cameras have high resolution, but when fixed parameters are set, the problem of inconsistent focus often occurs, and systems are not suitable for portable data collection occasions; there are also a few studies [39] using portable cameras such as Logitech HD Pro camera, but it is prone to problems of motion blur and inconsistent exposure. The hyperspectral camera [36] is composed of a spectrometer, a CCD detector, a lateral movement device, data acquisition and control module, etc., which can obtain the spectral characteristics of the tongue image, but does not perform color correction. The different types of cameras result in significant differences in the quality of the tongue images obtained, and these inconsistent tongue images are generally not compatible with each other. This has created obstacles to the process of computerized TCM tongue diagnosis. In the future, with the advancement of acquisition technology and color correction algorithms, the

performance of the tongue image acquisition system will be further improved.

Tongue image segmentation and effective part extraction are key tasks in the tongue image diagnosis process. The classic tongue segmentation methods can be roughly divided into four categories: thresholding tongue segmentation techniques [55, 161], edge detection tongue segmentation methods [57, 162], graph theory-based methods [163, 164], and active contour model based methods [54, 165]. These methods can produce tongue segmentation results to a certain extent, but there are still some challenges. For example, these methods are sensitive to the diversity of lighting and cluttered background and cannot accurately separate the tongue and the lips. In addition, these methods are generally slow in running time. Deep learning methods have made many breakthroughs in the field of tongue image segmentation. Some studies [69, 166, 167] try to apply deep CNN classification to tongue image segmentation, and the results are better than some traditional tongue segmentation methods. However, these methods generally require additional preprocessing, such as brightness discrimination and image enhancement. In addition, there are few tongue image datasets with segmentation labels, which makes the entire segmentation process complicated. In future work, real-time and accurate tongue image segmentation will still be a focus of research to address some challenges in tongue image segmentation. In particular, the tongue surface is not smooth with a complex structure for pixel-level labeling or segmentation. In some cases, the tongue cracks and boundaries have high similarity to tongue coating around the crack.

Tongue classification tasks can be divided into two main categories. The first is a classification based on the clinical characteristics of the tongue, such as the color of the tongue coating, the distribution of the tongue coating, tooth marks, tongue cracks, and tongue shape. The second is to classify diseases that can be reflected by the clinical attributes of the tongue, such as gastritis, appendicitis, and breast cancer. In the early research, machine learning methods such as support vector machines were applied to classify tongue images. Before training these classifiers, feature extraction is an important step. Many studies have tried to use different methods to extract features such as tongue color, texture, and shape to improve the accuracy of the classifier, and some studies have optimized the structure of the classifier, which greatly improves the accuracy of tongue image classification. With the introduction of deep learning, CNNs have significantly increased the number of parameters in the classifier and subsequently enhanced the accuracy of the classifier. After that, some

more complex networks such as VGG, Resnet, Inception were proposed, and the performance of tongue image classification was further improved. However, more research is needed to further improve the accuracy of tongue image classification. In the next few years, tongue classification also needs significant improvement for broad applications. Tongue classification and tongue diagnosis on mobile devices represent a general trend of TCM-assisted tongue diagnosis, but the data quality from mobile devices is limited. It requires bigger datasets and better machine learning models to address these issues.

The computerized tongue diagnosis system provides a certain level of convenience for both physicians and patients. Physicians can refer to the results of the tongue diagnostic system to aid in their diagnostic process; patients can upload their tongue pictures via their cell phones to obtain an assessment of their health status. Although many studies of computerized TCM diagnostic systems have been published, limited by the hardware technologies and computational methods, almost all of these computational diagnosis systems can only simulate one of the four diagnostic methods in TCM to assist in disease detection. According to the four diagnostic theories TCM, other data (e.g., pulse, etc.) of the patient are also necessary to make a more accurate assessment of health status. Combining tongue diagnosis with other data such as pulse diagnosis is an inevitable trend in computerized TCM diagnosis.

## Conclusion

In this review, we systematically summarized a variety of studies in computerized tongue diagnosis, including hardware for acquiring tongue images, the general process in computerized tongue diagnosis, methods for tongue color correction, tongue classification for disease prediction, and tongue diagnosis systems. Moreover, we present a system for the classification of TCM constitutions based on tongue images. In addition, we analyzed the current challenges of computerized TCM tongue diagnosis and provided an outlook on future research trends in TCM tongue diagnosis. We hope this review will provide a useful reference and guidance for researchers in computational tongue diagnosis, and other researchers can gain a quick overview of the field and borrow ideas for related studies.

**Acknowledgments:** We would like to thank Yu Xia for the technical assistance.

**Research funding:** This work is partially supported by the Paul K. and Diane Shumaker Endowment Fund to Dong Xu.

## References

1. Tang J-L, Liu B-Y, Ma K-W. Traditional Chinese medicine. The Lancet 2008;372:1938–40.
2. Zhang GG, Lee W, Bausell B, Lao L, Handwerger B, Berman B. Variability in the traditional Chinese medicine (TCM) diagnoses and herbal prescriptions provided by three TCM practitioners for 40 patients with rheumatoid arthritis. J Alternative Compl Med 2005;11:415–21.
3. Zhang D, Zhang H, Zhang B. Tongue image analysis. Singapore: Springer; 2017.
4. Kaptchuk TJ. Chinese medicine: the web that has no weaver. Toronto, Canada: Random House; 2000.
5. Jiang B, Liang X, Chen Y, Ma T, Liu L, Li J, et al. Integrating next-generation sequencing and traditional tongue diagnosis to determine tongue coating microbiome. Sci Rep 2012;2:1–15.
6. Ellis A, Wiseman N. Fundamentals of Chinese medicine. Brookline, MA: Paradigm Publications; 1995.
7. Kirschbaum B. Atlas of Chinese tongue diagnosis. Seattle, WA: Eastland Press; 2000.
8. Li C. Chinese medicine diagnostics, New Century 4th ed. Beijing: China Press of Traditional Chinese Medicine; 2016.
9. Hsu P-C, Wu H-K, Huang Y-C, Chang H-H, Lee T-C, Chen Y-P, et al. The tongue features associated with type 2 diabetes mellitus. Medicine 2019;98:e15567.
10. Han S, Yang X, Qi Q, Pan Y, Chen Y, Shen J, et al. Potential screening and early diagnosis method for cancer: tongue diagnosis. Int J Oncol 2016;48:2257–64.
11. Cao HYB. Distribution characteristics of TCM syndrome types in acute ischemic stroke and correlation with tongue image. Clin J Tradit Chinese Med 2021;33:1312–6.
12. Chen Y, Yuan H, Hui Y, Zhang XZ, Liu Y. Distribution of TCM syndrome types and tongue images in patients with failed Helicobacter pylori eradication based on propensity score matching. J Basic Chinese Med 2021;27:986–9.
13. Luo J, Zhang L, Chen J, Hu Q, Zhang Y, Tao Q. Tongue appearances of patients with primary Sjögren's syndrome and their correlations with syndromes. China J Tradit Chinese Med Pharm 2021;36:3653–6.
14. Chen H, Xu X, Zhou Y, Hu J. Discussion on the characteristics and significance of tongue manifestation in 8260 patients with primary insomnia. China J Tradit Chinese Med Pharm 2021;36:2971–3.
15. Anastasi JK, Chang M, Quinn J, Capili B. Tongue inspection in TCM: observations in a study sample of patients living with HIV. Med Acupunct 2014;26:15–22.
16. Li N, Zhang D, Wang K, Zhu Y. Tongue diagnostic. Beijing: Academy Press; 2011.
17. Zhang Q, Zhou J, Zhang B. Computational traditional Chinese medicine diagnosis: a literature survey. Comput Biol Med 2021;133:104358.
18. Gong Y, Chen H, Pu J, Lian Y, Chen S. Quantitative investigation on normal pathological tongue shape and correlation analysis between hypertension and syndrome. China J Tradit Chinese Med Pharm 2005;20:730–1.

19. Liu M-A, Xu J-P, Zhao Y, Liu Z. The clinical research of glossoscopy of acute cerebrovascular disease. *J Emerg Tradit Chin Med* 2008;11:38.
20. Pae E-K, Lowe AA. Tongue shape in obstructive sleep apnea patients. *Angle Orthod* 1999;69:147–50.
21. Lo L-C, Cheng T-L, Chiang JY, Damdinsuren N. Breast cancer index: a perspective on tongue diagnosis in traditional Chinese medicine. *J Tradit Complementary Med* 2013;3:194–203.
22. Lo L-C, Chen C, Chiang JY, Cheng T-L, Lin H-J, Chang H-H. Tongue diagnosis of traditional Chinese medicine for rheumatoid arthritis. *Afr J Tradit Complementary Altern Med* 2013;10:360–9.
23. Hsu Y-C, Chen Y-C, Lo L-C, Chiang JY, editors. *Automatic tongue feature extraction. 2010 international computer symposium (ICCS2010)*. IEEE; 2010.
24. Lo L-C, Hou MC-C, Chen Y-L, Chiang JY, Hsu C-C, editors. *Automatic tongue diagnosis system. 2009 2nd international conference on biomedical engineering and informatics*. IEEE; 2009.
25. Chiu C-C. A novel approach based on computerized image analysis for traditional Chinese medical diagnosis of the tongue. *Comput Methods Progr Biomed* 2000;61:77–89.
26. Yue X-Q, Liu Q. Analysis of studies on pattern recognition of tongue image in traditional Chinese medicine by computer technology. *Zhong xi yi jie he xue bao J Chinese Integrat Med* 2004;2:326–9.
27. Jung CJ, Jeon YJ, Kim JY, Kim KH. Review on the current trends in tongue diagnosis systems. *Integrat Med Res* 2012;1:13–20.
28. Cai Y, editor. *A novel imaging system for tongue inspection. IMTC/2002 proceedings of the 19th IEEE instrumentation and measurement technology conference (IEEE Cat No 00CH37276)*. IEEE; 2002.
29. Wang Y, Zhou Y, Yang J, Xu Q, editors. *An image analysis system for tongue diagnosis in traditional Chinese medicine. International conference on computational and information science*. Springer; 2004.
30. Jiang L, Xu W, Chen J, editors. *Digital imaging system for physiological analysis by tongue colour inspection. 2008 3rd IEEE conference on industrial electronics and applications*. IEEE; 2008.
31. Cibin N, Franklin SW, Nadu T. Diagnosis of diabetes mellitus and NPDR in diabetic patient from tongue images using LCA classifier. *Int J Adv Res Trends Eng Technol* 2015;2:57–62.
32. Zhi L, Zhang D, Yan J-Q, Li Q-L, Tang Q-L. Classification of hyperspectral medical tongue images for tongue diagnosis. *Comput Med Imag Graph* 2007;31:672–8.
33. Zuo W, Wang K, Zhang D, Zhang H, editors. *Combination of polar edge detection and active contour model for automated tongue segmentation. Third international conference on image and graphics (ICIG'04)*. IEEE; 2004.
34. Sun Y, Luo Y, Zhou C-L, Xu J-T, Zhang Z. A method based on split-combining algorithm for the segmentation of the image of tongue. *J Image Graph* 2003;8:1395–9.
35. Healey G, Slater D. Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions. *IEEE Trans Geosci Rem Sens* 1999;37:2706–17.
36. Yamamoto S, Tsumura N, Nakaguchi T, Namiki T, Kasahara Y, Terasawa K, et al. Regional image analysis of the tongue color spectrum. *Int J Comput Assist Radiol Surg* 2011;6:143–52.
37. Lo L-C, Chen Y-F, Chen W-J, Cheng T-L, Chiang JY. The study on the agreement between automatic tongue diagnosis system and traditional Chinese medicine practitioners. *Evid Based Compl Alternative Med* 2012;2012:1–9.
38. Lu Y, Li X, Zhuo L, Zhang J, Zhang H, editors. *Dccn: a deep-color correction network for traditional Chinese medicine tongue images. 2018 IEEE international conference on multimedia & expo workshops (ICMEW)*. IEEE; 2018.
39. Zhuo L, Zhang P, Qu P, Peng Y, Zhang J, Li X. A K-PLSR-based color correction method for TCM tongue images under different illumination conditions. *Neurocomputing* 2016;174: 815–21.
40. Qi Z, Tu L-P, Chen J-B, Hu X-J, Xu J-T, Zhang Z-F. The classification of tongue colors with standardized acquisition and ICC profile correction in traditional Chinese medicine. *BioMed Res Int* 2016; 2016:1–9.
41. Kim K, Do J, Ryu H, Kim J, editors. *Tongue diagnosis method for extraction of effective region and classification of tongue coating. 2008 First workshops on image processing theory, tools and applications*. IEEE; 2008.
42. Jang J, Kim J, Park K, Park S, Chang Y, Kim B, editors. *Development of the digital tongue inspection system with image analysis. Proceedings of the second joint 24th annual conference and the annual fall meeting of the biomedical engineering society engineering in medicine and biology*. IEEE; 2002.
43. Zhang H, Wang K, Zhang D, Pang B, Huang B, editors. *Computer aided tongue diagnosis system. 2005 IEEE engineering in medicine and biology 27th annual conference*. IEEE; 2006.
44. Al-Amri SS, Kalyankar NV. Image segmentation by using threshold techniques. *ArXiv preprint arXiv:10054020* 2010.
45. Tremeau A, Borel N. A region growing and merging algorithm to color segmentation. *Pattern Recogn* 1997;30:1191–203.
46. Al-Amri SS, Kalyankar N, Khamitkar S. Image segmentation by using edge detection. *Int J Comput Sci Eng* 2010;2:804–7.
47. Muthukrishnan R, Radha M. Edge detection techniques for image segmentation. *Int J Comput Sci Inf Technol* 2011;3:259.
48. Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. *Int J Comput Vis* 1988;1:321–31.
49. Pang B, Zhang D, Wang K. The bi-elliptical deformable contour and its application to automated tongue segmentation in Chinese medicine. *IEEE Trans Med Imag* 2005;24:946–56.
50. Pang B, Wang K, Zhang D, Zhang F, editors. *On automated tongue image segmentation in Chinese medicine. Object recognition supported by user interaction for service robots, Quebec City, QC, Canada*. IEEE; 2002, 1:616–9 pp.
51. Wu J, Zhang Y, Bai J, editors. *Tongue area extraction in tongue diagnosis of traditional Chinese medicine. 2005 IEEE engineering in medicine and biology 27th annual conference*. IEEE; 2006.
52. Yu S, Yang J, Wang Y, Zhang Y, editors. *Color active contour models based tongue segmentation in traditional Chinese medicine. 2007 1st international conference on bioinformatics and biomedical engineering*. IEEE; 2007.
53. Shi M, Li G-Z, Li F, Xu C, editors. *A novel tongue segmentation approach utilizing double geodesic flow. 2012 7th international conference on computer science & education (ICCSSE)*. IEEE; 2012.

54. Shi M, Li G, Li F. C 2 G 2 FSNAKE: automatic tongue image segmentation utilizing prior knowledge. *Sci China Inf Sci* 2013; 56:1–14.
55. Zhang L, Qin J. Tongue-image segmentation based on gray projection and threshold-adaptive method. *Chinese J Tissue Eng Res* 2010;14:1638.
56. Wang X, Zhang B, Yang Z, Wang H, Zhang D. Statistical analysis of tongue images for feature extraction and diagnostics. *IEEE Trans Image Process* 2013;22:5336–47.
57. Li Q, Xue Y, Wang J, Yue X. Automated tongue segmentation algorithm based on hyperspectral image. *J Infrared Millimeter Waves Chinese Ed* 2007;26:77.
58. Liang C, Shi D, editors. *A prior knowledge-based algorithm for tongue body segmentation. 2012 international conference on computer science and electronics engineering*. IEEE; 2012.
59. Wei C, Wang C, Huang S. *Using threshold method to separate the edge, coating and body of tongue in automatic tongue diagnosis. The 6th international conference on networked computing and advanced information management*. IEEE; 2010.
60. Wei B, Shen L, Wang Y, Wang Y, Wang A, Zhao Z. A digital tongue image analysis instrument for Traditional Chinese Medicine. *Zhongguo yi Liao qi xie za zhi Chinese J Med Instrum* 2002;26: 164–6, 9.
61. Yue H, Changjiang L, Lansun S. Digital camera based tongue manifestation acquisition platform, 5th ed. Beijing: World Science and Technology-Modernization of Traditional Chinese Medicine; 2007, 9:102–5 pp.
62. Zhang J, Hu G, Zhang X, editors. *Extraction of tongue feature related to TCM physique based on image processing. 2015 12th international computer conference on wavelet active media technology and information processing (ICCWAMTIP)*. IEEE; 2015.
63. Li W, Luo J, Hu S, Xu J, Zhang Z, editors. *Towards the objectification of tongue diagnosis: the degree of tooth-marked. 2008 IEEE international symposium on IT in medicine and education*. IEEE; 2008.
64. Li X, Yang D, Wang Y, Yang S, Qi L, Li F, et al., editors. *Automatic tongue image segmentation for real-time remote diagnosis. 2019 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE; 2019.
65. Li X, Yang T, Hu Y, Xu M, Zhang W, Li F, editors. *Automatic tongue image matting for remote medical diagnosis. 2017 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE; 2017.
66. Rother C, Kolmogorov V, Blake A. “GrabCut” interactive foreground extraction using iterated graph cuts. *ACM Trans Graph* 2004;23:309–14.
67. Levin A, Lischinski D, Weiss Y. A closed-form solution to natural image matting. *IEEE Trans Pattern Anal Mach Intell* 2007;30: 228–42.
68. Arun K. *Transactions on pattern analysis and machine intelligence*. IEEE 1987;PAMI-9:698–770.
69. Lin B, Xie J, Li C, Qu Y, editors. *Deeptongue: tongue segmentation via resnet. 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE; 2018.
70. Huang Y, Lai Z, Wang W, editors. *TU-Net: a precise network for tongue segmentation. Proceedings of the 2020 9th international conference on computing and pattern recognition*; 2020.
71. Zhou C, Fan H, Li Z. Tonguenet: accurate localization and segmentation for tongue images using deep neural networks. *IEEE Access* 2019;7:148779–89.
72. Zhou J, Zhang Q, Zhang B, Chen X. TongueNet: a precise and fast tongue segmentation system using U-Net with a morphological processing layer. *Appl Sci* 2019;9:3128.
73. Zhu J, Styler W, Calloway I. A CNN-based tool for automatic tongue contour tracking in ultrasound images. *ArXiv preprint arXiv:190710210* 2019.
74. Tang H, Wang B, Zhou J, Gao Y, editors. *DE-net: dilated encoder network for automated tongue segmentation. 2020 25th international conference on pattern recognition (ICPR)*. IEEE; 2021.
75. Huang X, Zhang H, Zhuo L, Li X, Zhang J. TISNet-enhanced fully convolutional network with encoder-decoder structure for tongue image segmentation in Traditional Chinese Medicine. *Comput Math Methods Med* 2020;2020:1–13.
76. Xu Q, Zeng Y, Tang W, Peng W, Xia T, Li Z, et al. Multi-task joint learning model for segmenting and classifying tongue images using a deep neural network. *IEEE J Biomed Health Informatics* 2020;24:2481–9.
77. Yuan W, Liu C, editors. *Cascaded CNN for real-time tongue segmentation based on key points localization. 2019 IEEE 4th international conference on big data analytics (ICBDA)*. IEEE; 2019.
78. Zhang H-K, Hu Y-Y, Wang L-J, Zhang W-Q, Li F-F. Computer identification and quantification of fissured tongue diagnosis. 2018 IEEE international conference on bioinformatics and biomedicine (BIBM).
79. Chen F, Xia C, Sui J, Wang Y, Peng Q. Extraction of tongue crack based on gray level and texture. *DEStech Trans Comput Sci Eng* 2018;11–21. <https://doi.org/10.12783/dtcse/csse2018/24477>.
80. Xue Y, Li X, Cui Q, Wang L, Wu P, editors. *Cracked tongue recognition based on deep features and multiple-instance SVM. Pacific Rim conference on multimedia*. Springer; 2018.
81. Chang W-H, Chu H-T, Chang H-H, editors. *Tongue fissure visualization with deep learning. 2018 conference on technologies and applications of artificial intelligence (TAAI)*. IEEE; 2018.
82. Chang W-H, Wu H-K, Lo L-C, Hsiao WW, Chu H-T, Chang H-H. Tongue fissure visualization by using deep learning – an example of the application of artificial intelligence in traditional medicine; 2020.
83. Peng J, Li X, Yang D, Zhang Y, Zhang W, Zhang Y, et al. Automatic tongue crack extraction for real-time diagnosis. 2020 IEEE international conference on bioinformatics and biomedicine (BIBM), Seoul, Korea (South). IEEE; 2020:694–9 pp.
84. Wang X, Zhang D. An optimized tongue image color correction scheme. *IEEE Trans Inf Technol Biomed* 2010;14:1355–64.
85. Wang X, Zhang D. A new tongue colorchecker design by space representation for precise correction. *IEEE J Biomed Health Informatics* 2013;17:381–91.
86. Zhang H-Z, Wang K-Q, Jin X-S, Zhang D, editors. *SVR based color calibration for tongue image. 2005 International conference on machine learning and cybernetics*. IEEE; 2005.
87. Xu X, Zhuo L, Zhang J, Shen L. Research on color constancy under open illumination conditions. *J Electron* 2009;26:681.
88. Sharma G, Bala R. *Digital color imaging handbook*. Boca Raton, Florida: CRC Press; 2017.



89. Kang HR. Color technology for electronic imaging devices. Bellingham, Washington: SPIE press; 1997.
90. Hong G, Luo MR, Rhodes PA. A study of digital camera colorimetric characterization based on polynomial modeling. *Color Res Appl* 2001;26:76–84.
91. Ilie A, Welch G, editors. *Ensuring color consistency across multiple cameras. Tenth IEEE International Conference on Computer Vision (ICCV'05)*. IEEE; 2005, vol 1.
92. McCamy CS, Marcus H, Davidson JG. A color-rendition chart. *J Appl Photogr Eng* 1976;2:95–9.
93. Hu M-C, Lan K-C, Fang W-C, Huang Y-C, Ho T-J, Lin C-P, et al. Automated tongue diagnosis on the smartphone and its applications. *Comput Methods Progr Biomed* 2019;174:51–64.
94. Hu M-C, Cheng M-H, Lan K-C. Color correction parameter estimation on the smartphone and its application to automatic tongue diagnosis. *J Med Syst* 2016;40:1–8.
95. Li CH, Yuen PC. Regularized color clustering in medical image database. *IEEE Trans Med Imag* 2000;19:1150–5.
96. Zhang B, Nie W, Zhao S. A novel Color Rendition Chart for digital tongue image calibration. *Color Res Appl* 2018;43:749–59.
97. Zhuo L, Zhang J, Dong P, Zhao Y, Peng B. An SA–GA–BP neural network-based color correction algorithm for TCM tongue images. *Neurocomputing* 2014;134:111–6.
98. Wei B. The research on color reproduction and texture morphological analysis of TCM tongue analysis. Beijing: Beijing University of Technology; 2004.
99. Rosipal R, Trejo LJ. Kernel partial least squares regression in reproducing kernel Hilbert space. *J Mach Learn Res* 2001;2:97–123.
100. Zhang J, Yang Y, Zhang J. A MEC-BP-Adaboost neural network-based color correction algorithm for color image acquisition equipments. *Optik* 2016;127:776–80.
101. Jiayun S, Chunming X, Ke Y, Zhang Y, Yiqin W, Haixia Y, et al., editors. *Tongue image color correction method based on root polynomial regression. 2019 IEEE 8th joint international information technology and artificial intelligence conference (ITAIIC)*. IEEE; 2019.
102. Zhu W, Zhou C, Xu D, Xu J, editors. *A multi-feature CBIR method using in the traditional Chinese medicine tongue diagnosis. 2006 first international symposium on pervasive computing and applications*. IEEE; 2006.
103. Huang B, Wu J, Zhang D, Li N. Tongue shape classification by geometric features. *Inf Sci* 2010;180:312–24.
104. Zhang H, Zhang B, editors. *Disease detection using tongue geometry features with sparse representation classifier. 2014 International conference on medical biometrics*. IEEE; 2014.
105. Zhang B, Zhang H. Significant geometry features in tongue image analysis. *Evid.-Based Complementary Altern. Med: eCam* 2015.
106. Wu K, Zhang D. Robust tongue segmentation by fusing region-based and edge-based approaches. *Expert Syst Appl* 2015;42:8027–38.
107. Wu J, Zhang B, Xu Y, Zhang D. Tongue image alignment via conformal mapping for disease detection. *IEEE Access* 2019;8:9796–808.
108. Gao Z, Po L, Jiang W, Zhao X, Dong H, editors. *A novel computerized method based on support vector machine for tongue diagnosis. 2007 third international IEEE conference on signal-image technologies and internet-based system*. IEEE; 2007.
109. Gao Z, Cui M, Lu G, editors. *A novel computerized system for tongue diagnosis. 2008 international seminar on future information technology and management engineering*. IEEE; 2008.
110. Kanawong R, Obafemi-Ajayi T, Ma T, Xu D, Li S, Duan Y. Automated tongue feature extraction for ZHENG classification in traditional Chinese medicine. *Evid Based Compl Alternative Med* 2012;2012:1–14.
111. Li S, Zhang Z, Wu L, Zhang X, Li Y, Wang Y. Understanding ZHENG in traditional Chinese medicine in the context of neuro-endocrine-immune network. *IET Syst Biol* 2007;1:51–60.
112. Zhang B, Wang X, You J, Zhang D. Tongue color analysis for medical application. *Evid Based Compl Alternative Med* 2013;2013:1–11.
113. Pang B, Zhang D, Wang K. Tongue image analysis for appendicitis diagnosis. *Inf Sci* 2005;175:160–76.
114. Zhang B, Kumar BV, Zhang D. Detecting diabetes mellitus and nonproliferative diabetic retinopathy using tongue color, texture, and geometry features. *IEEE Trans Biomed Eng* 2013;61:491–501.
115. Su W, Xu Z-Y, Wang Z-Q, Xu J-T. Objectified study on tongue images of patients with lung cancer of different syndromes. *Chin J Integr Med* 2011;17:272–6.
116. Ding J, Cao G, Meng D, editors. *Classification of tongue images based on doublet SVM. 2016 International symposium on system and software reliability (ISSSR)*. IEEE; 2016.
117. Dalal N, Triggs B, editors. *Histograms of oriented gradients for human detection. 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. IEEE; 2005.
118. Li X, Zhang Y, Cui Q, Yi X, Zhang Y. Tooth-marked tongue recognition using multiple instance learning and CNN features. *IEEE Trans Cybern* 2018;49:380–7.
119. Zhang X, Xu X, Cai Y, editors. *Tongue image classification based on the TSVM. 2009 2nd international Congress on image and signal processing*. IEEE; 2009.
120. Jiao Y, Zhang X, Zhuo L, Chen M, Wang K, editors. *Tongue image classification based on Universum SVM. 2010 3rd international conference on biomedical engineering and informatics*. IEEE; 2010.
121. Li X, Shao Q, Wang J, editors. *Classification of tongue coating using Gabor and Tamura features on unbalanced data set. 2013 IEEE international conference on bioinformatics and biomedicine*. IEEE; 2013.
122. Daugman JG. Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Trans Acoust Speech Signal Process* 1988;36:1169–79.
123. Manjunath BS, Ma W-Y. Texture features for browsing and retrieval of image data. *IEEE Trans Pattern Anal Mach Intell* 1996;18:837–42.
124. Huang W, Yan Z, Xu J, Zhang L, editors. *Analysis of the tongue Fur and tongue features by naive Bayesian classifier. 2010 International conference on computer application and system modeling (ICASM 2010)*. IEEE; 2010.
125. Xu J, Tu L, Ren H, Zhang Z, editors. *A diagnostic method based on tongue imaging morphology. 2008 2nd international conference on bioinformatics and biomedical engineering*. IEEE; 2008.

126. Tang Y, Sun Y, Chiang JY, Li X. Research on multiple-instance learning for tongue coating classification. *IEEE Access* 2021;9: 66361–70.
127. Hou J, Su H-Y, Yan B, Zheng H, Sun Z-L, Cai X-C, editors. *Classification of tongue color based on CNN. 2017 IEEE 2nd international conference on big data analysis (ICBDA)*. IEEE; 2017.
128. Huo C-M, Zheng H, Su H-Y, Sun Z-L, Cai Y-J, Xu Y-F, editors. *Tongue shape classification integrating image preprocessing and Convolution Neural Network. 2017 2nd Asia-pacific conference on intelligent robot systems (ACIRS)*. IEEE; 2017.
129. Fu S, Zheng H, Yang Z, Yan B, Su H, Liu Y, editors. *Computerized tongue coating nature diagnosis using convolutional neural network. 2017 IEEE 2nd international conference on big data analysis (ICBDA)*. IEEE; 2017.
130. Meng D, Cao G, Duan Y, Zhu M, Tu L, Xu D, et al. Tongue images classification based on constrained high dispersal network. *Evid Based Compl Alternative Med* 2017;2017:1–12.
131. Song C, Wang B, Xu J. *Classifying tongue images using deep transfer learning. 2020 5th International conference on computational Intelligence and Applications (ICCIA)*. IEEE; 2020.
132. Ma J, Wen G, Wang C, Jiang L. Complexity perception classification method for tongue constitution recognition. *Artif Intell Med* 2019;96:123–33.
133. Lo L-C, Cheng T-L, Chen Y-J, Natsagdorj S, Chiang JY. TCM tongue diagnosis index of early-stage breast cancer. *Compl Ther Med* 2015;23:705–13.
134. Kim J, Son J, Jang S, Nam D-H, Han G, Yeo I, et al. Availability of tongue diagnosis system for assessing tongue coating thickness in patients with functional dyspepsia. *Evid Based Compl Alternative Med* 2013;2013. <https://doi.org/10.1155/2013/348272>.
135. Wu T-C, Lu C-N, Hu W-L, Wu K-L, Chiang JY, Sheen J-M, et al. Tongue diagnosis indices for gastroesophageal reflux disease: a cross-sectional, case-controlled observational study. *Medicine* 2020;99. <https://doi.org/10.1097/MD.00000000000020471>.
136. Ryu I, Siio I, editors. *TongueDx: a tongue diagnosis for health care on smartphones. Proceedings of the 5th augmented human international conference*; 2014.
137. Hu M-C, Zheng G-Y, Chen Y-T, Lan K, editors. *Automatic tongue diagnosis using a smart phone. 2014 IEEE international conference on systems, man, and cybernetics*; 2014.
138. Wang J, Li Y, Ni C, Zhang H, Li L, Wang Q. Cognition research and constitutional classification in Chinese medicine. *Am J Chin Med* 2011;39:651–60.
139. Wang J, Wang Q, Li L, Li Y, Zhang H, Zheng L, et al. Phlegm-dampness constitution: genomics, susceptibility, adjustment and treatment with traditional Chinese medicine. *Am J Chin Med* 2013;41:253–62.
140. Liu L-C, Liu X-S, Wang T, Liu X, Yu H-L, Zou C, et al., editors. *The study of the constitution, mucosal inflammation, Chinese medicine syndrome types and clinical pathology in IgA nephropathy. 2014 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE; 2014.
141. Wang Q. *Constitutional doctrine of TCM*. Beijing: People's Medical Publishing House; 2005.
142. Qi W. Classification and diagnosis basis of nine basic constitutions in Chinese medicine. *J Beijing Univ Tradit Chinese Med* 2005;28:1.
143. Wang Q, Zhu Y. *Classification and determination of constitution in TCM*. Beijing: Zhongguo Zhongyiyao Chubanshe; 2009.
144. Zhang J, Hou S, Wang J, Li L, Li P, Han J, et al. Classification of traditional Chinese medicine constitution based on facial features in color images. *J Tradit Chinese Med Sci* 2016;3: 141–6.
145. Lin F, Xiahou J, Xu Z. TCM clinic records data mining approaches based on weighted-LDA and multi-relationship LDA model. *Multimed Tool Appl* 2016;75:14203–32.
146. Peng L, Lyu C, Qian L. Research of tongue features in Chinese medicine constitution classification and feasibility of differentiating tongue to test constitution. *J Tradit Chin Med* 2017;58:1002–4.
147. Wang X, Zhang D. A high quality color imaging system for computerized tongue image analysis. *Expert Syst Appl* 2013;40: 5854–66.
148. Liu J, Hu X, Tu L, Cui J, Li J, Bi Z, et al. Study on the syndrome characteristics and classification model of non-small cell lung cancer based on tongue and pulse data. *JMIR Med Inform* 2021. <https://doi.org/10.21203/rs.3.rs-355613/v1>.
149. Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. *ArXiv preprint arXiv: 171204621* 2017.
150. Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. Albuementations: fast and flexible image augmentations. *Information* 2020;11:125.
151. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12:2825–30.
152. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 2012;25:1097–105.
153. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al., editors. *Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition*; 2015.
154. He K, Zhang X, Ren S, Sun J, editors. *Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016.
155. Hu J, Shen L, Sun G, editors. *Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition*; 2018.
156. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst* 2019;32:8026–37.
157. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z, editors. *Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016.
158. Available from: <http://www.itongue.cn/>.
159. Wong W, Huang S. Studies on externalization of application of tongue inspection of TCM. *Eng Sci* 2001;3:78–82.
160. Jiang Y, Chen J, Zhang H. Computerized TCM tongue diagnosis system. *Chin J Integr Tradit West Med* 2000;20:66–8.
161. Li D, Wei Y. Tongue image segmentation method based on adaptive thresholds. *Comput Technol Dev* 2011;21:63–5.
162. Fu Z, Li X, Li F. Tongue image segmentation based on snake model and radial edge detection. *J Image Graph* 2009;14: 688–93.
163. Yu-ke W, Peng F, Gui Z. Application of improved GrabCut method in tongue diagnosis system. *Transd Microsyst Technol* 2014;33:157–60.

164. Chen S, Fu H, Wang Y. Application of improved graph theory image segmentation algorithm in tongue image segmentation. *Jisuanji Gongcheng yu Yingyong Comput Eng Appl* 2012;48:201–3.
165. Guo J, Yang Y, Wu Q, Su J, Ma F, editors. *Adaptive active contour model based automatic tongue image segmentation. 2016 9th International Congress on image and signal processing, BioMedical engineering and informatics (CISP-BMEI)*. IEEE; 2016.
166. Li J, Xu B, Ban X, Tai P, Ma B, editors. *A tongue image segmentation method based on enhanced HSV convolutional neural network. International conference on cooperative design, visualization and engineering*. Springer; 2017.
167. Qu P, Zhang H, Zhuo L, Zhang J, Chen G, editors. *Automatic tongue image segmentation for traditional Chinese medicine using deep neural network. International conference on intelligent computing*. Springer; 2017.