Contents lists available at ScienceDirect

Digital Chinese Medicine

journal homepage: http://www.keaipublishing.com/dcmed

Research status and prospect of tongue image diagnosis analysis based on machine learning

XU Jiatuo*, JIANG Tao, LIU Shi

School of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

ARTICLE INFO

Article history
Received 15 January 2024
Accepted 15 March 2024
Available online 25 March 2024

Keywords
Tongue image
Shape and texture feature
Deep learning
Intelligent diagnosis
Tongue diagnosis

ABSTRACT

Image-based intelligent diagnosis represents a trending research direction in the field of tongue diagnosis in traditional Chinese medicine (TCM). In recent years, machine learning techniques, including convolutional neural networks (CNNs) and Transformers, have been widely used in the analysis of medical images, such as computed tomography (CT) and nuclear magnetic resonance imaging (MRI). These techniques have significantly enhanced the efficiency and accuracy of decision-making in TCM practices. Advanced artificial intelligence (AI) technologies have also provided new opportunities for the research and development of medical equipment and TCM tongue diagnosis, resulting in improved standardization and intelligence of the tongue diagnostic procedures. Although traditional image analysis methods could transform tongue images into scientific and analyzable data, recognizing and analyzing images that capture complicated tongue features such as tooth-marked tongue, tongue spots and prickles, fissured tongue, variations in coating thickness, tongue texture (curdy and greasy), and tongue presence (peeled coating) continues posing significant challenges in contemporary tongue diagnosis research. Therefore, the employment of machine learning techniques in the analysis of tongue shape and texture features as well as their applications in TCM diagnosis is the focus of this study. In the study, both traditional and deep learning image analysis techniques were summarized and analyzed to figure out their value in predicting disease risks by observing the tongue shapes and textures, aiming to open a new chapter for the development and application of AI in TCM tongue diagnosis research. In short, the combination of TCM tongue diagnosis and AI technologies, will not only enhance the scientific basis of tongue diagnosis but also improve its clinical applicability, thereby advancing the modernization of TCM diagnostic and therapeutic practices.

1 Introduction

Tongue diagnosis is a distinctive method in traditional Chinese medicine (TCM) and is widely utilized in clinical practice due to its non-invasive, simple, and specific advantages. Tongue diagnosis was first recorded in *Inner Canon of Huangdi* (*Huang Di Nei Jing*,《黄帝内经》), and by the 13th century, specialized books on tongue diagnosis had been published, highlighting its long-standing significance in TCM. Tongue images offer physicians rich diagnostic information, playing a crucial role in indicating the progression and prognosis of diseases. Before

*Corresponding author: XU Jiatuo, E-mail: xjt@fudan.edu.cn.

Peer review under the responsibility of Hunan University of Chinese Medicine.

DOI: 10.1016/j.dcmed.2024.04.002
Citation: XU JT, JIANG T, LIU S. Research status and prospect of tongue image diagnosis analysis based on machine learning. Digital Chinese Medicine, 2024, 7(1): 3-12.

the advent of scientific tools, tongue diagnosis faced several issues such as over-subjectivity by physicians, a lack of unified standards, and challenges in quantification.

Since the 1990s, engineering and computer technologies have been employed to digitize and analyze tongue images [1-3], ensuring standardized image collection, correction of tongue colors, and quantitative analysis of tongue features. In recent years, intelligent tongue diagnosis primarily involves tongue image segmentation and feature classification. However, performing quantitative analysis on tongue images that display complicated tongue features, such as tooth-marked tongue, spots and prickles, fissured tongue, variations in puffiness and thinness, as well as diverse coatings like thin, thick, curdy, greasy, peeled, moist, dry, rough, and tender, remains a significant challenge that urgently needs to be addressed. Figure 1 listed common shape and texture features of the tongue, such as peeled coating, fissured tongue, tooth-marked tongue, spots and prickles, and curdy and greasy coating. With the advent of machine learning, particularly deep learning, new technological approaches have been developed for the intelligent analysis of tongue images. These advancements are expected to significantly enhance the efficiency and accuracy of analyzing tongue image features. This paper systematically reviews the current research progress in the analysis of tongue image features with machine learning techniques and their clinical applications, and suggested directions for further development in the field of intelligent tongue diagnosis.

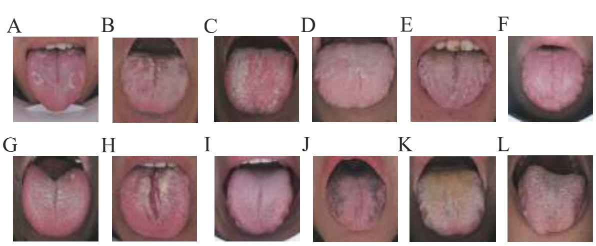

Figure 1 Complex shape and texture features of the tongue.
A – D, peeled coating. A, B, and E – J, fissured tongue. D – F, and I – L, tooth-marked tongue. I, K, and L, spots and prickles. A – D, E, G – I, K, and L, curdy and greasy coatings.

2 Early research on tongue shape and texture features

In the early stage of quantitative research on tongue shape features, researchers utilized algorithms such as color space transformation, light source variation, and grayscale co-occurrence matrices to extract features from tongue images, achieving promising results. ZHU et al. [4] employed the Douglas-Peucker method to extract numerical features of tooth marks on the tongue, with an accuracy rate of 80.00%. WANG et al. [5] segmented and registered tongue images collected under standard white and pure green light sources, successfully extracting prickles from green light tongue images with an accuracy rate of 88.47%. GUO [6] proposed an approach for analyzing the shape of the tongue body based on three-dimensional (3D) tongue images. Geomagic Stereo was used to fit and repair the 3D point cloud model of the tongue, enabling the calculation of the tongue body volume V. The ratio of the volume parameter V to the body surface area Mt was used to distinguish between enlarged and thin tongues. CAO et al. [2] employed the optimal linear fusion method and the AdaBoost algorithm to analyze tough and delicate tongues. It was discovered by comparison that the AdaBoost fusion method based on k-nearest neighbor classifier, which extracted fusion features including color, one-dimensional fractal dimension, and grayscale difference statistics, was more effective for identifying tongue toughness, achieving a recognition rate of 90.34%. To address issues in quantitative analysis, YANG et al. [7] proposed an approach for extracting cracks based on kernel false-color transformation. This approach involved using kernel false-color transformation to generate false-color images of tongue crack patterns and extracting cracks from the transformed gradient image using the lag threshold method. The correct detection ratio for extracting cracks from 200 colorful images using this method was 82.00%. YANG et al. [8] proposed a multi-index objective discrimination detection method for detecting tooth marks by combining the Graham scanning method and the Douglas-Peucker algorithm, which distinguished the presence, depth, and quantity of tooth marks on the tongues. The algorithm achieved an accuracy rate of 80.86% in distinguishing tooth marks and 80.00% in counting the number of tooth marks.

In the study of tongue coating texture, classical methods have successfully achieved preliminary quantification of tongue coating texture parameters through various algorithms, including Gabor wavelet and grayscale co-occurrence matrix. LIU et al. [9] employed an improved wavelet transform and grayscale co-occurrence matrix to extract texture features of the tongues, and then used the Relief feature selection approach to obtain the feature vector of tongue image textures, achieving a classification accuracy rate of 84.50%. Quantitative analysis of classified tongue images was achieved by calculating the mean square error and the coefficient of determination with the use of manual annotations as a reference. QU et al. [10] preprocessed tongue images and employed the Gabor wavelet transform to extract roughness features, grayscale co-occurrence matrix features, and Gabor wavelet transform features of the tongue coating. Although the recognition accuracy rate of the library for support vector machine (LIBSVM) classification for identifying rotten coating exceeded 85.00%, using roughness features to identify greasy coating did not yield satisfactory results. XIE [11] developed a dichotomous reflectance model and brightness gradient to differentiate between bright spot areas formed by mucus and by normal water films. This approach involved calculating the ratio of the bright spot area to the overall tongue coating area and the average brightness of the bright spot area to the maximum brightness of the tongue image. It successfully obtained a moistening coefficient, enabling the recognition and quantification of tongue image moistening. Currently, quantitative research on shape and texture of tongue images remains superficial, and the recognition accuracy rate is relatively low. This is particularly true for quantitative analysis concerning the calculation of the number of tooth marks, the size of peeled coating area, and the severity of cracks. Hence, in-depth investigation in the field seems necessary.

3 Research on machine learning for the extraction of shape and texture from tongue images

The convolutional neural networks (CNNs), a representative of deep learning technology, have applied extensively in medical research. Notably, *Nature* published an article in 2017 documenting the successful application of deep CNNs for the intelligent and accurate diagnosis of skin cancer [12]. In 2020, American researchers introduced an artificial intelligence (AI) system for the screening of breast cancer [13], outperforming six radiologists in image interpretation. American researchers have summarized studies on integrating imaging omics and AI for predicting cancer outcomes [14]. At present, the main areas of research in deep learning for tongue diagnosis include tongue image segmentation, tongue color classification, and tongue image texture recognition. The application of deep learning methods, has greatly improved the efficiency and accuracy of tongue image segmentation and analysis of tongue texture features. Regardless, how to effectively apply deep learning technologies to the study of TCM tongue diagnosis remains a central focus in the contemporary research and development of diagnostic methods in TCM.

3.1 Application of machine learning for tongue image segmentation

The advent of machine learning has provided an efficient solution to the challenges of tongue image segmentation, overcoming previous issues such as susceptibility to interference and low robustness. This represents an improvement in the efficiency and accuracy of tongue diagnosis, marking a notable advancement in the field. LIN et al. [15] developed an end-to-end deep learning model, Deep Tongue, based on the residual network (ResNet). The Mean Intersection over Union (MIoU) of the model was over 94.00%, significantly enhancing both the segmentation accuracy and computational speed in comparison to previous models. ZHOU et al. [16] proposed a semi-supervised CNN segmentation network model named “Semi Tongue”, which demonstrated higher segmentation accuracy and stronger robustness in comparison to network models such as U-Net [17], Deep Tongue, and Tongue-Net [18]. SONG et al. [19] developed an improved residual-based tongue image segmentation algorithm, residual attention feature fusion (RAFF)-Net, which achieved a MIoU of 97.85% and an F1 score of 97.73%. TANG et al. [20] constructed an improved tongue image segmentation model using DeepLabV3+ as the backbone network, enhanced with residual modules, the Vision Transformer, and the efficient channel attention (ECA) attention mechanism. The MIoU of this model achieved an accuracy rate of 95.00% with the self-built dataset and an accuracy rate of 94.10% with the public dataset BioHit. Currently, various deep learning algorithms have been employed for tongue image segmentation, but their performance is still influenced by factors such as the availability of training samples and preprocessing approaches. In the future, the segmentation results can be further enhanced by improving the network structure.

3.2 Application of deep learning for analyzing the shape and texture from tongue images

At present, deep learning is extensively applied in the field of tongue image recognition. In the tongue shape recognition research, the adoption of mainstream models such as Faster Regions with CNN (R-CNN), visual geometry group (VGG), and ResNet has markedly enhanced the accuracy and efficiency of recognizing tongue textures such as fissures, tooth marks, and pricks in comparison to traditional methods. It is noteworthy that deep learning algorithm models are evolving rapidly, transitioning from CNNs to Transformers, and now into the era of large models. Researchers have introduced iterative algorithm network models into the quality analysis of tongue images, continually improving accuracy with each advancement. Figure 2 illustrates the model architecture of shape and texture of tongue [21–23].

3.2.1 Analysis on deep learning techniques for tongue shape (i) Fissured tongue. Deep learning techniques are extensively utilized in crack extraction. TANG et al. [24] constructed a CNN for the recognition of tongue image features from a multi-label and multi-task perspective, achieving simultaneous analysis of tongue color, coating color, cracks, and tooth marks with multiple labels. The classification speed achieved by the CNN was 0.034 s per

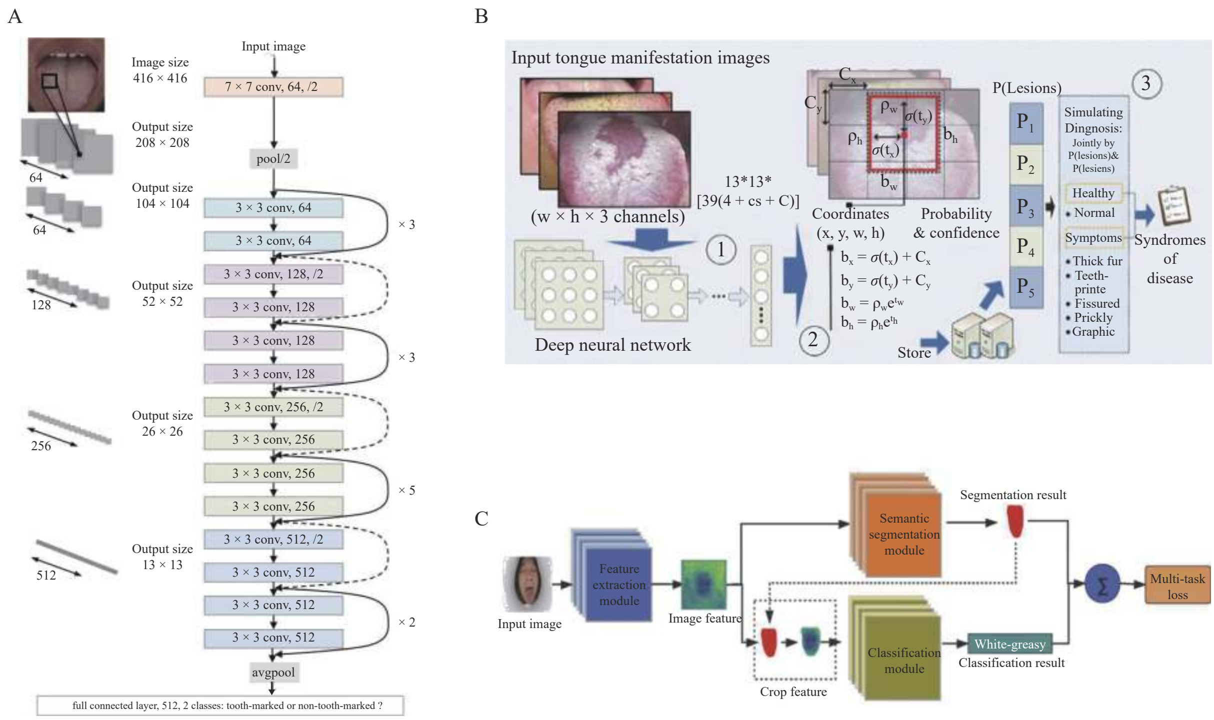

Figure 2 Models for the recognition of tongue shapes and textures based on deep learning techniques

A, visualization of the ResNet34 model structure for identifying tooth marks [21]. B, two-stage deep transfer learning model for multi-labelled shape and texture in tongue images [22]. C, framework of multi-task joint learning model for tongue segmentation and classification of greasy coating [23].

image, greatly shortening the time required to classify tongue image features in comparison to other techniques. LIU et al. [25] utilized the deep learning Faster R-CNN algorithm in combination with a fine-tuning technique and transfer learning to construct a model capable of recognizing cracks and tooth marks in tongue images. The macroscopic accuracies of the model were 0.910 and 0.977, and the recall rate was 0.860. SUN et al. [26] significantly enhanced the average pixel accuracy of images by combining an improved U-Net network structure with a focal loss function. Using the Inception V4+ filter response normalization (FRN) network structure to classify images of cracked tongues, the accuracy rate realized was 87.50%. YAN et al. [27] proposed a deep learning-based image segmentation network, segmentation-based deep learning (SBDL), which was composed of two components: a segmentation network and a decision network. In experiments on fissure extraction and classification, the accuracy of the SBDL network was 95.20%, and its performance on MIoU and frames per second (FPS) was also better than network models such as Mask R-CNN, DeepLabV3+, and U-Net.

(ii) Tooth-marked tongues. In recent years, the application of deep learning has significantly advanced research in the analysis of tooth marks in tongue diagnosis. Early experiments by LI et al. [28] utilized a combination of multi-instance support vector machine (SVM) and VGG 16 to identify tooth marks, with an average accuracy of 72.70%. Although the efficiency of this approach was greatly improved compared to classical methods, its accuracy still required further improvement. WANG et al. [21] utilized the highly robust ResNet structure to recognize tooth-marked tongues, with ResNet34 achieving an overall accuracy rate of 90.50% for recognizing tooth-marked tongues. Building on ResNet34, ZHOU et al. [29] proposed a weakly supervised network for detecting tooth marks that incorporated spatial pyramid pooling (SPP) to enable end-to-end training of the system. The SPP model proved to be a satisfactory solution for selecting candidate regions, with an accuracy rate of 91.97%. In addition, YAN et al. [30] employed the DeepLabV3+ deep learning framework in conjunction with the convex hull algorithm to identify tooth marks in tongue images. Afterward, a random forest algorithm was applied to further refine the classification model for tooth-marked tongues, achieving an overall accuracy rate of 93.00%. In the field of TCM tongue image diagnosis, the introduction of machine learning and the widespread application of deep learning have significantly enhanced the accuracy and efficiency of analyzing tongue images and tooth marks.

(iii) Tongue spots and prickles. Traditional approaches typically analyze tongue spots and prickles by examining features within the grayscale and color space. In recent years, researchers have introduced deep learning techniques that utilize color space, significantly enhancing the efficiency of spot and prickle recognition. PENG et al. [31] utilized a deep CNN to classify tongue spots and prickles, constructing a network model that included multi-scale feature map generation, candidate region search, and target region recognition. By deploying a classification network to detect and measure the pixels of the target area, segmentation of tongue spots and prickles was achieved, with a recognition accuracy rate of 84.30% for tongue spots. YANG et al. [32] implemented fully connected neural networks combined with transfer learning to extract features of tongue images such as spots and prickles, resulting in the development of models encompassing Inception V3 + two-layer neural network (2NN) and Inception V3 + three-layer neural network (3NN). These models achieved accuracies rates of 90.30% and 93.98%, respectively. WANG et al. [33] incorporated the deep learning-based Swin Transformer into the analysis of spots and prickles on the tongue, leveraging the Transformer's attention mechanism to effectively perform segmentation of the target area. Subsequently, a manually labeled dataset was employed to train a Faster R-CNN model, culminating in a final model accuracy of 88.46%.

(iv) Enlarged or thin tongue. Currently, researchers have introduced deep learning techniques into the analysis of enlarged and thin tongues; however, the research in this area remains somewhat superficial and requires further exploration and development. WEN et al. [34] applied deep learning techniques to analyze the features of enlarged and thin tongues, integrating a K-Block module based on DenseNet with the gradient descent algorithm to develop a classification model. The team successfully utilized their model to achieve precise classification among normal tongue body, fat tongue body, enlarged tongue body with tooth marks, and thin tongue body. ZHANG et al. [35] employed the Mask R-CNN model to detect and segment the tongue body in tongue images. They enhanced the model's capabilities by using the Laplace operator to extract the contour of the tongue body, a color distance algorithm to discern color features, and the Tamura algorithm to extract texture features. By measuring feature parameters, the enlarged and thin tongues were successfully distinguished.

3.2.2 Analysis on deep learning techniques for tongue coating (i) Curdy or greasy coating. Previous researchers have analyzed curdy and greasy tongue coatings using approaches such as grayscale co-occurrence matrix, Tamura texture roughness, and wavelet analysis, which characterize the coarse and greasy particles of curdy coatings as well as the fine and dense particles of greasy coatings. More recently, deep learning techniques have also been introduced to improve the analysis of tongue coatings. LI et al. [36] constructed a ShuffleNetV2 network and developed a fuzzy classification model for identifying greasy features in tongue images. This was achieved by training the network with preprocessed, high-confidence samples. WANG et al. [37] utilized ResNet18, ResNet34, and ResNet50 to build network models for recognizing greasy coatings on tongues. Additionally, they employed Grad-CAM to visualize the specific locations of target areas in tongue images, with an accuracy realized by ResNet34 of 0.880 and a precision of 0.883. Compared with classical methods, tongue image analysis based on deep learning has markedly enhanced the efficiency of analyzing greasy tongue coatings. ZHANG et al. [22] proposed a two-stage transfer learning strategy capable of achieving multi-objective detection of tongue features such as thickness, cracks, tooth-marks, peeling, and prickles. This strategy resulted in average detection accuracies rates of 91.00%, 92.00%, 88.00%, 84.00%, and 76.00% for each feature, respectively.

(ii) Thick or thin coating. The thickness of the tongue coating reflects the progression of diseases and the presence of pathogens. Classical techniques employ algorithms such as color space analysis, wavelet transforms, grayscale co-occurrence matrices, and support vector machines to analyze the thickness of tongue coatings. In recent years, the efficiency of CNNs in tongue image analysis has significantly improved. XU et al. [23] designed a multi-task model specifically to analyze both the color and thickness of tongue coatings in tongue images.

(iii) Peeled coating. In clinical practice, peeled coatings are commonly referred to as geographic tongues. At present, research on analyzing and recognizing peeled coatings remains in its early stages, although there have been initial attempts to apply deep learning techniques to this field. LI et al. [38] constructed a CNN model capable of recognizing multiple tongue features, with ResNet34 as the backbone for the feature classification network and ImageNet for the training network. The model achieved a remarkable classification accuracy rate of 98.90% in identifying peeled coatings.

In summary, deep learning has been extensively applied in the analysis of textures in tongue images. Compared with classical methods, it has significantly enhanced both the accuracy and efficiency of tongue image analysis. However, research on the analysis of certain tongue features, such as tender and geographic tongues, remains relatively underexplored. In addition, emphasizing local texture features is important, for it is crucial not to overlook global information. In the future, researchers should consider exploring more machine learning methods, enhancing the quality control of tongue image databases [44, 45], and further exploring and applying the intelligent analysis systems for tongue diagnosis using explainable AI technology [46-48]. Figure 3 and Table 1 list the target detection and visualization results of tongue shape and texture features.

Table 1 Comparison of accuracies between classical image analysis methods and deep learning methods for tongue image recognition

Tongue feature	Method		Result
Curdy and greasy coating	Classical	Rosenfeld and tamura texture roughness algorithms [39]	Accuracy: 83%
	Deep learning model	ResNet34 [37]	Accuracy: 87.8%; F1 score: 87.8%
		U-Net & ResNet50 [23]	F1 score: 93.02%
Thin and thick coating	Classical	Improved analysis method for wavelet transform and grayscale co-occurrence matrix [40]	Accuracy: 81.29% (max)
	Deep learning model	Inception V3 + 3NN [32]	Accuracy: 93.98%
Tough and delicate tongue	Classical	Classifier fusion technology [2]	Recognition rate of fused features: 90.34%
	Deep learning model	ResNet50 [41]	Accuracy: 89.3%
		ResNet101 [41]	Accuracy: 91%
Enlarged and thin tongue	Classical	BP neural network [42]	Result error rate: 0.018
	Deep learning model	K-CNN [34]	Classification of fat and thin tongue body: 32.45%
Tongue spots and prickles	Classical	RGB range and grayscale mean [43]	Accuracy: 77.10%
		Auxiliary light source [5]	Accuracy: 88.47%
	Deep learning model	Multiscale CNN [31]	Accuracy: 84.30%
		Inception V3 + 3NN [32]	Accuracy: 93.98%
		Swin transformer [24]	Accuracy: 99.79%
Tooth-marked tongue	Classical	Douglas-Puck method [4]	Accuracy: 80%
		Morphological feature extraction [8]	Accuracy: 80%
	Deep learning model	Multi-task convolutional neural network [24]	Accuracy (image-based): 98.94%; accuracy (category-based): 98.33%
		Inception V3 + 3NN [32]	Accuracy: 93.98%
		Faster R-CNN & fine-tune [25]	Accuracy: 86%
		VGG16 [28]	Accuracy: 89.41% (original tongue image); accuracy: 90.96% (tongue region)
		ResNet34 [21]	Accuracy: 90.50% (original tongue image); accuracy: 91.47% (tongue region)
Fissured tongue	Classical	Nuclear pseudocolor transformation [7]	Accuracy: 82%
	Deep learning model	Faster R-CNN & fine-tune [25]	Accuracy: 96%
		Multi-task convolutional neural network [24]	Accuracy (image-based): 98.94%; accuracy (category-based): 98.33%
		Inception V3 + 3NN [32]	Accuracy: 93.98%

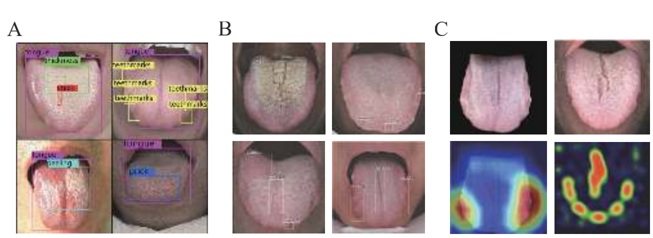

Figure 3 Detection results of tongue shape and texture features

A, detection results of tongue shape and texture features via a two-stage transfer learning strategy [22]. B, multi-labelled detection results based on deep learning techniques [49]. C, visualization of tooth marks and fissures on the tongues [21].

4 Clinical diagnosis based on tongue shape and texture features

The texture features of tongue images provide valuable pathological and physiological insights, playing a crucial role in strengthening the diagnosis and treatment processes with significant clinical implications. In the field of intelligent TCM diagnosis and treatment research, incorporating of tongue texture features as key parameters in disease diagnosis, alongside laboratory physical and chemical indicators, has significantly enhanced the accuracy of disease diagnoses.

4.1 Diagnosis of diabetes based on tongue shape and texture features

LI et al. [50] utilized the TDAS tongue image analysis system developed by the Intelligent Diagnostic Performance Laboratory of Shanghai University of Traditinal Chinese Medicine to extract the texture features from tongue images. Key parameters analyzed included contrast (CON), angular second moment (ASM), entropy (ENT) and mean value (MEAN), offering quantitative assessments of the features. Subsequently, they employed a Vision Transformer Network to build a diagnostic model that integrated deep learning techniques for a comprehensive analysis of tongue image features. By conducting a thorough analysis of tongue texture, color, and additional features from the tongue image, the model successfully classified diabetes-associated tongue features with precision. In addition, LI et al. [51] also built a prediction model for diabetes risk employing the ResNet50 backbone network framework, enabling the prediction of diabetes in both late and early stages.

4.2 Diagnosis of metabolic diseases based on tongue shape and texture features

In studies on metabolic diseases, the integration of features from tongue images has markedly strengthened diagnostic accuracy [52, 53] and proven to be valuable in medical checkups and disease screening [49]. YAO et al. [54] associated tongue image samples with body mass index (BMI) indices, extracted features from these tongue images, and obtained feature vectors to determine the color of tongue texture and coating. Subsequently, these feature vectors were employed for training a prediction model, which was utilized in conjunction with BMI indices for predicting the risks of fatty liver. DAI et al. [55] employed the generative adversarial network Tongue-generative adversarial network (GAN) to extract features from tongue images. They constructed a machine learning-based diagnostic model for metabolic associated fatty liver disease (MAFLD) diagnosis by integrating microbial characteristics of the tongue coating, and selected features from the tongue images, BMI, and liver function indicators such as alanine aminotransferase (ALT) and aspartate aminotransferase (AST). The model, which employed the XGBoost, achieved an impressive accuracy rate of 96.39%.

4.3 Diagnosis of cancer and other diseases based on tongue shape and texture features

WANG et al. [37] constructed a greasy coating recognition model using ResNet34 and explored the application of transfer learning algorithms to diagnose COVID-19. LIN et al. [56] found that the pathological changes reflected by the tongue of patients with idiopathic membranous nephropathy, along with their laboratory test results, might explain changes in clinical indicators. HAN et al. [57] explored the factors influencing changes in tongue coating thickness from the perspective of tongue coating microorganisms, and discussed the potential of using tongue coating and its microorganisms as biomarkers for the early diagnosis of cancer. Meanwhile, researchers constructed a deep learning diagnostic model for gastric cancer using tongue image features, confirming that tongue images can serve as reliable approaches for diagn osing gastric cancer [58].

5 Opportunities and challenges

Image-based intelligent diagnosis represents a crucial advancement in the modernization of TCM tongue diagnosis. Compared with classical image analysis methods, deep learning techniques have greatly improved the accuracy and efficiency of diagnosis based on tongue image features, and significantly facilitated the intelligent analysis process of tongue images. However, the following unaddressed issues remain: (i) there is currently no established mechanism for sharing and circulating the extensive amount of tongue diagnosis image data collected by various devices, and there is a need to create a large-scale and standardized tongue diagnosis dataset with classified and calibrated image samples; (ii) there is also a need to improve the accuracy of comprehensive intelligent analysis methods for tongue image features and enhance the generalization ability and interpretability of these methods; (iii) additionally, the complex mapping relationships between the quantitative features of tongue images, disease risks, and syndrome diagnosis need to be elucidated. Future research can target the following areas.

5.1 Establishment of a shared classified and calibrated tongue diagnosis dataset

Currently, multiple centers are collaborating to create an extensive database of tongue images [58]. Establishing a sample library for major diseases encompassing classified and calibrated datasets based on industry consensus is important components for advancing AI in tongue diagnosis. The issue of inconsistent charge-coupled-device (CCD) imaging equipment and light sources poses significant challenges, greatly restricting the generalization of AI for tongue diagnosis. Therefore, there is an urgent need to develop intelligent tongue diagnosis technologies that are adaptable to various devices and situations. A public tongue image dataset for intelligent diagnosis is beneficial for further improving the robustness and generalization of diagnostic models [20]. Research on self-supervised representation learning methods for developing a color difference correction model could standardize the intelligent analysis of tongue diagnosis images collected across various devices, lighting conditions, and environments.

5.2 Introduction to new deep learning methods for improving analysis accuracy and interpretability

Multi-label classification is crucial for analyzing the features of tongue images. By ensuring the quality of tongue image assessments, explainable AI technology has significantly enhanced both the accuracy and visualization of tongue image analysis [59]. Gradient-weighted Class Activation Mapping (Grad-CAM) [48] has been applied in tongue diagnosis research. Transformer can perform global interaction on image features. Combining Transformers and CNNs to preserve local information and the engagement in global interaction is an effective approach to optimize the existed models for image feature extraction.

5.3 Clinical value of tongue diagnosis based on human phenotypic omics data

Phenotype refers to the diverse characteristics of living organisms. Data from tongue diagnosis fall into the category of macroscopic phenotypic data. Based on the establishment of large-scale and high-quality cohorts of major diseases, the future deep learning models for tongue diagnosis can combine multi-modal and cross-scale human phenotypic omics data to reveal the dynamic evolutionary patterns of the onset of diseases and the progression of TCM patterns across multiple levels and dimensions [60].

TCM tongue diagnosis has evolved from a simple approach to a sophisticated, data-driven, and intelligent approach. The advent of deep learning has elevated tongue image feature analysis to an unprecedented level of sophistication and accuracy. Tongue shape and texture features from tongue images hold substantial physiological and pathological significance. Utilizing deep learning as a tool can significantly enhance the accuracy and efficiency of tongue diagnosis, thereby advancing the development of intelligent diagnostic systems for tongue image interpretation.

Fundings

National Natural Science Foundation of China (82104738), Capacity Building of Local Colleges and Universities under the Shanghai Municipal Science and Technology Commission (21010504400), and Space Station Engineering Aerospace Medical Experiment Project (HYZHXM05001).

Competing interests

The authors declare no conflict of interest.

References

[1] XU JT, SUN Y, ZHANG ZF, et al. Analysis and discrimination of tongue texture characteristics by difference statistics. Academic Journal of Shanghai University of Traditional Chinese Medicine, 2003, 17(3): 55–58.

[2] CAO ML, ZHANG XF, SHEN LS. Application survey of information combination in the toughness and tenderness of tongue manifestation recognition. Beijing Biomedical Engineering, 2006, 25(6): 644–648.

[3] SHI Z, ZHOU CL. Fissure extraction and analysis of image of tongue. Computer Technology and Development, 2007, 17(5): 245–248, 253.

[4] ZHU MLM, LU P, XIA CM, et al. Research on douglas-peucker method in feature extraction from 55 cases of tooth-marked tongue images. Chinese Archives of Traditional Chinese Medicine, 2014, 32(9): 2138–2140.

[5] WANG XM, WANG RY, GUO D, et al. A research about tongue-prickled recognition method based on auxiliary light source. Chinese Journal of Sensors and Actuators, 2016, 29(10): 1553–1559.

[6] GUO D. The research of tongue objectiveness based on binocular stereo vision. Tianjin: Tianjin University, 2018.

[7] YANG ZH, ZHANG D, LI NM. Kernel false-colour transformation and line extraction for fissured tongue image. Journal of Computer-Aided Design & Computer Graphics, 2010, 22(5): 771–776.

[8] YANG JX, HAN D, DONG XM, et al. Objectification of tooth-marked tongue in Chinese medicine based on morphological feature extraction. Laser & Optoelectronics Progress, 2022, 59(11): 365–373.

[9] LIU B, HU GQ, ZHANG XF, et al. An improved automatic description method of tongue coating thickness in Chinese medicine. Beijing Biomedical Engineering, 2018, 37(2): 157–163.

[10] QU TT, XIA CM, WANG YQ, et al. Recognition of greasy or curdy tongue coating based on gabor wavelet transformation. Computer Applications and Software, 2016, 33(10): 162–166.

[11] XIE T. A new approach to the tongue-image segmentation and moistening analysis based on image processing. Shanghai: East China University of Science and Technology, 2017.

[12] ESTEVA A, KUPREL B, NOVOA RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 2017, 542(7639): 115–118.

[13] MCKINNEY SM, SIENIEK M, GODBOLE V, et al. International evaluation of an AI system for breast cancer screening. Nature, 2020, 577(7788): 89–94.

[14] BERA K, BRAMAN N, GUPTA A, et al. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. Nature Reviews Clinical Oncology, 2022, 19(2): 132–146.

[15] LIN BQ, XIE JW, LI CH, et al. Deeptongue: tongue segmentation via resnet. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018: 1035–1039.

[16] ZHOU CG, FAN HY, ZHAO W, et al. Reconstruction enhanced probabilistic model for semisupervised tongue image segmentation. Concurrency and Computation: Practice and Experience, 2020, 32(22). doi: 10.1002/cpe.5844.

[17] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234–241.
[18] ZHOU CG, FAN HY, LI ZY. Tonguenet: accurate localization and segmentation for tongue images using deep neural networks. IEEE Access, 2019, 7: 148779–148789.
[19] SONG HB, HUANG ZH, FENG L, et al. RAFF-Net: an improved tongue segmentation algorithm based on residual attention network and multiscale feature fusion. Digital Health, 2022, 8: 20552076221136362.
[20] TANG Y, TAN DQ, LI HX, et al. RTC_TongueNet: an improved tongue image segmentation model based on DeepLabV3. Digital Health, 2024, 10: 20552076241242773.
[21] WANG X, LIU JW, WU CY, et al. Artificial intelligence in tongue diagnosis: using deep convolutional neural network for recognizing unhealthy tongue with tooth-mark. Computational and Structural Biotechnology Journal, 2020, 18: 973–980.
[22] ZHANG X, CHEN ZK, GAO J, et al. A two-stage deep transfer learning model and its application for medical image processing in Traditional Chinese Medicine. Knowledge-Based Systems, 2022, 239: 108060.
[23] XU Q, ZENG Y, TANG WJ, et al. Multi-task joint learning model for segmenting and classifying tongue images using a deep neural network. IEEE Journal of Biomedical and Health Informatics, 2020, 24(9): 2481–2489.
[24] TANG YP, WANG LR, HE X, et al. Classification of tongue image based on multi-task deep convolutional neural network. Computer Science, 2018, 45(12): 255–261.
[25] LIU M, WANG XT, ZHOU L, et al. Study on extraction and recognition of traditional chinese medicine tongue manifestation: based on deep learning and migration learning. Journal of Traditional Chinese Medicine, 2019, 60(10): 835–840.
[26] SUN CH. Research on tongue cleft extraction technology based on deep learning. Beijing: North China University of Technology, 2023.
[27] YAN JJ, CAI JX, XU Z, et al. Tongue crack recognition using segmentation based deep learning. Scientific Reports, 2023, 13(1): 511.
[28] LI XQ, ZHANG Y, CUI Q, et al. Tooth-marked tongue recognition using multiple instance learning and CNN features. IEEE Transactions on Cybernetics, 2019, 49(2): 380–387.
[29] ZHOU JG, LI SX, WANG XS, et al. Weakly supervised deep learning for tooth-marked tongue recognition. Frontiers in Physiology, 2022, 13: 847267.
[30] YAN JJ, LI DX, GUO R, et al. Research on classification model of teeth-printed tongue based on two level classifier. China Journal of Traditional Chinese Medicine and Pharmacy, 2022, 37(4): 2181–2185.
[31] PENG CD, WANG L, JIANG DM, et al. Establishing and validating a spotted tongue recognition and extraction model based on multiscale convolutional neural network. Digital Chinese Medicine, 2022, 5(1): 49–58.
[32] YANG JD, ZHANG P. Tongue image classification method based on transfer learning and fully connected neural network. Academic Journal of Naval Medical University, 2018, 39(8): 897–902.
[33] WANG XZ, LUO SY, TIAN GH, et al. Deep learning based tongue prickles detection in traditional Chinese medicine. Evidence-Based Complementary and Alternative Medicine, 2022, 2022: 5899975.
[34] WEN KZ, WEI YK. Fine classification of fat and thin tongue based on deep convolution neural network. Modern Computer, 2020(34): 87–90, 105.
[35] ZHANG J, ZENG D, HE L, et al. Tongue recognition method based on image segmentation: CN202310570065.9. 2023-08-15.
[36] LI XG, FANG ZY, ZHOU L. A fuzzy classification method based on data and model collaborative update for traditional Chinese medicine tongue image rotten and greasy features: CN202310068918.9. 2023-05-16.
[37] WANG X, WANG XR, LOU YN, et al. Constructing tongue coating recognition model using deep transfer learning to assist syndrome diagnosis and its potential in noninvasive ethnopharmacological evaluation. Journal of Ethnopharmacology, 2022, 285: 114905.
[38] LI JW, ZHANG ZD, ZHU XL, et al. Automatic classification framework of tongue feature based on convolutional neural networks. Micromachines, 2022, 13(4): 501.
[39] WEI BG, SHEN LS, CAI YH, et al. Research on curdy and greas tongue fur analysis for traditional Chinese medicine. Acta Electronica Sinica, 2003, 31(S1): 2083–2086.
[40] ZHANG K, ZHANG HL, JIN S, et al. Analysis plumpness and slenderness of tongue based on the neural net. China Journal of Traditional Chinese Medicine and Pharmacy, 2014, (10): 3111–3114.
[41] YAN JJ, CHEN BC, GUO R, et al. Tongue image texture classification based on image inpainting and convolutional neural network. Computational and Mathematical Methods in Medicine, 2022, 2022: 6066640.
[42] ZHANG K, ZHANG HL, JIN S, et al. Analysis plumpness and slenderness of tongue based on the neural net. Chinese Imaging Journal of Integrated Traditional and Western Medicine, 2014, 29(10): 3111–3114.
[43] XU JT, ZHANG ZF, SUN Y, et al. Recognition of acantha and ecchymosis in tongue pattern. Academic Journal of Shanghai University of Traditional Chinese Medicine, 2004, 18(4): 38–40.
[44] JIANG T, HU XJ, YAO XH, et al. Tongue image quality assessment based on a deep convolutional neural network. BMC Medical Informatics and Decision Making, 2021, 21(1): 147.
[45] XIAN HM, XIE YY, YANG ZZ, et al. Automatic tongue image quality assessment using a multi-task deep learning model. Frontiers in Physiology, 2022, 13: 966214.
[46] LOH HW, OOI CP, SEONI S, et al. Application of explainable artificial intelligence for healthcare: a systematic review of the last decade (2011-2022). Computer Methods and Programs in Biomedicine, 2022, 226: 107161.
[47] PANWAR H, GUPTA PK, SIDDIQUI MK, et al. A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. Chaos, Solitons, and Fractals, 2020, 140: 110190.
[48] JIANG HY, XU J, SHI RJ, et al. A multi-label deep learning model with interpretable grad-CAM for diabetic retinopathy classification. Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual International Conference, 2020: 1560–1563.
[49] JIANG T, LU Z, HU XJ, et al. Deep learning multi-label tongue image analysis and its application in a population undergoing

routine medical checkup. Evidence-Based Complementary and Alternative Medicine, 2022, 2022: 3384209.
[50] LI J, HUANG JB, JIANG T, et al. A multi-step approach for tongue image classification in patients with diabetes. Computers in Biology and Medicine, 2022, 149: 105935.
[51] LI J, YUAN P, HU XJ, et al. A tongue features fusion approach to predicting prediabetes and diabetes with machine learning. Journal of Biomedical Informatics, 2021, 115: 103693.
[52] LI J, CHEN QG, HU XJ, et al. Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques. International Journal of Medical Informatics, 2021, 149: 104429.
[53] JIANG T, GUO XJ, TU LP, et al. Application of computer tongue image analysis technology in the diagnosis of NAFLD. Computers in Biology and Medicine, 2021, 135: 104622.
[54] YAO SK, DUAN SJ, CHEN JL, et al. A prediction model for fatty liver based on tongue features and BMI index. 2019.
[55] DAI SX, GUO XJ, LIU S, et al. Application of intelligent tongue image analysis in Conjunction with microbiomes in the diagnosis of MAFLD. Heliyon, 2024, 10(7): e29269.
[56] LIN RY, YU HY, QIN JY, et al. Association between tongue coating thickness and clinical characteristics among idiopathic membranous nephropathy patients. Journal of Ethnopharmacology, 2015, 171: 125–130.
[57] HAN SW, CHEN Y, HU J, et al. Tongue images and tongue coating microbiome in patients with colorectal cancer. Microbial Pathogenesis, 2014, 77: 1–6.
[58] YUAN L, YANG L, ZHANG SC, et al. Development of a tongue image-based machine learning tool for the diagnosis of gastric cancer: a prospective multicentre clinical cohort study. EClinicalMedicine, 2023, 57: 101834.
[59] VAN DER VELDEN BHM, KUIJF HJ, GILHUIJS KGA, et al. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Medical Image Analysis, 2022, 79: 102470.
[60] YUAN CC, WANG J, SHU B, et al. The concept of traditional Chinese medicine phenomics and the construction of related research system. Journal of Traditional Chinese Medicine, 2022, 63(5): 407–411.

基于机器学习的舌象形质诊断分析研究现状与展望

许家佗*, 江涛, 刘实

上海中医药大学中医学院, 上海 201203, 中国

【摘要】基于图像的智能化诊断是中医舌诊现代化研究的重要方向。近年来，以卷积神经网络（CNNs）、Transformers 等深度学习为代表的机器学习方法被广泛应用于电子计算机断层扫描（CT）、核磁共振成像（MRI）等医学影像图像分析领域，使得临床决策更加高效和精准。先进的人工智能技术也为中医舌诊医疗器械研发和数字化中医舌诊方法创造了新的机遇，促进了中医舌诊的标准化和智能化。经典图像分析方法实现了舌象的颜色数据化表达，但对于复杂的舌象形质特征如齿痕、点刺、裂纹、厚薄、腐腻、剥苔等的综合识别分析仍是当前舌诊研究面临的瓶颈问题。本文从舌象形质特征的智能分析与病证诊断应用等方面展开论述，归纳了经典的图像分析方法与深度学习方法的研究现状，梳理了舌象特征在临床疾病风险预测中的应用情况，提出了人工智能舌诊技术的机遇挑战和发展方向。总之，传统中医舌诊与人工智能技术结合，将有效提升中医舌诊的科学内涵，提升舌诊临床普适应用，推动中医诊疗模式的现代化发展。

【关键词】舌象图像；形质特征；深度学习；智能诊断；舌诊