Development and validation of tongue imaging-based radiomics tool for the diagnosis of insomnia degree: a two-center study

Rui Yu1 3*, Ze-Kun Jiang4*, Rong Shao1, Qian Yan1, Li-Juan Zhou1, Ting-Rui Zhang5, Ying-Chun Sun1 1

1Department of Gynecology, Qingdao Hizer Hospital Affiliated of Qingdao University (Qingdao Traditional Chinese Medicine Hospital), Qingdao 266000, China. 2Department of Traditional Chinese Medicine, Jiaozhou Traditional Chinese Medicine Hospital, Qingdao 266000, China. 3College of Computer Science, Sichuan University, Chengdu 610000, China. 4Adult Traditional Chinese Medicine Department, Qingdao Women and Children’s Hospital, Qingdao 266000, China. 5College of Traditional Chinese Medicine, Shandong University of Traditional Chinese Medicine, Jinan 250000, China. 6Sichuan University - Pittsburgh Institute, Sichuan University, Chengdu 610000, China.

*These authors contributed equally to this work and are co-first authors for this paper.

Corresponding to: Ying-Chun Sun, Department of Gynecology, Qingdao Hizer Hospital Affiliated of Qingdao University (Qingdao Traditional Chinese Medicine Hospital), Fuxin Road, Shibei District, Qingdao 266000, China. E-mail: sunny8223@126.com.

Author contributions
RY: conceptualization, investigation, data collection, visualization, writing-original draft, writing-review & editing. ZK: conceptualization, methodology, visualization, writing-original draft, writing-review & editing. RS: conceptualization, data curation, validation, formal analysis. QY: formal analysis, investigation, supervision. LZ: data collection, formal analysis. TZ: formal analysis, writing-review & editing. YS: conceptualization, data curation, validation, project administration, supervision, resources, writing-original draft, writing-review & editing. All authors read and approved the submitted version.

Competing interests
The authors declare no conflicts of interest.

Acknowledgments
This work was financially supported by the Shandong Province Traditional Chinese Medicine Science and Technology Project (M-2022012). The authors thank all the patients for their willingness to cooperate with our study. The authors would like to thank all the reviewers who participated in the review and MJEditor (www.mjeditor.com) for its linguistic assistance during the preparation of this manuscript.

Peer review information
Medical Data Mining thanks all anonymous reviewers for their contribution to the peer review of this paper.

Abbreviations
TCM, traditional Chinese medicine; TIR, tongue imaging-based radiomics; ISI, Insomnia Severity Index; AI, artificial intelligence; DICOM, Digital Imaging and Communications in Medicine; ROI, region of interest; LASSO, least absolute shrinkage and selection operator; IMc, informational measure of correlation.

Citation

Executive editor: Xin-Yun Zhang.
Received: 24 August 2023; Accepted: 01 November 2023;
Available online: 10 November 2023.
© 2024 By Author(s). Published by TMR Publishing Group Limited. This is an open access article under the CC-BY license. (https://creativecommons.org/licenses/by/4.0/)

Abstract
Background: Traditional Chinese medicine (TCM) is commonly used for the diagnosis and treatment of insomnia, with tongue diagnosis being particularly important. The aim of our study was to develop and validate a novel tongue imaging-based radiomics (TIR) method for accurately diagnosing insomnia severity. Methods: This two-center analysis prospectively enrolled 399 patients who underwent tongue imaging between July and October 2021 and divided them into primary and validation cohorts by study center. Here, we referred to the Insomnia Severity Index (ISI) standard and the degree of insomnia was evaluated as absent, subthreshold, moderate, or severe. For developed the TIR diagnostic tool, a U-Net algorithm was used to segment tongue images. Subsequently, seven imaging features were selected from the extracted high-throughput radiomics features using the least absolute shrinkage and selection operator algorithm. Then, the final radiomics model was developed in the primary cohort and tested in the independent validation cohort. Finally, we assessed and compared the diagnostic performance differences between TCM tongue diagnosis and our TIR diagnostic tool with the ISI gold standard. The confusion matrix was calculated to evaluate the diagnostic performance. Results: Seven tongue imaging features were selected to build the TIR tool, with showing good correlations with the insomnia degree. The TIR method had an accuracy of 0.798, a macro-average sensitivity of 0.78, a macro-average specificity of 0.906, a weighted-average sensitivity of 0.798, and a weighted specificity of 0.916, showing a significantly better performance compared to the average performance of three experienced TCM physicians (mean accuracy of 0.458, P < 0.01). Conclusions: The preliminary study demonstrates the potential application of TIR in the diagnosis of insomnia degree and measurement of sleep health. The integration of quantitative imaging analysis and machine learning algorithms holds promise for advancing both of TCM and precision sleep medicine.

Keywords: insomnia; tongue image; radiomics; machine learning; traditional Chinese medicine
Introduction

Insomnia, a serious public health concern [1], is defined as difficulty falling asleep, difficulty maintaining sleep, or poor sleep quality despite adequate opportunities for sleep accompanied by some form of daytime dysfunction [2]. The prevalence of insomnia is approximately 10%–20%, with chronic insomnia accounting for approximately 50% of cases [3]. The etiology and physiology of insomnia are closely related to genetic, environmental, behavioral, and physiological factors [4, 5]. The sleep quality of patients with insomnia is usually assessed by the Insomnia Severity Index (ISI) [6]. ISI, developed by Morin [7] from Canada, is a reliable and effective tool to assess the severity of insomnia, which may be better than other tools [8].

In modern medicine, we look at insomnia scientifically from different perspectives, but we can also learn lessons from traditional medicine, and quantify and scientifically design experiential knowledge. We can find that in the traditional medicine of many countries and nations, there are views on the diagnosis of insomnia [9–13]. Traditional Chinese medicine (TCM) believes that the Heart is the key organ involved in insomnia and it is finally manifested as the balance or imbalance of the Yin (in Chinese philosophy, the female, latent, passive principle, characterized by dark, cold, wetness, passivity, disintegration, etc.) and Yang (in Chinese philosophy, the masculine, active and positive principle, characterized by light, warmth, dryness, activity, etc.) of functional state [13]. So TCM doctors can observe and judge patients’ insomnia symptoms by various manifestations of Yin and Yang imbalance, such as changes in tongue status [14].

At Chinese medicine hospitals, TCM physicians can diagnose insomnia through tongue diagnosis. From the perspective of TCM, the tongue status is closely related to human physiology and pathology [15]. Tongue imaging indicates valuable information about the disease state, which are used to diagnose various diseases, reported in many past studies [16–21]. These experimental results lead us to believe that in future research, new techniques may be used to quantify tongue information for more precision clinical diagnosis, especially for insomnia.

Artificial intelligence (AI) has been applied to public health and could bring about unprecedented advances in medical practice. The use of AI in medicine reduces the workload and working time of doctors and improves the accuracy of outcomes, thus contributing to precision medicine [22]. AI has been successfully applied to TCM [23–25] and tongue imaging diagnosis [26–30]. However, to date, the intelligent classification of insomnia based on tongue imaging has not been reported. The radiomics technology, an emerging AI technology, has made remarkable achievements in various medical imaging applications [31–34], which can be applied to develop tongue imaging biomarkers for insomnia diagnosis.

Therefore, this study hypothesizes that tongue images contain valuable information for diagnosing the severity of insomnia. The aim of our study was to develop and test a tongue imaging-based radiomics (TIR) tool to evaluate the degree of insomnia. This approach could accurately predict the insomnia degree and provide clinical decision-making references for TCM doctors, promoting individualized medical processing in TCM and sleep medicine.

Materials and methods

Ethical considerations and patient selection

This two-center study was conducted in accordance with the tenets of the Declaration of Helsinki (as revised in 2013). The study protocol was approved by institutional review boards of Qingdao Higer Hospital and Qingdao Women and Children’s Hospital (2022HC088LS011). All participants provided informed consent.

We enrolled a total of 411 patients underwent tongue imaging at Qingdao Higer Hospital (hospitals 1, n = 304) and Qingdao Women and Children’s Hospital (hospital 2, n = 107) between July and October 2021. They were all informed participants in the research project and underwent insomnia diagnosis. Inclusion criteria were: (1) age ≥ 18 years; (2) ability to fill out the ISI questionnaire independently; and (3) ability to stick the tongue out. Exclusion criteria were: (1) refusal to be photographed (n = 7); (2) too small an area of the tongue sticking out (n = 2); and (3) an unclear tongue image (n = 3). After applying these criteria, 399 patients were prospectively enrolled and divided into two independent cohorts: a primary cohort, including 300 patients from hospital 1 who would undergo the radiomics analysis and modeling, and an external validation cohort, including 99 patients from hospital 2 who would undergo the independent clinical evaluation. Figure 1 shows the flowchart of the two-center radiomics study, including tongue image acquisition, image preprocessing, U-Net segmentation, feature extraction and selection, and machine modeling and validation.

Figure 1 Overview of the study workflow, including tongue image acquisition, image preprocessing, U-Net segmentation, feature extraction and selection, and machine modeling and validation. LASSO, least absolute shrinkage and selection operator.
Tongue image acquisition
All tongue images were acquired by two TCM doctors using a special camera equipment with a 48-million-pixel. Before image acquisition, to obtain accurate tongue imaging information and eliminate the false tongue findings caused by various operating factors, we manually inspected the tongue before imaging. In order to obtain accurate information from the tongue diagnosis, care must be taken to exclude false tongue images caused by various manipulation factors. It mainly includes the following three aspects. (1) Improper tongue extension: such as tongue extension especially hard, tongue body tension curl; or tongue extension time is too long, will affect the tongue body blood circulation and cause tongue color change, or tongue moss compact change, or dry humidity changes. (2) Food and drug influence: certain food drugs will make the tongue stained, called dyed moss. Such as drinking milk, soy milk, coconut milk, etc. can make the tongue moss white and thick; consumption of peanuts, melon seeds, and beans. Walnuts, etc., can stain the tongue yellow; various dark brown food, drugs, or eat sour plum, long-term smoking, etc., can make the tongue dyed gray, or black. (3) Oral environment: tooth mutilation can cause ipsilateral tongue moss thickening, dentures and dental veneers can leave teeth marks on the tongue; open mouth breathing during sleep can make the tongue moss thicken and dry, etc. The tongue abnormalities caused by these factors cannot be taken as signs of the body’s case. Images of the entire tongue when sticking out were obtained under abundant natural light. The tongue image data of 399 patients were obtained under this imaging standard, ensuring the replicability of the study.

Degree of insomnia
The degree of insomnia was evaluated using ISI and classified as absent, subthreshold, moderate, or severe [7, 35]. ISI comprises seven items, and each item is scored on a 5-point Likert scale. The total score ranges from 0 to 28, and a higher total score implies a worse degree of insomnia. A total score of 0–7 indicates absence of insomnia, 8–14 indicates subthreshold insomnia, 15–21 indicates moderate insomnia, and 22–28 indicates severe insomnia. We used the Chinese version of ISI in clinical practice, which was a common tool in clinical teaching.

Image preprocessing and U-Net segmentation
We converted the tongue images into the Digital Imaging and Communications in Medicine (DICOM) format and added patient information into the DICOM data helping for patient data management [36]. When conducting the analysis, we only process the tongue image information of the patients, and all other patients’ information remains invisible, thus ensuring the privacy of clinical data. Subsequently, to define the region of interest (ROI) of the tongue, we trained a tongue segmentation model using U-Net algorithms. All images were normalized to 512 × 512 to feed the deep learning neural network. The U-Net algorithm was implemented by the open code (https://github.com/zhixuhao/unet), as described by Ronneberger et al. [37]. Here, we used 200 tongue images with real masks for U-Net model training, and the other 50 images with real masks for tested showing the mean Dice of 0.92. Finally, the trained U-Net model performed image segmentation for other tongue images and an experienced TCM doctor determined the final ROIs. The AI segmentation model could accelerate the tongue image processing pipeline in future works, thus improving diagnostic efficiency.

Radiomics analysis
Radiomics analysis included radiomics feature extraction and selection. First, the image preprocessing was implemented as follows: the interpolator was siktBSpline, the normalize scale was 100, the bin width was 25, and the voxel array shift was 300. From each ROI on the tongue images, we extracted 14 shape features, 18 first-order histogram features, and 73 texture features using the Python Pyradiomics Library (version 3.0.1; Computational Imaging and Bioinformatics Lab, Harvard Medical School) [38]. Most of them conform to the Imaging Biomarker Standardization Initiative, and the detailed information was provided in Pyradiomics documentation (https://pyradiomics.readthedocs.io/en/latest/index.html). Then, all radiomics features data were standardized to facilitate feature selection and modeling. Finally, the least absolute shrinkage and selection operator (LASSO) algorithm [39] was applied to select the relevant features. The detailed parameters were as follows: the alpha was 1, the max iter was 5000, the selection was cyclic. The important features were selected and assigned correlation coefficients as weights, while irrelevant features were assigned zero weight. Figure 2 shows the details of the radiomics feature selection.

Machine learning modeling and validation
LASSO logistic regression assigned weight to the selected features. Subsequently, the radiomics signature was developed using these features in the primary cohort and tested in the independent validation cohort. Here, we verified differences in distribution of the radiomics signatures across all degrees of insomnia. Prediction results of the signature in the primary and validation cohorts were displayed by a confusion matrix. The accuracy, precision, sensitivity, specificity, and F1 score were also calculated.

Figure 2 Radiomics feature selection. (A) Least absolute shrinkage and selection operator coefficient profiles of the radiomics features. (B) Weight coefficient histogram of the selected radiomics features. GLCM, gray-level co-occurrence matrix; GLSZM, gray-level size zone matrix.
Statistical analysis
All machine learning procedures and statistical analyses were performed using Python (versions 3.5 and 3.6). Categorical variables were compared using the chi-square or Fisher’s exact test. Continuous variables were compared using the Mann-Whitney U or Student’s t-test. The confusion matrix, accuracy, precision, sensitivity, specificity, and F1 score were calculated for prediction evaluation. The DeLong test was performed to compare differences in predictions. A P-value < 0.05 was considered to indicate statistical significance.

Results
Baseline characteristics
Table 1 shows the baseline characteristics of the patients. In the primary cohort, we included a total of 300 participants (mean age: 57.24 years; 204 females and 96 males), while in the validation cohort, there were 99 participants (mean age: 50.27 years; 94 females and 5 males). Based on the ISI, the primary cohort comprised 98 patients with no insomnia, 95 with subthreshold insomnia, 82 with moderate insomnia, and 25 with severe insomnia. In the validation cohort, there were 32, 50, 13, and 4 patients with no, subthreshold, moderate, and severe insomnia, respectively.

Table 1 Baseline information of the primary and validation cohorts

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All patients (n = 399)</th>
<th>Primary cohort (n = 300)</th>
<th>Validation cohort (n = 99)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (range), years</td>
<td>51.98 (18–88)</td>
<td>57.24 (21–88)</td>
<td>50.27 (18–86)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>101 (25.31%)</td>
<td>96 (32.00%)</td>
<td>5 (5.05%)</td>
</tr>
<tr>
<td>Female</td>
<td>298 (74.69%)</td>
<td>204 (68.00%)</td>
<td>94 (94.95%)</td>
</tr>
<tr>
<td>Degree of insomnia, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No insomnia</td>
<td>130 (32.58%)</td>
<td>98 (32.67%)</td>
<td>32 (32.32%)</td>
</tr>
<tr>
<td>Subthreshold</td>
<td>145 (36.34%)</td>
<td>95 (31.67%)</td>
<td>50 (50.51%)</td>
</tr>
<tr>
<td>Moderate</td>
<td>95 (23.81%)</td>
<td>82 (27.33%)</td>
<td>13 (13.13%)</td>
</tr>
<tr>
<td>Severe</td>
<td>29 (7.27%)</td>
<td>25 (8.33%)</td>
<td>4 (4.04%)</td>
</tr>
</tbody>
</table>

Table 2 Selected radiomics features with coefficient weights in the LASSO selection

<table>
<thead>
<tr>
<th>Features</th>
<th>Type</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imc2</td>
<td>GLCM</td>
<td>0.085</td>
</tr>
<tr>
<td>Maximum 2D diameter column</td>
<td>Shape</td>
<td>0.044</td>
</tr>
<tr>
<td>Elongation</td>
<td>Shape</td>
<td>0.014</td>
</tr>
<tr>
<td>Energy</td>
<td>First-order</td>
<td>0.006</td>
</tr>
<tr>
<td>Cluster shade</td>
<td>GLCM</td>
<td>−0.003</td>
</tr>
<tr>
<td>Gray-level variance</td>
<td>GLSZM</td>
<td>−0.030</td>
</tr>
<tr>
<td>Skewness</td>
<td>First-order</td>
<td>−0.033</td>
</tr>
</tbody>
</table>

Imc, informational measure of correlation; GLCM, gray-level co-occurrence matrix; GLSZM, gray-level size zone matrix; LASSO, least absolute shrinkage and selection operator.

Valuable radiomics features
Our selected valuable radiomics features are provided in Table 2. LASSO selection yielded two shape features (maximum 2D diameter column and elongation), two first-order histogram features (energy and skewness), and three texture features (informational measure of correlation (Imc) 2 and cluster shade from the gray-level co-occurrence matrix, gray-level variance from the gray-level size zone matrix). All the selected radiomics features had good correlations with the degree of insomnia. Among the four most correlated features, Imc2 and maximum 2D diameter column were positively correlated with the severity of insomnia, while the skewness and gray-level variance are primarily negatively correlated.

Model performance and validation
Figure 3 shows the radiomics signature distribution in the validation cohort. The signature showed statistical significance in diagnosing insomnia and determining all degrees of insomnia (P < 0.05). Figure 4 shows the confusion matrix in the primary and validation cohorts. And the detailed performance indicators are shown in Table 3. The diagnostic accuracies of TIR model for no, subthreshold, moderate, and severe insomnia were 0.808, 0.838, 0.96, and 0.99, respectively.

Figure 3 Radiomics signature distribution in the validation cohort. (A) Distribution in no (blue), subthreshold (orange), moderate (green), and severe insomnia (red) groups. (B) Distribution in health subjects (blue) and patients with insomnia (orange).
Comparison of the TIR tool with TCM physicians
The comparison of TIR and TCM doctors are shown in Table 4. Here, doctors A, B, and C were a resident physician, an attending physician, and an associate physician, with 3-year, 10-year, and more than 10-year clinical experience, respectively. The TIR tool showed an accuracy of 0.798, a macro-average sensitivity of 0.78, a macro-average specificity of 0.906, a weighted-average sensitivity of 0.798, and a weighted specificity of 0.916. Figure 5 shows the confusion matrix of TCM doctors’ diagnostic performance. The doctors showed an accuracy of 0.458, a macro-average sensitivity of 0.54, a macro-average specificity of 0.792, a weighted-average sensitivity of 0.458, and a weighted specificity of 0.709. The average diagnostic performance of the three doctors was significantly poorer compared to that of the radiomics signature (P < 0.01).

Finally, we provided eight representative examples of successful prediction by the TIR method (Figure 6), which showed the same viewpoint as the diagnosis of TCM doctors.

Discussion
In this study, we firstly explored the TIR analysis in assessing insomnia severity utilizing a dataset from two distinct centers. Through this endeavor, we successfully developed and validated a radiomics signature derived from processed tongue images, enabling intelligent classification of insomnia as defined by ISI scores. The experimental outcomes showed the accuracy and robustness of this approach, and the TIR can serve as a swift adjunctive tool for aiding TCM diagnosis.

In addition to building the radiomics model, this study also found other interesting findings. Seven radiomics features were related to the severity of insomnia. Imc2, maximum 2D diameter column, elongation, and energy, positively correlated with the severity of insomnia, while the skewness, gray-level variance, and cluster shade negatively correlated with the severity of insomnia. Here, Imc2 measures the correlation between different regions of image. A higher Imc2 value indicates stronger correlation between different regions in the image, implying higher texture complexity. As for the maximum 2D diameter column and elongation, they naturally describe the shape information of the tongue. Energy reflects grayscale level distribution uniformity and stability; higher values indicate a more uniform distribution. Skewness describes the degree of skewness in the grayscale level distribution. Positive values indicate skew towards higher grayscale levels, while negative values indicate skew towards lower levels. Gray-level variance quantifies the variability in grayscale levels within an image and the higher values indicate greater variation in grayscale levels, signifying uneven grayscale distribution. Cluster shade is a statistical measure describing the distribution of grayscale levels in an image, which reflects how grayscale levels tend to cluster together. By comprehending the mathematical essence underlying radiomic features and integrating the outcomes of correlation analysis, we can preliminarily observe that heightened complexity in texture information within tongue images, coupled with robust inter-regional correlations and a uniform grayscale distribution, typically signifies a more severe degree of insomnia. Conversely, tongues exhibiting substantial grayscale distribution disparities and non-uniformity tend to belong to healthy individuals. From a TCM perspective, these features detected by AI corresponded to the surface cracks, shape, and coating of the tongue, may reflecting the imbalance of Yin and Yang in the body [40].

In TCM diagnosis, cracked tongue/fat tongue is an important sign of TCM tongue diagnosis and observation, and an important indicator of TCM diagnosis of insomnia. However, due to the uneven clinical level of TCM doctors, the clinical diagnosis of insomnia will be greatly different. Therefore, there is no objective standard for the diagnosis of insomnia degree in TCM. Through deep learning and radiomics analysis, AI can quantitatively analyze tongue features and establish a model to diagnose the degree of insomnia, which may replace TCM doctors to make diagnosis, eliminate the subjectivity of TCM diagnosis, and form the standardization of TCM diagnosis of insomnia, which may be a great progress of TCM tongue diagnosis.

In addition, the performance of AI is quicker than that of TCM doctors, thereby saving time and labor and improving efficiency. AI can also help doctors with less experience in obtaining diagnoses. Thus, the addition of radiomics and machine learning can provide an objective standardization method to determine the degree of insomnia.

In this study, although TIR exhibited markedly superior diagnostic

Table 3 Diagnostic performance of the TIR model in the validation cohort

<table>
<thead>
<tr>
<th>Degree of insomnia</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>F1 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>No insomnia</td>
<td>0.808</td>
<td>0.741</td>
<td>0.625</td>
<td>0.870</td>
<td>0.678</td>
</tr>
<tr>
<td>Subthreshold</td>
<td>0.838</td>
<td>0.804</td>
<td>0.900</td>
<td>0.977</td>
<td>0.849</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.960</td>
<td>0.846</td>
<td>0.846</td>
<td>0.846</td>
<td>0.846</td>
</tr>
<tr>
<td>Severe</td>
<td>0.990</td>
<td>1.000</td>
<td>0.750</td>
<td>1.000</td>
<td>0.857</td>
</tr>
</tbody>
</table>

TIR, tongue imaging-based radiomics.
Table 4 Comparison of performance results between the TIR tool and TCM doctors

<table>
<thead>
<tr>
<th>Reader</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Precision*</th>
<th>Sensitivity*</th>
<th>Specificity*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine</td>
<td>0.798</td>
<td>0.848</td>
<td>0.780</td>
<td>0.906</td>
<td>0.798</td>
<td>0.798</td>
<td>0.916</td>
</tr>
<tr>
<td>Doctor A</td>
<td>0.350</td>
<td>0.325</td>
<td>0.389</td>
<td>0.763</td>
<td>0.397</td>
<td>0.350</td>
<td>0.702</td>
</tr>
<tr>
<td>Doctor B</td>
<td>0.450</td>
<td>0.438</td>
<td>0.617</td>
<td>0.782</td>
<td>0.423</td>
<td>0.450</td>
<td>0.677</td>
</tr>
<tr>
<td>Doctor C</td>
<td>0.575</td>
<td>0.641</td>
<td>0.613</td>
<td>0.831</td>
<td>0.565</td>
<td>0.575</td>
<td>0.747</td>
</tr>
<tr>
<td>Average</td>
<td>0.458</td>
<td>0.468</td>
<td>0.540</td>
<td>0.792</td>
<td>0.462</td>
<td>0.458</td>
<td>0.709</td>
</tr>
</tbody>
</table>

*, the performance index was calculated using the macro-average method; **, the performance index was calculated using the weighted-average method; TIR, tongue imaging-based radiomics; TCM, traditional Chinese medicine.

Figure 5 The confusion matrix of three TCM doctors for insomnia diagnosis. Reader 1–3 were a resident physician, an attending physician, and an associate physician, respectively. TCM, traditional Chinese medicine.

Figure 6 Representative examples of successful prediction by using TIR method. TIR, tongue imaging-based radiomics.

performance compared to three TCM doctors (a resident physician, an attending physician, and an associate physician), particularly surpassing a TCM doctor with more than 10-year clinical experience, some certain factors did influence the diagnostic proficiency of the TCM doctors. TCM diagnosis traditionally encompasses observation, listening, questioning, and pulse-taking. Relying solely on tongue image examination constitutes observation alone, potentially overlooking other diagnostic cues. As a result, in real clinical settings, TCM doctors often achieve diagnostic accuracies significantly higher than those presented here. Our study only involved both AI model and doctors diagnosing based on tongue images, hence the substantial advantage also demonstrated by the AI model. This is not a negation of TCM diagnosis, but rather a reflection and adjustment.

The introduction of this new imaging technology has many advantages compared to the traditional method of having patients fill out the ISI questionnaire. On the one hand, not all patients can use the ISI questionnaire, as it requires active cooperation from the patients to complete. In contrast, the tongue imaging-based diagnostic tool can diagnose the severity of insomnia for any patient, expanding the diagnostic scope. On the other hand, using the ISI questionnaire takes a considerable amount of time, and in regions with a large number of patients, its implementation may not be easy. The AI-driven TIR tool only takes a few seconds to provide diagnostic results and decision support, which will significantly save medical time and optimize medical resources.

Limitations of the study includes the following aspects. First, the
Conclusion

In conclusion, we have successfully developed and assessed a TIR method to determine the degree of insomnium. The TIR tool showed better inspection performance compared to TCM doctors and may help provide more reliable guidance for clinicians.

References

4. The tongue features associated with chronic kidney disease. Medicine (Baltimore) 2021;100(9):25037. Available at: https://doi.org/10.1155/2021/4441192


