Scilight

Journal of Machine Learning and Information Security
https://www.sciltp.com/journals/jmlis

Article

Check for updates

Prescription Recommendation Based on Multiview Learning with Latent Dirichlet Allocation

Keju Chen [1,2,3,†], Yun Zhang,[1,2,3,†], Li Zhong [1,2,3] and Yongguo Liu [1,2,3,*]

[1] Knowledge and Data Engineering Laboratory of Chinese Medicine, School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China
[2] Innovation Center for Electronic Information & Traditional Chinese Medicine, University of Electronic Science and Technology of China, Chengdu 610054, China
[3] Intelligent Digital Media Technology Key Laboratory of Sichuan Province, University of Electronic Science and Technology of China, Chengdu 610054, China
* Correspondence: ygliu_uestc@163.com
† These authors contributed equally to this work.

How To Cite: Chen, K.; Zhang, Y.; Zhong, L.; et al. Prescription Recommendation Based on Multiview Learning with Latent Dirichlet Allocation. *Journal of Machine Learning and Information Security* **2026**, *2*(3), 16. https://doi.org/10.53941/jmlis.2026.100016

Received: 8 May 2026
Revised: 1 July 2026
Accepted: 7 July 2026
Published: 31 July 2026

Abstract: Tongue diagnosis, as a key process in Traditional Chinese Medicine (TCM), enables clinicians to assess bodily conditions by visually inspecting the tongue, based on which appropriate prescriptions are formulated with corresponding TCM theory. However, conventional tongue diagnosis relies heavily on the subjective expertise of clinicians, which is labor-intensive, prompting the development of deep learning-based approaches to automate the process. To further explore the application of deep learning in prescription recommendation from tongue images, we introduce a Prescription Recommendation method based on Latent Dirichlet Allocation (PR-LDA). The proposed method begins by extracting tongue image features using a multi-path convolutional neural network enhanced with dense connections, which effectively captures rich and multi-scale visual information. These features are then processed through multiple fully-connected layers to predict relevant medications and their dosages, leveraging similarity measures between the input tongue features and those associated with existing prescriptions. Additionally, Latent Dirichlet Allocation (LDA) is incorporated to uncover latent topic structures within prescriptions, thereby improving accuracy by integrating image features with textual prescription data. Experimental evaluations on our tongue-prescription dataset demonstrate that PR-LDA achieves Precision, Recall, and F1-score values of 0.3212, 0.2447, and 0.2777, respectively.

Keywords: tongue image; latent dirichlet allocation; prescription recommendation; multi-view learning

1. Introduction

Traditional Chinese Medicine (TCM) represents a significant component of China's culture, with a long history of thousands of years. And the four diagnostic methods (observation, listening and smelling, inquiry, and palpation) are commonly employed to identify diseases. Based on individual conditions, appropriate clinical treatment is determined, leading to the formulation of personalized prescriptions based on the TCM theory [1].

In observation, tongue diagnosis, a key element of observation, involves assessing the state of the body by examining the tongue. Traditional tongue diagnosis involves clinicians inspecting the patient's tongue appearance, e.g., coating [2], color [3], texture [4], and other characteristics [5], integrating clinical experience to provide diagnostic suggestions [6,7].

In traditional Chinese medicine, prescription recommendations are closely related to disease diagnosis. Tongue-based prescription recommendation refers to the process of observing tongue features from patients and constructing a personalized prescription that conforms to the laws of traditional Chinese medicine with appropriate dosage. However, this process is influenced by the practitioner's skill level and clinical experience, leading to strong subjectivity

and poor reproducibility, underscoring the need for objectification in tongue diagnosis. To overcome the inherent subjectivity of traditional tongue diagnosis, computerized tongue image analysis typically follows a pipeline encompassing image acquisition, preprocessing, feature extraction, and downstream diagnostic modeling [8–10]. Tongue image analysis models have progressed from classic CNN architectures (e.g., FCN, DeepLabv3+, U-Net) to more sophisticated designs that incorporate multi-scale feature fusion and residual connections. Recently, Transformer-based models have further pushed the frontier of model performance and cross-dataset generalization, while consistently pursuing higher robustness against illumination variations, complex backgrounds, and inter-individual morphological differences. This trend reflects a shift from traditional handcrafted feature engineering and support vector machines to end-to-end deep learning models that jointly address detection, segmentation, and the multi-level classification of diseases, syndromes, and tongue symptoms. In the preprocessing stage, quality control mechanisms have been developed to automatically filter the tongue region in images based on deep learning models [9], which addresses the issues of poor adaptability to images collected from diverse sources in traditional methods. Jia et al. [11] proposed a rapid and precise tongue image segmentation network, which utilized style transfer to alleviate the few-shot sample problem, achieving an IoU of 98.03%, with a frame rate of 75.35 FPS. Chen et al. [12] proposed a multi-label attribute detection framework, in which a multi-branch network is designed to capture the correlation between color, shape, region, and tongue coating quality, achieving multi-label synchronous recognition of 8 surface attributes of tongue images. After the preprocessing stage, computer vision has been extensively applied to non-invasive health assessment. For instance, Jiang et al. [13] employed Mask R-CNN and Faster R-CNN to extract tongue color , coating, morphology and texture feature, and combined these with physiological markers in seven machine learning models, achieving the accuracy of 81.7% for diagnosing NAFLD. Similarly, Wang et al. [14] used 16 expert-annotated tongue attributes (color, shape, coating, etc.) alongside anthropometric indicators to build logistic regression, which reached AUC = 0.93 and accuracy = 86.98%, further confirming the feasibility of tongue image features for non-invasive early NAFLD screening. For metabolic dysfunction-associated steatotic liver disease (MASLD), Deng et al. [15] proposed a joint analysis framework for tongue image feature extraction and rRNA sequencing of oral microbiome, combining five machine learning methods to distinguish different syndromes of MASLD patients, achieving a classification performance with an AUC of 0.939 and an accuracy of 85%. Hu et al. [16] integrated tongue image features with tongue coating microbiome 16S sequencing data in MASLD syndrome classification task. Five machine learning models were employed, achieving an AUC of 0.871 and an accuracy of 79.1%. In the context of diabetes, color descriptor-based fusion strategies across multiple color spaces (RGB, HSV, Lab) have been developed to capture subtle chromatic variations in tongue images, enabling effective classification of diabetic patients [17]. These advancements underscore computer vision as a supportive tool for objective, non-invasive health assessment, providing a solid foundation for subsequent downstream diagnostic tasks.

Furthermore, the research on automatic prescription recommendation is of great significance for a deeper understanding of the principles of traditional Chinese medicine prescriptions and assisting clinicians in diagnosis. The end-to-end prescription recommendation method based on tongue image utilizes deep learning models to integrate feature of tongues and prescriptions, to achieve personalized prescription recommendation through machine learning algorithms. Zhao [18] proposed a prescription recommendation framework based on a dual-stream Vision Transformer (ViT) with multi-label classification, which utilizes GCN (Graph Convolutional Networks) to capture the relationships between herbs and tongue images. But prediction of dosage information was left for future work. Zhuang [19] proposed TCM-KLLaMA, a prescription recommendation LLM integrated with a traditional Chinese medicine knowledge graph, achieving a reduction of hallucination and improving accuracy. However, it primarily relies on textual descriptions of tongue features rather than original tongue images, which may neglect the complicated original tongue feature. Hu [20] utilized a neural framework consisting of single/double-path convolutional and multiple fully connected layers to achieve automatic generation of traditional Chinese medicine prescriptions. But it fails to process the critical dosage information in prescriptions, leading to a significant gap between generated and actual prescriptions.

To better imitate the process of clinicians in tongue diagnosis and explore the effectiveness of deep learning in tongue-based prescription proposing, we designed a method of multi-path convolutional neural network that captures feature of lingual and sublingual images together. And we introduced dosage prediction based on tongue images, to address the limitation of above mentioned method. Furthermore, to capture the compatibility relationship between traditional Chinese medicines and the correlation between prescription and original tongue features, latent dirichlet allocation was integrated in our method, which facilitates discovery of potential therapies of similar tongue feature. Our method for prescription recommendation makes the generated prescriptions more compatible with TCM theory. The contributions of this paper are as follows: (1) A novel TCM prescription recommendation

method (PR-LDA) based on LDA is proposed, addressing the limitation of existing methods that consider only tongue-prescription relationship and neglect the correlation between traditional Chinese medicine prescription and original tongue features; (2) A multi-path convolutional neural network is constructed to extract features from both the lingual and sublingual of the tongue, with an added dose regression module to generate prescription dosage information; (3) By integrating an LDA module, therapy-specific prescriptions are generated, promoting diversity of predicted medications and discovery of latent topics in prescriptions; (4) We conduct comprehensive experiments on a collected tongue-prescription dataset, demonstrating that PR-LDA achieves Precision, Recall, and F1-score values of 0.3212, 0.2447, and 0.2777, respectively.

2. Materials and Methods

2.1. Overview Architecture of PR-LDA

The overview architecture of PR-LDA model is shown in Figure 1, including four components: (1) Tongue feature extraction: Lingual and sublingual images are processed through a multi-path convolutional neural network with dense connections to extract discriminative features; (2) Medication information prediction: Specific medication within a prescription is obtained by measuring medication-tongue similarity; (3) Dose regression: A regression head is incorporated to predict medication dosages; leveraging dose constraint information available in the tongue image prescription dataset; (4) Prescription topic modeling: The LDA-based module discovers latent treatment topics from historical prescriptions. This effectively captures the correlations among medications, enabling the model to capture medication associations between prescriptions.

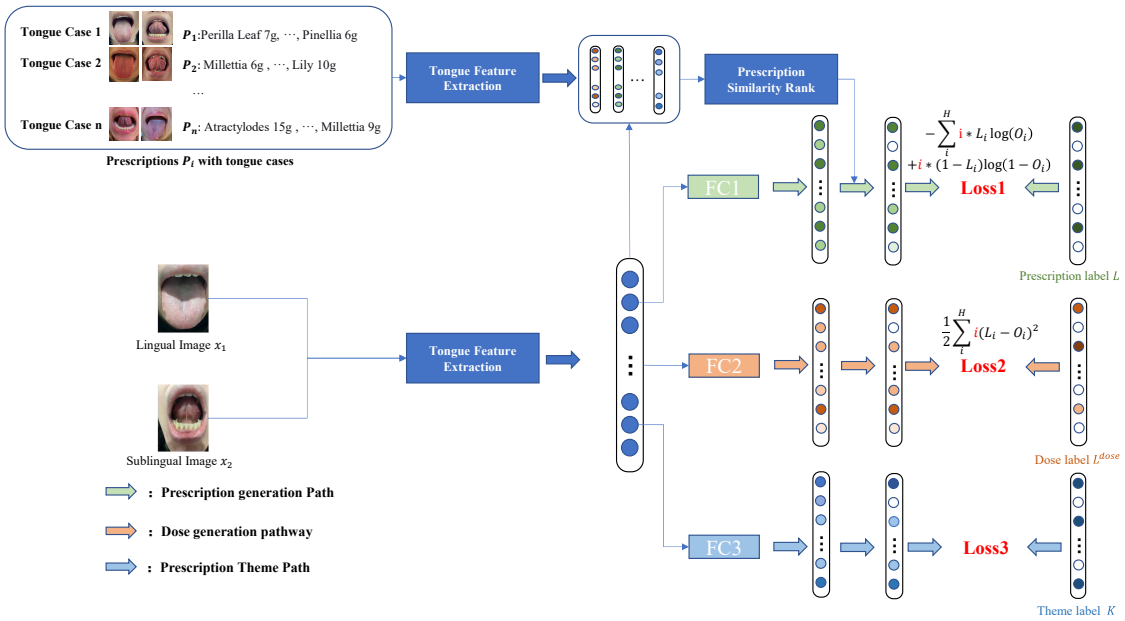


Figure 1. The architecture of the PR-LDA.

2.2. Tongue Feature Extraction

As shown in Figure 2, a multi-path convolutional network is introduced to extract the tongue feature. The lingual image and the sublingual image are taken as inputs for two separate paths. the features extracted from each path are fused to obtain the fused features of the image.

Each path is responsible for extracting features of lingual or sublingual images separately. To facilitate feature sharing between the two paths, dense connections are introduced. Specifically, the input to each convolutional block is formed by concatenating the outputs of all previous layers from both paths, meaning dense connections occur not only between layers within the path, but also between layers in different paths. The process for the two paths are shown in Equations (1) and (2).

$$I_l^1 = [I_{l-1}^1, I_{l-1}^2, I_{l-2}^1, I_{l-2}^2, ..., I_1^1, I_1^2] \tag{1}$$

$$I_l^2 = [I_{l-1}^2, I_{l-1}^1, I_{l-2}^2, I_{l-2}^1, ..., I_1^2, I_1^1] \tag{2}$$

In these equations, I represents the tongue feature in the feature extraction network, l represents the l-th layer of the network, $[\cdot, \cdots, \cdot]$ denotes the feature concatenation operation, while I_L^1 and I_L^2 represent the lingual feature vector and the sublingual vein feature vector output by the last layer, respectively, with L represents the number of layers of the network. Subsequently, these two feature vectors are concatenated to form the final tongue image feature F for prescription construction. The process is shown in Equation (3), where $[\cdot, \cdot]$ represents the concatenation operation of feature:

$$F = [I_L^1, I_L^2] \tag{3}$$

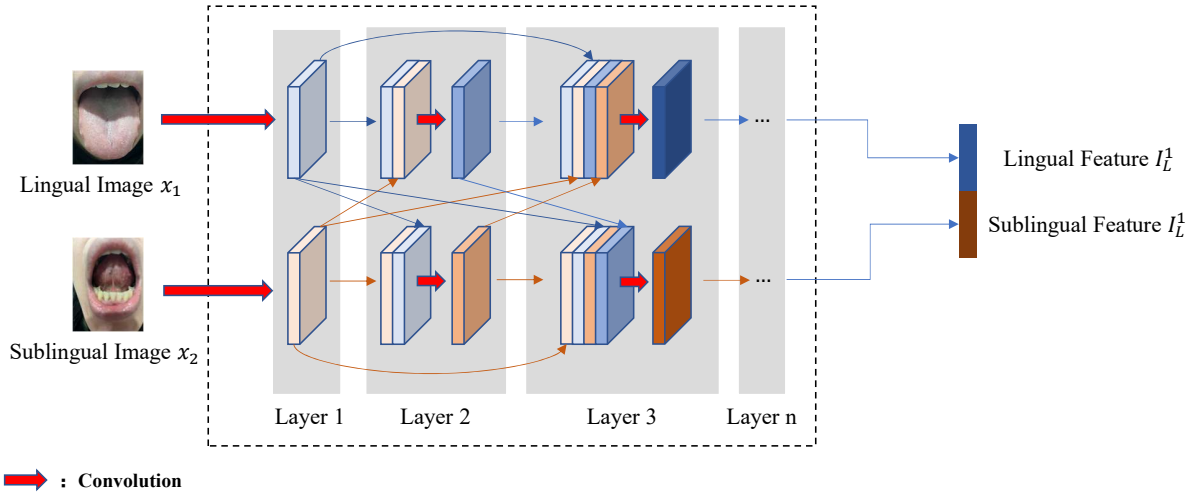


Figure 2. Structure of tongue feature extraction network.

2.3. Prescription Generation Based on Tongue Similarity

Tongue image similarity measures the similarity between states of different patients by comparing their tongue image features. This module aims to recommend medications by measuring the similarity between tongue images. It identifies a set of similar cases from the dataset and leverages the frequently occurring herbs in these cases to guide prescription generation. Specifically, the multi-path convolutional network is used to the tongue image dataset to extract the feature vector F of the input lingual and sublingual image. The cosine similarity between feature of input images and all features in the current dataset is computed as defined in Equation (4):

$$Sim_j = \frac{F \cdot F_j}{\|F\| \cdot \|F_j\|} \tag{4}$$

where F_j denotes the feature vector of the j-th tongue image in the current tongue image dataset. The operator $\cdot$ denotes vector dot multiplication, and $\|\cdot\|$ denotes the length of the computed vector. Subsequently, the calculated similarities are sorted, and the top-H prescriptions corresponding to the most similar tongue images are selected to form a similar prescription set. The probability S_i of the i-th medication appearing in the set of traditional Chinese medicine prescriptions. The calculation is shown in Equation (5):

$$S_i = \frac{h_i}{H} \tag{5}$$

where h_i represents the frequency of the i-th medication appearing in the top-H similar prescription set. An initial prescription vector F' is generated from the image feature vectors F through a fully connected layer followed by batch normalization (BN), as shown in Equation (6):

$$F' = BN(FC_1(F)) \tag{6}$$

where FC_1 denotes a fully connected layer, and each element in F' vector is normalized to the range $[0, 1]$. The final output prescription O is obtained with prescription vector F' constrained by medication probability vector S, as shown in Equation (7):

$$O = S \cdot F' \tag{7}$$

Finally, the binary cross-entropy loss is adopted to quantify the discrepancy between the generated prescription O and the ground-truth prescription label L This is formally defined in Equation (8):

$$Loss_1 = -\frac{1}{N}\sum_{i=1}^{N}\left(L_i \log\left(O_i\right) + \left(1 - L_i\right)\log\left(1 - O_i\right)\right) \tag{8}$$

where N represents the total number of classes of medications in dataset, O_i represents the probability of the i-th medication appearing in the output label, and L_i represents ground-truth label (0 or 1) in L.

2.4. Dose Generation Based on Tongue Features

In contrast to medication category prediction, medication dosage is heavily influenced by patient-specific demographic factors such as gender, age, weight, and height. These factors are the core basis for individualized dose adjustment in clinical practice. Accordingly, in the dosage regression module, we explicitly incorporate personal demographic information to generate patient-specific dose recommendations that balance therapeutic efficacy and safety. Therefore, in the process of dosage generation, personal demographics are integrated into the generation of medication dosage. Specifically, the demographic features (gender, age, weight, height) are normalized to form a 1×4-dimensional feature vector M. The normalization for each feature M_i is as shown in Equation (9):

$$M_i = \frac{M_i - \min\left(M_i\right)}{\max\left(M_i\right) - \min\left(M_i\right) + \epsilon} \tag{9}$$

where ϵ represents a smoothing constant term. The demographic feature vector M is then projected to the same dimensionality as the tongue image feature vector F through a fully connected layerat, resulting in M'. Then, the projected demographic features M' and the tongue image features F are fused via element-wise multiplication to produce a combined feature vector F_1. And the initial dose information O^{dose} is generated from the fused feature vector F_1 through another fully connected layer and normalization. The process is shown from Equation (10) to Equation (12):

$$M' = BN(FC(M)) \tag{10}$$

$$F_1 = F \cdot M' \tag{11}$$

$$O^{dose} = BN\left(FC_2(F_1)\right) \tag{12}$$

where each value in the O^{dose} vector is normalized to the range $[0, 1]$. To ensure clinical safety, the generated doses are constrained within the official ranges specified in the Pharmacopoeia of People's Republic of China [21]. These ranges are standardized based on the toxicity and active components of each herb. We statistically derive the applicable dosage interval for each herb from existing registered prescriptions by removing outliers (the highest and lowest values). The constraint is shown in Equation (13):

$$O_i^{dose} \in \left\{O^{dose} \middle| \min(O_i') \leq O_i^{dose} \leq \max\left(O_i'\right)\right\} \tag{13}$$

where O_i^{dose} represents the dose of i-th medication in the output O^{dose}, and O_i' denotes the dose of i-th medication in existing registered prescriptions. Any generated dose exceeding the maximum is clipped to $\max(O_i')$, and any dose below the minimum is set to $\min(O_i')$. Finally, the mean squared error loss between the generated dose O^{dose} and the ground-truth dose label L^dose is computed. The loss function $Loss_2$ is shown in Equation (14):

$$Loss_2 = \frac{1}{N}\sum_{i=1}^{N}\left(L_i^{dose} - O_i^{dose}\right)^2 \tag{14}$$

where N denotes the number of classes of medications in the prescription, O_i^{dose} is the predicted dose information of the i-th medication, and L_i^{dose} is its corresponding true dose in L^{dose}.

2.5. LDA-Based Therapy Topic Modeling

We incorporate an LDA module to capture the underlying treatment themes from tongue image features, assisting in the generation of medications and dosage information for prescriptions. Specifically, the generated tongue image feature vector F is transformed into an initial therapy information $O^{therapy}$ via a fully connected layer, as shown in Equation (15).

$$O^{therapy} = FC_3(F) \tag{15}$$

The set of TCM therapies is represented as $K = [k_1, k_2, ..., k_m]$, where k_i denotes a specific used therapy. The LDA model employs Gibbs sampling to infer the topic assignments. In the process, the probability of assigning therapy k to the i-th medication h_i in prescription P is given by Equation (16):

$$p\left(z_{pi} = k|\vec{z}_{\neg i}\right) \propto \frac{n_{pk} + \alpha}{\sum_{t \in K}\left(n_{pt}\right) + |K| * \alpha} \times \frac{n_{khi} + \beta}{\sum_{j=1}^{N} n_{khj} + |L| * \beta} \tag{16}$$

where α and β are prior parameters of Dirichlet, and z_{pi} is the therapy assigned to medication h_i in prescription P. $\vec{z}_{\neg i}$ denotes the therapy assignments for all medications, except the current h_i. n_{khi} is the count of the medication h_i is assigned to therapy k, and n_{pk} is the count of the medication h_i in prescription P that has been assigned to k treatment. L represents the label of the current sample's prescription, N represents the number of traditional Chinese medicine, and $\propto$ represents a proportional relationship. In this process, the therapy distribution for prescription P is shown in Equation (17):

$$p_{P \to k}\left(z_P = k|x = P\right) = \frac{n_{Pk} + \alpha}{\sum_{j=1}^{N} n_{khj} + |L| * \beta} \tag{17}$$

To align the generated prescription with realistic therapy, an auxiliary LDA loss $Loss_3$ is introduced, defined as the Kullback-Leibler (KL) divergence between the model's therapy distribution and the reference distribution from real prescriptions:

$$Loss_3 = \frac{1}{|K|} * \sum_{k \in K} p(z = k|X, O^{therapy}) \log \frac{p\left(z = k\middle|X, O^{therapy}\right)}{P_{G \to k}\left(z_G = k|x = G\right)} \tag{18}$$

where X represents the auxiliary convolution module parameters, G represents a real prescription, k denotes therapy topic. z_G denotes a certain therapy in the real prescription. And $p\left(z = k\,|X, O^{therapy}\right)$ denotes the probability that z equals a certain therapy k under the given parameter X and all therapy sets $O^{therapy}$.

3. Results

3.1. Dataset

The tongue-prescription dataset used in this study comprised records of the tongue images of patients during clinical diagnosis, patient demographics information and the corresponding prescriptions. Data inclusion criteria were defined as follows: (1) Patients who meet the requirements of TCM tongue diagnosis and undergo the diagnostic process. (2) Patients with clear TCM syndrome types, defined as a definite syndrome diagnosis recorded by the TCM practitioners during the clinical consultation without disagreement in the review process. Cases with ambiguous or controversial syndrome types were excluded. (3) Patients who sign the informed consent form. All tongue images were captured immediately after the face-to-face clinical diagnosis. Each photograph was reviewed by the TCM practitioners to confirm that it represented the visual appearance observed during the consultation. Images of poor quality (e.g., incomplete tongue presentation, blurred regions, poor illumination, or severe occlusion) were excluded from the dataset. Experienced TCM practitioners annotated the characteristics of the tongue image, the corresponding TCM syndrome types and prescriptions for each tongue image. The characteristics of the tongue image include the shape, color (e.g., pale red), texture (e.g., teeth-marked), coat (e.g., thick), and vein (e.g., thin and short) of the tongue to provide explicit semantic features alongside the raw image. In addition, the practitioners manually labeled the lingual and sublingual areas with bounding boxes. To eliminate background noise from the surrounding environment, the lingual and sublingual regions were cropped based on the bounding boxes, and the extracted tongue regions were retained for subsequent model training. Prior to model training, all cropped images were uniformly resized to a fixed resolution. To enhance data diversity and improve model generalization, data augmentation techniques were applied, including random cropping and horizontal flipping. The prescriptions in this dataset were determined by the TCM practitioners based primarily on their interpretation of the tongue presentation and supplemented by their clinical experience, rather than being derived from a full four-diagnostic procedure that includes inquiry, listening, and palpation. A total of 1452 clinical cases were collected between December 2021 and January 2022, with a mean patient age of 28.52 years. For each patient, lingual and sublingual images were collected, resulting in a total of 2904 tongue images.

3.2. Implementation Details

The experimental setup and evaluation metrics are described as follows: tongue image prescription dataset is randomly split into 70% for training, 20% for validation, and the remainder for testing. Adam is used as the optimization strategy. All experiments were conducted on a workstation equipped with an Intel Core i7-13700 CPU and an NVIDIA GeForce RTX 4060 GPU, with 16 GB of RAM. Model performance was evaluated using Precision, Recall, and F1-score as the primary metrics.

3.3. Experiments on Hyperparameters

The hyperparameters of PR-LDA are discussed separately, including learning rate, number of feature extraction layers, weights of different loss (medication information loss weight α_{pre}, dose loss weight β_{pre}, therapy loss weight γ_{pre}). To identify the optimal hyperparameters, experiments were conducted while other parameters held constant.

(1) The experimental results of the learning rate are shown in Table 1. The model achieved optimal results on the test set with a learning rate of 0.003. When the learning rate is set too high, the model fails to converge properly due to unstable updates. And an overly low learning rate results in slow convergence and susceptibility to poor local minima. (2) The experimental results of number of feature extraction layers are shown in Table 2. The results indicate that the 5-layer configuration yields the optimal performance, while models with fewer layers (3 or 4) or an additional layer (6) consistently underperform. The results suggest that the model with fewer than 5 layers (3 or 4) suffered from insufficient capacity and underfitting, failing to capture the data's complexity, while a deeper 6-layer network may lead to overfitting. (3) The recommendation of prescriptions depends on the prediction of medication, dosage, and prescription therapy. In order to verify the influence of different losses in prescription generation, hyperparameter experiments on loss values were conducted. The experimental results are shown in Table 3. It can be seen that when the ratios of α_{pre}, β_{pre}, and γ_{pre} are set to 2, 2, and 1, Precision, Recall, and F1-score reached the optimal indicators of 0.3212, 0.2447, and 0.2777, respectively. Therefore, it can be inferred that in the process of prescription generation, the medication information and dosage information of the prescription play a major role, while the therapy information serves as an auxiliary factor.

Table 1. Experimental results on PR-LDA learning rate.

Learning Rate	Precision	Recall	F1-score
0.001	0.2868(±0.0098)	0.2448(±0.0152)	0.2641(±0.0192)
0.002	0.3058(±0.0157)	0.2435(±0.0198)	0.2711(±0.0124)
0.003	0.3212(±0.0069)	0.2447(±0.0059)	0.2777(±0.0089)
0.004	0.3197(±0.0104)	0.2384(±0.0128)	0.2731(±0.0156)
0.005	0.2415(±0.0159)	0.2313(±0.0197)	0.2363(±0.0135)
0.006	0.1989(±0.0106)	0.2292(±0.0132)	0.2130(±0.0183)

Table 2. Experimental results on PR-LDA feature extraction network layers.

Feature extraction layers	Precision	Recall	F1-score
3	0.2517(±0.0151)	0.1866(±0.0165)	0.2143(±0.0129)
4	0.2957(±0.0109)	0.2279(±0.0122)	0.2574(±0.0122)
5	0.3212(±0.0069)	0.2447(±0.0059)	0.2777(±0.0089)
6	0.3047(±0.0077)	0.2398(±0.0169)	0.2684(±0.0098)

Table 3. Experimental results on PR-LDA loss weight.

α_{pre}	β_{pre}	γ_{pre}	Precision	Recall	F1-score
1	1	1	0.3055(±0.052)	0.2253(±0.0148)	0.2593(±0.0134)
2	1	1	0.3159(±0.0126)	0.2398(±0.0052)	0.2726(±0.0104)
1	2	1	0.3015(±0.0081)	0.2125(±0.0076)	0.2493(±0.0124)
1	1	2	0.2613(±0.0108)	0.2071(±0.0118)	0.2311(±0.0080)
1	2	2	0.2851(±0.0134)	0.2277(±0.0117)	0.2532(±0.0063)
2	1	2	0.3011(±0.0063)	0.2358(±0.0082)	0.2645(±0.0120)
2	2	1	0.3212(±0.0069)	0.2447(±0.0059)	0.2777(±0.0089)

3.4. Comparative Experiment

To verify the effectiveness of PR-LDA, it will be compared with the following methods:

(1) TCMPR [22]: This method constructs a knowledge graph of traditional Chinese medicine symptoms using clinical case data, comprising five types of entities and five types of relations. By integrating meta-path techniques with the knowledge graph, it enhances the understanding of inter-entity relationships. It proposes a symptom word mapping method based on sub-networks, which maps symptom terms to an extended set containing richer contextual information, and uses the embedded representation of this set as the feature representation of the original symptom words. TCMPR builds upon the SSTM approach, constructing patient symptom vectors from original symptom terms and the SSTM mapping, and applies a CNN framework for model training. The final model outputs predicted probabilities for each herb and generates recommended prescriptions.

(2) RPTIA [23]: This approach combines collaborative filtering and content-based filtering. It uses an autoencoder to extract tongue feature vectors from tongue images and herb embeddings from digitally encoded herb representations. A neural network is then applied for collaborative filtering, ultimately outputting recommended prescriptions.

(3) DGSCAM [24]: Based on a seq2seq architecture, this model integrates traditional Chinese medicine knowledge to automatically generate herbal prescriptions. The encoder incorporates a dual-branch information extraction module to guide a bidirectional hybrid encoder in encoding symptom information. Within this module, a cross-concept knowledge source guidance branch aids in extracting potential associations between prescriptions and symptoms. In the decoder, considering the compatibility principles of traditional Chinese medicine, a task-limited knowledge base matching module is introduced. It computes the matching degree between symptoms and a predefined knowledge base to obtain base prescriptions and construct a candidate herb pool. Additionally, a multi-dimensional complementary attention module is proposed to combine attention over the candidate herb pool and the general herb library, thereby improving the quality of generated prescriptions.

The results of comparative experiment are shown in Table 4. All models involved in the comparative experiments were trained and evaluated on this same dataset to ensure a fair comparison. The experimental results demonstrate that PR-LDA achieves superior performance in Precision and F1-score over all comparative methods on the test dataset. This improvement can be attributed to the introduction of the tongue image similarity constraint, which effectively guides the prescription generation process and significantly reduces the inclusion of erroneous herbs. However, PR-LDA achieves a Recall of 0.2447, slightly lower than TCMPR's 0.2692. This is an expected trade-off, as the model employs a conservative thresholding strategy to strictly prevent the generation of toxic herbs, leading to more cautious predictions and an inherent tendency to favor precision over recall.

Table 4. Comparative experimental results.

Model	Precision	Recall	F1-score
TCMPR	0.1989(±0.0127)	0.2692(±0.0152)	0.2288(±0.0093)
RPTIA	0.3121(±0.0077)	0.2251(±0.0126)	0.2616(±0.0065)
DGSCAM	0.3039(±0.0097)	0.2404(±0.0063)	0.2684(±0.0070)
PR-LDA	0.3212(±0.0069)	0.2447(±0.0059)	0.2777(±0.0089)

3.5. Ablation Study

To validate the individual contributions of prescription similarity information and prescription therapy information on prescription generation, we conducted an ablation study on the PR-LDA. The experiment separately examines the impact of incorporating a similarity-based recommendation strategy for existing prescriptions and the effect of adding a therapy information module on the model's performance. The results are summarized in Table 5. As shown in the table, the similarity ranking strategy significantly improves the precision of prescription generation. Furthermore, the inclusion of therapy information also contributes positively to the overall performance, leading to improved quality of prescription recommendation.

Table 5. Ablation experiment results on PR-LDA.

Model	Precision	Recall	F1-score
w/o ranking strategy	0.3107(±0.0094)	0.2355(±0.0070)	0.2679(±0.0093)
w/o LDA	0.2925(±0.0092)	0.2189(±0.0062)	0.2504(±0.0084)
PR-LDA	0.3212(±0.0069)	0.2447(±0.0059)	0.2777(±0.0089)

To investigate the impact of the feature fusion operation between lingual and sublingual representations, we conducted a controlled ablation study comparing several fusion variants. All experiments were performed under identical training configurations, with the sole variation being the strategy used to fuse lingual feature and sublingual feature. The evaluated strategies are as follows: Concatenation, Element-wise Addition, Fixed Weighted Sum (lingual:sublingual=0.6:0.4), Fixed Weighted Sum (lingual:sublingual=0.4:0.6), Bidirectional Cross-Attention Fusion. In the fixed weighted sum variants, the numbers denote the relative weights assigned to the lingual and sublingual features, respectively. The Bidirectional Cross-Attention Fusion applies a lightweight cross-attention mechanism that uses lingual feature as the query and sublingual feature as key and value to compute a sublingual-conditioned representation, and vice versa. The final fused representation is obtained by summation of the two cross-attention representations. The results of this ablation are reported in Table 6.

Element-wise summation, which directly adds the two feature vectors, yields a notable drop in Precision (0.2926) and F1-score (0.2552), indicating that direct addition may discard specific details of lingual and sublingual features. The 0.6:0.4 weighting (favoring lingual features) achieves higher Precision than summation, but shows lower Recall and F1-score, while the inverted 0.4:0.6 weighting leads to a sharp decline across all metrics. This performance drop may be attributed to the higher noise level in the sublingual images. Because the sublingual region is often difficult to fully expose, the sublingual image frequently contains unrelated regions such as teeth and lips. The bidirectional cross-attention mechanism, which dynamically reweights the two feature sets through learned queries, achieves slightly higher performance than the 0.6:0.4 weighted fusion but remains below concatenation. This indicates that the cross-attention operation introduces additional learnable parameters that are difficult to optimize stably on a limited dataset of 1452 cases, potentially leading to overfitting or suboptimal attention distributions. Furthermore, the dense connections between the two convolutional paths in the PR-LDA already enable feature sharing between lingual and sublingual feature vectors. Thus, introducing a cross-attention fusion mechanism brings limited improvement. In summary, the straightforward concatenation strategy proves to be the most effective, as it avoids both the information loss inherent in additive fusion and the overfitting risk of attention-based fusion. This empirical finding justifies maintaining the independence of lingual and sublingual features through concatenation in this scenario.

Table 6. Ablation experiment results on feature fusion strategy.

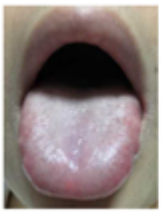
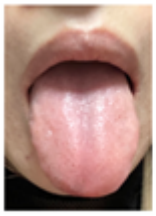
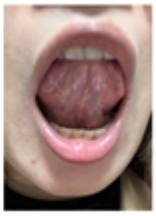
Fusion Strategy	Precision	Recall	F1-score
Concatenation	0.3212(±0.0069)	0.2447(±0.0059)	0.2777(±0.0089)
Summation	0.2926(±0.0053)	0.2264(±0.0083)	0.2552(±0.0061)
Weighted Fusion(0.6:0.4)	0.3082(±0.0065)	0.2115(±0.0046)	0.2508(±0.0079)
Weighted Fusion(0.4:0.6)	0.2612(±0.0132)	0.1947(±0.0104)	0.2230(±0.0116)
Cross-Attention	0.3129(±0.0067)	0.2382(±0.0092)	0.2704(±0.0102)

3.6. Case Study

The PR-LDA model can recommend prescriptions based on the features in tongue body images and combined with prescription datasets. This section shows the prescription recommendation ability of the model through two cases shown in Table 7. In terms of generating the number of traditional Chinese medicines, the proposed model generates fewer medicinal herbs than the classic ground-truth prescriptions, and is more inclined to generate key traditional Chinese medicines; In terms of generating the dosages of traditional Chinese medicines, the dose information of the generated prescriptions falls within the commonly used dose range of this type of medication. For example, in Case 1, although the amount of astragalus predicted was 20g, which was less than that of the traditional Chinese medicine prescription, the dose range was still within the common dose range of the prescription library. In addition, in both Case 1 and Case 2, the generated prescriptions tend to output common Chinese medicines in the prescription library, such as citron fruit and poria, even if these medications do not appear in the traditional Chinese medicine prescription.

Overall, the generated prescriptions have significant similarities in the use of traditional Chinese medicine with the prescriptions provided by Chinese medicine practitioners, indicating that the prescriptions generated by PR-LDA can be helpful in assisting clinicians in prescribing medication.

Table 7. Cases of prescription recommendation based on PR-LDA.

Case	Tongue Images	Prescription Sources	Results
Case 1	 	Ground-truth	Astragalus 30 g, Cornus 10 g, Malt 9 g, Medicated Leaven 9 g, Cinnamon 6g, Hawthorn 9 g, Chinese Angelica 10 g, Pinellia 6 g, Poria 20 g, Prepared Rehmannia 15 g, Wolfberry 10 g, Eucommia 10 g, Dried Tangerine Peel 6 g, Chicken Gizzard Lining 10 g, Atractylodes 10 g
		PR-LDA	Astragalus 20 g, Medicated Leaven 9 g, Cinnamon 6 g, Cornus 12 g, Citron Fruit 9 g, Poria 15 g, Amomum Fruit 9 g, Pinellia 6 g, Chinese Angelica 10 g, Dried Tangerine Peel 6 g, Aucklandia Root 6 g
Case 2	 	Ground-truth	Prepared Licorice 3 g, Hyacinth Bean 15 g, Ginseng 6 g, Sanguisorba Root 6 g, Immature Bitter Orange 6 g, Pinellia 6 g, Patrinia Herb 9 g, Dry Ginger 6 g, Coptis 3 g
		PR-LDA	Ginseng 9 g, Poria 12 g, Prepared Licorice 3 g, Dry Ginger 6 g, Jujube 6 g, Chinese Yam 9 g, Pinellia 6 g, Coptis 3 g

4. Discussion

While the proposed PR-LDA achieves competitive performance, several limitations should be noted. The dataset was constructed by selecting patients with clear TCM syndrome types to ensure definitive prescription labels for supervised learning. This strict inclusion criterion, while beneficial for model training, introduces selection bias. Patients with ambiguous syndromes, which are common in clinical practice, were excluded. Consequently, the reported prediction accuracy may not directly translate to more complex real-world scenarios. Future studies should investigate methods to handle syndrome uncertainty and validate the model on broader, unselected clinical cohorts. Second, the tongue images in this study were collected by multiple TCM practitioners during routine clinical practice without a unified acquisition device or standardized light source, and no dedicated color correction procedure was applied. This lack of standardized hardware and color calibration may introduce inter-image variations in color, brightness, and texture representation, potentially limiting the model's ability to learn subtle, quantitative chromatic features—such as the degree of redness or swelling—in a fully consistent manner. Future work should adopt a calibrated imaging system with a fixed illuminant and color calibration target to improve the objectivity of tongue image analysis.

5. Conclusions

In this paper, we propose a deep learning-based method for prescription recommendation in Traditional Chinese Medicine. While existing models often neglect the correlation between TCM prescription formulation and tongue image features and lack the ability to predict dosage of medications, PR-LDA is proposed to address these limitations. PR-LDA integrates four key components: tongue feature extraction based on multi-path convolutional neural network with dense connections, medication prediction based on tongue image similarity, dosage prediction based on tongue features, and LDA-based therapy topic modeling. By incorporating TCM knowledge from pharmacopoeia, the model ensures the generated prescriptions are more theoretically sound and clinically applicable within the TCM theory. To validate the model's performance, a series of experiments were conducted on a tongue-prescription dataset, including parameter analysis, comparative experiments, ablation study and case study. The results show that PR-LDA achieved Precision, Recall, and F1-scores of 0.3212, 0.2447, and 0.2777, respectively. Furthermore, case studies demonstrate that the predicted prescriptions exhibit strong correspondence with classical formulas in the dataset regarding both medication types and dosages.

However, while the model considers inter-medication relationships, it exhibits a tendency to recommend commonly used medications from the prescription database, primarily because such medications are associated with lower prediction error. Similarly, the generated dosages often align with the most frequent ranges observed in the data. Consequently, future work could involve introducing a reward that encourages the recommendation of less frequent but potentially suitable prescriptions, and more thoroughly integrating patient-specific characteristics to increase the probability of generating such personalized formulas.

Author Contributions

K.C.: software, validation, writing—original draft preparation; Y.Z.: investigation, methodology, conceptualization; L.Z.: data curation, visualization; Y.L.: supervision, writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research was supported in part by the Science & Technology Fundamental Resources Investigation Program under grant 2022FY102002, the Sichuan Science and Technology Program under grants 2023YFS0325, and the Natural Science Foundation of Sichuan under grant 2024NSFSC0717.

Institutional Review Board Statement

This study was approved by the Ethics Committee of the University of Electronic Science and Technology of China (approval code: 106142025031133093).

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

Data are available on request to the correspondence author.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used Deepseek-V4 to polish the language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Wang, X. *Fundamentals of Traditional Chinese Medicine*; Shanghai Scientific & Technical Publishers: Shanghai, China, 1999.
2. Tang, Y.; Sun, Y.; Chiang, J.Y.; et al. Research on Multiple-Instance Learning for Tongue Coating Classification. *IEEE Access* **2021**, *9*, 66361–66370.
3. Zhuo, L.; Zhang, J.; Dong, P.; et al. An SA-GA-BP Neural Network-Based Color Correction Algorithm for TCM Tongue Images. *Neurocomputing* **2014**, *134*, 111–116.
4. Li, J.; Yuan, P.; Hu, X.; et al. A Tongue Features Fusion Approach to Predicting Prediabetes and Diabetes with Machine Learning. *J. Biomed. Inform.* **2021**, *115*, 103693.
5. Li, X.; Zhang, Y.; Cui, Q.; et al. Tooth-Marked Tongue Recognition Using Multiple Instance Learning and CNN Features. *IEEE Trans. Cybern.* **2019**, *49*, 380–387.
6. Pang, B.; Zhang, D.; Wang, K. Tongue Image Analysis for Appendicitis Diagnosis. *Inf. Sci.* **2005**, *175*, 160–176.
7. Li, J.; Zhang, B.; Lu, G.; et al. Body Surface Feature-Based Multi-Modal Learning for Diabetes Mellitus Detection. *Inf. Sci.* **2019**, *472*, 1–14.
8. Xu, J.; Jiang, T.; Liu, S. Research Status and Prospect of Tongue Image Diagnosis Analysis Based on Machine Learning. *Digit. Chin. Med.* **2024**, *7*, 3–12.
9. Liu, Q.; Li, Y.; Yang, P.; et al. A Survey of Artificial Intelligence in Tongue Image for Disease Diagnosis and Syndrome Differentiation. *Digit. Health* **2023**, *9*, 20552076231191044.
10. Cui, Y.; Ni, Y.; Wei, G. Review of Deep Learning in Tongue Image Segmentation. *Comput. Eng. Appl.* **2026**, *62*, 29–46.
11. Jia, G.; Cui, Z.; Fei, Q. QA-TSN: QuickAccurate Tongue Segmentation Net. *Knowl.-Based Syst.* **2025**, *307*, 112648.

12. Chen, Y.; Ho, S.S.; Xu, C.; et al. Dr. Tongue: Sign-Oriented Multi-Label Detection for Remote Tongue Diagnosis. In Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025 ; pp. 2302–2310.
13. Jiang, T.; Guo, X.; Tu, L.; et al. Application of Computer Tongue Image Analysis Technology in the Diagnosis of NAFLD. *Comput. Biol. Med.* **2021**, *135*, 104622.
14. Wang, R.; Chen, J.; Duan, S.; et al. Noninvasive Diagnostic Technique for Nonalcoholic Fatty Liver Disease Based on Features of Tongue Images. *Chin. J. Integr. Med.* **2024**, *30*, 203–211.
15. Deng, J.; Dai, S.; Liu, S.; et al. Clinical Study of Intelligent Tongue Diagnosis and Oral Microbiome for Classifying TCM Syndromes in MASLD. *Chin. Med.* **2025**, *20*, 78.
16. Hu, R.; Deng, J.; Tu, L.; et al. Integration of Tongue Image Features and Tongue Coating Microbiome for Differentiating Dampness Patterns in MASLD. *Front. Endocrinol.* **2026**, *17*, 1851610.
17. Zhang, N.; Jiang, Z.; Li, J.; et al. Multiple Color Representation and Fusion for Diabetes Mellitus Diagnosis Based on Back Tongue Images. *Comput. Biol. Med.* **2023**, *155*, 106652.
18. Zhao, Z.; Qiang, Y.; Yang, F.; et al. Two-Stream Vision Transformer Based Multi-Label Recognition for TCM Prescriptions Construction. *Comput. Biol. Med.* **2024**, *170*, 107920.
19. Zhuang, Y.; Yu, L.; Jiang, N.; et al. TCM-KLLaMA: Intelligent Generation Model for Traditional Chinese Medicine Prescriptions Based on Knowledge Graph and Large Language Model. *Comput. Biol. Med.* **2025**, *189*, 109887.
20. Hu, Y.; Wen, G.; Luo, M.; et al. Fully-Channel Regional Attention Network for Disease-Location Recognition with Tongue Images. *Artif. Intell. Med.* **2021**, *118*, 102110.
21. Pharmacopoeia Commission of People's Republic of China. *Pharmacopoeia of People's Republic of China*, 2025th ed.; China Medical Science Press: Beijing, China, 2025. (In Chinese)
22. Dong, X.; Zheng, Y.; Shu, Z.; et al. TCMPR: TCM Prescription Recommendation Based on Subnetwork Term Mapping and Deep Learning. *BioMed Res. Int.* **2022**, *2022*, 4845726.
23. Wen, G.; Wang, K.; Li, H.; et al. Recommending Prescription via Tongue Image to Assist Clinician. *Multimed. Tools Appl.* **2021**, *80*, 14283–14304.
24. Hou, J.; Song, P.; Zhao, Z.; et al. TCM Prescription Generation via Knowledge Source Guidance Network Combined with Herbal Candidate Mechanism. *Comput. Math. Methods Med.* **2023**, *2023*, 3301605.