Automatic Construction of Chinese Herbal Prescriptions From Tongue Images Using CNNs and Auxiliary Latent Therapy Topics
Yang Hu, Guihua Wen, Huiqiang Liao, Changjun Wang, Dan Dai, and Zhiwen Yu, Senior Member, IEEE

Abstract—The tongue image provides important physical information of humans. It is of great importance for diagnoses and treatments in clinical medicine. Herbal prescriptions are simple, noninvasive, and have low side effects. Thus, they are widely applied in China. Studies on the automatic construction technology of herbal prescriptions based on tongue images have great significance for deep learning to explore the relevance of tongue images for herbal prescriptions, it can be applied to healthcare services in mobile medical systems. In order to adapt to the tongue image in a variety of photographic environments and construct herbal prescriptions, a neural network framework for prescription construction is designed. It includes single/double convolution channels and fully connected layers. Furthermore, it proposes the auxiliary therapy topic loss mechanism to model the therapy of Chinese doctors and alleviate the interference of sparse output labels on the diversity of results. The experiment use the real-world tongue images and the corresponding prescriptions and the results can generate prescriptions that are close to the real samples, which verifies the feasibility of the proposed method for the automatic construction of herbal prescriptions from tongue images. Also, it provides a reference for automatic herbal prescription construction from more physical information.

Index Terms—Convolution channels, neural networks, prescriptions construction, therapy topics, tongue images.

I. INTRODUCTION

CHINA has a vast and increasing population. Making full use of medical resources to meet the needs of the patients is always an urgent concern [1]. Healthcare informatics [2], [3], especially the rapid development of mobile medical services, can help immobile and mobile people obtain cheap and high-quality medical services [4]–[6].

In China, herbal prescriptions contain low side effects and provide a simple, cheap, natural, and noninvasive treatment or health preservation option [7]. The tongue image is a potentially important body sign in clinical diagnosis and treatment. Therefore, there have already been extensive studies on the collection [8], [9], processing [10], [11] and analysis [12], [13] of tongue images, which can assist in diagnosis and treatment. Zhang et al. [14] integrated the color, texture, and geometric features of tongue images to detect diabetes mellitus and nonproliferative diabetic retinopathy. Li et al. [15] combined the color, texture, and geometric features of the tongue image, facial parts, and sublingual information to detect diabetes and early glycemia. However, previous studies usually require a professional tongue image acquisition device, these devices usually do not appear in some community hospitals with poor conditions. The previous studies also lack robustness to the requirements of the camera, shooting angle, illumination, and other conditions, which is not conducive to expand on mobile terminals to convenient for carrying. Therefore, the research on the feature coding of low-quality tongue images and the auxiliary construction of herbal prescriptions has research significance and application value.

Deep learning based on neural networks has been widely applied in medical informatics [16], [17], especially in medical image processing [18]–[21] in recent years. It covers medical image preprocessing [22], [23], medical image classification [24], [25] and other directions and provides great technical support for auxiliary diagnoses and treatments. Tang et al. [26] encoded the multiscale medical image features using sparse autoencoders with different receptive field sizes and generated the feature maps for classification using the convolution operation. Zhang et al. [27] enhanced the multiscale feature ensembles and utilization efficiency and integrated the improved attention mechanism to read the images and generate diagnostic reports.

Also, artificial intelligence plays an important role in medicine recommendations [28], medical simulation [29], and clinical decision support [30]. Hoang et al. [31] used social media as the information source to mine sequences of drugs and adverse effects that signal detrimental prescription cascades. Long and Yuan [34] described an accurate,
computationally efficient, and scalable algorithm to construct the medication history timeline that can inform clinical decision making. Yu et al. pioneered a traditional Chinese medicine (TCM) knowledge graph that integrated terms, databases, and other resources. It provides services, such as visualization, knowledge retrieval, and recommendations; thus, this system has great significance to the studies of medical clinical decision support in China [35].

In this paper, we intend to use convolutional networks to model the relationship between tongue images and Chinese herbal prescriptions. We also introduce multitask structure and amalgamate the main and auxiliary features with features fusion strategy, as in [32] and [33].

Furthermore, we develop an assistant prescription system based on this paper, which is also applied in several basic community hospitals. It provides tips and references for interns or young doctors in prescribing, to improve their efficiency. The final prescriptions is decided by doctors.

To the best of our knowledge, there has not been any work using tongue images to construct herbal prescription. In order to robustly model the tongue images of patients in various photographic environments and accurately excavate the association between tongue information and herbal prescriptions (as shown in Fig. 1), the following main work is carried out in this paper.

1) A neural framework of single/double convolution channels and multiple fully connected layers is applied to encode the tongue features and generate the herbal prescription (without considering the dose).
2) Considering the restrictive effect of extremely sparse output labels on the diversity of generated prescriptions, the auxiliary loss of latent therapy topics is proposed to improve the precision and stimulate diversity in prescription construction.
3) We conduct experiments and verify the convolution network for tongue feature encoding and herbal prescription construction. We compare the effects of single/double channel models and the auxiliary topic loss mechanism for accuracy, completeness, and diversity of prescription construction.

The rest of this paper is organized as follows. In Section II, we detail the materials and methods. Sections III and IV illustrate, analyze, and discuss the results of all prescription construction models. We conclude this paper in Section V.

II. MATERIALS AND METHOD

A. Collection of Tongue Images and Prescriptions

All of the patients’ tongue images are manually taken from the outpatient departments of the cooperating hospital, on the basis of the signed agreements (the number is 9585 persons/times). The shooting tools are digital cameras or smart phones with a more universal image quality.

Handwritten Chinese herbal prescriptions were also taken by volunteers and then entered into the database by professionals. The prescription doses were temporarily masked, and only the names of the herbs were kept.

B. Architecture of Neural Prescription Construction

1) Prescription Construction Task Description: All utilized Chinese herbal medicines were included in a unified dictionary \( H = \{ h_1, h_2, \ldots, h_n \} \). After we extracted the tongue image features, the task of herbal prescription construction can be seen as an ensemble binary classification task. It needs the tongue image feature vector \( f_z \) as an input and outputs the decision vector \( P_z = [p_1, p_2, \ldots, p_n] \), whose dimension \( n \) is equivalent to the size of the herbal medicine dictionary. The value of element \( p_i \) in \( P_z \) can only be “1” or “0,” which indicates whether the corresponding herb \( h_i \) should be included in the prescription construction result.

2) Convolutional Channels and MLP Encoding Layers: The convolution network has been successfully applied to the processing and analyses of medical images [20], [21], [36], [37]. As the inputs of tongue image \( X_i \) are \( (224 \times 224 \times 3) \) matrices with the RGB color channel, the deep convolutional network is used to extract the latent features from the original RGB image matrix, and the multilayer convolutional structure can be defined as a convolutional channel.

The convolutional channel is constructed by several duplicate modules that include convolutional and max-pooling layers. The convolution operation parameterized by the kernel weight matrices \( k \) uses the kernel filters to scan the image matrices and reconstruct the feature maps \( \hat{C} \). The convolution operations in the convolutional channels are abstracted as the following functions with ReLU activation functions:

\[
C(X, k) = \text{ReLU}(\text{Conv}(X, k)).
\]

With the aim of capturing the most representative features and reducing the computational complexity, max-pooling operations are used to down-sample the feature maps \( \hat{C} \), as follows:

\[
\hat{C}(X, k) = \text{Max}(C(X, k)).
\]

Two fully connected layers parameterized by \( W_i \) and \( W_o \) are used to further encode features as follows:

\[
f_z(\hat{C}, [W_i, W_o]) = \text{ReLU} \left( \text{ReLU} \left( \hat{C}, W_i \right), W_o \right).
\]
each herb as follows:

\[ P(X, \theta) = \phi(f_c, W_c) \]

\[ = \left[ p(h_1|X, \theta), p(h_2|X, \theta), \ldots, p(h_n|X, \theta) \right] \quad (4) \]

where \( \theta = \{k, W_i, W_o, W_z\} \) is the set of all parameters.

Then, the loss function is defined as the mean of the binary cross-entropy for every herb with ground truth \( G = [g_1, g_2, \ldots, g_n] \), where \( g_i \in \{0, 1\} \), and the objective minimizes the loss by optimizing all parameters in set \( \theta \), as follows:

\[ J(\theta) = D_b(P(X, \theta), G) \]

\[ = \frac{1}{n} \sum_{i=1}^{n} -g_i \log(p(h_i|X, \theta)) \]

\[ - (1 - g_i) \log(1 - p(h_i|X, \theta)) \quad (5) \]

\[ \theta^* = \{k^*, W_i^*, W_o^*, W_z^*\} = \arg \min_\theta D_b(P(X, \theta), G). \quad (6) \]

In many circumstances, Chinese doctors can directly prescribe herbs based on their rich experience, but sometimes they still need to draw support from the diagnosis of the patient’s condition to guide the direction of the prescription. Based on this consideration, an auxiliary convolutional channel is added to the prescription construction model. It is used to simulate the diagnosis process and provide more information for the construction of herbal prescriptions.

The structure of the auxiliary convolutional channel is similar to the main convolution channel with less convolutional kernels. It outputs the auxiliary feature maps \( \tilde{C}_{aux} \) as follows:

\[ \tilde{C}_{aux}(X, k_{aux}) = \max(C_{aux}(X, k_{aux})) \]

\[ = \max(\text{ReLU(Conv}_{aux}(X, k_{aux}))). \quad (7) \]

Two fully connected layers encode the feature maps from the main and auxiliary convolutional channels and merge them into another fully connected layer. They are parameterized by \( W_i, W_{aux}, \) and \( W_o \), and use the ReLU activation functions as follows:

\[ f_i(\tilde{C}, W_i) = \text{ReLU}(\tilde{C}, W_i) \quad (8) \]

\[ f_{aux}(\tilde{C}_{aux}, W_{aux}) = \text{ReLU}(\tilde{C}_{aux}, W_{aux}) \]

\[ f_c(\tilde{C}, W_c) = \text{ReLU}(\tilde{C}, W_c) \]

where \( \oplus \) is the concatenation operator.

The sigmoid function is still used as the activation function of the output layer, which is the same as (4). The set of parameters is updated as \( \theta = \{k, k_{aux}, W_i, W_{aux}, W_o, W_z\} \). Then, the objective function is as follows:

\[ \theta^* = \{k^*, k_{aux}^*, W_i^*, W_{aux}^*, W_o^*, W_z^*\} = \arg \min_\theta D_b(P(X, \theta), G). \quad (11) \]

The architecture of the neural prescription construction model is shown in Fig. 2. It consists of one or two convolutional channels and the MLP encoding layers. The outputs of the last layer are sampled from the herb dictionary \( H \) through the threshold \( t_h \) to obtain the final prescription.

3) Auxiliary Topic Distribution Ground Truth of Prescription: The “0,” which accounts for the vast majority of output labels and may suppress the confidence of the neural network to produce more diverse and correct herbs. Moreover, besides the real prescriptions ground truth, more auxiliary ground truths are needed to help the network framework model the TCM therapies.

Considering that the prescription does not have obvious context, it can be regarded as a textual model based on a “bag of words.” The latent Dirichlet allocation (LDA) model is a classical topic modeling approach based on a bag of words, and it has been successfully applied in the modeling of prescriptions [38].

Using only the training set to model the LDA prescription topics does not affect the primitiveness and objectivity of the test data. Assume that each topic is regarded as a TCM therapy.

The list of TCM therapies is denoted as \( K = [k_1, k_2, \ldots, k_m] \), and the Dirichlet prior parameters are \( \alpha \) and \( \beta \). The LDA model uses Gibbs sampling [42] to estimate the probability that a TCM therapy \( k \) is assigned to the \( i \)th herb \( h_i \) in a prescription \( P \). This probability in each iteration of Gibbs
of the therapy assignment for prescription \( P \) been assigned to therapy \( k \) is given by

\[
\begin{align*}
  p(z_P = k | \bar{z} - i) &\propto \frac{n_{pk} + \alpha}{\sum_{i \in K} n_{pi} + |K| \cdot \alpha} \\
  &\times \frac{n_{kh_i} + \beta}{\sum_{j=1}^{|H|} n_{kh_j} + |H| \cdot \beta}
\end{align*}
\]

(12)

where \( z_P \) is the therapy assignment of herb \( h_i \) in prescription \( P \); \( \bar{z} - i \) is all herbs’ therapy assignments excluding the current one \( h_i \); \( n_{kh_i} \) is the number of times herb \( h_i \) has been assigned to therapy \( k \); and \( n_{pk} \) is the number of times herbs in \( P \) have been assigned to therapy \( k \). Then, the probability distribution of the therapy assignment for prescription \( P \) is given by

\[
p_{P \rightarrow k}(z_P = k \mid X = P) = \frac{n_{pk} + \alpha}{\sum_{i \in K} n_{pi} + |K| \cdot \alpha}
\]

(13)

The Kullback–Leibler (KL) divergence [43] can be used to measure the gap between the LDA therapy distributions of generated prescriptions and real prescriptions. Also, the therapy distribution divergence between a real prescription and itself is set as 0 as follows:

\[
\hat{D}_{KL}(P, G) = \frac{1}{|K|} \sum_{k \in K} p_{P \rightarrow k}(z_P = k \mid X = P) \log \frac{p_{P \rightarrow k}(z_P = k \mid X = P)}{p_{G \rightarrow k}(z_G = k \mid X = G)}
\]

(14)

\[
\hat{D}_{KL}(G, G) = 0.
\]

(15)

The auxiliary convolutional channel [shown as (7) and (9)] outputs the predicted therapy distribution using a fully connected output layer with weight \( W_{z_{aux}} \) and the softmax activation function as follows:

\[
P_{aux}(X, \theta_{aux}) = \text{Softmax}(f_{aux}, W_{z_{aux}})
\]

(16)

where \( f_{aux} \) is the features encoded by the auxiliary convolutional channel as (7) and (9) and the parameter set for the auxiliary output is denoted as \( \theta_{aux} = \{k_{aux}, W_{aux}, W_{z_{aux}}\} \). The loss function of the auxiliary output is given by

\[
J(\theta_{aux}) = \hat{D}_{KL}(P(X, \theta_{aux}), G) + \lambda \cdot \hat{D}_{KL}(P(X, \theta_{aux}), G)
\]

(17)

Correspondingly, the loss function of the main output is similar to (4) and parameterized by the main parameter set \( \theta_{main} = \{k, k_{aux}, W_i, W_{aux}, W_o, W_z\} \) as follows:

\[
J(\theta_{main}) = D_b(P(X, \theta_{main}), G).
\]

(18)

In this way, the LDA topic distribution ground truths of prescriptions are joined to optimize the weights of the auxiliary convolutional channel. This will help the double-output neural network to better model the doctors’ therapies, as shown in Fig. 3.

The loss function of the double-output network is integrated by the loss of the auxiliary and main outputs, and the coefficient \( \lambda \) is used to regulate the proportion of auxiliary loss. Then, the union set \( \theta = \theta_{main} \cup \theta_{aux} \) in the final objective function is the set of all parameters to be optimized as follows:

\[
J(\theta_{main}, \theta_{aux}) = D_b(P(X, \theta_{main}), G) + \lambda \cdot \hat{D}_{KL}(P(X, \theta_{aux}), G)
\]

(19)

\[
\theta^* = \theta_{main}^* \bigcup \theta_{aux}^*
\]

\[
= \arg \min_{\theta} D_b(P(X, \theta_{main}), G) + \lambda \cdot \hat{D}_{KL}(P(X, \theta_{aux}), G).
\]

(20)

4) Data Augmentation for Tongue Images: In the real-world use cases, data collection of patients’ tongue images is extremely expensive. Therefore, it is necessary to make the most out of the limited data.
Fig. 4. Deeper prescription construction structures with pretraining on ImageNet. (a) Deep convolutional network with 1 pipeline, which is trained from scratch. (b) Deep convolutional network with 1 pipeline and is pretrained with ImageNet, in which, the “Block” refers to the tiny block, it only contains half number of filters each layer than layers in regular block. (c) Deep convolutional network with double pipeline and is pretrained with ImageNet, in which, the auxiliary pipeline also outputs the auxiliary therapy topics.

| Parameters           | Explanations                                                                 | Values |
|----------------------|-----------------------------------------------------------------------------|--------|
| rotation_range       | A value in degrees (0-180) within which pictures are randomly rotated        | 25     |
| width_shift_range    | Ranges (as a fraction of total width or height) within which to randomly translate pictures vertically or horizontally | 0.05   |
| height_shift_range   |                                                                                     | 0.05   |
| shear_range          | A value for randomly applying shearing transformations                        | 0.05   |
| zoom_range           | A value for randomly zooming inside pictures                                  | 0.2    |
| horizontal_flip      | A flag for randomly flipping half of the images horizontally                 | True   |
| fill_mode            | The strategy used for filling in newly created pixels                       | “nearest” |
| others               | (default)                                                                   |        |

The image data generator is used for the data augmentation of tongue images, which can randomly copy and transform the original tongue images. In each iteration, the image generator randomly selects some images from the original images. It then randomly transforms and returns them to the training set. The augmented images keep their original labels. In order to concurrently ensure image quality and data diversity, some restrictive parameters for the image augmentation are set, as shown in Table I.

5) Weights Transfer With Pretraining on ImageNet: Moreover, we plan to use the deeper networks as the backbones, to further improve the performance of auto prescription construction. In addition, in order to alleviate the over-fitting caused by nonmassive training data, we pretrain the deeper prescriptions construction models on ImageNet2012 [49] dataset and transfer the weights to training on tongue images.

As shown in Fig. 4, which give the architectures of auto prescription construction models, with deeper backbones, such as VGG16, VGG19 [39], and ResNet50 [41]. We also applied the auxiliary channel design for therapy topics on these deeper networks, implement it with another auxiliary pipeline, which is same structured with the main pipeline and only has half number of filters. All of these settings strictly follow the original design on shallow networks.

So, the main objects of comparison are still the following: 1 pipeline models; 2 pipelines models; and the models with 2 pipelines and auxiliary therapy topics outputs.

In the step of pretraining, all 2 pipelines models are trained as a whole with the shared inputs of ImageNet dataset and without the auxiliary output branch.

### III. RESULTS

#### A. Datasets and Preprocessing

The real-world “tongues-prescriptions” experiment dataset that was collected by volunteers contains 9585 tongue images and their corresponding prescriptions. Since the tongue images and prescriptions are very precious, pseudo fivefold cross-validation is used in the experiment. More specifically, 500 samples are randomly selected each time as the test set, the others construct the training set (10% of them are set as the validation set), and the 5 selected test sets do not contain any of the same samples with each other.

The training set can be augmented by the image generator described in Section II-B4. In each iteration, the image generator randomly transforms 64 original images and returns them to the training set. By repeating this operation 200 times, we get 12800 augmented tongue images.

Over all the prescriptions, there are 566 types of herbs, and the training sets and test sets share one vocabulary for all herbs.

The preprocessing for the tongue images and corresponding prescriptions includes the following two steps.

1) Enable the models to process the directly captured tongue images and only perform slight preprocessing for some low-quality tongue images, including tongue positioning and cutting of nontongue regions.

2) Normalization of the names of herbs. Each herb retains only one unique appellation, and the others are removed.
Additional information about the datasets is listed in Table II.

### B. Experimental Setup

1) **Experimental Workflow:** To the best of our knowledge, there has not been any other work that strictly explores the same task as this paper. The baseline method is set as directly using the random forest multilabel image classifier on herbs.

The experiment will also examine the effectiveness of the prescription construction ability of bi-convolutional channels, auxiliary therapy topics and image augmentation. Brief introduction of the comparison methods are as follows.

- **RF-Baseline:** Implement the multilabel classification on herbs and tongue images using the multilabel supported random forest classifier [44].

- **1-CNNs + MLP:** The neural prescription construction model contains only one convolutional channel and fully connected encoding layers. Also, the sigmoid function is used in the output layer.

- **2-CNNs + MLP:** It is like the previous model, and the only difference is that it contains two convolutional channels.

- **2-CNNs + MLP + AUX_LDA:** Based on the model “2-CNNs + MLP,” the therapies are modeled by introducing the auxiliary ground truths of the prescription topics, and the auxiliary convolution channel is optimized.

- **1-CNNs + MLP With Data Augmentation:** It is like the model “1-CNNs + MLP,” but it uses the augmented dataset.

- **2-CNNs + MLP With Data Augmentation:** It is like the model 2-CNNs + MLP, but it uses the augmented dataset.

- **2-CNNs + MLP + AUX_LDA With Data Augmentation:** It is like the model “2-CNNs + MLP + AUX_LDA,” but it uses the augmented dataset.

2) **Parameter Setting of Neural Networks:** In order to take into account the performance of the model and the consumption of computing resources, the parameters of the neural prescription construction models are set by the experience of many experimental iterations. The main parameters are listed in Table III.

3) **Evaluation Metrics:** To measure the similarity between the generated prescription and the real prescription, the metrics precision similarity $p_{\text{sim}}$, recall similarity $r_{\text{sim}}$, and intersection-over-union rate $\text{IoU}\_\text{sim}$ are set as follows:

\[
p_{\text{sim}} = \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{nc_i}{np_i} , \quad r_{\text{sim}} = \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{nc_i}{ng_i} , \quad \text{IoU}\_\text{sim} = \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{nc_i}{np_i + ng_i - nc_i}
\]

where $nc_i$ is the number of identical herbs in the generated prescription and the real prescription of the $i$th sample; $np_i$ and $ng_i$ are the respective numbers of herbs in the generated prescription and the real prescription of $i$th sample; and $N_t$ is the number of test samples. In addition, $\text{IoU}\_\text{sim}$ cannot be directly deduced from $p_{\text{sim}}$ and $r_{\text{sim}}$ at the macro level.

Sometimes, the above three metrics cannot perfectly evaluate the performance of prescription construction. Some additional metrics are taken into account, which are used to indicate the completeness of the prescription and the exactitude of the therapies. They are as follows.

- $nb_p$: The average number of herbs that appear in the generated prescriptions.

- $nb_c$: The average number of correct herbs that appear in the generated prescriptions.

- $nb_d$: The absolute value of the average quantity difference of herbs that appear in the generated prescriptions and the real prescriptions.

- $kl_t$: The KL divergence of topic distributions from the generated prescription to the real prescription.
C. Model Evaluation

1) Training Process and Therapy Topics Distributions:
Figs. 5 and 6 show the trend of validation losses with regard to the training process of the neural prescription construction models. From Fig. 5, we see that the variations and the minimum values of the validation losses for the different models are very close. However, the training process with data augmentation can obtain relatively lower validation loss for each model, and the 2-CNNs + MLP + AUX_LDA model obtains the lowest validation loss regardless of whether data augmentation is used. In particular, the optimal performance of the negative log-likelihood for the 2-CNNs + MLP + AUX_LDA model without data augmentation is close to the 1-CNNs + MLP model after using data augmentation.

The green lines in Figs. 5(c) and 6(c) show a fairly similar trend to the validation loss, which indicates the integration of the negative log-likelihood of prescriptions and the auxiliary loss of therapy topics. Therefore, the learning objective of the auxiliary output is consistent with the generation of the prescription, which can promote the optimization of the prescription construction models.

The blue shadows in Figs. 5 and 6 show that the training of the model will converge in the 20th to 30th round. The wider shadow reveals the greater fluctuation of the convergence time in the fivefold experiments, and the convergence time of the training with data augmentation is generally more stable. Moreover, in training processes without data augmentation, the 2-CNNs + MLP + AUX_LDA model is most stable. On another note, the confidence field of each model reveals that the training of the 2-CNNs + MLP + AUX_LDA model has smaller undulations in all cases.

Fig. 11, in the supplementary material, shows the complete topic distributions that are (all compressed to 2-dim by online dictionary learning function [45] in the fivefold experiments for all neural prescription construction models. For a more convenient observation, Fig. 7 illustrates the comparative examples of the therapy topic distributions from the 1-CNNs + MLP without data augmentation and 2-CNNs + MLP + AUX_LDA with data augmentation models. The figure clearly exhibits the differences in therapy distributions between an ordinary result and an optimal result.

The therapy topic distribution of the real samples shows some concentration. This means that there are some frequently used herbs in the real prescriptions, and the prescriptions composed of some seldom-used herbs are sparsely distributed in the marginal area. For the 1-CNNs + MLP without data augmentation model, the topic distribution of its generated samples partially fit the real samples, but they are more concentrated, which reveals that these generated prescriptions are more frequent for the use of common herbs. Moreover, a part of these results is gathered in the marginal regions that are not fit for the real samples, and most of the outlier real samples are not fitted. In contrast, for the 2-CNNs + MLP + AUX_LDA with data augmentation model, although the topic distribution of the generated samples is still a little more concentrated than the real samples, it appears quite homogeneous and well fitting, regardless of whether it is in the concentrated area of the real samples or the slightly sparse area. Even in
Fig. 7. Comparative examples of the topic distributions (all compressed to 2-dim) of various models’ prescription construction results. (a) 1-CNNs + MLP without data augmentation and (b) 2-CNNs + MLP + AUX_LDA with data augmentation. The red and green dots represent the generated and real samples, respectively.

As the supplement to Table IV, Table V shows that the performances of all neural prescription construction models are better than the RF-baseline, which shows the result of the collapse of the metric $\text{IoU}_{\text{sim}}$. When data augmentation is not used, the performance of the 1-CNNs + MLP, 2-CNNs + MLP, and 2-CNNs + MLP + AUX_LDA models is increasing. The improvement of the 2-CNNs + MLP + AUX_LDA model on the $p_{\text{sim}}$ and $r_{\text{sim}}$ metrics is more obvious, and its performance is even better than the 1-CNNs + MLP model with data augmentation. When data augmentation is used, the performances of various models is generally enhanced due to the increased training samples. Of note, the AUX_LDA mechanism has a significant effect on the model’s performance improvement and the 2-CNNs + MLP + AUX_LDA with data augmentation model achieves the best performance on the main metrics.

In some conditions, part of the main metrics cannot well reflect the model’s prescription construction ability. For example, too few and coincidental correct herbs in the generated prescription will lead to the false high $p_{\text{sim}}$, and too many herbs in a generated prescription may lead to the false high $r_{\text{sim}}$.

As the supplement to Table IV, Table V shows that the average size $nb_p$ of the generated prescriptions is less than the real prescriptions. The majority of the “0”s in the output labels inhibit the confidence of the prescription construction model, and they prefer to prescribe less herbs rather than take the risk of prescribing the wrong herb. By contrast, the size of the prescription produced by the 2-CNNs + MLP + AUX_LDA with the data augmentation model is the closest to the Real. Furthermore, the AUX_LDA mechanism that encourages the models to prescribe more herbs can also effectively improve the number of correct herbs $nb_c$. As described in Section III-C1 with the LDA topic distributions of the prescriptions, the 2-CNNs + MLP + AUX_LDA model can also acquire the better performance on metric $kl_t$ regardless of the use of data augmentation.

Tables VIII and IX more vividly illustrate the prescription construction results of various models in the experiment. It can be seen that the 2-CNNs + MLP model can more accurately prescribe herbs than the 1-CNNs + MLP model. Furthermore, the 2-CNNs + MLP + AUX_LDA model can accurately prescribe less-common herbs, even if a few wrong herbs are occasionally prescribed, such as “Sculellaria barbata” and “Semen zizyphi spinosae” (in the sample of Table VIII and the sample of Table IX). Interestingly, some of the wrong herbs prescribed by the 2-CNNs + MLP + AUX_LDA model have similar benefits as some missing correct herbs, like “Herba Hedyotis” and “Sculellaria barbata” (in the sample of Table VIII).

3) Artificial Evaluation: We also introduce the artificial evaluation for auto-generated prescriptions, to check the Chinese pharmacological logic of them. We invited more than 10 doctors from Department of TCM in Guangdong General Hospital, to conduct artificial evaluation for auto-generated prescriptions of 500 samples. The main content of the artificial evaluation is checking the rationality of the compatibility logic of the prescription, it mainly includes two factors: 1) whether the herbs in same prescription constitute some common-pairs and 2) whether there is any conflict in same prescription. As shown in Table VI, the green boxes mark the common-pairs of herbs, which indicate that these herbs are commonly used together to play a mutually beneficial role, and the red frames mark the conflict-pairs, which mean that these herbs are not suitable for use together.

Although the logical evaluation of the prescription is solely based on the subjective judgement of the doctors, it is basically based on the following rules:

$$S_l(p_i) = S_{\text{pos}}(p_i) + S_{\text{neg}}(p_i)$$

where $S_l(\cdot)$ refers to the total score of logic evaluation of a prescription; $S_{\text{pos}}(\cdot)$ and $S_{\text{neg}}(\cdot)$ are the scores of common herb
TABLE IV
COMPARISON RESULTS FOR THE PRESCRIPTION CONSTRUCTIONS OF VARIOUS MODELS WITH THE MAIN METRICS. HIGHER VALUES OF THESE METRICS INDICATE A BETTER PERFORMANCE OF PRESCRIPTION CONSTRUCTION

| Model                                      | \( p_{sim} \) (%) | \( r_{sim} \) (%) | \( IoU_{sim} \) (%) |
|--------------------------------------------|-------------------|-------------------|---------------------|
| RF-baseline                                | 39.83 ± 1.17      | 28.96 ± 0.96      | 21.89 ± 1.27        |
| 1-CNNNs + MLP                             | 43.26 ± 1.06      | 32.89 ± 1.18      | 28.14 ± 1.06        |
| 2-CNNNs + MLP                             | 44.39 ± 1.35      | 33.55 ± 1.30      | 28.61 ± 1.35        |
| 2-CNNNs + MLP + AUX_LDA                   | 45.28 ± 0.64      | 35.62 ± 1.36      | 29.72 ± 0.64        |
| 1-CNNNs + MLP with data augmentation      | 45.27 ± 0.63      | 34.55 ± 1.42      | 28.97 ± 0.63        |
| 2-CNNNs + MLP with data augmentation      | 45.52 ± 0.97      | 35.84 ± 1.13      | 30.62 ± 0.97        |
| 2-CNNNs + MLP + AUX_LDA with data augmentation | 46.77 ± 1.03      | 37.17 ± 0.88      | 32.23 ± 1.03        |

TABLE V
COMPARISON RESULTS FOR THE PRESCRIPTION CONSTRUCTIONS OF VARIOUS MODELS WITH ADDITIONAL METRICS. THE CLOSER TO THE REAL VALUE OF \( nb_P \), THE HIGHER IS THE VALUE OF \( nb_C \), AND LOWER VALUES OF \( nb_D \) AND \( kl_t \) INDICATE BETTER PRESCRIPTION CONSTRUCTION PERFORMANCE

| Model                                      | \( nb_P \) | \( nb_C \) | \( nb_D \) | \( kl_t \) |
|--------------------------------------------|------------|------------|------------|------------|
| Real                                       | 13.85      | -          | -          | -          |
| RF-baseline                                | 7.18       | 2.76       | 8.67       | 3.41       |
| 1-CNNNs + MLP                             | 9.53       | 4.28       | 4.94       | 2.79       |
| 2-CNNNs + MLP                             | 9.64       | 4.43       | 4.81       | 2.74       |
| 2-CNNNs + MLP + AUX_LDA                   | 10.61      | 4.99       | 4.04       | 2.59       |
| 1-CNNNs + MLP with data augmentation      | 10.30      | 4.57       | 3.95       | 2.61       |
| 2-CNNNs + MLP with data augmentation      | 11.14      | 5.12       | 3.11       | 2.54       |
| 2-CNNNs + MLP + AUX_LDA with data augmentation | 11.55      | 5.53       | 2.68       | 2.39       |

TABLE VI
EXAMPLES OF LOGIC PAIRING IN HERBAL PRESCRIPTIONS. THE RED MARKS REFER TO THE CONFLICT-PAIRS AND THE GREEN MARK INDICATE THE COMMON COOPERATIVE PAIRS

| Generated prescription                      | Common-pairs                | Conflict-pairs                | Common-pairs                |
|--------------------------------------------|-----------------------------|-------------------------------|-----------------------------|
| 白芷 白芍 白术 甘草 黄芩 | 白芷 白芍 白术 甘草 黄芩 | 白芷 白芍 白术 甘草 黄芩 | 白芷 白芍 白术 甘草 黄芩 |
| (Bupleurum, Radix paeoniae alba) Cyperus Rotundus, Fructus aurantii, Codonopsis pilosula, Rhizome of lagenhead atractylodes, Cuttle-bone, Pinellia Rhizoma Prepared by Ginger, Caulis perillae acutae, Folia perillae acutae) |
| 甘草 甘草 甘草 甘草 甘草 | 甘草 甘草 甘草 甘草 甘草 | 甘草 甘草 甘草 甘草 甘草 | 甘草 甘草 甘草 甘草 甘草 |
| (Cassia twig Liquorice, Codonopsis pilosula, Radix Pseudostellariae, Astragalus membranaceus, Monekshoad Mother of pearl) |

As listed in Table VII, the artificial evaluation scores of the generated prescriptions are still less than that of the corresponding real prescriptions, the main difference is that the generated prescription seldom contains the common herb pair. Then, we can also notice that the mechanism of AUX_LDA can alleviate this problem quite well, the model of 2-CNNs + MLP + AUX_LDA can generate more common-pairs of herbs while avoiding taboo conflicts, thus, the models of 2-CNNs + MLP + AUX_LDA have a relatively higher artificial evaluation scores than other models. Fig. 8 also shows the artificial evaluation scores (herbs logic check scores) of all compared models, for 500 samples. Further, we use linear discriminant analysis (Linear-DA) to reduce the dimension of artificial evaluation score of different models, then metric the distance of each tested model from label prescription. Actually, we want the artificial evaluation score of generated prescriptions can be as high as possible. However, due to the score of the generated prescriptions are commonly lower than that of the real prescriptions, the smaller distance from the real prescriptions indicates the better result. As listed in last column of Table VII, the models using the mechanism of AUX_LDA get the smallest distance from real prescription.

Moreover, we also analyze the score distributions of artificial evaluation, as shown in Fig. 9. The artificial evaluation score distributions is roughly obeying the Gaussian distribution. Where the blue part indicates the label prescriptions from veteran doctors and the orange part refers to the generated prescription from proposed models, axis X is the line of score and axis Y means the probability on corresponding score. We can notice that the distribution of the label prescription score is usually more than that of the generated prescription score in higher score area. Also, we can see the score distribution of...
| Label | Herb pair | Taboo avoidance | Total score | Linear-DA distance |
|-------|-----------|----------------|-------------|--------------------|
| 1-CNNs + MLP | 0.792 | 5.866 | 6.658 | — |
| 2-CNNs + MLP | 0.060 | 3.086 | 3.146 | 4.248 |
| 2-CNNs + MLP + AUX_LDA | 0.052 | 3.296 | 3.348 | 3.358 |
| 1-CNNs + MLP with data augmentation | 0.616 | 4.770 | 5.386 | 1.831 |
| 2-CNNs + MLP with data augmentation | 0.064 | 3.378 | 3.442 | 3.157 |
| 2-CNNs + MLP + AUX_LDA with data augmentation | 0.076 | 3.640 | 3.716 | 3.664 |

4) Results of Deeper Models With Weights Transfer: The generated prescription is most similar with that of the label prescription when using the mechanism of AUX_LDA.

4) Results of Deeper Models With Weights Transfer: Table X lists the results of deeper prescription construction models with pretraining on ImageNet2012 [49], the upper half part indicates the results without data augmentation, and the lower part indicates the results with data augmentation on tongue image datasets. From this table, we can notice that the transferred weights from ImageNet do provide impressive performance improvements, for deeper models in prescription construction, the overfitting phenomenon has also been alleviated to some extent. In addition, with pretrained weights, the deep models can converge faster (within nearly 30 epochs). Last but not least, the design of auxiliary therapy topics framework can further improve the performance of prescription generation, than typical double pipeline deeper models, which means the proposed auxiliary therapy topics module is also effective, on deeper backbones.

IV. DISCUSSION

Tongue images provide important bodily information and can serve as an important basis for clinical diagnosis and treatment. In this paper, neural network models are used to model the relationship between tongue images and Chinese herbal prescriptions. Then, we automatically construct the herbal prescriptions. After that we verified its validity through experiments. The performance of the proposed method is outstanding in a very small number of similar studies [48]. However, the tongue image is not the only basis for a doctor to make a diagnosis and treatment. It is reasonable to believe that integrating more body sign information can further improve the performance of automatic prescription construction, and this paper provides a reference for tongue images.

The model of the double convolution channel can improve the precision of prescription construction, and the auxiliary therapy topic output mechanism can further improve the precision and diversity of the generated prescription. In addition, we fuse the features of double channels before output, in order to retain the original information of the two channels as more as possible and avoid unnecessary loss of features, we choose the strategy of “concatenation” to merge the data flows.

We further validate the effectiveness of the proposed design of auxiliary therapy topic, on deeper backbones, like VGG [39] and residual network [41].
The auxiliary therapy topic output mechanism is sensitive to a few hyper-parameters, and we can determine the optimal selection of these parameters by several enumerated experiments.

Although the experiment in this paper used a dataset of Chinese patients’ tongue images and Chinese herbal prescriptions, the proposed model can be easily extended to other datasets and application scenarios.

For two reasons, we do not take the dose into consideration in this paper.

1) The current study is very basic, there is still much room for improving the performance of herbs recommendation.

2) The application set up of this paper is to provide herbal tips for the rookie doctors, so, using or not using of one herb seems more important than its dosage. Although
TABLE IX
PRESCRIPTION CONSTRUCTION EXAMPLES OF VARIOUS MODELS WITH DATA AUGMENTATION. THE GREEN WORDS REPRESENT THE CORRECT HERBS (ALSO APPEAR IN THE CORRESPONDING REAL PRESCRIPTION) IN GENERATED PRESCRIPTION, WHILE THE RED WORDS AND BLACK WORDS IN BOXES REPRESENT THE WRONG HERBS (DO NOT APPEAR IN THE CORRESPONDING REAL PRESCRIPTION) AND THE MISSING HERBS (APPEAR IN THE CORRESPONDING REAL PRESCRIPTION BUT NOT APPEAR IN THE GENERATED PRESCRIPTION)

| Prescription | 1-CNNs + MLP with data augmentation | 2-CNNs + MLP with data augmentation | 2-CNNs + MLP + AUX_LDA with data augmentation |
|--------------|-------------------------------------|-------------------------------------|-----------------------------------------------|
| Real         | 菊花 枸杞 白术 黄连 山药 天麻 蝉甲 甘草 玉米 白花蛇舌草 梨花 石斛 紫花前胡 天山雪莲 天山雪莲 平枝蓬 红豆杉 | 菊花 枸杞 白术 黄连 山药 天麻 蝉甲 甘草 玉米 白花蛇舌草 梨花 石斛 紫花前胡 天山雪莲 平枝蓬 红豆杉 | 菊花 枸杞 白术 黄连 山药 天麻 蝉甲 甘草 玉米 白花蛇舌草 梨花 石斛 紫花前胡 天山雪莲 平枝蓬 红豆杉 |
| 1-Pipeline VGG16 | 44.28 ± 1.08 | 53.92 ± 1.22 | 29.01 ± 0.85 |
| 2-Pipelines VGG16 | 44.78 ± 1.09 | 53.52 ± 1.22 | 29.01 ± 0.85 |
| 2-Pipelines VGG16 + AUX_LDA | 46.02 ± 0.97 | 53.99 ± 1.01 | 30.56 ± 0.86 |
| 1-Pipeline VGG19 | 44.37 ± 1.05 | 33.60 ± 1.07 | 29.57 ± 0.93 |
| 2-Pipelines VGG19 | 45.38 ± 1.01 | 33.45 ± 1.02 | 29.76 ± 0.78 |
| 2-Pipelines VGG19 + AUX_LDA | 45.96 ± 0.80 | 36.16 ± 0.65 | 30.74 ± 0.86 |

TABLE X
COMPARISON PRESCRIPTION CONSTRUCTIONS RESULTS FOR THE PROPOSED MODELS USING DEEPER BACKBONES WITH THE MAIN METRICS. HIGHER VALUES OF THESE METRICS INDICATE A BETTER PERFORMANCE OF PRESCRIPTION CONSTRUCTION

| Description | p_sim (%) | r_sim (%) | IoU_sim (%) |
|-------------|-----------|-----------|-------------|
| 1-Pipeline VGG16 | 44.28 ± 1.08 | 53.92 ± 1.22 | 29.01 ± 0.85 |
| 2-Pipelines VGG16 | 44.78 ± 1.09 | 53.52 ± 1.22 | 29.01 ± 0.85 |
| 2-Pipelines VGG16 + AUX_LDA | 46.02 ± 0.97 | 53.99 ± 1.01 | 30.56 ± 0.86 |
| 1-Pipeline VGG19 | 44.37 ± 1.05 | 33.60 ± 1.07 | 29.57 ± 0.93 |
| 2-Pipelines VGG19 | 45.38 ± 1.01 | 33.45 ± 1.02 | 29.76 ± 0.78 |
| 2-Pipelines VGG19 + AUX_LDA | 45.96 ± 0.80 | 36.16 ± 0.65 | 30.74 ± 0.86 |

V. CONCLUSION
This paper attempts to apply neural network models to fulfill the automatic construction of Chinese herbal prescriptions. To improve the accuracy and diversity of prescription construction, the double convolutional channel frame is applied and the auxiliary therapy topic output mechanism is proposed. Then, the abilities of the proposed models are verified in experiments on the automatic construction of Chinese herbal prescriptions.

In future work, we plan to do the following.

1) Collect more data (not only for tongue images) and then verify and improve the performance of the proposed models on larger and richer datasets.
2) Absorb more medicine transcendental knowledge or ontology [35], [47] to guide the neural prescription construction models.
3) Invite experienced Chinese doctors to provide a more professional artificial score for the generated prescription, which would introduce the online reinforcement learning mechanism into the basic system.

ACKNOWLEDGMENT
The authors would like to thank the anonymous reviewers for their constructive advice. They are also grateful for the hard works of the data collection volunteers.
Yang Hu received the M.A.Eng. degree from the Kunming University of Science and Technology, Kunming, China, in 2016. He is currently pursuing the Ph.D. degree with the South China University of Technology, Guangzhou, China. His current research interests include neural network and biomedical information processing.

Guihua Wen received the Ph.D. degree in computer science and engineering from the South China University of Technology, Guangzhou, China. He is currently a Professor and a Doctoral Supervisor with the School of Computer Science and Technology, South China University of Technology, where he is also the Professor in Chief with Data Mining and Machine Learning Laboratory and the Director of the Artificial Intelligence Chinese Medicine Engineering Research Institute. His current research interests include cognitive affective computing, machine learning, and data mining.

Huiqiang Liao is currently pursuing the master’s degree with the College of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His current research interests include image processing and deep learning.

Changjun Wang received the Ph.D. degree from the Shanghai University of Traditional Chinese Medicine, Shanghai, China. He is currently a Professor and a Doctoral Supervisor with the School of Medicine of South China University of Technology. He is also the Director with the Department of Traditional Chinese Medicine, Guangdong General Hospital, Guangzhou, China.

Dan Dai received the M.A.Eng. degree from the Kunming University of Science and Technology, Kunming, China, in 2016. She is currently pursuing the Ph.D. degree with the South China University of Technology, Guangzhou, China. Her current research interests include machine learning and biomedical information processing.

Zhiwen Yu (S’06–M’08–SM’14) received the Ph.D. degree from the City University of Hong Kong, Hong Kong, in 2008. He is a Professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China, and an Adjunct Professor with Sun Yat-sen University, Guangzhou. He has published over 100 referred journal papers and international conference papers, including the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, the IEEE TRANSACTIONS ON NANOBIOSCIENCE, Information Sciences, Pattern Recognition, Bioinformatics, and SIGKDD. His current research interests include data mining, machine learning, bioinformatics, and pattern recognition. Dr. Yu is a Senior Member of ACM, IRSS, China Computer Federation, and Chinese Association for Artificial Intelligence.