Multimedia Tools and Applications (2019) 78:35665–35688
https://doi.org/10.1007/s11042-019-08118-7

Convolutional herbal prescription building method from multi-scale facial features

Huiqiang Liao[1] · **Guihua Wen**[1] · **Yang Hu**[1] · **ChangJun Wang**[2]

Received: 30 November 2018 / Revised: 10 June 2019 / Accepted: 13 August 2019 /
Published online: 14 October 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

In Traditional Chinese Medicine (TCM), facial features are important basis for diagnosis and treatment. A doctor of TCM can prescribe according to a patient's physical indicators such as face, tongue, voice, symptoms, pulse. Previous works analyze and generate prescription according to symptoms. However, research work to mine the association between facial features and prescriptions has not been found for the time being. In this work, we try to use deep learning methods to mine the relationship between the patient's face and herbal prescriptions (TCM prescriptions), and propose to construct convolutional neural networks that generate TCM prescriptions according to the patient's face image. It is a novel and challenging job. In order to mine features from different granularities of faces, we design a multi-scale convolutional neural network based on three-grained face, which mines the patient's face information from the organs, local regions, and the entire face. Our experiments show that convolutional neural networks can learn relevant information from face to prescribe, and the multi-scale convolutional neural networks based on three-grained face perform better.

Keywords Convolutional neural networks · Face · Prescription · Traditional Chinese medicine

---

✉ Guihua Wen
crghwen@scut.edu.cn

Huiqiang Liao
huiqiangliao@163.com

Yang Hu
superhy199148@hotmail.com

ChangJun Wang
gzwchj@126.com

[1] School of Computer Science and Engineering in South China University of Technology, Guangzhou, China

[2] Department of Traditional Chinese Medicine in Guangdong General Hospital, Guangzhou, China

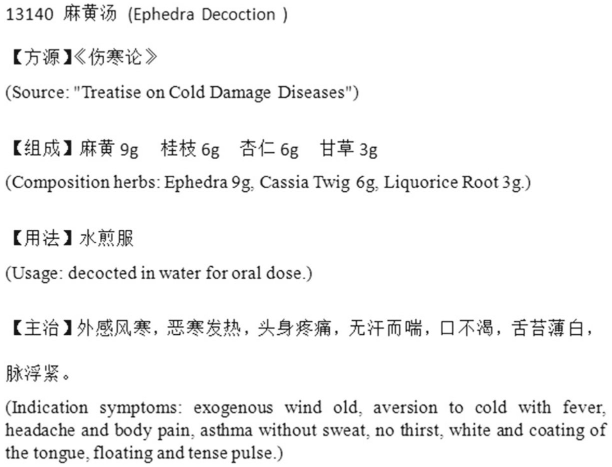
1 Introduction

TCM (Traditional Chinese Medicine) was developed through thousands of years of empirical testing and refinement, and played an important role in health maintenance for the Chinese ancient people [7]. It is a theoretical system that is gradually formed and developed through long-term medical practice. TCM has the advantages of convenience, cheap and low side effects, and is suitable for use in hospitals, even in community hospitals with poor conditions.

Prescription in TCM consists of a variety of herbs, which is the main way to treat diseases for thousands of years. In the long Chinese history, a lot of prescriptions have been invented to treat diseases and more than 100,000 have been recorded [24]. An example of a prescription in Dictionary of Traditional Chinese Medicine Prescriptions is given in Fig. 1 [22, 38].

There are four important diagnostic methods in TCM: Observing, Listening, Inquiry, Pulse feeling. Observing understands the state of health or disease through objective observation of all visible signs and effluents of the whole body and part of the body. Face diagnosis is a common method of observing, which can understand the pathological state of various organs in the body by observing changes in facial features [39]. Face appearance signals information about an individual [16]. The face is rich with capillaries, which is like a mirror that reflects the physiological pathology of humans. From the view of TCM, the characteristics of the various regions of the face represent the health status of various internal organs of the human body. The doctor can judge the physical condition of the patient by observing the facial features of the patient.

Computer aided diagnosis (CAD) based on artificial intelligence (AI) is an extremely important research field in intelligent healthcare [6]. According to a survey, deep learning algorithms, especially convolutional neural networks(CNN), have been widely used in various fields of medical image processing in recent years due to their excellent performances in the field of computer vision, such as disease classification, lesion detection, and substructure segmentation [20]. From the 306 papers reviewed in this survey, it is evident that



13140 麻黄汤 (Ephedra Decoction)

【方源】《伤寒论》

(Source: "Treatise on Cold Damage Diseases")

【组成】麻黄 9g　桂枝 6g　杏仁 6g　甘草 3g

(Composition herbs: Ephedra 9g, Cassia Twig 6g, Liquorice Root 3g.)

【用法】水煎服

(Usage: decocted in water for oral dose.)

【主治】外感风寒，恶寒发热，头身疼痛，无汗而喘，口不渴，舌苔薄白，

脉浮紧。

(Indication symptoms: exogenous wind old, aversion to cold with fever, headache and body pain, asthma without sweat, no thirst, white and coating of the tongue, floating and tense pulse.)

Fig. 1 An example of a prescription in Dictionary of Traditional Chinese Medicine Prescriptions. The Composition herbs are the most important in a prescription, and generally the prescriptions means composition herbs

deep learning has pervaded every aspect of medical image analysis [20]. End-to-end convolutional neural network training has become a good choice for medical image processing tasks.

However, to the best of our knowledge, there has not been research work that mines the relationship between the patient's face and TCM prescriptions. In realistic TCM, doctors prescribe through features of face, tongue, pulse, voice, and symptoms. Using face images to generate TCM prescription is of great significance to assist doctors in the diagnosis and treatment. Especially for some young doctors, the generated prescriptions can give them some references. It can recommend prescriptions to doctors. After making some modifications, doctors can apply them to practice. It saves treatment costs compared to directly prescribe from scratch and improve the efficiency of the doctor's prescribing. A large number of data samples can be used to learn the relevant information of patient's face and TCM prescription. Learning to how to prescribe through the patient's diagnosis data can provide a reference for TCM doctors to observe and diagnose patients.

In this paper, we propose to use deep learning (convolutional neural network) to prescribe (TCM prescriptions) based on the patient's face image. The main work is as follows:

1. A conventional convolutional neural network was designed to encode the patient' face image features and generate TCM prescriptions.
2. Considering different facial organs(eyes, nose, mouth) and regions(cheeks, chin) represent the status of internal organs(heart, liver, spleen, lungs and kidney) in various parts of the human body, a multi-scale convolutional neural network based on three-grained face is proposed to extract feature of facial organs, facial regions and entire face to learn to generate TCM prescriptions.
3. We conduct experiments to verify the effectiveness of convolutional neural network for face feature encoding and prescription generation.

The rest of this paper is organized as follows. In Section 2, we discuss some related work on TCM prescriptions and medical image processing. Section 3 illustrates task description and methodology. Section 4 elaborates and analysis the experiment results. We have some discussions in Section 5 and conclude this paper in Section 6.

2 Related work

Deep learning in medical image processing Deep learning and convolutional neural network have become popular topics in medical image processing. There are already a lot of research works that apply deep learning to medical image processing. In terms of disease classification, there are studies on breast cancer image classification [2, 8], lung pattern classification [1], Alzheimer's disease classification [12]. In the detection of lesion targets and diseases, there are cancerous tissue recognition [28], detecting cardiovascular disease [31], predicting pancreatic cancer [25], and melanoma recognition [40]. In the segmentation of organs and substructures, there are studies on skin lesion segmentation [23, 42], microvasculature segmentation of arterioles [17], tumor segmentation [47]. In addition, there are many other applications, such as studies of visual attention of patients with Dementia [5], diagnosis of cirrhosis stage [36], constructing brain maps [46].

TCM prescriptions On the other hand, some work has been devoted to the study of TCM prescriptions. Some studies analyzed and explored TCM prescriptions and discovered the regularity [21, 34, 45, 48, 49]. Some studies used topic model to discover prescribing patterns [37, 38]. There are other studies such as TCM medicine classification and recognition [9, 33], knowledge graph [32, 41] for TCM.

In practice, TCM doctors can judge the health of various internal organs of the body by observing the patient's face. Combining other characteristics, doctors can give TCM prescription based on their knowledge. Our work is to try to simulate and learn this process. Using deep learning techniques, we can learn how to prescribe from a large amount of medical data. At the current stage, the medical data from which we learn are the patient's face images and corresponding prescriptions. The study is of great importance to assist doctors to diagnosis and treat.

3 Methodology

3.1 Data collection and task description

The data set used in our study are collected from cooperative hospitals. After preprocessing, there are 9,653 data pairs totally. Each data pair contains a patient's face image and a corresponding TCM prescription.

All Chinese herbal medicines are included in a unified dictionary $H = \{h_1, h_2, ..., h_n\}$. The i-th element h_i in H represents the i-th Chinese herbal medicines, and there are n Chinese herbal medicines. In our dataset, n is 559. Each TCM prescription can be represented by a vector $y = [y_1, y_2, ..., y_n]$. The element y_i in y can only be 1 or 0, indicating whether the Chinese herbal medicine is prescribed. Each patient's face image is represented by a pixel matrix x, and the size of x is 224x224x3. X represents all face images in dataset and Y represents all prescriptions in dataset.

The task of this paper is to input a patient's face image (pixel matrix x) and output the patient's corresponding prescription y. The prescription y is a multi-label vector. In fact, the task is a multi-label learning. Multi-label learning studies the problem where each example is represented by a single instance while associated with a set of labels simultaneously [44].

3.2 Construction of conventional convolutional neural network

Deep convolutional neural networks are widely used in the field of image processing. It can extract potential features from the original pixel matrix with RGB color channels for use in various image tasks such as classification, detection, segmentation. Classical convolutional neural network structures include AlexNet [19], VGGNet [26], GoogleNet [29], ResNets [11, 35, 43], DenseNet [14] and SENet [13].

The convolutional neural network used for prescription generation is composed of several convolutional modules and fully connected layers. Each convolutional module includes a convolutional layer and a pooling layer. In order to extract features from the image, the convolutional layer uses some convolutional kernels to scan image matrix to reconstruct a feature map C. A convolutional kernel is a weight matrix. We use K to represent it. The above operation can be abstracted as a function with the relu [10] activation function:

$$C(x, K) = ReLU(Conv(x, K)). \tag{1}$$

In order to extract more important features and reduce the computational complexity, the max-pooling layer is used to downsample the feature map C, which can be represented by the following function (the parameters of max-pooling layer are omitted):

$$\hat{C}(x, K) = Max(C(x, K)). \tag{2}$$

Three consecutive convolution and pooling operations can be abstracted into the following function:

$$\hat{C}^3(x, K) = \hat{C}(\hat{C}(\hat{C}(x, K))). \tag{3}$$

In order to encode features, several fully connected layers are usually connected to the end of several convolution modules. The weight parameters of the fully connected layer layer are denoted by $W1$. An operation of fully connected layer (with a relu activated function) can be abstracted as the following function:

$$f(\hat{C}^3, W1) = ReLU(FC(\hat{C}^3, W1)). \tag{4}$$

The last layer is the output layer, which is a fully connected layer with sigmoid activation function. The weight is represented by $W2$. It outputs the probability of whether each Chinese herbal medicine is prescribed, which can be abstracted as the following function:

$$\begin{aligned} P(x, \Theta) &= sigmoid(FC(f, W2)) \\ &= [P(h_1|x, \Theta), ..., P(h_n|x, \Theta)]. \end{aligned} \tag{5}$$

$\Theta = \{K, W1, W2\}$ is the set of all parameters, and the convolutional kernel K for each convolutional operation described above is different.

The loss function of the convolutional neural network is designed as the average value of multiple cross-entropy. Each cross-entropy measures the difference between the probability of prescribing of each Chinese herbal medicine $P(h_i|x, \Theta)$ and actual output y_i. The neural network minimizes the loss function by optimizing all parameters using stochastic gradient descent [3], which can be abstracted as the following functions(m is the size of the dataset):

$$\begin{aligned} J(\Theta, x) = \tfrac{1}{n} \sum_{i=1}^{n} [-y_i log(P(h_i|x, \Theta)) - \\ (1 - y_i) log(1 - P(h_i|x, \Theta))]; \end{aligned} \tag{6}$$

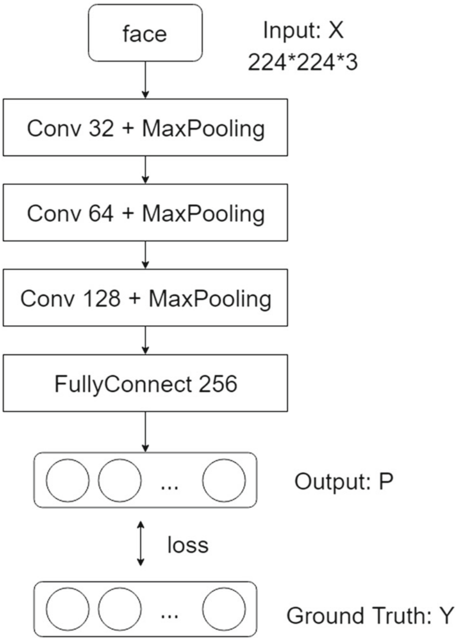
$$\Theta^* = \underset{\Theta}{arg\,min} \frac{1}{m} \sum_{j=1}^{m} J(\Theta, x_j). \tag{7}$$

The structure of the convolutional neural network is shown in Fig. 2. It contains three convolution modules for extracting features, a fully connected layer for coding features, and the final output layer. All the sizes of convolution kernels are 3x3. The input of the network is the face image matrix x of the patient, and the size is 224x224x3. The number of elements in the output layer is n, the size of the Chinese herbal medicine dictionary H, and each unit represents the probability that a certain Chinese herbal medicine is prescribed. The number of dimensions of the real output y is n, and each value is 0 or 1, indicating whether to prescribe. The loss is the average cross-entropy loss calculated from the network output P and the real output Y. P contains the probabilities of being prescribed for all Chinese herbal medicine in dictionary H. Finally, according to dictionary H, a final prescription is obtained by sampling from P through a probability threshold t.

3.3 Construction of multi-scale convolutional neural network based on three-grained face

Different regions of a face image have different local statistics [30]. Taigman et al. [30] use locally connected layers, which like a convolutional layer but every location in the

Fig. 2 A conventional convolutional neural network for generating TCM prescriptions



feature map learns a different set of filters, to deal with this problem. However, the use of local layers greatly increases the parameters of the model. Only a large amount of data can support this approach, so instead of doing this, we extract features of different facial regions using different small convolutional networks.

According to TCM, the characteristics of various regions of the face represent the health of various internal organs of the human body. In order to encode the features of each region of the face more efficiently, the paper proposes a multi-scale convolutional neural network based on three-grained face. The “three-grained” refers to the organ block, the local region block, and the face block. Each block extracts characteristics of the face area from different granularities. The organ block includes the left eye, right eye, nose, and mouth. The local region block includes the left cheek, right cheek, and chin. The face block means the entire face. The network is expected to extract and encode more effective facial features from different granularities, thereby improving the effectiveness of prescription generation.

In the data preprocessing stage, the patient’s face is segmented to obtain different region images of the face. An example of different region images after cutting the face [15] is given in Fig. 3. The sizes of different regions images are reduced. The organ block images X_{organ} includes a left-eye image X_{o-1}, a right-eye image X_{o-2}, a nose image X_{o-3}, a mouth image X_{o-4}, and their sizes are 56x56x3. The local region block images X_{region} includes a left cheek image X_{r-1}, a right cheek image X_{r-2} and a chin image X_{r-3}, and their sizes are 112x112x3. The face block means to the entire face X_{face}, and the size of face image is 224x224x3.

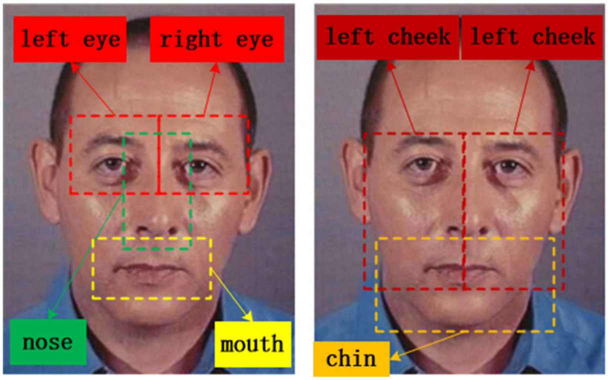


Fig. 3 Different organ and local region images of the segmented face. **Left:** The organ block images. **Right:** The local region block images

3.3.1 Extracting feature of facial organ

Firstly, feature extraction is performed on the organ block. After convolution of four organ block images, concatenate the four feature maps. The operation can be abstracted as the following functions:

$$C_{o-i} = C(X_{o-i}, K), i = \{1, 2, 3, 4\}; \tag{8}$$

$$Concat_o = Concat(\ C_{o-1}, C_{o-2}, C_{o-3}, C_{o-4}). \tag{9}$$

In the field of computer vision applications, there is often not enough data, and the overfitting of models easily occur. Usually, dropout [27] is used to prevent overfitting. Dropout randomly discards neural units during training phase. This prevents units from co-adapting too much and force the network to learn more robust features. It reduces the size of the network during the training phase and gets a number of more streamlined networks that have similar integration effects [27]. After dropout the above feature map $Concat_o$, a convolution operation is performed again to obtain a feature map C_o, which extracts features of organ block. The above operations can be abstracted as the following function:

$$C_o = C(Concat_o, K). \tag{10}$$

3.3.2 Extracting feature of facial local region

Secondly, feature extraction is performed on the local region block. After convolution and max-pooling of the three local region block images, concatenate the three local region block feature maps together with the feature map extracted by the organ block. The above operation can be abstracted as the following functions:

$$C_{r-i} = \hat{C}(X_{r-i}, K), i = \{1, 2, 3\}; \tag{11}$$

$$Concat_{o\_r} = Concat(\ C_o, C_{r-1}, C_{r-2}, C_{r-3}). \tag{12}$$

After dropout the above feature map $Concat_{o\_r}$, convolution and max-pooling operations are performed to extract features to obtain a feature map C_{o_r} (fuses the features of the organ block and local region block). The above operation can be abstracted as the following function:

$$C_{o\_r} = \hat{C}(Concat_{o\_r}, K); \tag{13}$$

3.3.3 Extracting feature of entire face

Finally, feature extraction is performed on the face block. After several convolution and max-pooling of the entire face, concatenate the face block feature map together with the feature map C_{o_r}. The above operation can be abstracted as the following function:

$$C_{face} = \hat{C}(\hat{C}(\hat{C}(X_{face}, K))); \tag{14}$$

$$Concat_{o\_r\_f} = Concat(C_{o\_r}, C_{face}). \tag{15}$$

After dropout the above feature map $Concat_{o\_r\_f}$, two fully connected layers are used to encode feature to get the final features (fuse the features of organ block, region block and face block). The above operation can be abstracted as the following function, where $W3$ and $W4$ are the weights of the fully connected layers.

$$C_{o\_r\_f} = f(f(Concat_{o\_r\_f}, W3), W4) \tag{16}$$

3.3.4 Training based on three-grained face features (organ, local region, entire face)

The convolutional neural network has three output layers. The first output P_{organ} uses the feature map C_o, which extracts the features of organ block, to predict. The second output P_{region} uses the feature map C_{o_r}, which extracts the features of organ block and region block, to predict. The third output P_{face} uses the final feature $C_{o\_r\_f}$, which extracts the features of organ block, region block and face block, to predict. The above operation can be abstracted as the following function, where W_{o1}, W_{o2} and W_{o3} represent the weights of output layers.

$$P_{organ} = f(C_o, W_{o1}) \tag{17}$$

$$P_{region} = f(C_{o\_r}, W_{o2}) \tag{18}$$

$$P_{face} = f(C_{o\_r\_f}, W_{o3}) \tag{19}$$

P_{organ}, P_{region}, and P_{face} denote the probabilities of being prescribed for all Chinese herbal medicine in dictionary H. Among them, P_{face} is the main output of the neural network, which is the decision output of the final generation. P_{organ} and P_{region} are auxiliary outputs, which are used to assist the training of the entire network. The final loss is addition of three losses, which are calculated by P_{organ}, P_{region}, and P_{face} and the real output Y. We use stochastic gradient descent to optimize the parameters so that the final loss is minimized. The loss functions are as follow, where Θ denote the set of all parameters of the neural network and n means the dimension of each real prescription y.

$$J1(\Theta) = \frac{1}{n}\sum_{i=1}^{n}[\ -Y_i log(P_{organ}) - (1 - Y_i)log(1 - P_{organ})] \tag{20}$$

$$J2(\Theta) = \frac{1}{n}\sum_{i=1}^{n}[\ -Y_i log(P_{region}) - (1 - Y_i)log(1 - P_{region})] \tag{21}$$

$$J3(\Theta) = \frac{1}{n}\sum_{i=1}^{n}[\ -Y_i log(P_{face}) - (1 - Y_i)log(1 - P_{face})] \tag{22}$$

$$\Theta^* = arg \min_{\Theta} J1(\Theta) + J2(\Theta) + J3(\Theta). \tag{23}$$

The multi-scale convolutional neural network based on three-grained face structure is shown in Fig. 4, in which the sizes of the input organ block images are 56x56x3, and the sizes of the input region block images are 112x112x3, the size of the input face block image is 224x224x3. All the sizes of convolution kernels are 3x3.

The network is divided into three parts. The first part extracts the features of organ block to obtain output P_{organ}. The second part extracts the features of region block and then merges them with the features of the organ block to continue to extract feature to get the output P_{region}. The third part extracts the features of face block and then merges them with the features of the organ block and region block to continue to extract feature to get the

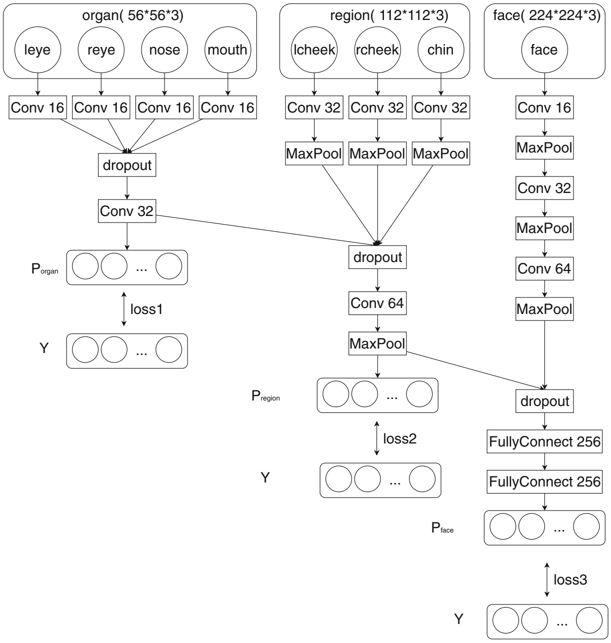

organ(56*56*3)
leye
reye
nose
mouth
region(112*112*3)
lcheek
rcheek
chin
face(224*224*3)
face
Conv 16
Conv 16
Conv 16
Conv 16
Conv 32
Conv 32
Conv 32
Conv 16
MaxPool
MaxPool
MaxPool
MaxPool
dropout
Conv 32
Conv 32
MaxPool
dropout
Conv 64
P_organ
loss1
Y
Conv 64
MaxPool
MaxPool
P_region
dropout
loss2
FullyConnect 256
Y
FullyConnect 256
P_face
loss3
Y

Fig. 4 A multi-scale convolutional neural network based on three-grained face for generating TCM prescriptions

output P_{face}. The three outputs denote the probabilities of being prescribed for all Chinese herbal medicine in dictionary H. The loss used to train the entire network is addition of three losses, which are calculated by P_{organ}, P_{region}, P_{face} and the real output Y. Finally, the final generated prescription is obtained by sampling from the output P_{face} through the probability threshold t.

3.4 Data augmentation

In the real world, patient's medical data is precious and difficult to collect. Therefore, the data collected from the patient's faces and prescriptions are very limited. Due to the limited data set, it is easy to cause the model to overfit, which is one reason for not choosing an overly complex network. Data augmentation is an effective way to cope with not enough data. It can reduce overfitting of the model and improve the model's predictive performance.

In order to make full use of limited data, data augmentation is performed. The "data augmentation" randomly extracts some of the original patient's face images, then randomly transforms the images (such as rotation, zoom) and then saves the image as a new patient's face image. The original patient's prescription are used as the prescription labels of the new patient's face image. Data augmentation can increase the size and diversity of the data set. The sample size of the original data set is 9653. After data augmentation, the data set size increases to 18,463. Some parameters used in data augmentation are shown in Table 1.

4 Experiment

4.1 Dataset

"Face image - TCM prescription" dataset is collected from some cooperative hospitals. Due to the limited collection conditions, the collected raw data have a certain noise. For example, there are different medicine names but exactly they are the same medicine. After some preprocessing, the experimental dataset is obtained. One of the preprocessing is face detection. Although we have tried to ensure the image quality when collecting face images: let the patient face the camera, put the face as accurately and clearly as possible in the middle position, and fill the entire image, there is still more or less background noise, so we should use face detection to reduce noise. Firstly we used dlib library [18] to detect the approximate position of the face. Because the bounding box of face is small, the image may lose some information if dlib give an inaccurate detection. And the bounding box doesn't contain forehead. Therefore, we increase the bounding box by doing a certain percentage of expansion. The final detection effect can be seen from face images in Table 8.

Table 1 Parameters of data augmentation

Parameters	Values	Explanations
rotation_range	25	A value in (0-180) within which images are randomly rotated
width_shift_range	0.05	Translate horizontally by a certain percentage of width
height_shift_range	0.05	Translate vertically by a certain percentage of height
zoom_range	0.2	The zoom range is [1-0.2, 1+0.2]
horizontal_flip	True	A flag for randomly flipping half of the images

Table 2 Face images information

Parameters	Values
Number of face images	9653
Number of face images after data augmentation	18463
Size of face images	224x224x3
Size of local region block images(left cheek, right cheek, chin)	112x112x3
Size of organ block images(left-eye, right-eye, nose, mouth)	56x56x3
Size of test set selected in 5-fold cross-validation	500

The size of experimental dataset D_{origin} is 9653. After data augmentation, the size of dataset is increased to 18463, and the dataset is denoted as D_{aug}. In order to train multi-scale convolutional neural network based on three-grained face, the face images are segmented into different face areas: eyes, nose, mouth, cheeks, and chin. The specific description of the dataset is shown in Tables 2 and 3.

In order to enhance the accuracy and persuasiveness of the experimental results, we use 5-fold cross-validation method to train and evaluate model: repeatedly performs training for five times and 500 samples are taken as test set for each time(conventional approach should divide data into five equal parts, each equal part is taken as the test set for each time. Only 500 samples are taken as test set for each time here due to the limited dataset). The 500 test samples taken for each time do not overlap. The average of five evaluation results is used as the final evaluation result.

4.2 Experimental setup

According to conventional convolutional neural network, multi-scale convolutional neural network based on three-grained face, and data augmentation, five models are run for TCM prescription generation, briefly described as follows:

Random forest (baseline): Random forest [4] classifier is used to generate TCM prescriptions. The features are face images matrix and the labels are multi-label vectors representing the TCM prescriptions.

Conventional CNN: Construct a CNN as described in Section 3.2 to train according to face images and TCM prescriptions to obtain a model for generating TCM prescriptions. The experimental data set used is D_{orgin}.

Conventional CNN_{aug}: The method is the same as conventional CNN, but the experimental data set used is D_{aug}.

Table 3 Prescription information

Parameters	Values
Number of prescriptions	9653
Number of Chinese herbal medicine species	559
Maximum length of TCM prescriptions	28
Minimum length of TCM prescriptions	2
Average length of TCM prescriptions	14
Average times each Chinese herbal medicine appears in prescriptions	240
Proportion of Chinese herbal medicine that appear in more than 100 prescrptions	36%

Table 4 Parameters of neural networks

Parameters	Values
The size of convolutional kernel	3x3
Numbers of convolutional kernel in conventional CNN	32,64,128
Numbers of convolutional kernel in multi-scale CNN based on three-grained face	16,32,64
The size of max-pooling	2x2
Dropout rate	0.4
Number of neurons in the fully connected layer	256
Activation of convolutional layer and fully connected layer	relu
Number of output layer	559
Activation of output layer	sigmoid
Learning rate	0.01
Batch size	64
Number of epochs	300

Multi-scale CNN based on three-grained face: Construct a CNN as described in Section 3.3 to train to obtain a model for generating TCM prescriptions according to images of different face regions and TCM prescriptions. The experimental data set used is D_{orgin}.

Multi-scale CNN_{aug} based on three-grained face: The method is the same as multi-scale CNN based on three-grained face, but the experimental data set used is D_{aug}.

The structure and some parameters of the conventional CNN and multi-scale CNN based on three-grained face have been described in Sections 3.2 and 3.3. The more specific parameters are shown in Table 4. The optimization algorithm is SGD (stochastic gradient descent), and learning rate decay is 1e-6, and momentum is 0.9.

4.3 Evaluation metrics

In order to measure the similarity between the generated TCM prescription and the actual TCM prescription, the indicators precision, recall, and f-score are set as shown in the following formulas. n_true_i denotes the number of Chinese herbal medicine appearing in both the i-th generated prescription and the i-th real prescription. $n\_predict_i$ denotes the number of Chinese herbal medicine appearing in the i-th generated prescription. n_real_i denotes the number of Chinese herbal medicine appearing in the i-th real prescription. $precision_i$ measures the how the Chinese herbal medicines are precise in generated prescription, and $recall_i$ measures the how the Chinese herbal medicines are complete in generated prescription. $f1\_score_i$ (f_score_i) is the harmonic mean of $precision_i$ and $recall_i$, neutralizing these two indicators.

$$\boldsymbol{precision_i} = \frac{n\_true_i}{n\_predict_i} \tag{24}$$

$$recall_i = \frac{n\_true_i}{n\_real_i} \tag{25}$$

$$f1\_score_i = \frac{2 * precision_i * recall_i}{precision_i + recall_i} \tag{26}$$

The indicators are calculated for each sample generated by the model, and then averaged to obtain the indicators used to evaluate the quality of the model:

$$precision = \frac{1}{m}\sum_{i=1}^{m} prescision_i, \tag{27}$$

$$recall = \frac{1}{m}\sum_{i=1}^{m} recall_i, \tag{28}$$

$$f1\_score = \frac{1}{m}\sum_{i=1}^{m} f1\_score_i, \tag{29}$$

where m is the size of the dataset. Test set is used to evaluate the model and the size is 500.

For each example x_i, $f1\_score_i$ is the harmonic mean of $precision_i$ and $recall_i$. But note that $precision$, $recall$, $f1\_score$ are averages, so $f1\_score$ is not the harmonic mean of $precision$ and $recall$.

4.4 Results and analysis

4.4.1 Training process

In order to prevent overfitting, the model uses data augmentation, dropout methods. In addition, the strategy “EarlyStopping” is also used in the experiment. During training, a certain percentage of data is divided from the training set as a validation set for training observations. The proportion used in the experiment is 0.1. The 10% of training data is used as a validation set that does not participate in training. During the training process, observe the loss of the model on the validation set. After the validation set loss is no longer declining, wait for a certain number of iterations (we use 10 in the experiment) to stop the training. This can prevent the model from overfitting the training set and make a better prediction of the test set.

Take one of the training results in the 5-fold cross-validation. The changes of the training set and the validation set’s loss during the training process are shown in Figs. 5 and 6. It can be seen that although the number of epochs is 300(ensure sufficient number of iterations), training is usually stopped at about 30-70 iterations, and the later iterations overfit in the training set. With data augmentation, compared to the conventional CNN, the relative gap

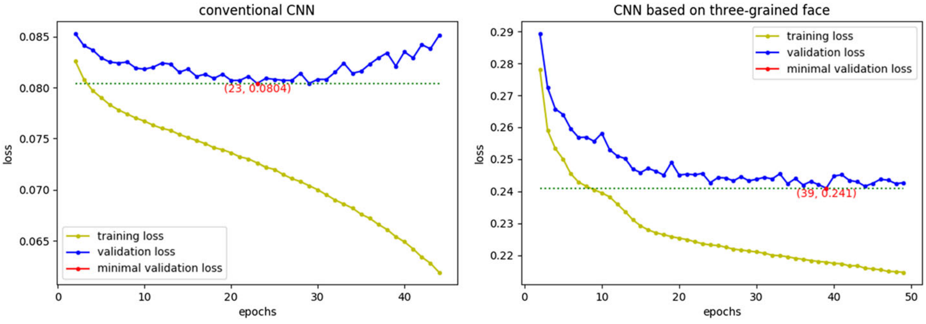


Fig. 5 Learning curve of conventional CNN and multi-scale CNN based on three-grained face prescriptions

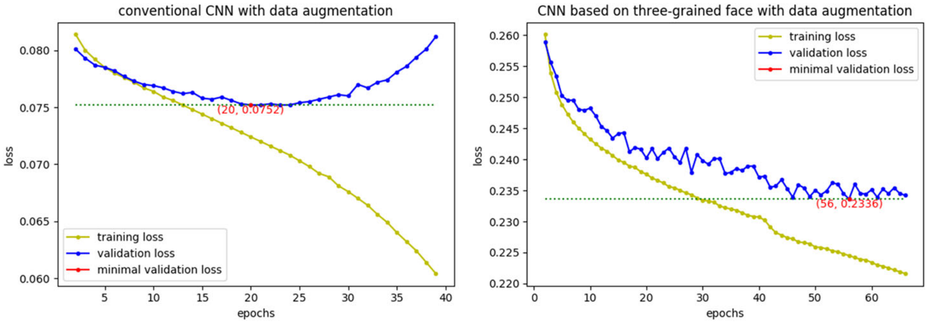


Fig. 6 Learning curve of conventional CNN and multi-scale CNN based on three-grained face (with data augmentation)

between the loss of the training set and the validation set in multi-scale CNN based on three-grained face is smaller, which indicates that the generalization ability of the multi-scale CNN based on three-grained face is relatively high.

4.4.2 Influence of threshold parameter

From the final output of the neural network, a series of probability values can be obtained. Finally, the outputs are 559 neurons, representing 559 Chinese herbal medicines. Finally, 559 corresponding probability values are obtained. The final prescription is predicted based on a threshold value t. The Chinese herbal medicine is prescribed if the probability of the Chinese herbal medicine is more than t.

One general choice for threshold is 0.5. Furthermore, when all the unseen instances in the test set are available, the threshold can be set to minimize the difference on certain multi-label indicator between the training set and test set [44]. As shown in Figs. 7 and 8, setting different thresholds, the final evaluation results will be different (the results in the figure are the average results of 5-fold cross validation). When a larger threshold is set, a higher precision will be obtained because the prescription generated by the model try to be as precise as possible without errors, and it prefer to give fewer medicines to prevent errors. When a smaller threshold is set, a higher recall is achieved because the

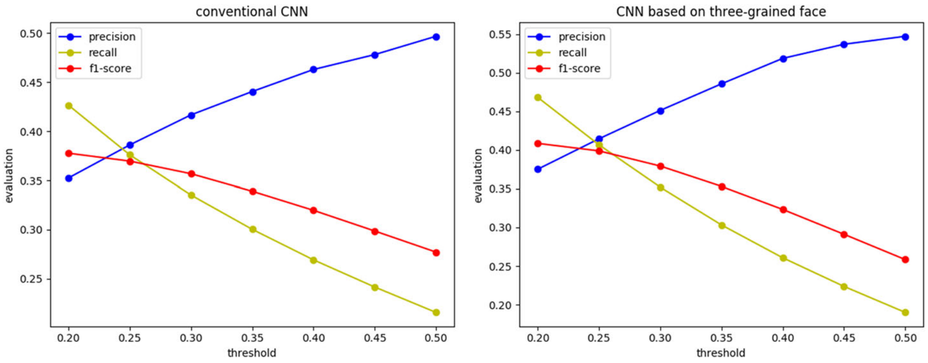


Fig. 7 Influence of threshold on conventional CNN and multi-scale CNN based on three-grained face

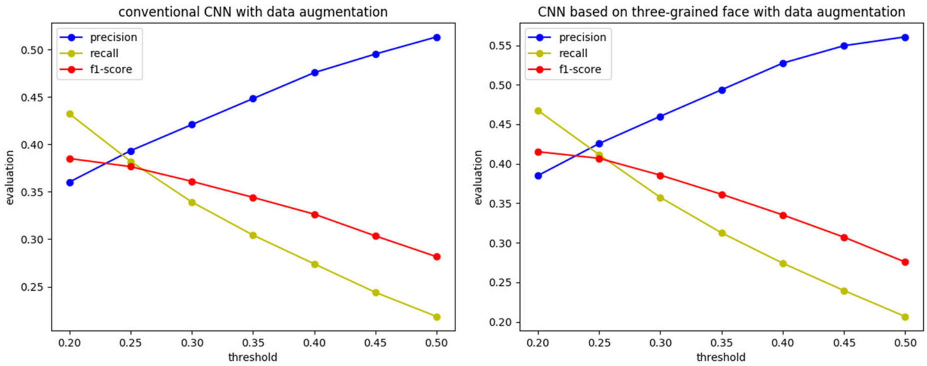


Fig. 8 Influence of threshold on conventional CNN and multi-scale CNN based on three-grained face (with augmentation)

prescription generated by the model attempted to be as complete as possible and at the expense of a certain of precision. The "f1_score" is the harmonic mean of precision and recall, which neutralizes the accuracy and completeness. Note that the f1_score shown in the experimental data is not harmonic mean of precision and recall, because the f1_score is an average. We choose 0.25 as the final threshold, because at this time the value of f1_score is high relatively, and the difference between precision and recall is small, which can ensure high precision and recall simultaneously.

4.4.3 Performance comparison

The experimental results of the five models are shown in Table 5. In order to enhance accuracy and persuasiveness of results, the evaluation results are averaged by 5 results, calculated by 5-fold cross validation methods. The values after "±" indicate the standard deviation of the 5 results.

Random forest is a ensemble learning technique, which should give good performances. However, it can be seen from the experimental results that the other four models improve the performances compared to the baseline classifier random forest, indicating that the convolutional neural network is better than the random forest in this task. The neural network can extract and represent useful features from large and complex data. There are a large number of original image features that need to be extracted and represented on the task, so using a convolutional neural network for image processing to build a model is a better choice.

Table 5 Experimental performances of different models

Model	Precision(%)	Recall(%)	f1-score(%)
Random forest - baseline	37.25 ± 2.12	35.50 ± 2.97	33.07 ± 1.96
Conventional CNN	38.60 ± 2.76	37.60 ± 1.53	36.69 ± 1.63
Conventional CNN_{aug}	39.34 ± 2.56	38.21 ± 1.92	37.66 ± 1.97
Multi-scale CNN based on three-grained face	41.44 ± 2.65	40.67 ± 1.70	39.88 ± 0.73
Multi-scale CNN_{aug} based on three-grained face	**42.56** ± 2.48	**41.15** ± 2.01	**40.69** ± 1.67

Best results are highlighted in bold

The performances of conventional CNN_{aug} are better than conventional CNN, and the performance of multi-scale CNN_{aug} based on three-grained face are better than the multi-scale CNN based on three-grained face. It can be seen that after using data augmentation, the models perform better because using data augmentation increases the size and diversity of the data, allowing the convolutional neural network to learn more knowledge when training. It can reduce overfitting of model.

The performances of multi-scale CNN based on three-grained face are better than conventional CNN, and the performances of multi-scale CNN_{aug} based on three-grained face are better than the conventional CNN_{aug}. A reasonable explanation for this result is that the multi-scale CNN based on three-grained face extracts features from different granularities(organs, local regions, and the entire face), and it can extract and utilize local features and global features more effectively.

As shown in Table 8, three samples were taken to show the actual predicted results. For each example, patient's face image and corresponding prescriptions are shown. The red bold type of Chinese herbal medicines indicate that it appears both in the real prescription and predicted prescription, from which we can seen the precision of model. The more red bold medicines, the higher the precision of model. The cyan bold type of Chinese herbal medicines indicate that it doesn't appear in the predicted prescription but appear in real prescription, from which we can seen the recall of model. The less cyan bold medicines, the higher the recall of model. It can be seen that the results of the model prediction have certain similarities with the actual prescriptions, which shows that the model has indeed learned something. In the four models, the results of multi-scale CNN_{aug} based on three-grained face(we omit the "multi-scale" just for neat alignment in Table 8) are the most precise and complete. It can be seen that for common Chinese herbal medicines, the prediction of the model will be more accurate, such as Radix Glycyrrhizae and Poria Cocos. For some unusual Chinese herbal medicines, the model cannot accurately predict, such as Perilla Stem and Curcuma Zedoary. A reasonable explanation for this phenomenon is that common Chinese herbal medicines always appear in the training samples, and the model can learn more useful distinguishable features from a large number of training data. However, it is rarely used for the unusual Chinese herbal medicines, which only occasionally used by a few patients. With a small amount of data, the model is difficult to learn. The model cannot find distinguishable features.

4.4.4 Effect of different image size

The input size of conventional CNN is 224x224. The "multi-scale" of CNN based on three-grained face means to the three-scale input 56x56, 112x112, 224x224, but the actual input size is still 224x224, which is the size of face image. We just input a 224x224 face image and the face image is segmented to 112x112 local region block images and 56x56 organ block images during preprocessing. So we say that the input size of $CNNs$ used above is 224x224. However, the size of the patient's face image in reality is uncertain.

In order to verify that the CNN models can adapt to various images of different sizes, we retrain the networks of different input sizes with D_{orgin} and get the experiment results, as shown in Table 6. The evaluation results(precision, recall, f1-score) are calculated by 5-fold cross validation methods. For multi-scale CNN based on three-grained(we omit the "multi-scale" just for neat alignment in Table 6), the image size means the size of face, and the size of local region block images is half of the size of face, the size of organ block images is half of the size of local region block images. The "average" in Table 6 means the average results of different image sizes(32, 56, 84, 112, 168, 224).

Table 6 Experimental performances of different image sizes

Image size	32x32	56x56	84x84	112x112	168x168	224x224	Average
Precision(%) of conventional CNN	42.14	42.54	41.55	41.95	41.40	38.60	41.36
Precision(%) of CNN based on three-grained face	40.77	41.38	41.42	42.24	42.00	41.44	41.54
Recall(%) of conventional CNN	38.76	39.10	39.34	38.41	37.70	37.60	38.49
Recall(%) of CNN based on three-grained face	37.97	38.37	38.70	40.16	40.59	40.67	39.41
f1-scores(%) of conventional CNN	38.97	39.42	39.27	38.83	38.18	36.69	38.56
f1-scores(%) of CNN based on three-grained face	37.88	38.28	38.66	39.93	40.04	39.88	39.11

It can be seen from Table 6 that the models obtain similar results for different sizes of input images, indicating the robustness of the models. From the average, multi-scale CNN based on three-grained face performs better than conventional CNN. In addition, for smaller image sizes(32, 56, 84), multi-scale CNN based on three-grained face is slightly worse than conventional CNN, but when the input image is relatively large(112, 168, 224), multi-scale CNN based on three-grained face is still excellent in the three evaluation indicators. We conjecture that multi-scale CNN based on three-grained face needs to capture three-grained face features and fine-grained features are difficult to mine when the image is small. But on average, multi-scale CNN based on three-grained face is still superior, and in reality it is unlikely to take too small patient's images. From the results of larger image sizes(112, 168, 224), the performances of multi-scale CNN based on three-grained face are all higher than conventional CNN.

In addition, as can be seen from the results in Table 6, in conventional CNN, the performances of 224x224 images are worse than 168x168. This is mainly because our dataset isn't very large, and the image features of 224x224 are too much (compared to 168x168). When the ratio of the number of features to the number of samples is relatively large, the model is easier to over-fitting. By comparison, in our multi-scale CNN based on three-grained face, although the model is more complicated than conventional CNN, it alleviates the over-fitting, which can be seen from that the performance gaps of 224x224 and 168x168 images become smaller in multi-scale CNN based on three-grained face, and even the recall of 224x224 is slightly higher than 168x168. This also illustrates the advantages of our model from another perspective. We speculate that this should be due to the nature of this model. The motivation for its design is based on the actual experience of Traditional Chinese Medicine, that is, to mine the facial features of the patient from the three granularities of the face, so that more comprehensive and useful information can be obtained.

4.4.5 Ablation study

To analyze the effects of the three branches on the results, we performed an ablation experiment to illustrate the importance of the three branches. We did several experiments with CNN based on three-grained face, CNN based on double grained face (remove one block), CNN based on single grained face (remove two block). The experimental results are shown in Table 7.

As we can see, the best model is CNN based on three-grained face. If the organ block or the region block is removed, there will be a slight decrease in performance. If both the organ block and the region block are removed, the performance will be reduced even more.

Table 7 Ablation study about removing different block

Model	Remove organ	Remove region	Remove face	Precision(%)	Recall(%)	f1-score(%)
CNN based on three-grained face	–	–	–	41.44	40.67	39.88
CNN based on double-grained face	√	–	–	40.90	39.28	38.89
CNN based on double-grained face	–	√	–	40.66	39.09	37.44
CNN based on single-grained face	√	√	–	38.60	37.60	36.69
CNN based on double-grained face	–	–	√	34.80	33.37	31.54
CNN based on double-grained face	√	–	√	33.18	32.65	31.34
CNN based on single-grained face	–	√	√	32.33	31.66	30.20

The phenomena illustrate the need for organ block and region block for our tasks. The model does learn some more detailed features from organ block and region block to help prediction.

Table 8 Real predicted results of different models

	real prescription	甘草 白芍 川芎 当归 茯苓 党参 白术 黄连 熟地黄 肉桂 厚朴 白芷 蔓荆子 (Radix Glycyrrhizae, Radix Paeoniae Alba, Rhizoma Chuanxiong, Angelica, Poria Cocos, Radix Codonopsis, Macrocephalae Rhizoma, Goldthread, Prepared Rehmannia Root, Cinnamon, Mangnolia Officinalis, Radix Angelicae Dahuricae, Fructus Viticis)
	conventional CNN	**甘草 茯苓 白术** 浙贝母 砂仁 蜈蚣子 **白芍 川芎 当归 党参 黄连 熟地黄 肉桂 厚朴 白芷 蔓荆子** (**Radix Glycyrrhizae, Poria Cocos, Macrocephalae Rhizoma,** Thunberg Fritillary Bulb, Fructus Amomi, Scolopendra, **Radix Paeoniae Alba, Rhizoma Chuanxiong, Angelica, Radix Codonopsis, Goldthread, Prepared Rehmannia Root, Cinnamon, Mangnolia Officinalis, Radix Angelicae Dahuricae, Fructus Viticis**)
	conventional CNN_{aug}	**甘草 茯苓 党参 白术** 法半夏 山药 **白芍 川芎 当归 黄连 熟地黄 肉桂 厚朴 白芷 蔓荆子** (**Radix Glycyrrhizae, Poria Cocos, Radix Codonopsis, Macrocephalae Rhizoma,** Rhizoma Pinellinae Praeparata, Dioscoreae Rhizoma, **Radix Paeoniae Alba, Rhizoma Chuanxiong, Angelica, Goldthread, Prepared Rehmannia Root, Cinnamon, Mangnolia Officinalis, Radix Angelicae Dahuricae, Fructus Viticis**)
	CNN based on three-grained face	**甘草 茯苓 党参 白术** 蜈蚣 **白芍 川芎 当归 黄连 熟地黄 肉桂 厚朴 白芷 蔓荆子** (**Radix Glycyrrhizae, Poria Cocos, Radix Codonopsis, Macrocephalae Rhizoma,** Scolopendra, **Radix Paeoniae Alba, Rhizoma Chuanxiong, Angelica, Radix Codonopsis, Goldthread, Prepared Rehmannia Root, Cinnamon, Mangnolia Officinalis, Radix Angelicae Dahuricae, Fructus Viticis**)
	CNN_{aug} based on three-grained face	**甘草 白芍 茯苓 党参 白术** 蜈蚣 **川芎 当归 黄连 熟地黄 肉桂 厚朴 白芷 蔓荆子** (**Radix Glycyrrhizae, Radix Paeoniae Alba, Poria Cocos, Radix Codonopsis, Macrocephalae Rhizoma,** Scolopendra, **Rhizoma Chuanxiong, Angelica, Radix Codonopsis, Goldthread, Prepared Rehmannia Root, Cinnamon, Mangnolia Officinalis, Radix Angelicae Dahuricae, Fructus Viticis**)
	real prescription	甘草 法半夏 茯苓 前胡 桔梗 薏苡仁 浙贝母 细辛 天麻 鳖甲 款冬花 莪术 炙麻黄 蜈蚣 白花蛇舌 (Radix Glycyrrhizae, Rhizoma Pinellinae Praeparata, Poria Cocos, Radix Peucedani, Platycodonis Radix, Coicis Semen, Fritillaria Thunbergii Miq, Asarum Sieboldi Mig, Gastrodiae Rhizoma, Trionycis Carapax, Flos Farfarae, Curcuma Zedoary, Fried Herba Ephedrae, Scolopendra, Herba Hedyotidis)
	conventional CNN	**甘草 浙贝母 天麻 鳖甲 蜈蚣 白花蛇舌** 柴胡 党参 酸枣仁 生地黄 红花 延胡索 山药 天山雪莲 半枝莲 **法半夏 茯苓 前胡 桔梗 薏苡仁 细辛 款冬花 莪术 炙麻黄** (**Radix Glycyrrhizae, Gastrodiae Rhizoma, Trionycis Carapax, Scolopendra, Herba Hedyotidis,** Radix Bupleuri, Radix Codonopsis, Semen Zizyphi Spinosae, dried Rehamnnia Root, Carthamus Tinctorious, Corydalis Rhizoma, Thunberg Fritillary Bulb, Dioscoreae Rhizoma, Saussureae Involucratae Herba, Scutellariae Barbatae Herba, **Rhizoma Pinellinae Praeparata, Poria Cocos, Radix Peucedani, Platycodonis Radix, Coicis Semen, Fritillaria Thunbergii Miq, Asarum Sieboldi Mig, Flos Farfarae, Curcuma Zedoary, Fried Herba Ephedrae**)
	conventional CNN_{aug}	**甘草 茯苓 天麻 鳖甲 蜈蚣 白花蛇舌** 党参 白术 山药 天山雪莲 半枝莲 **法半夏 前胡 桔梗 薏苡仁 浙贝母 细辛 款冬花 莪术 炙麻黄** (**Radix Glycyrrhizae, Poria Cocos, Gastrodiae Rhizoma, Trionycis Carapax, Scolopendra, Herba Hedyotidis,** Radix Codonopsis, Macrocephalae Rhizoma, Dioscoreae Rhizoma, Saussureae Involucratae Herba, Scutellariae Barbatae Herba, **Rhizoma Pinellinae Praeparata, Radix Peucedani, Platycodonis Radix, Coicis Semen, Fritillaria Thunbergii Miq, Asarum Sieboldi Mig, Flos Farfarae, Curcuma Zedoary, Fried Herba Ephedrae**)
	CNN based on three-grained face	**甘草 茯苓 薏苡仁 天麻 鳖甲 蜈蚣 白花蛇舌** 党参 天山雪莲 **法半夏 前胡 桔梗 浙贝母 细辛 款冬花 莪术 炙麻黄** (**Radix Glycyrrhizae, Poria Cocos, Coicis Semen, Gastrodiae Rhizoma, Trionycis Carapax, Scolopendra, Herba Hedyotidis,** Radix Codonopsis, Saussureae Involucratae Herba, **Rhizoma Pinellinae Praeparata, Radix Peucedani, Platycodonis Radix, Fritillaria Thunbergii Miq, Asarum Sieboldi Mig, Flos Farfarae, Curcuma Zedoary, Fried Herba Ephedrae**)
	CNN_{aug} based on three-grained face	**甘草 茯苓 薏苡仁 浙贝母 天麻 鳖甲 蜈蚣 白花蛇舌** 党参 天山雪莲 **法半夏 前胡 桔梗 细辛 款冬花 莪术 炙麻黄** (**Radix Glycyrrhizae, Poria Cocos, Coicis Semen, Fritillaria Thunbergii Miq, Gastrodiae Rhizoma, Trionycis Carapax, Scolopendra, Herba Hedyotidis,** Radix Codonopsis, Saussureae Involucratae Herba, **Rhizoma Pinellinae Praeparata, Radix Peucedani, Platycodonis Radix, Asarum Sieboldi Mig, Flos Farfarae, Curcuma Zedoary, Fried Herba Ephedrae**)
	real prescription	法半夏 茯苓 前胡 桔梗 防风 白芷 款冬花 紫菀 白前 百部 炙甘草 辛夷 紫苏梗 广藿香 蜜麻黄 (Rhizoma Pinellinae Praeparata, Poria cocos, Radix Peucedani, Platycodonis Radix, Radix Saposhnikoviae, Radix Angelicae Dahuricae, Flos Farfarae, Aster Tataricus Linn, Rhizoma Cynanchi Stauntonii, Radix Stemonae, Radix Glycyrrhizae Preparata, Flos Magnoliae, Perilla Stem, Herba Pogostemonis, Honey-fried Herba Ephedrae)
	conventional CNN	**防风 白芷 炙甘草** 白芍 陈皮 党参 枸杞子 首乌藤 **法半夏 茯苓 前胡 桔梗 款冬花 紫菀 白前 百部 辛夷 紫苏梗 广藿香 蜜麻黄** (**Radix Saposhnikoviae, Radix Angelicae Dahuricae, Radix Glycyrrhizae Preparata,** Radix Paeoniae Alba, Tangerine Peel, Radix Codonopsis, Fructus Lycii, Caulis Polygoni Multiflori, **Rhizoma Pinellinae Praeparata, Poria cocos, Radix Peucedani, Platycodonis Radix, Flos Farfarae, Aster Tataricus Linn, Rhizoma Cynanchi Stauntonii, Radix Stemonae, Flos Magnoliae, Perilla Stem, Herba Pogostemonis, Honey-fried Herba Ephedrae**)
	conventional CNN_{aug}	**防风 白芷 炙甘草 辛夷 广藿香** 麻黄 白芍 川芎 荆芥穗 豆蔻 **法半夏 茯苓 前胡 桔梗 款冬花 紫菀 白前 百部 紫苏梗 蜜麻黄** (**Radix Saposhnikoviae, Radix Angelicae Dahuricae, Radix Glycyrrhizae Preparata, Flos Magnoliae, Herba Pogostemonis,** Herba Ephedrae, Radix Paeoniae Alba, Rhizoma Chuanxiong, Herba Schizonepetae, Fructus Amomi Rotundus, **Rhizoma Pinellinae Praeparata, Poria cocos, Radix Peucedani, Platycodonis Radix, Flos Farfarae, Aster Tataricus Linn, Rhizoma Cynanchi Stauntonii, Radix Stemonae, Perilla Stem, Honey-fried Herba Ephedrae**)
	CNN based on three-grained face	**法半夏 茯苓 前胡 款冬花 紫菀 炙甘草** 陈皮 太子参 北沙参 **桔梗 防风 白芷 白前 百部 辛夷 紫苏梗 广藿香 蜜麻黄** (**Rhizoma Pinellinae Praeparata, Radix Peucedani, Flos Farfarae, Aster Tataricus Linn, Radix Glycyrrhizae Preparata,** Tangerine Peel, Poria Cocos, Radix Pseudostellariae, Radix Glehniae, **Poria cocos, Platycodonis Radix, Radix Saposhnikoviae, Radix Angelicae Dahuricae, Rhizoma Cynanchi Stauntonii, Radix Stemonae, Flos Magnoliae, Perilla Stem, Herba Pogostemonis, Honey-fried Herba Ephedrae**)
	CNN_{aug} based on three-grained face	**法半夏 茯苓 前胡 桔梗 防风 款冬花 紫菀 白前 百部 炙甘草 广藿香 蜜麻黄** 炒紫苏子 **白芷 辛夷 紫苏梗** (**Rhizoma Pinellinae Praeparata, Poria cocos, Radix Peucedani, Radix Saposhnikoviae, Flos Farfarae, Aster Tataricus Linn, Rhizoma Cynanchi Stauntonii, Radix Stemonae, Radix Glycyrrhizae Preparata, Herba Pogostemonis, Honey-fried Herba Ephedrae,** Radix Platycodi, Radix Glehniae, Fried Perilla Fruit, **Platycodonis Radix, Radix Angelicae Dahuricae, Flos Magnoliae, Perilla Stem**)

The red bold type of Chinese herbal medicines indicate that it appears both in the real prescription and predicted prescription. The cyan bold type of Chinese herbal medicines indicate that it doesn't appear in the predicted prescription but appear in real prescription. The common type of Chinese herbal medicines indicate that it appear in the predicted prescription but dosen't appear in real prescription

In addition, it can be clearly seen that if we remove the face block, the performance of the model will be greatly reduced, which means that the face block is the most important block. This is in line with the doctor's intuition and our motivation for model design, because when the doctor is treating, he first looks at the whole face, extracts features from the whole face to model, and then observes some details from organs and regions. So the face block is the most important feature sources, while the organ block and region block are complementary to get some of the detailed features.

5 Discussion

Our results show that convolutional neural networks are capable of mining the prescription information from patient's face images to generate prescription, and the multi-scale convolutional neural networks based on three-grained indeed can generate prescriptions that are closer to real prescriptions, as shown in the actual prediction results in Table 8 and the evaluation results in Tables 5 and 6. By building such a prescription generation system, the doctors can obtain recommended prescription, and then modify it, finally apply it to the actual treatment.

Generation of TCM prescriptions from face image using deep learning can provide us with a possible result. Although the predicted result is not an inevitable conclusion, it provides us with a choice, a kind of opinion for reference, which greatly reduces the blindness of work. In fact, in reality, different TCM doctors do not always give the same prescriptions to patients, and there may be multiple prescriptions for the same patient. It is possible that system-generated prescriptions can inspire doctors to develop new useful prescriptions.

6 Conclusion

In this paper, we propose to use convolutional neural network to generate TCM prescriptions according to the patient's face image. In order to more fully and effectively extract and utilize the features of the patient's face, we propose a multi-scale convolutional neural network based on three-grained face and compare it with the conventional convolutional neural network. In addition, we use data augmentation to increase the size and diversity of the data to improve the effect.

To the best of our knowledge, few people do the work to generate TCM prescriptions. Chinese herbal medicine is a medical asset accumulated by the Chinese ancient people's long-term practice. It is extremely rich and precious. It is of great significance to fully mine and learn information from the prescribing data of patients using deep learning technique.

In fact, when treating patients, doctors of TCM need to integrate multiple features (face, tongue, pulse, voice, symptoms) and their own experience to give solutions, which can overcome the limitations of using face images alone. Due to the limited data, in our preliminary research work we only consider to using patient's face image to generate TCM prescriptions. In the future work, we plan to collect more quantities, more types of patient data.

Acknowledgements This study was supported by the China National Science Foundation (60973083, 61273363), Science and Technology Planning Project of Guangdong Province (2014A010103009, 2015A020217002), and Guangzhou Science and Technology Planning Project (201504291154480, 2016040-20179, 201803010088).

References

1. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S (2016) Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans Med Imaging 35(5):1207–1216
2. Bayramoglu N, Kannala J, Heikkilä J (2016) Deep learning for magnification independent breast cancer histopathology image classification. In: 2016 23rd International conference on pattern recognition (ICPR), pp 2440–2445
3. Bottou L (2012) Stochastic gradient descent tricks. In: Neural networks: tricks of the trade. Springer, Berlin, pp 421–436. 9780201398298
4. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
5. Chaabouni S, Benois-pineau J, Tison F, Ben Amar C, Zemmari A (2017) Prediction of visual attention with deep cnn on artificially degraded videos for studies of attention of patients with dementia. Multimed Tools Appl 76(21):22527–22546
6. Chen M, Shi X, Zhang Y, Wu D, Guizani M (2017) Deep features learning for medical image analysis with convolutional autoencoder neural network. IEEE Trans Big Data, 1
7. Cheung F (2011) TCM: made in China. Nature 480:S82
8. Chougrad H, Zouaki H, Alheyane O (2018) Deep convolutional neural networks for breast cancer screening. Comput Methods Programs Biomed 157:19–30
9. Dehan L, Jia W, Yimin C, Hamid G (2014) Classification of Chinese herbal medicines based on SVM. In: 2014 International conference on information science, electronics and electrical engineering, vol 1, pp 453–456
10. Glorot X, Bordes A, Bengio Y (2011) Deep sparse rectifier neural networks. In: AISTATS '11: Proceedings of the 14th international conference on artificial intelligence and statistics, pp 315–323, 1502.03167
11. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778
12. Hon M, Khan NM (2017) Towards Alzheimer's disease classification through transfer learning. In: 2017 IEEE International conference on bioinformatics and biomedicine (BIBM), pp 1166–1169
13. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: The IEEE Conference on computer vision and pattern recognition (CVPR)
14. Huang G, Liu Z, v d Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: 2017 IEEE Conference on computer vision and pattern recognition (CVPR), pp 2261–2269
15. Jain V, Learned-Miller E (2010) Fddb: a benchmark for face detection in unconstrained settings. Tech. Rep UM-CS-2010-009. University of Massachusetts, Amherst
16. Jones AL (2018) The influence of shape and colour cue classes on facial health perception. Evol Hum Behav 39(1):19–29
17. Kassim YM, Prasath VBS, Glinskii OV, Glinsky VV, Huxley VH, Palaniappan K (2017) Microvasculature segmentation of arterioles using deep CNN. In: 2017 IEEE International conference on image processing (ICIP), pp 580–584
18. King DE (2009) Dlib-ml: a machine learning toolkit. J Mach Learn Res 10:1755–1758
19. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Advances In Neural Information Processing Systems, pp 1097–1105, 1102.0183
20. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. Med Image Anal 42:60–88
21. Liu B, Zhou X, Wang Y, Hu J, He L, Zhang R, Chen S, Guo Y (2012) Data processing and analysis in real-world traditional Chinese medicine clinical data: challenges and approaches. Stat Med 31(7): 653–660
22. Peng H (1996) Dictionary of traditional Chinese medicine prescriptions. People Health Press, Beijing
23. Peng Y, Wang N, Wang Y, Wang M (2019) Segmentation of dermoscopy image using adversarial networks. Multimed Tools Appl 78(8):10965–10981
24. Qiu J (2007) Traditional medicine: a culture in the balance. Nature 448:126
25. Sekaran K, Chandana P, Krishna NM, Kadry S (2019) Deep learning convolutional neural network (cnn) with gaussian mixture model for predicting pancreatic cancer. Multimedia Tools and Applications
26. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. Int Conf Learn Represent (ICRL) 1409:1556
27. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958. 1102.4807

28. Stanitsas P, Cherian A, Truskinovsky A, Morellas V, Papanikolopoulos N (2017) Active convolutional neural networks for cancerous tissue recognition. In: 2017 IEEE International conference on image processing (ICIP), pp 1367–1371
29. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: 2015 IEEE Conference on computer vision and pattern recognition (CVPR), pp 1–9
30. Taigman Y, Yang M, Ranzato M, Wolf L (2014) DeepFace: closing the gap to human-level performance in face verification. In: 2014 IEEE Conference on computer vision and pattern recognition, pp 1701–1708
31. Wang J, Ding H, Bidgoli FA, Zhou B, Iribarren C, Molloi S, Baldi P (2017) Detecting cardiovascular disease from Mammograms with deep learning. IEEE Trans Med Imaging 36(5):1172–1181
32. Weng H, Liu Z, Yan S, Fan M, Ou A, Chen D, Hao T (2017) A framework for automated knowledge graph construction towards traditional Chinese medicine. In: Siuly S, Huang Z, Aickelin U, Zhou R, Wang H, Zhang Y, Klimenko S (eds) Health information science. Springer International Publishing, Cham, pp 170–181
33. Weng JC, Hu MC, Lan KC (2017) Recognition of easily-confused TCM herbs using deep learning. In: Proceedings of the 8th ACM on multimedia systems conference, MMSys'17. ACM, New York, pp 233–234
34. Xie D, Pei W, Zhu W, Li X (2017) Traditional Chinese medicine prescription mining based on abstract text. In: 2017 IEEE 19th International conference on e-health networking, applications and services (Healthcom), pp 1–5
35. Xie S, Girshick R, Dollár P, Tu Z, He K (2017) Aggregated residual transformations for deep neural networks. In: 2017 IEEE Conference on computer vision and pattern recognition (CVPR), pp 5987–5995
36. Xu Z, Liu X, Cheng XE, Song JL, Zhang JQ (2017) Diagnosis of cirrhosis stage via deep neural network. In: 2017 IEEE International conference on bioinformatics and biomedicine (BIBM), pp 745–749
37. Yao L, Zhang Y, Wei B, Wang W, Zhang Y, Ren X, Bian Y (2015) Discovering treatment pattern in traditional Chinese Medicine clinical cases by exploiting supervised topic model and domain knowledge. J Biomed Inform 58:260–267
38. Yao L, Zhang Y, Wei B, Zhang W, Jin Z (2018) A topic modeling approach for traditional Chinese medicine prescriptions. IEEE Trans Knowl Data Eng 30(6):1007–1021
39. Yiqin W (2012) Objective application of TCM inspection of face and tongue. Chin Arch Tradit Chin Med 30(2):349–352
40. Yu L, Chen H, Dou Q, Qin J, Heng PA (2017) Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans Med Imaging 36(4):994–1004
41. Yu T, Li J, Yu Q, Tian Y, Shun X, Xu L, Zhu L, Gao H (2017) Knowledge graph for TCM health preservation: design, construction, and applications. Artif Intell Med 77:48–52
42. Yuan Y, Chao M, Lo YC (2017) Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance. IEEE Trans Med Imaging 36(9):1876–1886
43. Zagoruyko S, Komodakis N (2016) Wide residual networks. In: Wilson RCERH, Smith WAP (eds) Proceedings of the British machine vision conference (BMVC). BMVA Press, pp 87.1–87.12
44. Zhang ML, Zhou ZH (2014) A review on multi-label learning algorithms. IEEE Trans Knowl Data Eng 26(8):1819–1837
45. Zhang NL, Zhang R, Chen T (2012) Discovery of regularities in the use of herbs in traditional Chinese medicine prescriptions. In: Cao L, Huang JZ, Bailey J, Koh YS, Luo J (eds) New frontiers in applied data mining. Springer, Berlin, pp 353–360
46. Zhao Y, Dong Q, Chen H, Iraji A, Li Y, Makkie M, Kou Z, Liu T (2017) Constructing fine-granularity functional brain network atlases via deep convolutional autoencoder. Med Image Anal 42:200–211
47. Zhao X, Wu Y, Song G, Li Z, Zhang Y, Fan Y (2018) A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. Med Image Anal 43:98–111
48. Zheng G, Jiang M, Lu C, Lu A (2014) Prescription analysis and mining. Springer International Publishing, Cham, pp 97–109
49. Zhu X, Liu Y, Li Q, Zhang Y, Wen C (2019) Mining patterns of chinese medicinal prescription for diabetes mellitus based on therapeutic effect. Multimedia Tools and Applications

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Huiqiang Liao is currently a master candidate in the College of Computer Science and Engineering, South China University of Technology. His main research interests include deep learning and computer vision.

Guihua Wen received the Ph.D. degree in Computer Science and Engineering, South China University of Technology, and now is professor, doctoral supervisor at the School of Computer Science and Technology of South China University of Technology. His research area includes Cognitive affective computing, Machine Learning and data mining. He is also professor in chief of the data mining and machine learning laboratory at the South China University of Technology.

Yang Hu received the MA.Eng degree from Kunming University of Science and Technology in 2016, where he is currently pursuing the Ph.D degree in South China University of Technology, China. His research interests include neural network and biomedical information processing.

ChangJun Wang received the Ph.D. degree in Shanghai University of Traditional Chinese Medicine, and now is professor, doctoral supervisor at the School of Medicine of South China University of Technology. He is also the Director of the Department of traditional Chinese Medicine in Guangdong General Hospital.