

TPGen: Prescription Generation Using Knowledge-guided Translator

Chunyang Ruan

Shanghai International Studies University

Hua Luo (✉ loha_u@shisu.edu.cn)

Shanghai International Studies University

Yingpei Wu

Fudan University

Yun Yang

Longhua Hospital Shanghai University of Traditional Chinese Medicine

Research Article

Keywords: Prescription generation, Machine translation, Transformer, Medical decision

Posted Date: April 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-453032/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

TPGen: prescription generation using knowledge-guided translator

Chunyang Ruan¹, Hua Luo^{1*}, Yingpei Wu² and Yun Ying³

Abstract

Background: Prescriptions contain a lot of clinical information and play a pivotal role in the clinical diagnosis of Traditional Chinese Medicine (TCM), which is a combination of herb to treat the symptoms of a patient from decision-making of doctors. In the process of clinical decision-making, a large number of prescriptions have been invented and accumulated based on TCM theories. Mining complex and the regular relationships between symptoms and herbs in the prescriptions are important for both clinical practice and novel prescription development. Previous work used several machine learning methods to discover regularities and generate prescriptions but rarely used TCM knowledge to guide prescription generation and described why each herb is predicted for treating a symptom.

Methods: In this work, we employ a machine translation mechanism and propose a novel sequence-to-sequence (seq2seq) architecture termed TPGen to generate prescriptions. TPGen consisting of an encoder and a decoder is a well known framework for resolving the machine translation problem in the natural language processing (NLP) domain. We use the lite transformer and Bi-directional Gate Recurrent Units (Bi-GRUS) as a fundamental model in TPGen, and integrate TCM clinical knowledge to guide the model improvement termed TPGen+.

Results: We conduct extensive experiments on a public TCM dataset and clinical data. The experimental results demonstrate that our proposed model is effective and outperforms other state-of-the-art methods in TCM expert evaluation. The approach will be beneficial for clinical prescription discovery and diagnosis.

Keywords: Prescription generation; Machine translation; Transformer; Medical decision

Background

Traditional Chinese Medicine (TCM) is an important part of the Chinese nation's outstanding culture, and it has a long history of development. TCM has received increasing attention as a complementary approach or alternative to modern medicine for its poly-pharmacological effects [1, 2]. TCM prescriptions including the lung-cleaning and toxicity-excluding (LCTE) soup have played an important role in treating the ongoing COVID-19 in China [3, 4]. TCM prescriptions (generally called a formula) are made by combining existing crude herbs or minerals instead of developing new drugs. That is why TCM has achieved great success in response to recent epidemics such as SARS, H1N1, and Zika, and is playing a vital role in reducing the incidence rate and controlling the spread of COVID-19 [5]. An example TCM prescription in

dictionary of TCM Prescriptions [6] is given in Fig. 1. It has the herbs composition, proportion, usage, doses, and indication symptoms. Chinese electronic medical

【名称】	小柴胡汤剂
Prescription Name	Xiao Chaihu Decoction
【组成】	柴胡 (9克), 黄芩 (9克), 半夏 (6克), 炙甘草 (3克), 党参 (6克), 生姜片 (3片), 大枣 (3枚)
Composition Herbs	Bupleurum 9g, Scutellaria 9g, Pinellia 6g, Licorice 3g, Codonopsis 6g, Ginger 3 pieces, Jujube 3 pieces
【用法】	水煎服
Usage	Decocted in water for oral dose
【适应症】	寒热、心烦喜呕、口苦、咽干、目眩、舌苔薄白、弦脉
Indication Symptoms	Chill and fever alternation, Vexation and vomiting, Bitter taste, Dry throat, Dizzy, White and thin coating of the tongue, Stringy pulse

Figure 1: An example TCM prescription, including herbs composition, proportion, usage, doses and indication symptoms.

*Correspondence: loha_u@shisu.edu.cn

¹Department of Data Science and Big Data Technology, Shanghai International Studies University, Shanghai, China

Full list of author information is available at the end of the article

records (EMRs) contain much information about clinical diagnoses and treatment events [7]. As a special electronic medical record, in a prescription, herbs play different roles (jun-chen-zuo-shi) for treating different symptoms of a disease or syndrome, and regularities on the composition of the herbs in prescriptions and corresponding symptoms play a significant role for clinical treatment and novel prescription development [8]. Meanwhile, the induction of regularities from daily prescriptions is one of the main approaches for TCM doctor to define their clinical skills and empirical knowledge [9]. However, the complex pattern, i.e., symptoms are many-to-many related herbs, and the huge amount of clinical data make the detection of effective regularities difficult. This indicates that modeling the regularities that underlying matching patterns of herbs and symptoms and automatic generating prescriptions are necessary for the development of clinical modernization of TCM.

From the perspective of data, a data instance in a TCM prescription corpus consists of two parts, a set of symptoms, and a set of herbs, which means the herb combination can well cure the symptom set. To generalize to any symptom set, the prescription generation task focuses on modeling the relationships between symptoms and herbs, which is analogous to the native recommendation task that models the interactions between users and items. Artificial Intelligence (AI) is an emerging technology that could be used in learning features or interactions among lots of relationship-unknown objects [10,11]. With the development of machine learning in AI, the task of recognizing the pattern of prescriptions and automatic generating prescriptions has recently received the attention of the research community. At present, topic models have gained extensive attention in the above task. A common strategy used to study prescription generation is to regard each prescription as a "document" with several latent topics, i.e., treatment activities, in which symptoms and herbs are regarded as a group of "words" respectively [9,12,13]. It is worth noting that topic models can be readily extended if we have prior knowledge about specific elements [14–17], which can enhance the efficiency of prescribing regularities discovery [8,18]. Thanks to these advantages, topic models can not only recommend proper herbs but also explore their indications and compatibility [19]. Unfortunately, there are more herbs in common use and always exists high-order herb or symptom associations, which is computationally intractable for topic models to capture [20]. This makes them insufficient to be applied in the task of prescription generation.

Graph is an effective structure to model complex relation data, and has been successfully used in medical data analysis [21,22]. In addition to topic models,

graph embedding is a technique common in prescription generation, that is to obtain the low-dimensional representations of TCM entities, i.e., herbs and symptoms, and then calculate the relationships between them based on the embedding vector [?,23–26]. In recent years there has been considerable interest in prescription mining using graph-based deep learning techniques. Most of these approaches are focused on optimizing the Sequence to Sequence(Seq2Seq) model [27], Autoencoder model [28] and Recurrent Neural Network(RNN) [29] to do text-level or word-level representation learning for prescriptions. Then, researchers can use these representations to detect the associations between herbs and symptoms, and immediately employ machine learning methods, e.g. multi-label classification to automatically generate prescriptions [30]. However, the above deep learning models are initially designed for the euclidean space data and lack the interpretability and reasoning ability for the non-euclidean space graph data [31]. They still do not fully exploit the associations between herbs and symptoms and achieve promising prediction performance. Graph neural networks (GNNs) are the extension of neural networks on graph structure, which can capture both features of node and edge, and structures of graphs simultaneously. Due to its convincing performance and interpretability, GNNs have been widely applied in computational drug discovery and automatic diagnosis recently [32,33]. One of the earliest attempts to recommend herbs according to symptoms based GNN is the work SMGCN [31]. These work constructed many relationships networks and used GNN to learn symptoms and herbs embedding. At last, it employed multi-label loss to predict the probability score for all herbs with a symptom in the symptom set. This is a superior work for traditional deep learning-based model, but it ignore the TCM knowledge that "jun-chen-zuo-shi", which means the different herbs and symptoms have different weights. This limits the performance of SMGCN on prescription generation. Graph Attention Network (GAT), a novel GNN architecture, which recursively propagates the embeddings from a node's neighbors to obtain the node embedding, and adopts the attention mechanism to discriminate the importance of the neighbors [34]. Intuitively, we can utilize GAT to measure the relationships between herbs and symptoms reasonably.

On the other hand, inspired by the Chinese machine translation method [35] and the success of natural language processing (NLP) for relations predicting in bioinformatics [36], we propose to consider automatic prescription generation as a translation task. TCM Translator is the earliest to use machine translation model to complete the task of prescription gen-

eration. It is a good work but failed to comprehensively describe how a prescription is generated using TCM domain knowledge well. This makes the model lack of rationality and interpretability to a certain extent. For each prescription, a set of symptoms is regarded as a source sentence to be translated to the target sequence that is a group of herbs, via the RNN based Seq2Seq model, which achieves tremendous performance in the task of machine translation [37]. This model is consisting of an encoder and a decoder based on RNN which encodes the source sequence of tokens into the latent space and then iteratively decoded to each word in the target sentence. However, TCM prescriptions are not exactly like natural language sentences. Most prescriptions keep their own way of organizing the herbs and symptoms, which are highly correlated to each other but not strictly ordered. To overcome this problem in our work, we attempt to take Lite transformer [38], a lightweight transformer, as the encoder of sequence model without position encoding directly. When a doctor prescribes an herb, it is impossible to consider all the symptoms equally. The LSTM based sequence model has limitations in this condition. The key module in transformer is attention that endows our proposed method the ability of fully capturing the correlation of symptoms in different subspaces [20]. Moreover, transformer is a special case of GNN from the internal architecture of two models [1]. This characteristic can improve the performance of the model. Notably, doctors often ignore the herbs already used and the symptoms considered in the diagnosis. According to this clinical knowledge, we employ coverage mechanism [39] and masking mechanism [27] to simulate the process in prescription generation. To this end, we propose a new sequence model consisting of Lite Transformer and Bi-directional Gate Recurrent Units (Bi-GRU). The basic model termed TPGen, and we use a knowledge guided model termed TPGen+ that performs relationships modeling between symptoms and herbs for the prescription generation. We evaluate the proposed model on a real-world TCM dataset collecting from lung-cancer patients and public datasets. The experimental results demonstrate that TPGen is well-performed and capable of generating herbal prescriptions effectively and rationally, compared with state-of-art approaches for TCM prescription generation. The contributions of this paper can be summarized as follows:

- we denote prescription generation as a machine translation task and build a translator, i.e., transformer based seq2seq model to translate the symptoms to herbs.

- we adopt the Lite transformer as our encoder and use Bi-GRU as the decoder to predict the herbs sequentially in the task of prescription generation.
- We highlight the importance of TCM clinical knowledge and enhance the basic sequence model with coverage mechanism and masking mechanism, which improve the quality of herb prediction according to symptoms.
- We propose a feasible approach to assist TCM doctors to compose prescriptions with the deep neural network, which was approved by TCM doctors. This may make a profound impact on health care in real life.

The rest of the paper is organized as follows: Section II describes the problem definition and introduces the proposed TPGen for predicting prescription. Section III presents the experimental results for the comparison of baselines. Discussion and conclusion are given in Section IV.

1 Problem definition

The task of prescription generation aims to generate the herb set as the treatment to a specific symptom set. Given a large TCM prescription corpus that consists of K data samples, the i -th sample $(X^{(i)}, P^{(i)})$ contains a source text $x^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}\}$ that describes the symptoms of a patient, and $P^{(i)} = \{h_1^{(i)}, h_2^{(i)}, \dots, h_m^{(i)}\}$ represent m herbs that $h_1^{(i)}, h_2^{(i)}, \dots, h_m^{(i)}$ make up the prescription $P^{(i)}$. Hereafter, we use $S_n = \{s_1, s_2, \dots, s_n\}$ and $H_m = \{h_1, h_2, \dots, h_m\}$ represent the set of symptoms and the group of herbs respectively. In this way, each prescription is denoted by $P = \langle S_n, H_m \rangle$.

We view the set of symptoms S_n as a sequence of characters. Given S_n , our task is to gain the n -dimensional probability vector, where the value of dimension d denotes a probability that the d -th herb can cure S_n . This is achieved by a learned prediction function $y_{S_n} = f(S_n | H_m; \theta)$, where y_{S_n} represents the probability vector; H_m is the labels, i.e., herbs, to the specific symptoms. And, θ indicates trainable parameters of function f . The training procedure is shown in Fig. 2.

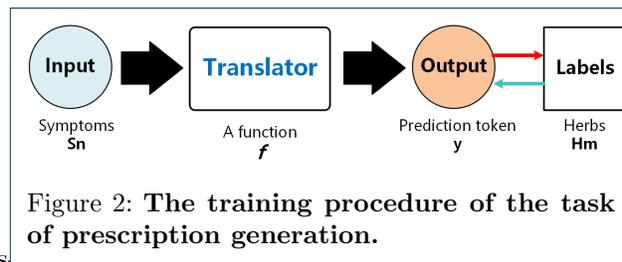


Figure 2: The training procedure of the task of prescription generation.

[1] <https://graphdeeplearning.github.io/post/transformers-are-gnns/>

Methods

Seq2seq model has proven to be superior to another neural network models in a wide range of natural language generation tasks [40], including machine translation that is related to our task. In this section, we first describe the definition of the TCM prescription generation, then we introduce how to apply translator model in the prescription prediction, at last we show how to enhance the model to generate more diverse herbs in the specific setting of this task by introducing attention and coverage mechanism. An overview of the basic model is shown in Fig. 3.

1.1 Fundamental translator

Sequence-to-sequence model was first proposed to solve machine translation problem [41]. The model consists of two parts, an Encoder and a Decoder. Seq2seq model predicts the next word according to the source statement and previous output every time, so its output can be regarded as a conditional language model. It is modeling to generate the optimal sequence that maximizes the conditional probability $p(y | x)$:

$$p_r(y | x) = \prod_{t=1}^n p_r(y_t | y_1, y_2, \dots, y_{t-1}, x) \quad (1)$$

we first consider represent the input each words, i.e., symptom as a embedding vector via word2vec [42]. Then, the encoder of Seq2seq converts the embedding vector of symptoms into a set of hidden representation $\hat{h} = \{\hat{h}_1, \hat{h}_2, \dots, \hat{h}_t\}$ in the latent space iteratively. Next, all latent feature about symptoms are aggregated into the context vector C , which acting as the representation of the whole source through a nonlinear function f :

$$C = \psi(\hat{h}_1, \hat{h}_2, \dots, \hat{h}_t) \quad (2)$$

where ψ can be softmax [43]. Afterwards, a decoder generates a variable-length sequence $y = (y_1, y_2, \dots, y_t)$ token by the corresponding label herbs h_1, h_2, \dots, h_m via a conditional language model:

$$\hat{s}_t = f(\hat{s}_{t-1}, c_t) \quad (3)$$

$$p_r(y_t | y_1, \dots, t, x) = g(\hat{s}_t) \quad (4)$$

where t is the hidden state of the decoder at the time step t . f can be a LSTM. The g represent a nonlinear functions, which can be a classifier. The classifier outputs the probabilities of all the herbs in the herb vocabulary. Although this fundamental translator can complete the task of prescription generation, it does not perform well due to the particularity of

TCM data. To solve this issue, we need TCM-clinical knowledge to guide the construction and train of this translator.

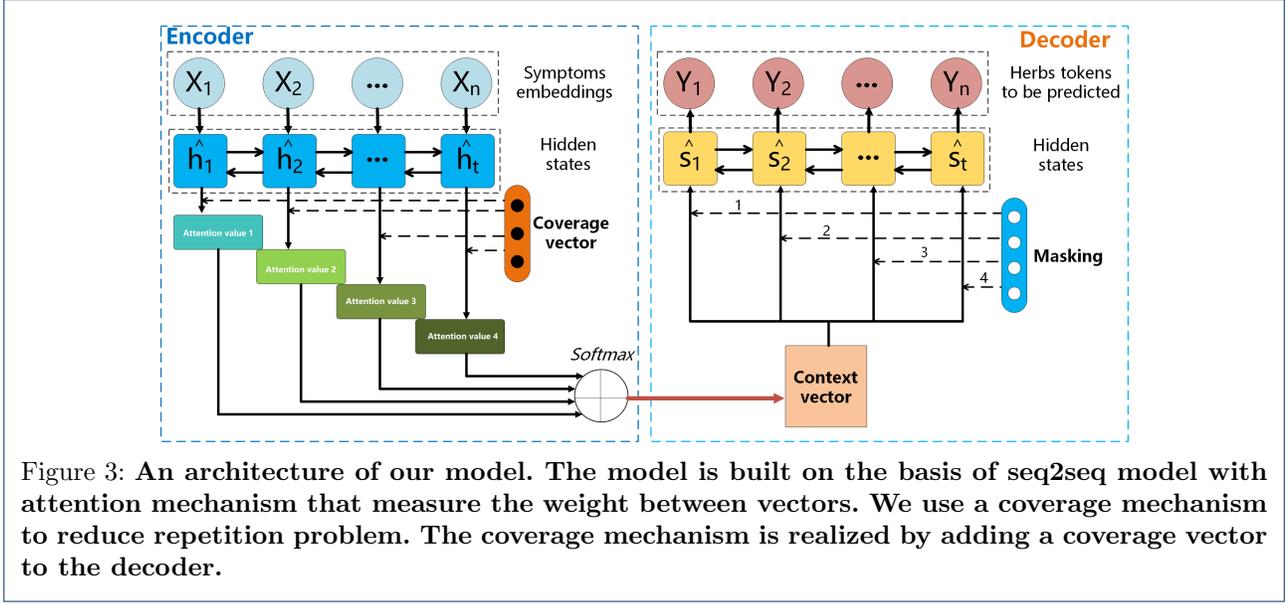
1.2 Knowledge-guided translator

Due to the lack of formalization and the individual differences of TCM doctors. TCM prescription is quite different from other natural language corpus. Guided by TCM experts, we summarize this knowledge and use it to optimize the proposed model, which is the biggest difference from the naive TCM translator [20].

- i. The two herbs in a pair would appear in the same prescription.
- ii. The order of the words is very important which is strictly restricted by syntax and grammar, while in prescriptions, the very last generated herb token is not directly related to the current one.
- iii. Different herb are often related to the same symptom of the patient.
- iv. There is no duplicated herb (label) in the TCM prescription generation task.

For the **knowledge-i**, we attempt to beam-search [44], optimized strategies for Seq2Seq model. According to **knowledge-ii** that the target sequence of the herbs are not in strict order. So, improvement of beam-search method is not obvious. To analyze above issue and consult TCM experts, we regard the **knowledge-i** and **knowledge-ii** the global and local contexts separately. Inspired by Lite Transformer [38], we utilize the scaled dot-product attention [45] and the convolutional module for text generation [46] side by side, encouraging them to have a different perspective of the token, i.e. herbs globally and locally. The architecture of Lite Transformer reduces the amount of parameters and calculation compared with traditional transformer model and achieve better efficiency. The encoder of Lite Transformer for our task is shown in Fig. 4. Moreover, **knowledge-iii** means the different symptoms and herbs contain different relationship. The weight calculated by attention can illustrate the relevance between the words, i.e., symptoms from the source sentence and the words, i.e., herbs from the target sentence. The scaled dot-product attention receives three inputs, a query Q , key K and the corresponding value V , which are all the embedding vectors from the observed symptoms S_n . Given a query Q and key K , the dot-product between them is first calculated and then divided by a scaling factor d model to guarantee the stable gradient in back-propagation of the model training process. A softmax function is applied after to obtain the attention distribution over values V are defined as follows:

$$\text{ATTN}(Q, K, V) = \psi \left(\frac{QK^T}{\sqrt{d}} \right) V \quad (5)$$



where ψ is a softmax function, $\sqrt{d'}$ is the dimension of Q , K and V embedding vectors. And, we use a 1-dimensional convolution kernel to get the local attention. Thereafter, we contact the above attention value and normalize it, which is formulated as follows:

$$C' = \text{Concat}(\text{ATTN}_{global}, \text{ATTN}_{local}) \quad (6)$$

$$C'_N = \text{Norm}(C' + X) \quad (7)$$

Then we utilize two 1-dimensional convolution layers with kernel size 1, followed the activation function \tanh :

$$\text{FFN}(x) = \tanh(xW_1 + b_2)W_2 + b_2 \quad (8)$$

where $W_1 \in \mathbb{R}^{d' \times 2}$ and $W_2 \in \mathbb{R}^{d'}$. We use the FFN and normalization layer to gain the final representation vectors C from encoder:

$$\tilde{C} = \text{FFN}(C'_N) \quad (9)$$

$$C = \text{Norm}(C'_N + \tilde{C}) \quad (10)$$

Embedding via attention, convolution and feed-forward network, the correlation and characteristic of symptoms can be captured and stored in low-dimensional representations in the latent space. These vectors bring a lot of valuable information to the individual body conditions and are deciphered to the suitable herbs sequentially via the following decoder.

According to **knowledge-ii** and **knowledge-iv**, Bidirectional Recurrent Networks with Gated Recurrent Unit (GRU) [47], simple structure and few parameters compared with LSTM, is used as the decoder to

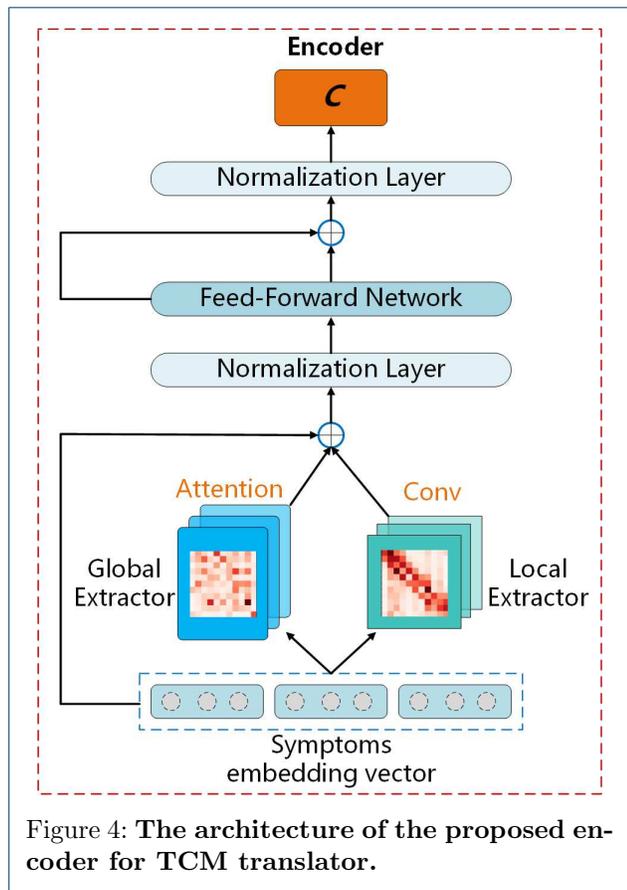
map the representation vectors to the labels one by one. Because this model can be better able to model long distance dependency and solve the problem of back bias. Fig. 5 gives the architecture of decoder in our TCM translator. The first GRU takes the context vector C as input to predict the first label through the softmax function, which is then embedded and used to the next label prediction. Repeat this process until all labels are generated. In each step, GRU is formulated as follows:

$$\begin{aligned} z^{(t)} &= \sigma(W^{(z)}x^{(t)} + U^{(z)}\hat{s}^{(t-1)}) \\ r^{(t)} &= \sigma(W^{(r)}x^{(t)} + U^{(r)}\hat{s}^{(t-1)}) \\ \tilde{s}^{(t)} &= \tanh(Wx^{(t)} + r^{(t)} \circ U\hat{s}^{(t-1)}) \\ \hat{s}^{(t)} &= (1 - z^{(t)}) \circ \tilde{s}^{(t)} + z^{(t)} \circ \hat{s}^{(t-1)} \end{aligned} \quad (11)$$

where $z^{(t)}$, $r^{(t)}$ and $\tilde{s}^{(t)}$ are update gate, reset gate and new memory cell respectively, $W^{(z)}$, $U^{(z)}$, $W^{(r)}$, $U^{(r)}$, $Wx^{(t)}$, $r^{(t)}$ and U are the parameter matrix to be solved. s_t is the output hidden state at the time t , and \circ represents element-wise multiplication.

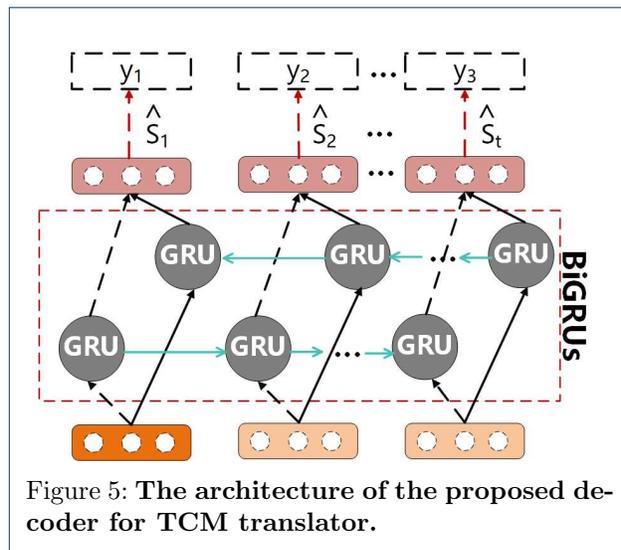
Follow **knowledge-iii** and **knowledge-iv** to instruct our decoder to output more diverse and sensible herbs token, we employ masking mechanism and coverage mechanism to optimize our model. According to masking mechanism, we first mask the probability of predicted token to zero before generating a new token in the inference stage. In other words, at time step t of decoding, the probability of token that have been already in the generated sequence would be manually degraded as follows:

$$pos = 0, \quad pos \in S_t \quad (12)$$



where pos is the index of the herb, S_t is the group of tokens that have already been generated until the time step t . Please note that masking mechanism influences the normal training process by reducing the punishment of repeated tokens in the inference stage.

The input words, i.e., symptoms, in machine translation, some words are over translated. That is to say, another translation is generated for a word that has been translated. This limit our model to achieve a one-to-one mapping between symptoms and herbs, which result in get multiple herbs in the prescription that treat the same symptom. This problem is known clinically as over-treatment. On the contrary, some symptoms have not been prescribed the corresponding herbs as a result of the lack of translation of words. Above problem of over-translation and under-translation has a great impact on the quality of predicted prescriptions, which can not meet the needs of clinical diagnosis. So, we use coverage mechanism to solve problem, which can help the decoder focus on the part that has not been shed much attention by feeding a fertility vector to the attention calculation. If the generated herbs from decoder is for specific symptoms, the coverage vector will lower the local attention weight of these symptoms, otherwise, it will be increased. So, Equation 3 is



modified by the following equation.

$$\begin{aligned} cov_t &= \tanh(W S_t + b) \\ s_t &= f(\hat{s}_{t-1}, c_t, cov_t) \end{aligned} \quad (13)$$

where cov_t is the coverage vector at the t -th time step of decoder. S_t is the representation of the generated tokens until the t -th time step. W is a learnable parameter matrix. Giving the cover vector to the GRU unit, our model can softly switch more probability to the herbs that have not been predicted. In this way, the model is encouraged to generate novel herbs rather than repeatedly predict the frequently observed ones, thus improve the recall of our model.

The encoder and decoder networks are trained jointly to maximize the conditional probability of the label, i.e., herbs.

Results&Discussion

In this section, we evaluate our proposed model on the benchmark. We first introduce the TCM datasets, experimental setup, metrics, and baselines. Then the experimental results are demonstrated in detail. Last, we will discuss the results.

1.3 Datasets

We verify our proposed model TPGen on a real-world dataset *TLG*, collected from lung cancer patients in clinic, and on a public dataset *CKCEST* [2] [8]. Our dataset consists of 10000+ TCM prescriptions targeted to lung cancer provided by the cooperative hospital. The number of the total sample is 10052, in which the dimension of input data, i.e., the symptom, is 189 and

[2]<https://github.com/yao8839836/PTM>

the number of the output labels, i.e., herbs, is 357. The statistics of the TCM datasets is summarized in Tab. 1. An example along with the dataset *TLD* is shown in Fig. 6. We also give the frequency-distribution of herbs in the clinical dataset *TLD* is shown in Fig. 1.

Table 1: **The statistics of the prescription datasets.**

Datasets	prescriptions	Symptoms	Herbs
TLD	10052	189	357
CKCEST	26360	360	753

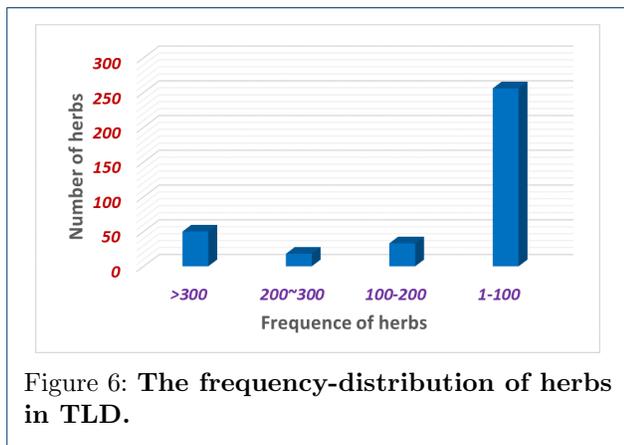


Figure 6: **The frequency-distribution of herbs in TLD.**

1.4 Experimental setup

We set the embedding size of both Chinese characters in the symptoms and the herb tokens to 512. We set the hidden state size of GRU to 300, and set the batch size to 50. We set the maximum length of the herb sequence to 22.

1.5 Results and Analysis

1.5.1 Quantitative evaluation

To test our proposed models as well as the baseline models, we evaluate these four models on the above datasets, *TLD* and *CKCEST*. Notably, in order to verify the combination of domain knowledge improve our model to a higher level actually, we split our model to basic model TPGen and knowledge fusion model TPGen+. Tab. 2 3 4 list the comparison results that recall, precision, and F1 score for each models get respectively. We implement our approach and the comparative methods using PyTorch. Our model is deployed in the server Linux 4.15.1-041501-generic x86 64 that hardware configuration including $4 \times$ CPU Intel(R) Core(TM) i9-7900X CPU@3.30GHz, $4 \times$ GPU NVIDIA Geforce GTX 1080 ti, 4T hard disk and $2 \times 64G$ memory.

In the training stage, multi-classification cross-entropy is taken as our loss function, which defined as follows:

$$Loss = - \sum_{i=0}^{C-1} y_i \log(p_i) = - \log(p_c) \quad (14)$$

where $p = [p_0, \dots, p_{C-1}]$ is a probability distribution, each p_i indicates the probability of sample belongs to class i ; $y = [y_0, \dots, y_{C-1}]$ is the one-hot representation of sample with label, if sample belong to the class i , $y_i = 1$, otherwise $y_i = 0$; c is the label. The Adam algorithm [48] is employed to train our model.

In order to obtain the reasonable performance comparison, we randomly split each dataset to three parts, 70% training set, 20% evaluation set and 10% testing set, is used to evaluate the performance with 5 repetitions. To evaluate the performance of our approach, we adopt the following three measures, *precision@k*, *recall@k* and *NDCG@k* [20] commonly used in the task of recommendation and prediction.

1.6 Baselines

We adopt the following approaches for comparison. They They are from topic models and deep learning models.

- **PTM** [8]: It proposes a novel topic model integrating knowledge to capture the correlations of symptoms, diseases, and herbs. We set the number of topics is 20 with 5 herbs predicted as best parameters set for comparison.
- **TCM Translator** [20]: It designs a novel Seq2Seq model based on original transformer and LSTM to characterizes the generative process of prescriptions.
- **SHR** [31] It use a series of GCNs to simultaneously learn the symptom embedding and herb embedding from the relationships of symptom-herb, and recommend herbs with syndrome induction from symptoms.

As shown, our proposed model TPGen+ performs the best compared to other baseline models. And, our model with knowledge also is close to the second best model performs TCM Translator. Two deep-learning based models performance exceeds the commonly used topic model approaches, PTM, with certain advantages, indicating that catching the relationships of between word and labels is necessary to make a multiple-labels classifier. On the other, the TPGen performs slightly better than TCM Translator and TPGen, in which herbs compatibility is included in calculation, which can help TPGen to know the label correlations and TCM knowledge effectively. PTM as a topic models easier integrate different information into computation, such as prior knowledge, which makes its performance has more room for improvement. TPGen+

Table 2: The $precision@k$ for each model on two TCM datasets.

Methods	TLD			CKCEST		
	precision%10	precision%20	precision%50	precision%10	precision%20	precision%50
PTM	0.273	0.228	0.204	0.266	0.205	0.159
PTM	0.275	0.225	0.203	0.269	0.226	0.163
TCM	0.278	0.211	0.202	0.271	0.223	0.193
Translator						
TPGen	0.273	0.252	0.200	0.270	0.241	0.194
TPGen+	0.291	0.263	0.221	0.287	0.251	0.208
%Improv. by +	6.11%	4.12%	9.50%	5.91%	4.01%	6.73%

Table 3: The $recall@k$ for each model on two TCM datasets.

Methods	TLD			CKCEST		
	recall%10	recall%20	recall%50	recall%10	recall%20	recall%50
PTM	0.192	0.314	0.435	0.186	0.292	0.409
SHR	0.193	0.317	0.436	0.183	0.291	0.411
TCM	0.194	0.320	0.443	0.187	0.302	0.412
Translator						
TPGen	0.191	0.321	0.439	0.185	0.297	0.409
TPGen+	0.202	0.341	0.450	0.197	0.312	0.416
%Improv. by +	5.41%	5.80%	2.41%	6.02%	4.81%	1.62%

performs better than another approach, which reflects that its ability for capturing the relationships between symptoms and herbs is much better than others.

Moreover, we test the efficiency of different model to evaluate the complexity and practicability of the algorithm. The average running time of different model under different sample size are shown in Fig. 7. As shown in Fig. 7, our proposed model got the best running time compared with others. PTM took less time than the TCM Translator model. The TCM Translator based on transformer and LSTM needs high computational power to achieve sufficient performance. But, our proposed model based on lite-transformer greatly reduces the running time, which improves the usability of the model for clinical diagnosis supporting.

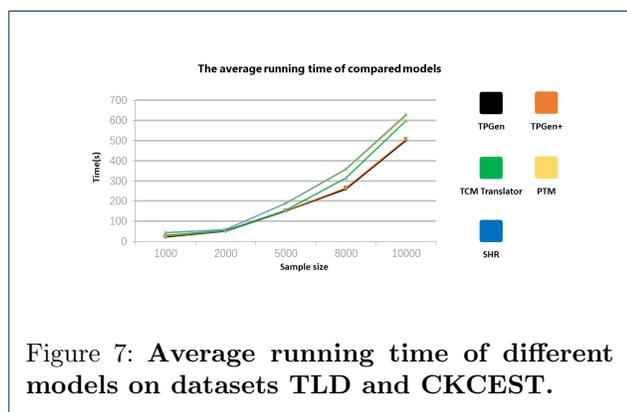
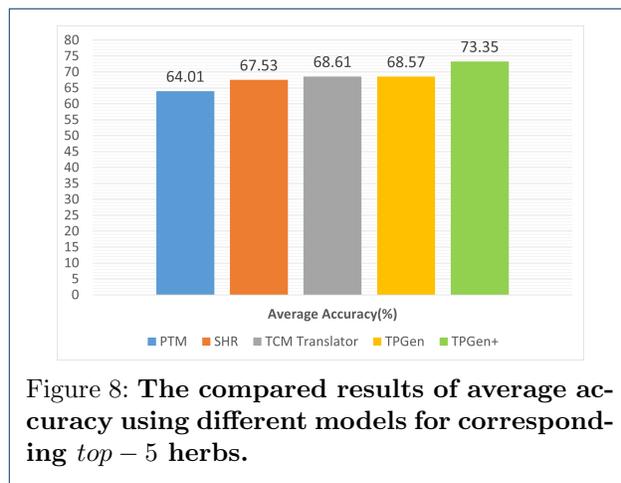


Figure 7: Average running time of different models on datasets TLD and CKCEST.

1.6.1 Qualitative evaluation

In order to show the different ability of relationships detecting more intuitively, we test the average accuracy that matching degree between the weight of the relationships given by all models and the result given by the TCM experts. TCM experts chose seven classical symptoms and calculated their corresponding $top-5$ herbs that the most relevant to them. As shown in Fig. 8, we found the results predicted by our model are closer to the benchmark given by the experts. The prediction results of PTM and TCM Translator are very close, but that lower than our model gave. And,

Figure 8: The compared results of average accuracy using different models for corresponding $top-5$ herbs.

we gave the thermodynamic chart of herb-symptom relationships from our proposed model are shown in Fig. 9. As shown in Fig. 9, the weights illustrate the relationships between the symptoms from the source

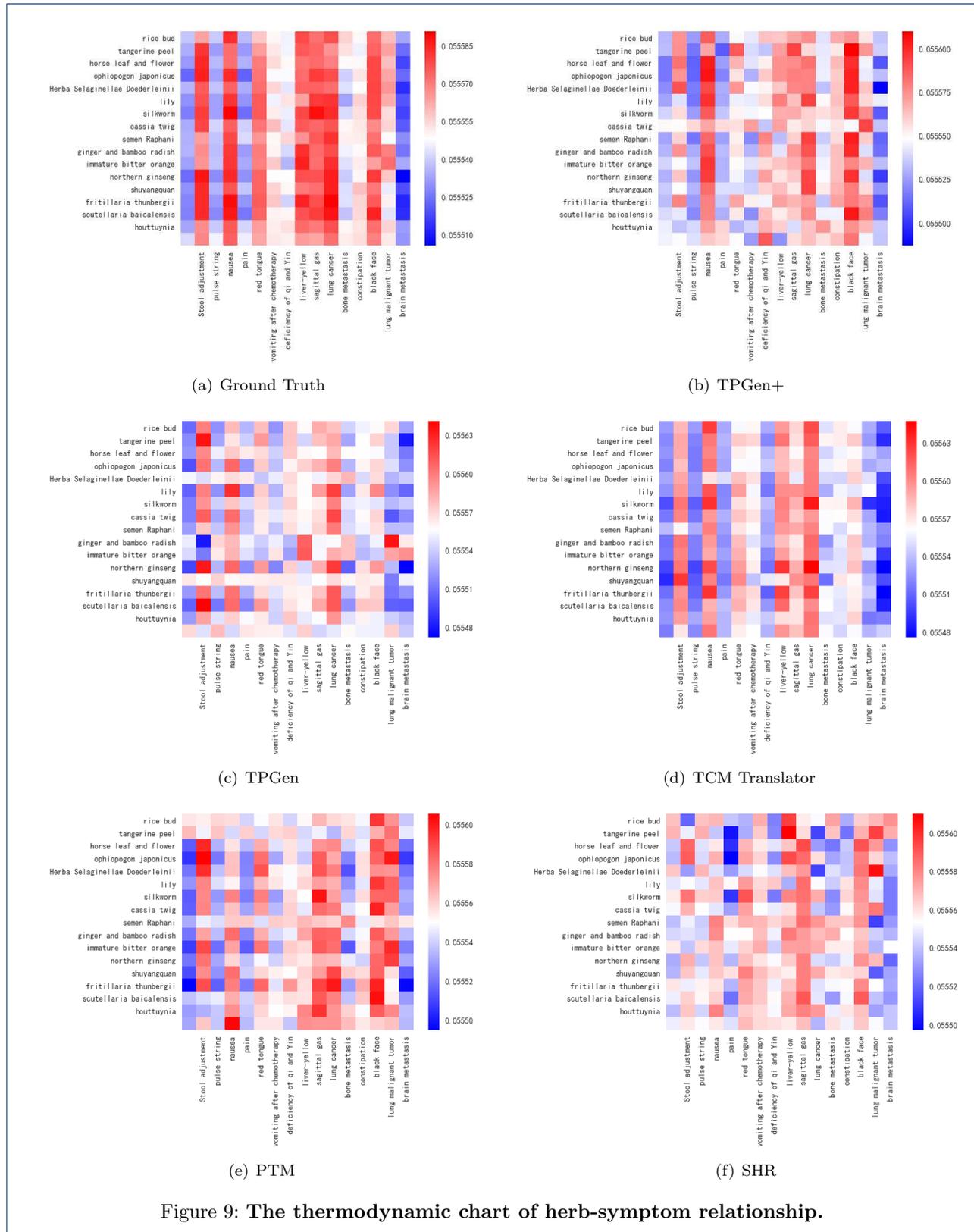


Figure 9: The thermodynamic chart of herb-symptom relationship.

Table 4: The $F - score@k$ for each model on two TCM datasets.

Methods	TLD			CKCEST		
	ndcg%10	ndcg%20	ndcg%50	ndcg%10	ndcg%20	ndcg%50
PTM	0.352	0.415	0.515	0.346	0.407	0.498
SHR	0.356	0.413	0.517	0.352	0.405	0.508
TCM	0.360	0.437	0.531	0.381	0.413	0.505
Translator						
TPGen	0.351	0.435	0.489	0.378	0.412	0.502
TPGen+	0.375	0.451	0.534	0.390	0.423	0.512
%Improv. by +	6.41%	3.52%	6.11%	3.01%	2.61%	1.93%

sequence and the herbs from the target sequence. The blue area in the bar represents the no-relevant, and the red area in the bar represents the relevant. In each area, a darker color, a larger weight. For example, the i -th symptom in the source sequence pays more attention to the j -th herb in the target sequence in the dark red area. It is easy to see from the visualization that the results of our proposed model are closer to the ground truth. In addition, our proposed and TCM Translator all perform better than the topic model based PTM. Because the transformer model use attention to encode words, which is the multiplication of several matrices, and the intermediate output is the weight coefficient similar to the vector compression. Using weight coefficient matrix from the program, we can see at any time which herbs are generated and which symptoms are considered emphatically. In short, the transformer model has a good explanation for the generation of TCM prescriptions. This is very important in the current deep learning research for clinic, because the biggest problem of many deep learning models is that they are insufficiently interpretable.

To summarize, we can confidently state that our proposed model has taken the symptom and label correlations into computation, and perfectly match the automatic prescription generation task. Moreover, it is easy to be extended in the future that any valuable information such as prior knowledge and experience can be injected into the whole architecture through embedding method.

1.6.2 Case Study

In this part, we design a case study to verify the power of our proposed prescription generation model. To better show the feasibility of our best model TPGen+ in TCM clinical applications, we gives a group of herbs prescribed by TCM clinical experts as ground truth, and we use the herbs predicted by TPGen+ to match it. The herbs prediction cases and ground truth are shown in Tab. 5, and the bold black font indicates the common herbs between the herbs predicted by TPGen+ and the ground truth in the Herbs column. We use the common metrics *Precision*, *Recall* and *F-1* to

Symptoms	TPGen+	Herbs
1. Profuse phlegm. 2. loose stools. 3. fat enlarged tongue. 4. slimy white fur. 5. slippery pulse	medicinal changium root. indian buead. liquorice. parched white atractylodes rhizome. pinellia tuber. magnolia officinalis. Chinese yam. fructus amomi. amomum cardamomum. baked ginger. fructus aurantii. golden thread. coconut. Chinese sage herb. doederlein's spikemoss herb. chicken's gizzard-membrane. charred fructus and scorched medicated leaven	Ground Truth medicinal changium root. indian buead. parched white atractylodes rhizome. semen dolichoris. magnolia officinalis. Chinese yam. doederlein's Spikemoss herb. Chinese sage herb. ternate buttercup root. fructus amomi. amomum cardamomum. fingered citron. charred fructus crataegi and scorched medicated leaven. liquorice
	Precision=0.785	Recall=0.687
		F1=0.733

Table 5: Actual predictions made by TPGen+ and ground truth matching.

evaluate the performance of TPGen+. As shown in Tab. 5, we can observe that the result of our model is very close to the reality. According to the herbal knowledge, the missing herbs actually have similar functions with the remaining ground-truth herbs and can be alternatives in clinical practice sometimes.

We give an example generated our proposed TPGen and TPGen+. Three clinical doctors score the two generated prescription. The score range is 0 to 10. Compared with the ground truth mapping, the experts' assessment is more reasonable and closer to reality. Given the symptom set, our proposed TPGen generates a group of herbs to cure the listed symptoms. In the prescription column, the bold black font indicates the common herbs between the TPGen and TPGen+. Input symptoms and output herbs are shown in the Tab. 6. *TPGen+* got a better perform than *TPGen*. From the results, it can be seen that after knowledge fusion, the performance of our model is improved really. For the first set of symptoms, *TPGen+* generate some new herbs. According to the herbal knowledge, the missing herbs actually have similar functions with the remaining herbs and can be alternatives in clinical practice. For the second set of symptoms, *TPGen+* replace four herbs from *TPGen* as other three herbs. Because these four herbs are more suitable for related symptoms than the other three herbs, which have the same effect of medicine. In summary, a case study is given to demonstrate that our proposed model can provide reasonable and suitable herbs, which show the ef-

Table 6: The difference and intersection herbs prescribed by our proposed *TPGen* and *TPGen+* according to two clinical symptoms set of lung tumor.

Symptoms	Prescriptions		Symptoms	Prescriptions	
	TPGen	TPGen+		TPGen	TPGen+
1.cough, 2.fever, 3.phlegm, 4.sticky pulse, 5.strings fur, 6.greasy tongue, 7.pale red	Raw Astragalus, Radix Glehniae, Radix Asparagi, Salviae Chinensis Selaginella, Houttuynia Cordata, Scutellaria Baicalensis, Hedyotis Diffusa, Fructus Ligustri Lucidi, Radix Ophiopogonis, Dried Orange Peel, Pinellia, Endothelium Corneum Gigeriae Galli, Poria Cocos, Radix Bupleuri, Bamboo Shavings, Artemisia Annua, Edible Tulip, Medicated Leaven, Semen Cuscutae, Gynostemma Pentaphyllum, Platycodon Grandiflorum, Herba Epimedii, Fritillaria thun-bergli, Yam Rhizome	Houttuynia Cordata, Scutellaria Baicalensis, Hedyotis Diffusa, Salviae Chinensis, Radix Glehniae, Rice Bud, Fructus Ligustri Lucidi, Radix Ophiopogonis, Dried Orange Peel, Selaginella, Pinellia, Semen Raphani, Endothelium Corneum Gigeriae Galli, Poria Cocos, Radix Bupleuri, Bamboo Shavings, Artemisia Annua, Edible Tulip, Green Jade Powder, Verbenae, Medicated Leaven, Semen Cuscutae, Gynostemma Pentaphyllum, Raw Astragalus, Radix Asparagi, Platycodon Grandiflorum, Rhizoma Imperata, Herba Epimedii, Chinese Yew, Buckwheat, Magnolia Officinalis, Rhizoma Atractylodis, Radix Stemonae, Fritillaria Thun-bergli, Rhizoma Fagopyri Cymos, Yam Rhizome	1.cough, 2.thin pulse, 3.thin white coating, 4.epigastric discomfort, 5.light tongue, 6.white phlegm	Radix Glehniae, Radix Ophiopogonis, Selaginella, Salviae Chinensis Dried Orange Peel, Pinellia, Houttuynia Cordata, Hedyotis Diffusa, Endothelium Corneum Gigeriae Galli, Poria Cocos, Edible Tulip, Atractylodes Macrocephala, Semen Cuscutae, Gynostemma Pentaphyllum, Almond, Raw Astragalus, Platycodon Grandiflorum, Herba Epimedii, Fritillaria Thun-bergli, Raw Rice Kernel, Yam Rhizome, Aster	Houttuynia Cordata, Hedyotis Diffusa, Salviae Chinensis Radix Glehniae, Radix Ophiopogonis, Selaginella, Pinellia, Endothelium Corneum Gigeriae Galli, Poria Cocos, Edible Tulip, Atractylodes Macrocephala, Semen Cuscutae, Gynostemma Pentaphyllum, Almond, Raw Astragalus, Platycodon Grandiflorum, Herba Epimedii, Taxus chinensis, Magnolia Officinalis, Amomum Tsao-ko
Score of TPGen=67.1	Score of TPGen+=79.6		Score of TPGen=75.3	Score of TPGen+=81.5	

fectiveness of our models in real-world applications for TCM clinic.

Conclusion

Generating prescriptions automatically is an important issue for TCM clinical future research. This paper presented a novel translator model guided by TCM knowledge for TCM prescription generation. It characterizes the generative process of prescriptions in TCM clinical knowledge. Our proposed model can discover the prescribing patterns in TCM. Furthermore, it can outperform several state-of-art approaches to generating herbs for a set of symptoms. Our method is helpful for clinical research and practice was verified by experiments and TCM experts.

In future work, we plan to incorporate more prescription information such as usage, contraindications, and dosage of herbs, and more domain knowledge such as specific syndrome for symptoms as prior knowledge into our model. And, we will improve the embedding quality of the TCM entities by adopting advanced techniques such as Temporal Convolutional Network [49] and BERT [50]. We hope our proposed model can lay a good baseline and encourage more researchers to pay attention on TCM intelligent diagnosis.

Acknowledgment

We gratefully acknowledge all reviewers for their thoughtful comments and efforts towards improv-

ing our manuscript and the clinical data providers, Pro.Tian and Dr.Yang, for their hard work.

Funding

This work was supported in part by the National Natural Science Foundation of China (No.61672161), Youth Research Fund of Shanghai municipal health and Family Planning Commission (No.2015Y0195), and Youth Research Fund of Shanghai International Studies University (No.2020114096).

Author details

¹Department of Data Science and Big Data Technology, Shanghai International Studies University, Shanghai, China. ²School of Computer Science, Fudan University, Shanghai, China. ³Department of Oncology, Longhua Hospital Shanghai University of Traditional Chinese Medicine, Shanghai, China.

References

- Cheung, F.: Tcm: made in china. *Nature* **480**(7378), 82–3 (2011)
- Wei, J., Wang, J., Zhu, Y., Sun, J., Xu, H., Li, M.: Traditional chinese medicine pharmacovigilance in signal detection: decision tree-based data classification. *BMC Medical Informatics Decis. Mak.* **18**(1), 19 (2018)
- Xu, X., Shi, Y., Wang, R., Ting Liu Jingming Xu, Q.S. Wen Mao: Home-based traditional chinese medicine nursing interventions for discharged patients with covid-19: A rapid review of chinese guidelines. *Integrative Medicine Research* (2020)
- Zhang, D.-h., Zhang, X., Peng, B., Deng, S.-q., Wang, Y.-f., Yang, L., Zhang, K.-z., Ling, C.-q., Wu, K.-l.: Network pharmacology suggests biochemical rationale for treating covid-19 symptoms with a traditional chinese medicine. *Communications Biology* **3**(1), 1–9 (2020)
- Zheng, Y.-J., Yu, S.-L., Yang, J.-C., Gan, T.-E., Song, Q., Yang, J., Karatas, M.: Intelligent optimization of diversified community prevention of covid-19 using traditional chinese medicine. *arXiv preprint arXiv:2007.13926* (2020)
- Hailan, B., Lingling, L.I., Wei, X.U., Tiange, Q.U., Hui, M.A., Xingwu, D., Hospital, D.: Data mining of psoriasis prescription for external use:

- Based on dictionary of traditional chinese medicine prescriptions. *Journal of Traditional Chinese Medicine* (2019)
7. Zhang, T., Wang, Y., Wang, X., Yang, Y., Ye, Y.: Constructing fine-grained entity recognition corpora based on clinical records of traditional chinese medicine. *BMC Medical Informatics and Decision Making* **20** (2020)
 8. Yao, L., Zhang, Y., Wei, B., Zhang, W., Jin, Z.: A topic modeling approach for traditional chinese medicine prescriptions. *IEEE Transactions on Knowledge and Data Engineering* **30**(6), 1007–1021 (2018)
 9. Zhang, X.-p., Zhou, X.-z., Huang, H.-k., Feng, Q., Chen, S.-b., Liu, B.-y.: Topic model for chinese medicine diagnosis and prescription regularities analysis: case on diabetes. *Chinese journal of integrative medicine* **17**(4), 307–313 (2011)
 10. Wang, Z., Li, L., Yan, J., Yao, Y.: Approaching high-accuracy side effect prediction of traditional chinese medicine compound prescription using network embedding and deep learning. *IEEE Access* **8**, 82493–82499 (2020)
 11. Goh, W.P., Tao, X., Zhang, J., Yong, J.: Feature-based learning in drug prescription system for medical clinics. *Neural Process. Lett.* **52**(3), 1703–1721 (2020)
 12. Yao, L., Zhang, Y., Wei, B., Wang, W., Zhang, Y., Ren, X., Bian, Y.: Discovering treatment pattern in traditional chinese medicine clinical cases by exploiting supervised topic model and domain knowledge. *Journal of biomedical informatics* **58**, 260–267 (2015)
 13. Ji, W., Zhang, Y., Wang, X., Zhou, Y.: Latent semantic diagnosis in traditional chinese medicine. *World Wide Web* **20**(5), 1071–1087 (2017)
 14. Xue, F., Hong, R., He, X., Wang, J., Qian, S., Xu, C.: Knowledge based topic model for multi-modal social event analysis. *IEEE Transactions on Multimedia* (2019)
 15. Gurcan, F., Cagiltay, N.E.: Big data software engineering: Analysis of knowledge domains and skill sets using lda-based topic modeling. *IEEE Access* **7**, 82541–82552 (2019)
 16. Zhang, Y., Mao, W., Zeng, D.: A non-parametric topic model for short texts incorporating word coherence knowledge. In: *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*, pp. 2017–2020 (2016)
 17. Wang, S., Huang, E.W., Zhang, R., Zhang, X., Liu, B., Zhou, X., Zhai, C.: A conditional probabilistic model for joint analysis of symptoms, diseases, and herbs in traditional chinese medicine patient records. In: *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 411–418 (2016). IEEE
 18. Ma, J., Zhang, Y.: Using topic model for intelligent computer-aided diagnosis in traditional chinese medicine. In: *2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS)*, pp. 504–507 (2016). IEEE
 19. Wang, S., Hu, Y., Tan, W., Wu, X., Chen, R., Cao, J., Chen, M., Wang, Y.: Compatibility art of traditional chinese medicine: from the perspective of herb pairs. *Journal of Ethnopharmacology* **143**(2), 412–423 (2012)
 20. Wang, Z., Poon, J., Poon, S.: Tcm translator: A sequence generation approach for prescribing herbal medicines. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2474–2480 (2019). IEEE
 21. Fan, J., Chow, T.W.S., Zhao, M., Ho, J.K.L.: Nonlinear dimensionality reduction for data with disconnected neighborhood graph. *Neural Process. Lett.* **47**(2), 697–716 (2018)
 22. Campanharo, A.S.L.O., Doescher, E., Ramos, F.M.: Application of quantile graphs to the automated analysis of EEG signals. *Neural Process. Lett.* **52**(1), 5–20 (2020)
 23. Huang, E.W., Wang, S., Zhang, R., Liu, B., Zhou, X., Zhai, C.: Parecat: Patient record subcategorization for precision traditional chinese medicine. In: *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 443–452 (2016)
 24. Ning, Wang, Peng, Li, Xiaochen, Hu, Kuo, Yang, Yonghong, and, Q.: Herb target prediction based on representation learning of symptom related heterogeneous network. *Computational Structural Biotechnology Journal* (2019)
 25. Chen, X., Ruan, C., Zhang, Y., Chen, H.: Heterogeneous information network based clustering for precision traditional chinese medicine. *BMC Medical Informatics Decis. Mak.* **19-S**(6), 264 (2019)
 26. Chen, L., Liu, X., Zhang, S., Yi, H., Lu, Y., Yao, P.: Efficacy-specific herbal group detection from traditional chinese medicine prescriptions via hierarchical attentive neural network model. *BMC Medical Informatics Decis. Mak.* **21**(1), 66 (2021)
 27. Li, W., Yang, Z., Sun, X.: Exploration on generating traditional chinese medicine prescription from symptoms with an end-to-end method. *arXiv preprint arXiv:1801.09030* (2018)
 28. Ruan, C., Ma, J., Wang, Y., Zhang, Y., Yang, Y.: Discovering regularities from traditional chinese medicine prescriptions via bipartite embedding model. In: *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pp. 3346–3352 (2019). AAAI Press
 29. Li, W., Yang, Z.: Distributed representation for traditional chinese medicine herb via deep learning models. *arXiv preprint arXiv:1711.01701* (2017)
 30. Chu, X., Sun, B., Huang, Q., Peng, S., Zhou, Y., Zhang, Y.: Quantitative knowledge presentation models of traditional chinese medicine (tcm): A review. *Artificial Intelligence in Medicine* **103**, 101810 (2020)
 31. Jin, Y., Zhang, W., He, X., Wang, X., Wang, X.: Syndrome-aware herb recommendation with multi-graph convolution network. In: *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, pp. 145–156 (2020). IEEE
 32. Sun, M., Zhao, S., Gilvary, C., Elemento, O., Zhou, J., Wang, F.: Graph convolutional networks for computational drug development and discovery. *Briefings in bioinformatics* **21**(3), 919–935 (2020)
 33. Yuan, Q., Chen, J., Lu, C., Huang, H.: The graph-based mutual attentive network for automatic diagnosis. In: *Bessiere, C. (ed.) Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pp. 3393–3399 (2020)
 34. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph Attention Networks. *International Conference on Learning Representations* (2018)
 35. Li, H., Sha, J., Shi, C.: Revisiting back-translation for low-resource machine translation between chinese and vietnamese. *IEEE Access* **8**, 119931–119939 (2020)
 36. Madani, A., McCann, B., Naik, N., Keskar, N.S., Anand, N., Eguchi, R.R., Huang, P.-S., Socher, R.: Progen: Language modeling for protein generation. *arXiv preprint arXiv:2004.03497* (2020)
 37. Aji, A.F., Bogoychev, N., Heafield, K., Sennrich, R.: In neural machine translation, what does transfer learning transfer? In: *Proceedings of the 2020 Annual Conference of the Association for Computational Linguistics* (2020). Association for Computational Linguistics
 38. Wu, Z., Liu, Z., Lin, J., Lin, Y., Han, S.: Lite transformer with long-short range attention. In: *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*
 39. Tu, Z., Lu, Z., Liu, Y., Liu, X., Li, H.: Modeling coverage for neural machine translation. *arXiv preprint arXiv:1601.04811* (2016)
 40. Palasundram, K., Sharef, N.M., Kasmiran, K.A., Azman, A.: Enhancements to the sequence-to-sequence-based natural answer generation models. *IEEE Access* **8**, 45738–45752 (2020)
 41. Bahar, P., Brix, C., Ney, H.: Towards two-dimensional sequence to sequence model in neural machine translation. In: *Riloff, E., Chiang, D., Hockenmaier, J., Tsujii, J. (eds.) Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pp. 3009–3015 (2018)
 42. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. In: *Bengio, Y., LeCun, Y. (eds.) 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings* (2013)
 43. Joulin, A., Cissé, M., Grangier, D., Jégou, H., *et al.*: Efficient softmax approximation for gpus. In: *International Conference on Machine Learning*, pp. 1302–1310 (2017). PMLR
 44. Cohen, E., Beck, C.: Empirical analysis of beam search performance degradation in neural sequence models. In: *International Conference on Machine Learning*, pp. 1290–1299 (2019)

45. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: *Advances in Neural Information Processing Systems*, pp. 5998–6008 (2017)
46. Akhter, M.P., Jiangbin, Z., Naqvi, I.R., Abdelmajeed, M., Mehmood, A., Sadiq, M.T.: Document-level text classification using single-layer multisize filters convolutional neural network. *IEEE Access* **8**, 42689–42707 (2020)
47. Lynn, H.M., Pan, S.B., Kim, P.: A deep bidirectional gru network model for biometric electrocardiogram classification based on recurrent neural networks. *IEEE Access* **7**, 145395–145405 (2019)
48. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (2015)
49. Zhou, D., Li, Z., Zhu, J., Zhang, H., Hou, L.: State of health monitoring and remaining useful life prediction of lithium-ion batteries based on temporal convolutional network. *IEEE Access* **8**, 53307–53320 (2020)
50. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Burstein, J., Doran, C., Solorio, T. (eds.) *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pp. 4171–4186 (2019)

Figures

【名称】	小柴胡汤剂
Prescription Name	Xiao Caihu Decoction
【组成】	柴胡 (9克), 黄芩 (9克), 半夏 (6克), 炙甘草 (3克), 党参 (6克), 生姜片 (3片), 大枣 (3枚)
Composition Herbs	Bupleurum 9g, Scutellaria 9g, Pinellia 6g, Licorice 3g, Codonopsis 6g, Ginger 3 pieces, Jujube 3 pieces
【用法】	水煎服
Usage	Decocted in water for oral dose
【适应症】	寒热、心烦喜呕、口苦、咽干、目眩、舌苔薄白、弦脉
Indication Symptoms	Chill and fever alternation, Vexation and vomiting, Bitter taste, Dry throat, Dizzy, White and thin coating of the tongue, Stringy pulse

Figure 1

Please see the Manuscript PDF file for the complete figure caption

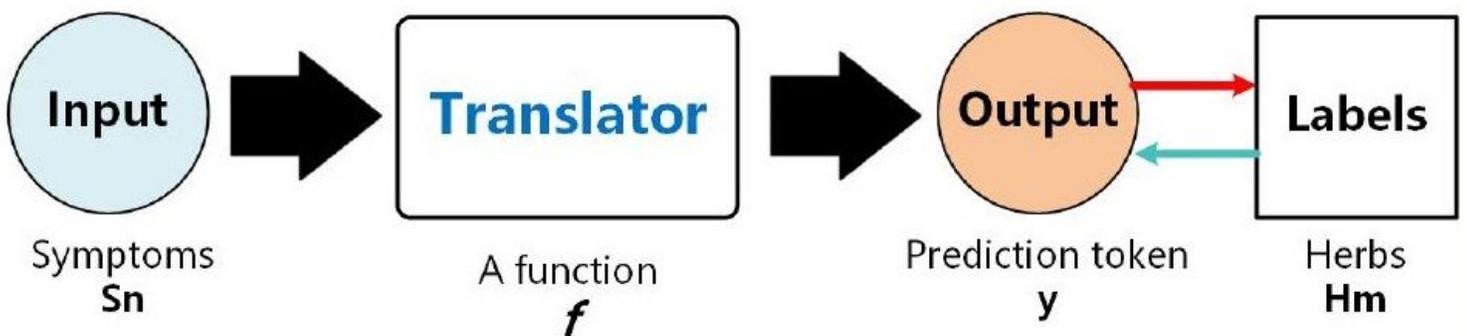


Figure 2

Please see the Manuscript PDF file for the complete figure caption

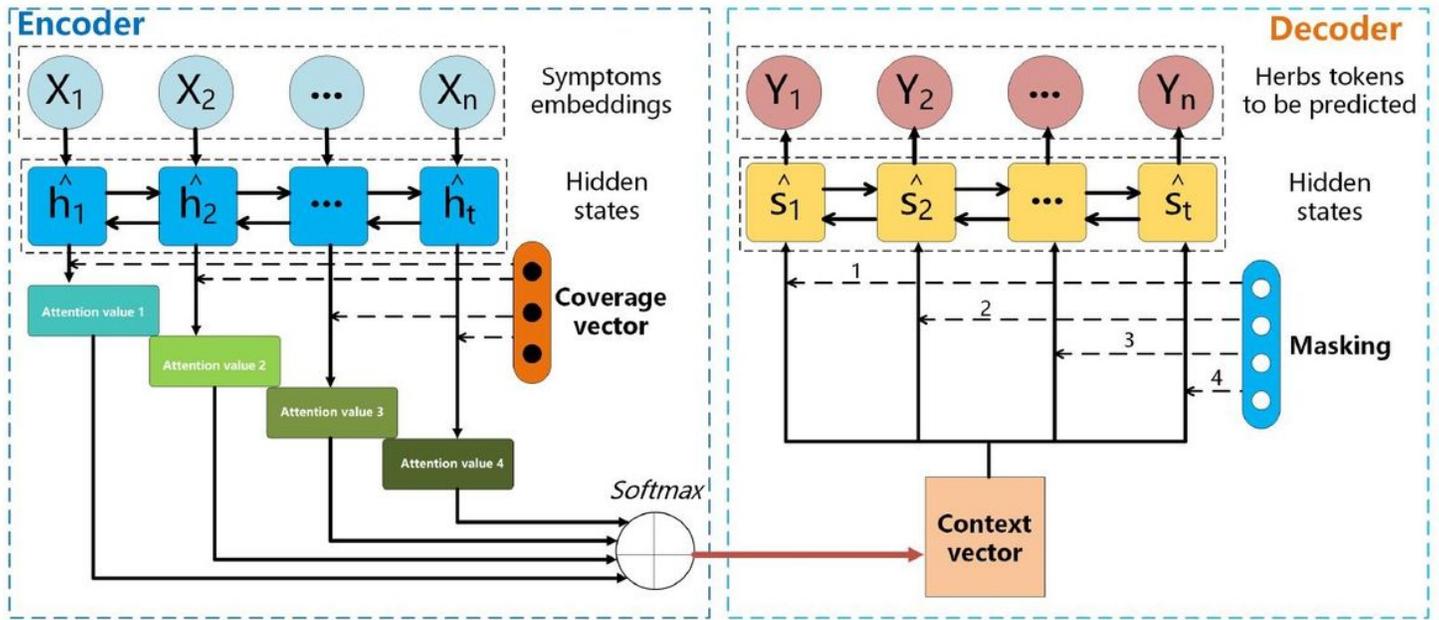


Figure 3

Please see the Manuscript PDF file for the complete figure caption

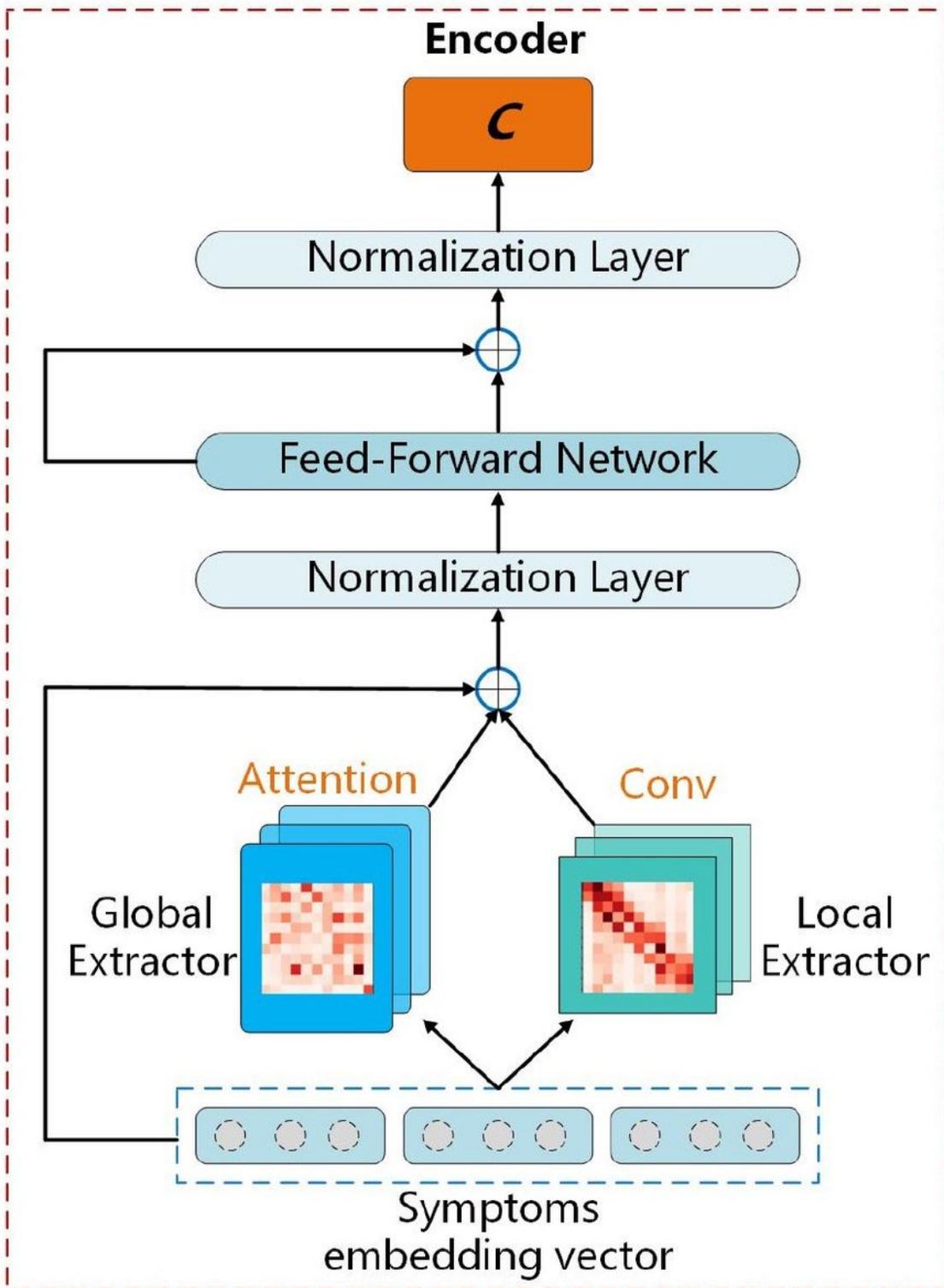


Figure 4

Please see the Manuscript PDF file for the complete figure caption

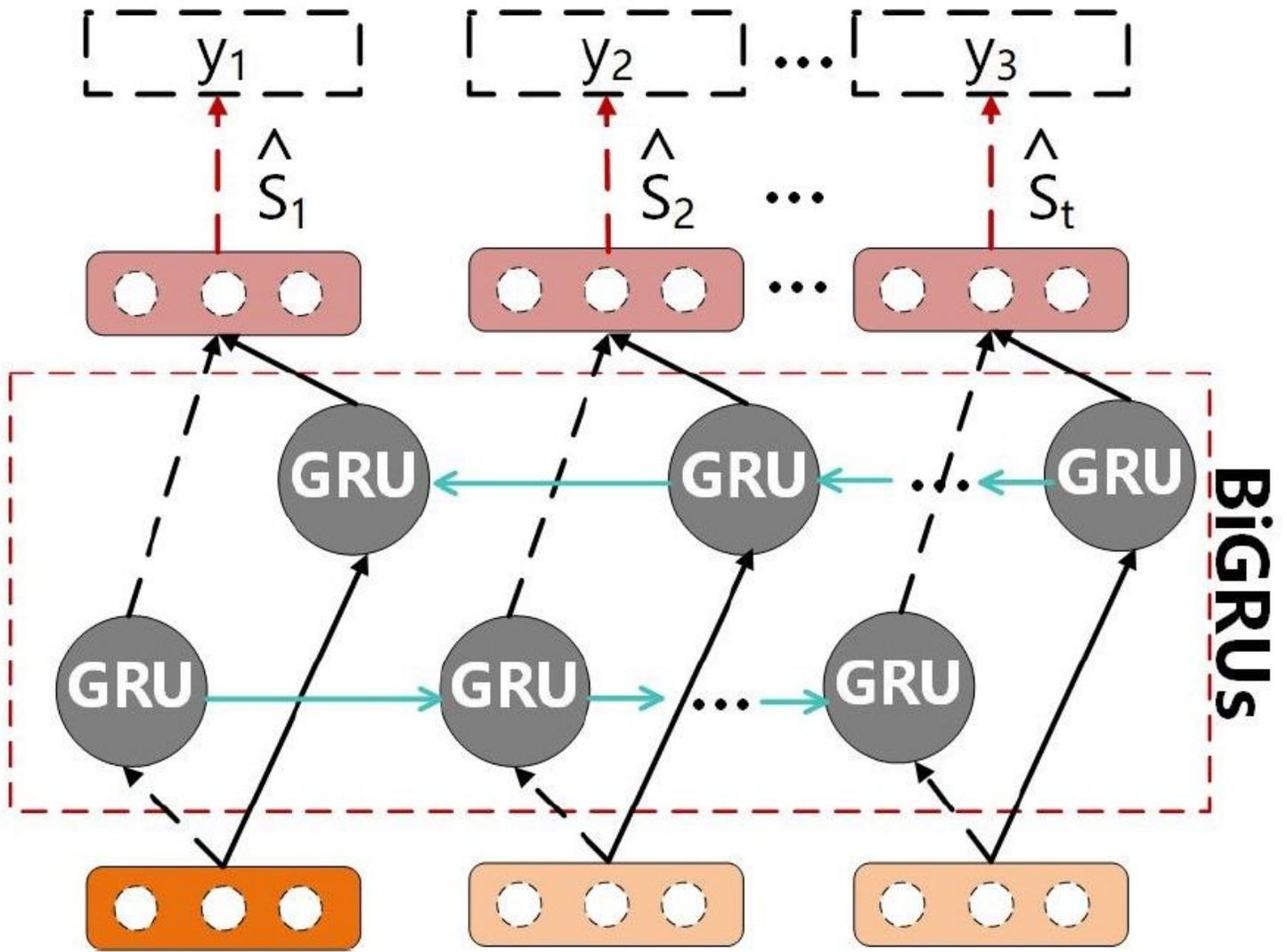


Figure 5

Please see the Manuscript PDF file for the complete figure caption

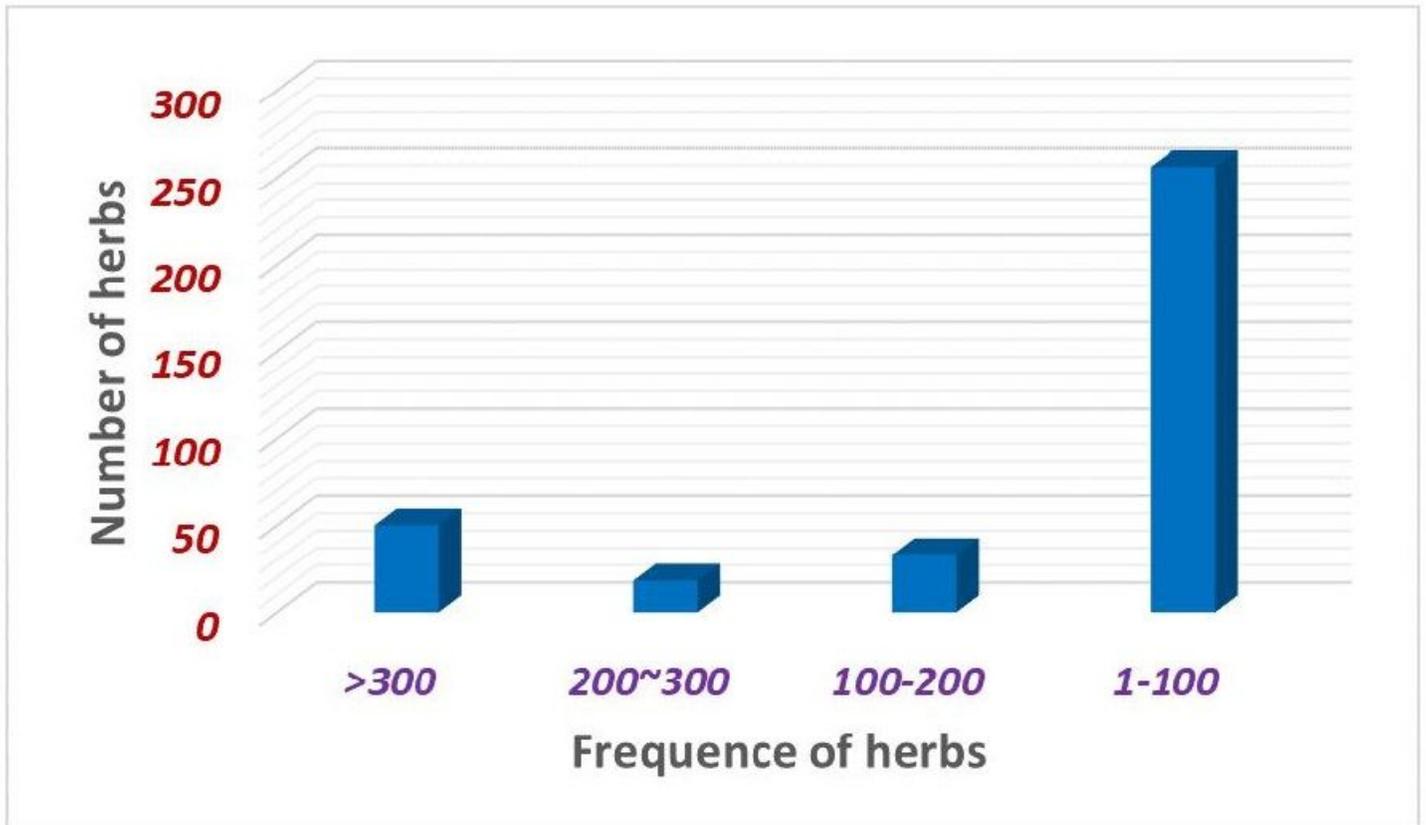


Figure 6

Please see the Manuscript PDF file for the complete figure caption

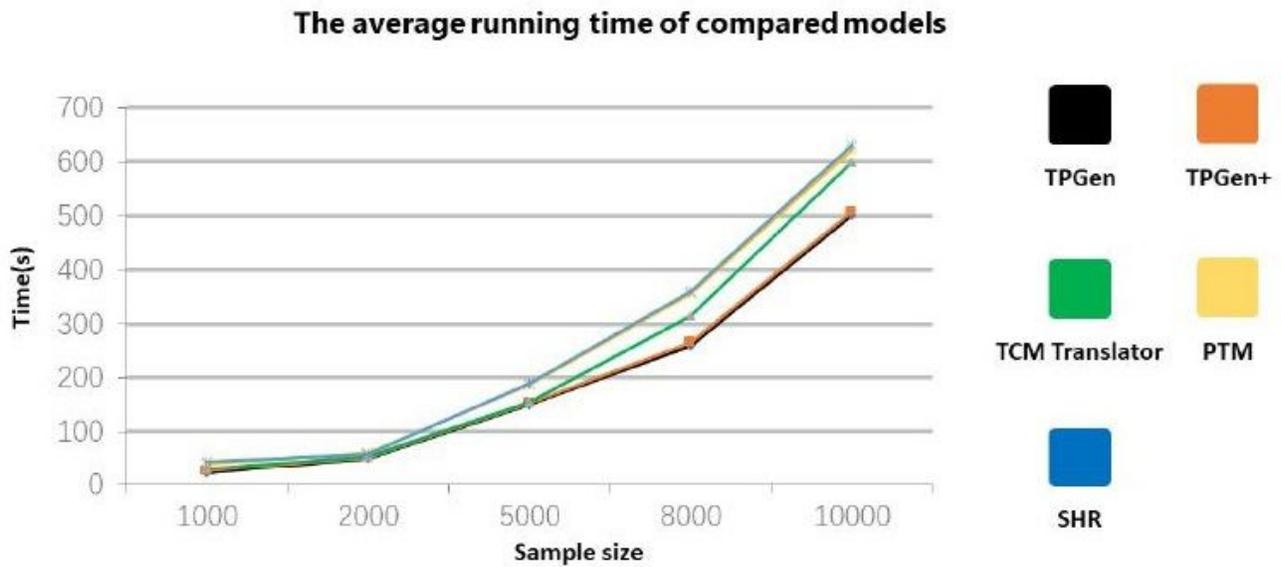


Figure 7

Please see the Manuscript PDF file for the complete figure caption

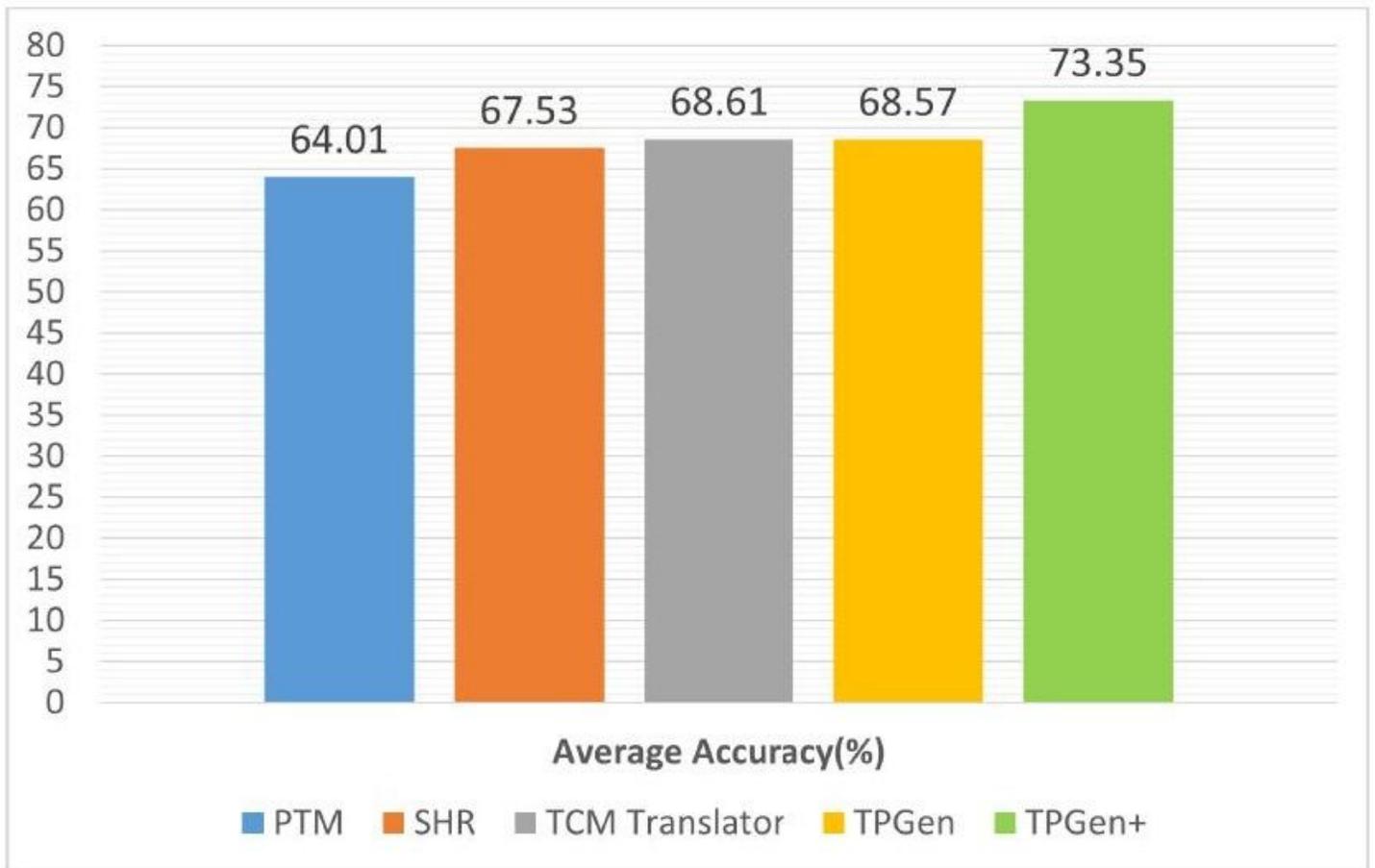


Figure 8

Please see the Manuscript PDF file for the complete figure caption

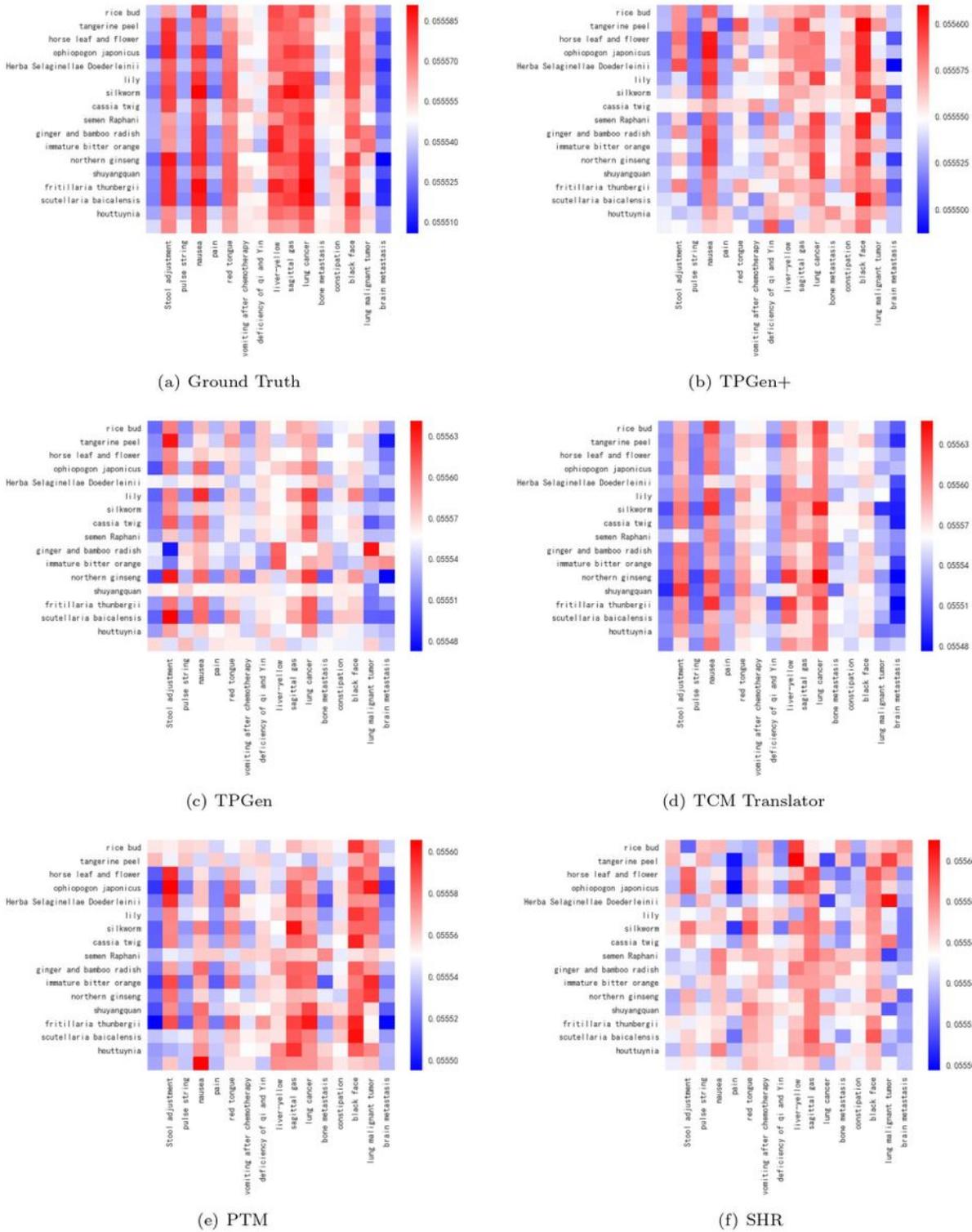


Figure 9

Please see the Manuscript PDF file for the complete figure caption