

TCM-FTP: Fine-Tuning Large Language Models for Herbal Prescription Prediction

Xingzhi Zhou*, Xin Dong[†], Chunhao Li*, Yuning Bai[§], Yulong Xu[¶], Ka Chun Cheung[‡], Simon See[‡], Xinpeng Song[†]

✉Runshun Zhang[†], ✉Xuezhong Zhou[†], and ✉Nevin L. Zhang*

*Department of Computer Science and Technology,

The Hong Kong University of Science and Technology, Hong Kong, China

{xzhoubl, chunhao.li}@connect.ust.hk, lzhang@cse.ust.hk

[†]Institute of Medical Intelligence, Beijing Key Lab of Traffic Data Analysis and Mining,

School of Computer Science & Technology, Beijing Jiaotong University, Beijing, 100044, China

{x_dong, xpsong, xzzhou}@bjtu.edu.cn

[‡]NVAITC, NVIDIA, Hong Kong, China

{chcheung, ssee}@nvidia.com

[§]Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, China

byn-1973@163.com, runshunzhang@139.com

[¶]School of Information Technology, Henan University of Chinese Medicine, Henan, China

flyxyl@126.com

Abstract—Traditional Chinese medicine (TCM) relies on specific combinations of herbs in prescriptions to treat symptoms and signs, a practice that spans thousands of years. Predicting TCM prescriptions presents a fascinating technical challenge with practical implications. However, this task faces limitations due to the scarcity of high-quality clinical datasets and the intricate relationship between symptoms and herbs. To address these issues, we introduce *DigestDS*, a new dataset containing practical medical records from experienced experts in digestive system diseases. We also propose a method, TCM-FTP (TCM Fine-Tuning Pre-trained), to leverage pre-trained large language models (LLMs) through supervised fine-tuning on *DigestDS*. Additionally, we enhance computational efficiency using a low-rank adaptation technique. TCM-FTP also incorporates data augmentation by permuting herbs within prescriptions, capitalizing on their order-agnostic properties. Impressively, TCM-FTP achieves an F1-score of 0.8031, surpassing previous methods significantly. Furthermore, it demonstrates remarkable accuracy in dosage prediction, achieving a normalized mean square error of 0.0604. In contrast, LLMs without fine-tuning perform poorly. Although LLMs have shown capabilities on a wide range of tasks, this work illustrates the importance of fine-tuning for TCM prescription prediction, and we have proposed an effective way to do that.

Index Terms—Large language models, Traditional Chinese medicine, Fine-tuning, Prescription prediction, Herb dosage prediction

We gratefully acknowledge the support of NVIDIA Corporation with the GPU computational resources used for this research.

This work was supported by the Natural Science Foundation of Beijing (No. L232033), the National Natural Science Foundation of China (No. 82374302), and the Science and Technology Innovation Project of China Academy of Traditional Chinese Medicine (No. CI12021A00513).

Xingzhi Zhou and Xin Dong have contributed equally to this work.

✉Runshun Zhang, Xuezhong Zhou, and Nevin L. Zhang are corresponding authors.

I. INTRODUCTION

Traditional Chinese Medicine (TCM) has been an indispensable part of healthcare for the Chinese population for thousands of years. TCM employs a wide range of practices, including herbal medicine, acupuncture, cupping therapy, and tuina massage [1]. Herbal medicine is the primary treatment modality of TCM. It has been shown to effectively treat the novel coronavirus (COVID-19), resulting in improved cure rates and reduced mortality [2], [3].

Herbal prescriptions require doctors to assess patient symptoms using the four diagnostic methods: observation (wang), listening and smelling (wen), questioning (wen), and pulse-taking (qie), guided by the principle of li-fa-fang-yao [4]. Training a doctor involves extensive experience and continuous practical feedback, resulting in a prolonged training period. This contributes to a shortage of TCM doctors and limited TCM resources. Developing an effective prescription predictor is an approach to alleviate challenges such as doctor shortages and the need for prolonged training periods.

Prescription prediction in TCM involves designing a computational system capable of predicting the appropriate prescription based on given symptoms. This system models the complex relationships between symptoms and herbs [5]. Language generation techniques show particular promise in TCM prescription prediction, treating prescription generation as a machine translation problem and solving it using a sequence-to-sequence (seq2seq) model [6]. After numerous attempts by various researchers [7]–[10], generative models have shown positive results in TCM prescription prediction. However, current models are beset with the following limitations:

- **Scarcity of clinical datasets.** Existing works often rely

on datasets derived from classical documents rather than high-quality datasets from clinical records, which introduces noisy information. This results in models that cannot effectively provide personalized prescription recommendations [11]. Moreover, there is currently a lack of high-quality clinical data for prescription prediction [12].

- **Sub-optimal Performance.** Existing models frequently yield results that do not meet expectations and may even mislead users due to their sub-optimal prediction accuracy [12]. However, relying solely on clinical data is insufficient to enhance model performance [13]. It is necessary to integrate more advanced language model techniques to improve the representation and encoding capabilities of prescription prediction models.
- **Lack of herb dosage.** Existing models overlook the crucial prediction of dosage weights, a critical component in TCM for effective disease treatment [5], [14]. Additionally, there is a lack of evaluation metrics specifically for herb dosage prediction.

Inspired by the robust predictive capabilities of large language models (LLMs), these challenges motivate us to develop a high-quality prescription dataset and leverage LLMs to construct an advanced prescription predictor. In this study, we introduce a high-quality prescription dataset *DigestDS* derived from clinical medical records and propose TCM-FTP (Fig. 1), a novel TCM prescription generation model based on LLMs. Our goal is to leverage the capabilities of LLMs to overcome the limitations of current methodologies. Specifically, we utilize a low-rank adaptation technique (LoRA) [15] for efficient LLM fine-tuning. To take advantage of the order-agnostic nature of herb prescriptions, we implement data augmentation by randomizing the sequence of herbs in the training data. We validated our model on *DigestDS*, achieving a Precision of 0.7951, a Recall of 0.8113, and an F1-score of 0.8031. This represents a significant improvement over the best performance achieved by previous methods.

Our main contributions are outlined as follows:

- We construct a TCM clinical diagnostic dataset *DigestDS* derived from practical records in real-world TCM clinical scenarios. This dataset serves as a foundational resource for understanding the nuanced relationships between symptoms and effective prescriptions in TCM.
- We introduce TCM-FTP, a novel LLM-driven method for generating TCM prescriptions. Our model excels at capturing the intricate interplay between symptoms and herbal prescriptions, marking a significant advancement in TCM prescription prediction. Additionally, we have designed the NMSE evaluation metric for dosage prediction. To the best of our knowledge, our work is the first in the prescription prediction field to design dosage prediction metrics tailored for real-world scenarios.
- We empirically validate TCM-FTP against existing models on our dataset, showcasing its superior performance in predicting herbs and dosages. Additionally, we conduct a thorough analysis of factors such as the impact of

foundational models, learning rates, the number of herb permutations, and the decoding parameters in inference. Expert case studies by TCM professionals further provide valuable insights into the adaptability and robustness of our innovative approach.

II. BACKGROUND AND SIGNIFICANCE

A. Problem Definition

We now formally define the prescription prediction problem. Given a prescription dataset $\mathcal{P}_{\text{train}}$, an element in $\mathcal{P}_{\text{train}}$ consists of a symptom description s and its corresponding prescription $\{h_i, w_i\}^{i \in [k]}$. Here, h denotes the herb name, w represents the herb dosage, and $[k]$ signifies a list ranging from 1 to k . Our objective is to train a model \mathcal{M} such that $\mathcal{M}(s)$ reproduce $\{h_i, w_i\}^{i \in [k]}$ accurately. For alignment with the language generation task, we concatenate $\{h_i, w_i\}^{i \in [k]}$ using a comma separator to form a single sentence. Concrete examples are presented in Fig. 1.

B. TCM Herbal Prescription Prediction

Research on TCM prescription prediction mainly falls into three categories: topic model-based, graph model-based, and language model-based approaches. Topic model-based approaches treat relationships between symptoms and herbs as that of documents and topics [16]–[18]. These approaches rely on statistical relationships, lacking a semantic understanding of symptoms. Graph-based approaches construct a medical knowledge graph to model the relationships between symptoms and herbs [14], [19], [20]. However, these approaches also lack consecutive semantic information regarding symptoms. Language model-based approaches are a more promising way to model the complicated relationships between symptoms and herbs.

Language model-based prescription prediction models take patient symptom descriptions as input and generate herbal prescriptions sequentially. TCM Translator [21] uses transformer architectures to distill context vectors from symptoms and LSTM [22] as the decoder. AttentiveHerb [7] employs a seq2seq [6] model with dual attention mechanisms to distinguish primary from secondary symptoms and map herb-symptom interactions using clinical data. Herb-Know [8] utilizes herb descriptions to model associations with symptoms and evaluates whether herb effects align with symptom descriptions. TCMBERT [9] integrates transfer learning by initially training an ALBERT [23] on TCM-related documents and fine-tuning an LSTM-based seq2seq model for context vector extraction. RoKEPG [10] incorporates additional herb knowledge for fine-tuning prescription prediction models.

C. Large Language Models in TCM

LLMs have made significant strides in various NLP tasks. Notable examples like ChatGPT and GPT-4.0 [24] have attracted considerable attention, although specifics about their architectures, parameters, and training strategies from OpenAI remain undisclosed. The emergence of open-sourced LLMs has sparked widespread interest among researchers. Prominent

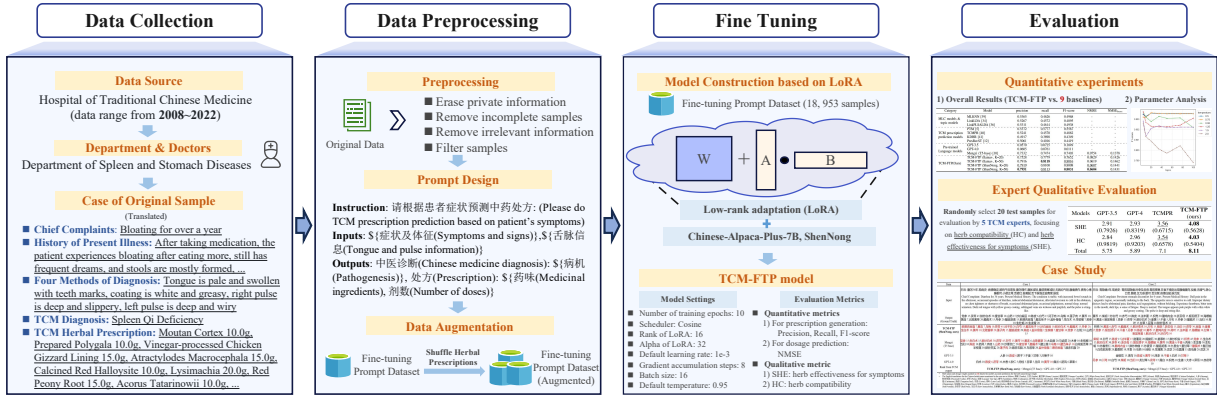


Fig. 1: **Workflow of the TCM-FTP.** Our work consists of four parts: "Data Collection" involves gathering and organizing raw data; "Data Processing" includes data preprocessing, prompt design, and integrating data augmentation to create a fine-tuning dataset; "Fine-tuning" utilizes the ShenNong LLM and LoRA technique to optimize the model; and "Evaluation" assesses the outcomes using both quantitative and qualitative evaluation metrics.

open-sourced models include LLaMA [25], Bloom [26], and ChatGLM [27], among others. While LLMs excel in standard NLP tasks, they often struggle with specialized domains requiring specific knowledge, such as medicine or finance. Supervised fine-tuning has become a standard approach to enhance LLMs for these domains by incorporating specialized knowledge.

In the context of TCM, several innovative models have been proposed. Bentsao [28] utilizes supervised fine-tuning on LLaMA, integrating structured and unstructured knowledge from CMeKG [29]. Huatuo [30] leverages both original and distilled data from ChatGPT, incorporating a Reinforcement Learning from Artificial Intelligence Feedback (RLAIF) mechanism and employing Proximal Policy Optimization (PPO) during fine-tuning. Zhongjing [31] applies continual pretraining to inject domain knowledge and uses supervised fine-tuning on a Chinese multi-turn medical dialogue dataset, complemented by Reinforcement Learning from Human Feedback (RLHF). ShenNong [32] builds on LLaMA with LoRA-based fine-tuning on an instructional dataset derived from ChatGPT and a traditional Chinese medicine knowledge graph, benefiting from a large-scale dataset of over 110,000 instructions.

D. Parameter Efficient Fine-Tuning

Parameter Efficient Fine-Tuning (PEFT) utilizes a small amount of parameters to fine-tune a large language model effectively. Assuming there is a pretrained model $f_{\theta}(y|x)$, PEFT seeks to adjust a limited number of parameters, $\Delta\theta$, such that $|\theta| \gg |\Delta\theta|$. In contrast to conventional fine-tuning, which updates all parameters, denoted as $|\theta| = |\Delta\theta|$, and requires significant computational resources, PEFT selectively updates a limited number of learnable parameters to achieve results comparable to complete fine-tuning.

PEFT encompasses three primary methodologies: adapter [33], p-tuning [34]–[36], and LoRA [15]. Adapter tuning [33] introduces PEFT by appending additional learnable modules,

named adapters, to every layer of the pretrained model, exclusively fine-tuning these adapters while maintaining the original parameters. In contrast, p-tuning leveraged prompt engineering techniques, fine-tuning learnable vectors as new tokens and inserting these tokens into the multi-head self-attention layer (MSA) [34]–[36]. In this work, we employ LoRA [15] to efficiently fine-tune LLMs on the target dataset.

III. MATERIALS AND METHODS

In this study, as illustrated in Fig. 1, we gather a high-quality prescription dataset *DigestDS* and propose a PEFT approach for prescription prediction TCM-FTP. *DigestDS* is comprised of practical medical records collected from specialists in digestive system disorders. TCM-FTP utilizes supervised fine-tuning to train the model on *DigestDS*, incorporating LoRA technique and an effective data augmentation technique, which involves permuting the herbs in the prescriptions.

A. Datasets

1) *Data collection:* We collect prescription data from practical patient records generated by specialists in TCM hospital¹ over a span from 2008 to 2022. The prescriptions specifically focus on digestion system disorders.

2) *Data processing:* Initially, we remove the incomplete data items and erase any privacy information. Subsequently, we exclude irrelevant information, retaining only essential information for prescription prediction. Specifically, we keep the chief complaint, medical history, and tongue-coating details from the symptom descriptions, as well as the names and dosages of herbs in the prescriptions.

3) *Data statistics:* We present the processed data statistics in Table I. Prescription items are randomly divided into training (90%) and testing (10%) datasets. The training dataset comprises 16,896 samples with an average of 23.92 herbs per prescription, while the test dataset includes 2,057 samples

¹Dr. Runshun Zhang

averaging 23.86 herbs per prescription. Fig. 2 illustrates the distributions of herb counts in the training and testing datasets.

TABLE I: **Statistics of training and test datasets** The table displays the median, mean, and standard deviation for the number of herbs per prescription. Category: Number of distinct herbs in the dataset.

dataset	size	category	median	mean	std
training	16,896	674	24	23.92	5.69
test	2,057	533	24	23.86	5.44

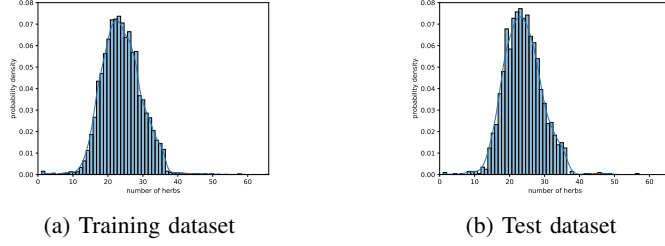


Fig. 2: **Distribution of the number of herbs in prescriptions** The blue curves represent kernel density estimates.

B. TCM-FTP

To better model the intricate relationships between symptoms and herbs, we propose TCM-FTP, which employs a pre-trained LLM coupled with an auto-regressive mechanism for TCM prescription prediction. For efficient fine-tuning of the LLM, we employ the LoRA technique [15], which optimizes the LLM with limited parameters and computational resources. To enhance the dataset, we permute the herbs in prescriptions to leverage the order-agnostic property inherent in TCM prescriptions.

Given a prescription dataset \mathcal{P}_{train} with $\{s, \{h_i, w_i\}^{i \in [k]}\}$, our primary goal is to use a language model to maximize the estimation of conditional probability:

$$\max_{\theta} \mathbb{E}_{s, \{h_i, w_i\}^{i \in [k]} \sim \mathcal{P}_{train}} \sum_i^k \log P((h_i, w_i) | h_{<i}, w_{<i}; s; \theta), \quad (1)$$

where s is the symptom description, k is the number of herbs in the prescription, and (h_i, w_i) represents the i -th (herb, dosage) pair with $i \in \{1, 2, \dots, k\}$. $h_{<i}$ and $w_{<i}$ represent a set of herbs and a set of weights with index less than i , respectively. θ represents our model parameters.

To leverage the auto-regressive mechanism in LLMs, we concatenate the herbs and dosages into one string. We denote concatenated string $\text{CAT}(\{h_i, w_i\}^{i \in [k]})$ as y . Consequently, our objective function becomes:

$$\max_{\theta} \mathbb{E}_{s, \{h_i, w_i\}^{i \in [k]} \sim \mathcal{P}_{train}} \sum_{j=1}^{|y|} \log P(y_j | \{y_{<j}\}; s; \theta), \quad (2)$$

where $\{y_{<j}\}$ denotes the set of the tokens from index 1 to index $j-1$ in y , and $|y|$ represents the number of tokens in y .

1) *Low-rank Adaptation*: We employ the LoRA [15] for PEFT to conserve computational resources. Beginning with a pre-trained LLM f_{θ} , we encounter two primary types of parametric functions within our model framework: linear and embedding functions. These functions are mathematically described as:

$$\mathcal{F}_{\text{linear}}(x) = W_l \cdot x, \quad (3)$$

$$\mathcal{F}_{\text{emb}}(x) = \text{EMB}(x; W_e), \quad (4)$$

where $\text{EMB}(x; \cdot)$ represents the embedding operator that selects the x -th column from the specified matrix, and W_l and W_e denotes the parameters for the linear and embedding functions, respectively. For fine-tuning, we introduce updates via low-rank matrices A and B , leading to modifications in the original functions:

$$\mathcal{F}_{\text{linear}}(x)' = \mathcal{F}_{\text{linear}}(x) + A_l \cdot B_l \cdot x, \quad (5)$$

$$\mathcal{F}_{\text{emb}}(x)' = \mathcal{F}_{\text{emb}}(x) + A_e \cdot \text{EMB}(x; B_e), \quad (6)$$

where A_l , B_l , A_e , and B_e are the learnable parameters associated with the linear and embedding functions. With the updates to the low-rank matrices represented by $\Delta\theta$, our goal is to maximize the expected conditional probability:

$$\max_{\Delta\theta} \mathbb{E}_{s, \{h_i, w_i\}^{i \in [k]} \sim \mathcal{P}_{train}} \sum_{y=\text{CAT}(\{h_i, w_i\}^{i \in [k]})}^{|y|} \log P(y_j | \{y_{<j}\}; s; \theta + \Delta\theta). \quad (7)$$

2) *Order-Agnostic Property*: Recognizing the order-agnostic characteristic of herbs in TCM prescriptions, we implement data augmentation by permuting the herbs in the prescriptions. Given a prescription sample $\{s, \{h_i, w_i\}^{i \in [k]}\}$, we define the permuted herb sequence as $\text{CAT}(\sqcup(\{h_i, w_i\}^{i \in [k]}))$. The sequence resulting from the permutation, $\sqcup(\{h_i, w_i\}^{i \in [k]})$, is represented as $\{h_{r_i}, w_{r_i}\}^{i \in [k]}$, where r_i denotes the indices after shuffling. After the herb permutation, our final goal becomes:

$$\max_{\Delta\theta} \sum_{t=1}^K \mathbb{E}_{s, \{h_i, w_i\}^{i \in [k]} \sim \mathcal{P}_{train}} \sum_{j=1}^{|y|} \log P(y_j | \{y_{<j}\}; s; \theta + \Delta\theta), \quad (8)$$

where K is the number of permutation times. t refers to different permutation index since permutation function \sqcup is nondeterministic at each run.

C. Baselines

To demonstrate the performance of our model, we compare it with the following baselines:

(1) Topic Models and Multi-label Classification Models:

- LinkLDA [37] uses Link Latent Dirichlet Allocation [38] to model relationships between symptoms and corresponding herbs.

- LinkPLSALDA [39] combines LDA [40] and PLSA [41] for topic modeling.
- MLKNN [42] is an enhanced K-nearest neighbor (KNN) approach for multi-label classification (MLC).

(2) TCM Prescription Prediction Models:

- PTM [5] excels in recommending herbs based on symptoms and predicting symptoms from herbs.
- TCMPR [43] utilizes sub-networks for symptom description and information extraction.
- KDHR [44] integrates herb properties and features using multi-level GCN layers.
- PresRecST [12] follows a systematic approach with three stages for clinical predictions in TCM.

(3) Pre-trained Language Models:

- Mengzi (T5-base) [45] is a T5-based [46] seq2seq model pre-trained on Chinese text and fine-tuned on the specific dataset as a competitive baseline.
- GPT-3.5 and GPT-4.0 [24]: We also include the pre-trained LLMs without fine-tuning as baselines.

D. Experimental setup

We implement the fine-tuning process with the Transformers package from Hugging Face². The number of training epochs is set to 10. A cosine scheduler is adopted, with the rank of LoRA being 16 and the alpha of LoRA at 32. We consider the number of herb permutations, K , as 20 and 50, respectively. The default learning rate is set at $1e-3$. We leverage the gradient accumulation technique to increase the batch size, with 8 accumulation steps and a batch size of 16. We employ two foundation models Chinese-Alpaca-Plus-7B³ and ShenNong⁴. Chinese-Alpaca-Plus-7B is a variant of LLaMA [25] with continual pre-training and supervised fine-tuning on Chinese corpus, denoted as LLaMA+ for simplicity. ShenNong is a further refinement of LLaMA+, fine-tuned with TCM instruction datasets. We use an 8-V100 GPU machine for fine-tuning and the running time for TCM-FTP ($K = 50$) is 146 hours. During the inference stage, We employ top-k and top-p combinations for decoding, setting the top-k to 50 and the top-p to 0.7, with a default temperature of 0.95.

E. Evaluation Metrics

For the proposed TCM-FTP, we evaluate it from both quantitative and qualitative perspectives, including the following evaluation metrics.

Quantitative Evaluation. For evaluation metrics, we use precision, recall, and F1-score as herb prediction metrics and designed NMSE for herb dosage evaluation.

- Precision: The proportion of correctly predicted herbs out of all predicted herbs, reflecting the accuracy of positive predictions.
- Recall: The proportion of correctly predicted herbs out of all herbs in the ground truth, indicating the model’s ability to identify all relevant items.

- F1-score: The harmonic mean of precision and recall, giving a balanced measure of overall accuracy.
- Normalized mean square error (NMSE): To tackle the problem above, we design this metric to evaluate the accuracy of predicted dosage by normalizing the squared differences using the original weights. For a given prescription p and a predicted prescription and \hat{p} , we suppose they are composed by a set of pairs of a herb and a dosage, $p = \{(h, w)\}$, where h, w refer to herb and dosage respectively, and $\hat{p} = \{(h', w')\}$ indicates the generated result from model. The calculation of NMSE is as follows,

$$NMSE = \frac{1}{Z} \sum_{\substack{(h,w) \in p, \\ (h',w') \in \hat{p}}} \mathbf{1}[h = h'] \left(\frac{w' - w}{w} \right)^2, \quad (9)$$

where Z is the number of correctly predicted herbs, and $\mathbf{1}[\cdot]$ is the indicator function. For evaluation, the average dosage of each herb from the training data is used as a baseline for dosage predictions. For herbs unseen in the training data, the dosage is predicted as the average dosage of all known herbs. This approach is referred to as $NMSE_{base}$ in the NMSE calculations.

Qualitative Evaluation. Existing quantitative evaluation metrics assess model quality based solely on sample labels, neglecting the compatibility and symptom-specific effectiveness of prescriptions. To comprehensively evaluate our model, we engaged five TCM experts to conduct an expert qualitative evaluation (EQE) of selected prescriptions generated by our model. Each doctor independently assessed the prescriptions for herbal effectiveness in treating symptoms (SHE) and herbal compatibility (HC), assigning scores on a scale of 0 to 5. Higher scores indicate greater effectiveness or compatibility.

IV. RESULTS AND DISCUSSION

A. Overall Results

The overall comparison results between TCM-FTP and the baselines are presented in Table II. The proposed TCM-FTP outperforms all baseline models in the herb prediction task on *DigestDS*, achieving an F1-score of 0.8016 using LLaMA+ as the foundation model and 0.8031 using ShenNong. This highlights the superior capability of TCM-FTP in modeling the intricate relationships between symptoms and herbs in prescriptions. Unlike previous approaches, TCM-FTP also includes herb dosage prediction, which is crucial in TCM due to the significant impact of dosage combinations. As shown in Table II, TCM-FTP achieves a much lower NMSE compared to the baseline using average statistics (0.0604 for TCM-FTP versus 0.0754 for Mengzi (T5-base) [45]). This enhances the practicality of TCM-FTP in clinical TCM prescription generation.

Our proposed TCM-FTP shows significant advantages over various baseline models across all aspects.

- Compared to the pre-trained language model Mengzi (T5-base) [45], TCM-FTP significantly enhances prediction

²<https://huggingface.co/docs/transformers>

³<https://github.com/ymcui/Chinese-LLaMA-Alpaca/tree/main>

⁴<https://huggingface.co/michaelwzhu/ShenNong-TCM-LLM>

TABLE II: **Prediction results of herbs and dosages in prescriptions** TCM-FTP is evaluated with different foundation models.

Category	Model	Precision	Recall	F1-score	NMSE	NMSE _{base}
MLC models & topic models	MLKNN [42]	0.5365	0.4626	0.4968	-	-
	LinkLDA [37]	0.5267	0.4572	0.4895	-	-
	LinkPLSALDA [39]	0.5311	0.4614	0.4938	-	-
TCM prescription prediction models	PTM [5]	0.5372	0.5777	0.5567	-	-
	TCMPR [43]	0.5241	0.4570	0.4882	-	-
	KDHR [44]	0.4917	0.3898	0.4349	-	-
	PresRecST [12]	0.5061	0.4016	0.4419	-	-
Pre-trained Language models	GPT-3.5	0.0570	0.0725	0.1049	-	-
	GPT-4.0	0.0605	0.0761	0.0111	-	-
	Mengzi (T5-base) [45]	0.7332	0.7474	0.7403	0.0754	0.1378
TCM-FTP(Ours)	TCM-FTP (llama+, K=20)	0.7528	0.7779	0.7652	0.0829	0.1426
	TCM-FTP (llama+, K=50)	0.7916	0.8118	0.8016	0.0619	0.1462
	TCM-FTP (ShenNong, K=20)	<u>0.7919</u>	0.8100	0.8008	<u>0.0607</u>	0.1441
	TCM-FTP (ShenNong, K=50)	0.7951	<u>0.8113</u>	0.8031	0.0604	0.1431

performance with improved precision (from 0.7332 to 0.7951), recall (from 0.7474 to 0.8113), and F1-score (from 0.7403 to 0.8031). Performances on GPT-3.5 and GPT-4.0 [24] were notably poor, indicating the limitations of general llms in tasks requiring specialized knowledge.

- In comparison to other TCM prescription prediction models, TCM-FTP outperforms PTM [5], which leads among these baselines, followed by TCMPR [43] in accuracy, with KDHR [44] and PresRecST [12] showing poorer performance. PTM, a topic model-based approach, is computationally complex, while TCMPR, KDHR, and PresRecST are graph-based models sensitive to graph-related factors. TCM-FTP excels these approaches in generating accurate herb recommendations due to advanced language modeling capability.
- Compared to multi-label classification (MLC) and topic model approaches, our model demonstrates superior performance. MLKNN [42] and two topic model methods perform similarly, whereas MLC and topic model-based approaches achieve lower performance due to computational intensity and limited consideration of inter-herb relationships. This underscores the proposed TCM-FTP proves more suitable for real-world herb prediction tasks, highlighting its enhanced performance over various baseline methods.

In summary, our approach exhibits significant advantages over various baseline methods and is better suited for real-world herb recommendation tasks. In addition, TCM knowledge embedded in foundation models enhances the fine-tuning process. Epoch-wise performance, shown in Fig. 3a for both loss and F1-score, indicates that although TCM-FTP (ShenNong, K=50) shows similar loss changes to TCM-FTP (LLaMA+, K=50), the F1-score line for TCM-FTP (ShenNong, K=50) consistency surpasses that of TCM-FTP (LLaMA+, K=50). This suggests that TCM knowledge significantly aids the fine-tuning process with the ShenNong model. Additionally, with a lower number of permutations K=20, TCM-FTP (ShenNong) achieves an F1-score of 0.8008, compared to 0.7652 for TCM-FTP (LLaMA+), highlighting

the advantage of having TCM knowledge embedded in the ShenNong model for modeling symptom-herb relationships.

B. Parameter Analysis

The number of permutation The herb permutation introduces the order-agnostic property to enhance the prediction performance. We present the results of TCM-FTP with varying numbers of herb permutations K in Table III. As K increases from 0 to 50, the f1-score improves gradually from 0.4885 to 0.8031, and the NMSE decreases from 0.1683 to 0.0604. This suggests the importance of the order-agnostic property as an effective inductive bias for prescription prediction.

TABLE III: **The impact of herb permutations** TCM-FTP is evaluated with different numbers of herb permutations, K .

K	Precision	Recall	F1-score	NMSE	NMSE _{base}
0	0.4640	0.5156	0.4885	0.1683	0.1509
1	0.5572	0.5765	0.5667	0.1269	0.1410
10	0.7748	0.7902	0.7824	0.0608	0.1391
20	0.7919	0.8100	0.8008	0.0607	0.1441
50	0.7951	0.8113	0.8031	0.0604	0.1431

Learning rates We validate the impact of learning rates by conducting experiments with K=10 and varying learning rates, as shown in Table IV. Higher learning rates correlate with improved precision, recall, and F1-score. This suggests that lower learning rates lead to underfitting, likely due to significant disparities between the pre-trained model parameters and the target parameters. Our objective is to fine-tune a task-specific generator without retaining the broad language capabilities of large language models (LLMs), necessitating substantial adjustments to achieve superior performance. However, excessively large learning rates cause model corruption, such as a learning rate of 0.0025.

TABLE IV: **Impact of learning rates** TCM-FTP (ShenNong, K=10) is fine-tuned with different learning rates.

lr	Precision	Recall	F1-score	NMSE	NMSE _{base}
5e-5	0.5766	0.5894	0.5830	0.1057	0.1238
1e-4	0.6395	0.6615	0.6503	0.0869	0.1282
5e-4	0.7571	0.7734	0.7652	0.0693	0.1475
7.5e-4	0.7707	0.7832	0.7769	0.0824	0.1394
1e-3	0.7748	0.7902	0.7825	0.0608	0.1391
2.5e-3	0.0003	0.0000	0.0000	0.0178	0.0400

Decoding paramters We investigate the robustness of decoding parameters by varying top-p, top-k, and temperature. Initially, in Fig. 3b, we set top-p to 0.7 and chart the f1-score across various top-k and temperature settings. The results show that the model’s performance is stable across different temperatures and top-k values, with only a slight decline when the temperature exceeds 0.95. Next, in Fig. 3c, with top-k fixed at 50, we graph the f1-score for different top-p and temperature values, observing that lower temperatures are more stable, leading to higher probabilities for the top predicted tokens, thus enhancing prediction accuracy. Lastly, in Fig. 3d, with the temperature set at 0.95, we map the f1-score for various top-p and top-k combinations, noting that lower top-k values offer more stability in the face of top-p variations.

C. Expert Qualitative Evaluation

Since conducting expert qualitative evaluation (EQE) is time-consuming and labor-intensive for doctors, this study randomly selected 20 data points from the test set for evaluation. As shown in Table V, the TCM-FTP model outperformed other baseline models across all metrics, achieving an average SHE score of 4.08 (standard deviation 0.5628) and an average HC score of 4.03 (standard deviation 0.5404). This indicates that TCM-FTP can generate effective prescriptions that adhere to TCM compatibility principles. In contrast, GPT-3.5 and GPT-4.0 [24] scored below 3 for both SHE and HC, with total scores of 5.75 and 5.89, respectively, significantly lower than TCM-FTP and TCMPR [43]. This demonstrates that GPT-3.5 and GPT-4.0 perform poorly in generating prescriptions that adhere to TCM principles. TCMPR performed relatively well, with SHE and HC scores of 3.56 (standard deviation 0.6715) and 3.54 (standard deviation 0.6578), respectively, but still fell short of TCM-FTP. Overall, TCM-FTP excelled in prescription generation, receiving higher approval from doctors and showing less score variability, indicating better stability and robustness.

TABLE V: **Comparative Results of expert evaluation.** The table records the mean scores (standard deviation) for all evaluated samples.

Models	GPT-3.5	GPT-4.0 [24]	TCMPR [43]	TCM-FTP(ours)
SHE	2.91 (0.7926)	2.93 (0.8319)	3.56 (0.6715)	4.08 (0.5628)
HC	2.84 (0.9819)	2.96 (0.9203)	3.54 (0.6578)	4.03 (0.5404)
Total	5.75	5.89	7.1	8.11

D. Case Study

In order to visually showcase the predictive performance and capabilities of our TCM-FTP, we obtain the predicted herb results formed by each model on the test set and had them evaluated by TCM experts. Fig. 4 illustrates the results for two cases, including the input (chief complaints, present medical history), output (actual prescriptions provided by the doctor), and the predicted prescription results from TCM-FTP (ShenNong) and Mengzi (T5 base) [45] models, as well as GPT-3.5 and GPT-4.0 [24] predictions. Herb names/dosage weights marked in red indicate that the model’s prediction match the actual prescription provided by the doctor. Additionally, professional evaluations of each model are presented, ranking them using the “>” symbol (with models ranked higher indicating better performance).

The results show that the TCM-FTP (ShenNong) models achieve strong prediction performance. The herb label predictions closely match the doctor’s prescriptions, indicating effective data fitting after fine-tuning. Additionally, the dosage predictions align well with the actual dosages provided by the doctor, matching clinically relevant herb dosages. In contrast, GPT-3.5 and GPT-4.0 were also tested on the same cases but produced comparatively poor predictions with significant deviations from the actual results. This likely stems from their lack of specialized TCM training, hindering their performance.

In summary, the proposed TCM-FTP (ShenNong) outperforms general models. This highlights the importance of fine-tuning large language models on high-quality, domain-specific data to fully leverage their capabilities in specialized fields.

V. CONCLUSION

To deal with the lack of high-quality datasets and to improve the performance in Traditional Chinese Medicine (TCM) prescription predictions, we build a TCM prescription dataset *DigestDS* from clinical records and propose TCM-FTP to fine-tune large language models to predict herbs with the corresponding dosages. We collect *DigestDS* from practical clinical records by focusing on digestion disorder diseases. TCM-FTP employs a low-rank adaptation for computational and storage efficiency and adapts a data augmentation by randomly permutating the order of herbs in prescriptions. The experimental results reveal the remarkable effectiveness of TCM-FTP, surpassing previous methods by large margins in precision, recall, and F1-score. Additionally, our method achieved the best results in NMSE, effectively forming accurate herb and dosage predictions. In future work, we will continue to incorporate domain knowledge into model construction to further enhance performance, aiming to develop a practically usable prescription prediction model.

REFERENCES

- [1] F. Cheung, “Tcm: Made in china,” *Nature*, vol. 480, no. 7378, pp. S82–S83, 2011.
- [2] M. Liu, Y. Gao, Y. Yuan, K. Yang, S. Shi, J. Zhang, and J. Tian, “Efficacy and safety of integrated traditional chinese and western medicine for corona virus disease 2019 (covid-19): a systematic review and meta-analysis,” *Pharmacological Res.*, vol. 158, p. 104896, 2020.

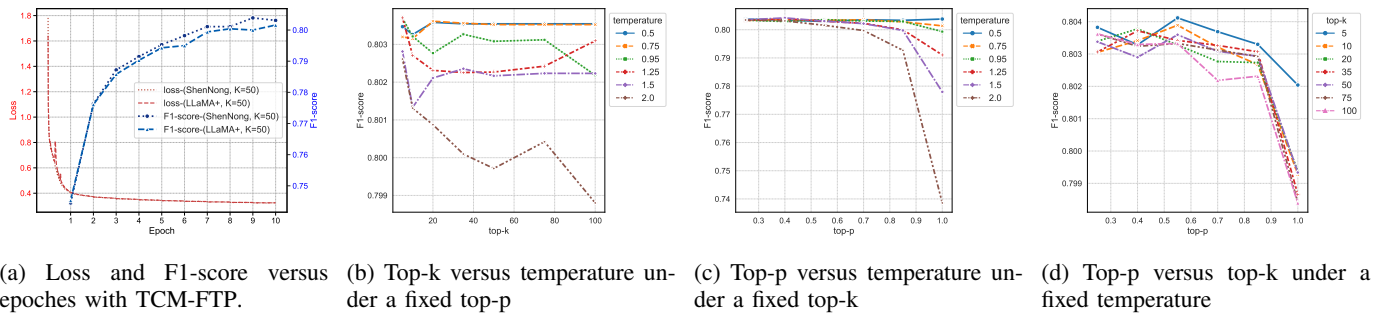


Fig. 3: Epoch-wise performance and decoding parameter analysis.

Item	Case 1	Case 2
Input	主诉: 腹泻36年。现病史: 病情稳定。肠鸣午后易发。腹胀减轻。腹部畏寒减轻。无胸闷气短。腹痛偶作。偶有心悸。睡眠可。小便正常。舌暗红。苔黄腻。舌下络脉迂曲青紫。脉弦。 Chief Complaint: Diarrhea for 36 years. Present Medical History: The condition is stable, with increased bowel sounds in the afternoon, occasional episodes of diarrhea, reduced abdominal distension, alleviated aversion to cold in the abdomen, no chest tightness or shortness of breath, occasional abdominal pain, occasional palpitations, normal sleep, normal urination. Dark red tongue with yellow greasy coating, sublingual veins are tortuous and purplish, and the pulse is string-like.	主诉: 胃隐痛4年。现病史: 胃脘部隐痛。时及后背。胃脘畏寒。饮食不慎则出现腹痛。腹泻。反酸。无暖气。烧心。口苦。唇暗。乏力感。眠可。舌淡紫。苔薄白。脉沉弦。 Chief Complaint: Persistent stomach discomfort for 4 years. Present Medical History: Dull pain in the epigastric region, occasionally radiating to the back. The epigastric area is sensitive to cold. Improper dietary choices lead to abdominal pain, diarrhea, acid regurgitation, without belching. Experience heartburn, bitter taste in the mouth, dark lips, a sense of fatigue. Sleep is normal. The tongue appears pale purple with a thin white and greasy coating. The pulse is deep and string-like.
Output (Ground Truth)	党参 15, 茯苓 45, 麸炒白术 30, 蜜甘草 10, 山药 15, 炒白扁豆 10, 陈皮 6, 白芍 15, 诃子肉 20, 乌梅 10, 莲子肉 15, 黄芩 10, 黄连 7, 北败酱草 20, 醋莪术 15, 丹参 20, 醋延胡索 15, 麸炒肉豆蔻 7, 醋五味子 10, 盐补骨脂 7, 生白术 10, 生地黄 7, 苦参 10, 牡蛎粉 20, 生牡蛎 20	黄芩 15, 柴胡 7, 炒白芍 10, 赤芍 10, 陈皮 10, 法半夏 15, 枳壳 30, 醋鸡内金 20, 浙贝母 15, 煅瓦楞子 30, 海螵蛸 30, 黄连 6, 制吴茱萸 1, 党参 15, 茯苓 30, 麸炒白术 20, 姜黄 15, 干姜 5, 川芎 15, 丹参 20, 醋莪术 15, 当归 10, 荷叶 15, 甘草 5, 豆蔻 10, 麸炒苍术 10
TCM-FTP (ShenNong, ours)	麸炒肉豆蔻 7, 黄莲 7, 乌梅 10, 茯苓 45, 诃子肉 20, 白芍 15, 醋五味子 10, 炒白扁豆 10, 麸炒白术 30, 醋莪术 15, 丹参 20, 生白术 10, 黄芩 10, 北败酱草 20, 莲子肉 15, 醋延胡索 20, 陈皮 6, 盐补骨脂 7, 生地黄 7, 蜜甘草 10, 党参 15, 甘松 10, 山药 15	枳壳 30, 黄连 6, 赤芍 10, 醋莪术 15, 麸炒苍术 10, 川芎 15, 柴胡 7, 浙贝母 15, 当归 10, 茯苓 30, 豆蔻 10, 姜黄 15, 党参 15, 煅瓦楞子 30, 干姜 5, 丹参 20, 陈皮 10, 黄芩 15, 醋鸡内金 20, 荷叶 15, 法半夏 15, 海螵蛸 30, 甘草 5, 制吴茱萸 1, 麸炒白术 20, 炒白芍 10
Mengzi (T5 base)	党参 15, 生白术 15, 麸炒白术 30, 茯苓 45, 白芍 15, 黄芩 10, 黄连 7, 北败酱草 20, 大枣 10, 乌梅 20, 木香 10, 生地黄 10, 当归 10, 陈皮 10, 防风 7, 肉桂 3, 山药 20, 炒葱白 20, 蜜甘草 7, 醋莪术 9, 醋三棱 9, 乌梅 10, 醋五味子 10, 白花蛇舌草 20, 半枝莲 10, 麸炒苍术 20, 莲子肉 20, 煅赤石脂 20, 车前草 20, 盐补骨脂 7, 煅肉豆蔻 7, 煅赤石脂 20	柴胡 10, 白芍 15, 陈皮 15, 法半夏 15, 旋覆花 20, 煅赭石 30, 姜厚朴 15, 麸炒枳实 15, 枳壳 45, 党参 15, 生白术 15, 麸炒苍术 30, 茯苓 45, 浙贝母 15, 煅瓦楞子 30, 海螵蛸 30, 黄芩 10, 黄连 6, 干姜 5, 肉桂 5, 生地黄 30, 生牡蛎 30, 炒酸枣仁 20, 醋鸡内金 15, 麸炒神曲 15, 炒麦芽 15, 醋延胡索 20, 九香虫 9, 蜜甘草 7, 醋莪术 9, 醋三棱 9, 白花蛇舌草 20, 醋香附 10, 木香 10, 乌药 10, 桂枝 10, 生蒲黄 20, 灵芝 20, 石菖蒲 15, 首乌藤 20, 灵芝 20
GPT-3.5	人参 10, 陈皮 6, 附子 3, 干姜 3, 甘草 3, 川椒子 10	金银花 15, 泽泻 10, 陈皮 6, 黄芩 10, 泽漆 10, 干姜 6, 白术 10, 甘草 5
GPT-4.0	白术 10, 陈皮 6, 茯苓 10, 木香 6, 砂仁 5, 肉桂 3, 甘草 5, 当归 10, 黄芩 10, 佩兰 6, 防风 6, 草果 6	丹参 10, 川芎 10, 白芍 10, 柴胡 10, 当归 10, 炙甘草 6, 茯苓 15, 薤白 10, 枳壳 10, 生姜 3, 大枣 4, 泽泻 10, 赤茯苓 10
Rank from TCM expert	TCM-FTP (ShenNong, ours) > Mengzi (T5 base) > GPT-4.0 > GPT-3.5	TCM-FTP (ShenNong, ours) > Mengzi (T5 base) > GPT-4.0 > GPT-3.5

Fig. 4: Two Specific Case Analyses. This figure presents two specific test data, including inputs, outputs, predictions from each model, and evaluations from experts. Herb names/dosage weights marked in red indicate the results consistent with the expert-prescribed outcomes.

[3] L. Ni, L. Chen, X. Huang, C. Han, J. Xu, H. Zhang, X. Luan, Y. Zhao, J. Xu, W. Yuan *et al.*, "Combating covid-19 with integrated traditional chinese and western medicine in china," *Acta Pharm. Sin. B.*, vol. 10, no. 7, pp. 1149–1162, 2020.

[4] Y. Huang, S. Wang, L. Wang, X. Yu, M. Jiang, J. Zhan, A. Lu, and G. Zheng, "Exploring the rules of li-fa-fang-yao on diabetes mellitus within traditional chinese medicine through text mining," in *ICCC*, 2012, pp. 1369–1373.

[5] L. Yao, Y. Zhang, B. Wei, W. Zhang, and Z. Jin, "A topic modeling approach for traditional chinese medicine prescriptions," *TKDE*, vol. 30, no. 6, pp. 1007–1021, 2018.

[6] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *NeurIPS*, vol. 27, 2014.

[7] Z. Liu, Z. Zheng, X. Guo, L. Qi, J. Gui, D. Fu, Q. Yao, and L. Jin, "Attentiveherb: a novel method for traditional chinese medicine prescription generation," *IEEE Access*, vol. 7, pp. 139 069–139 085, 2019.

[8] C. Li, D. Liu, K. Yang, X. Huang, and J. Lv, "Herb-know: Knowledge enhanced prescription generation for traditional chinese medicine," in *BIBM*, 2020, pp. 1560–1567.

[9] Z. Liu, C. Luo, D. Fu, J. Gui, Z. Zheng, L. Qi, and H. Guo, "A novel transfer learning model for traditional chinese medicine prescription generation from unstructured resources and knowledge," *Artif. Intell. Med.*, vol. 124, p. 102232, 2022.

[10] H. Pu, J. Mi, S. Lu, and J. He, "Rokepg: Roberta and knowledge enhancement for prescription generation of traditional chinese medicine," in *BIBM*, 2023, pp. 4615–4622.

[11] Y. Zhang, S. Liu, J. Xie, R. Liu, Y. Zhu, and Z. Bai, "Homogeneous symptom graph attentive reasoning network for herb recommendation," in *IJCNN*. IEEE, 2021, pp. 1–8.

[12] X. Dong, C. Zhao, X. Song, L. Zhang, Y. Liu, J. Wu, Y. Xu, N. Xu, J. Liu, H. Yu, K. Yang, and X. Zhou, "Presrecst: a novel herbal prescription recommendation algorithm for real-world patients with integration of syndrome differentiation and treatment planning," *JAMIA*, pp. 1268–1279, 2024.

[13] J. Liu, H. H. Zhuo, K. Jin, J. Yuan, Z. Yang, and Z. Yao, "Sequential condition evolved interaction knowledge graph for traditional chinese medicine recommendation," *arXiv preprint arXiv:2305.17866*, 2023.

[14] Y. Jin, W. Zhang, X. He, X. Wang, and X. Wang, "Syndrome-aware herb recommendation with multi-graph convolution network," in *ICDE*, 2020, pp. 145–156.

[15] E. J. Hu, y. shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," in *ICLR*, 2022.

[16] L. Yao, Y. Zhang, B. Wei, W. Zhang, and Z. Jin, "A topic modeling approach for traditional chinese medicine prescriptions," *TKDE*, vol. 30, no. 6, pp. 1007–1021, 2018.

[17] X. Wang, Y. Zhang, X. Wang, and J. Chen, "A knowledge graph enhanced topic modeling approach for herb recommendation," in *DASFAA*, 2019, pp. 709–724.

[18] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," *NeurIPS*, vol. 26, 2013.

[19] W. Zhao, W. Lu, Z. Li, H. Fan, Z. Yang, X. Lin, C. Li *et al.*, "Tcm herbal prescription recommendation model based on multi-graph convolutional network," *J. Ethnopharmacol.*, vol. 297, p. 115109, 2022.

[20] Y. Yang, Y. Rao, M. Yu, and Y. Kang, "Multi-layer information fusion based on graph convolutional network for knowledge-driven herb recommendation," *Neural Networks*, vol. 146, pp. 1–10, 2022.

[21] Z. Wang, J. Poon, and S. Poon, "Tcm translator: A sequence generation approach for prescribing herbal medicines," in *BIBM*, 2019, pp. 2474–2480.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

- [23] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," in *ICLR*, 2020.
- [24] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altschmidt, S. Altman, S. Anadkat *et al.*, "Gpt-4 technical report," *arXiv preprint arXiv:2303.08774*, 2023.
- [25] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, and G. Lample, "Llama: Open and efficient foundation language models," 2023.
- [26] B. Workshop, T. L. Scao, A. Fan, C. Akiki, E. Pavlick, S. Ilić, D. Hesslow, R. Castagné, A. S. Luccioni, F. Yvon *et al.*, "Bloom: A 176b-parameter open-access multilingual language model," *arXiv preprint arXiv:2211.05100*, 2022.
- [27] A. Zeng, X. Liu, Z. Du, Z. Wang, H. Lai, M. Ding, Z. Yang, Y. Xu, W. Zheng, X. Xia, W. L. Tam, Z. Ma, Y. Xue, J. Zhai, W. Chen, Z. Liu, P. Zhang, Y. Dong, and J. Tang, "GLM-130b: An open bilingual pre-trained model," in *ICLR*, 2023.
- [28] H. Wang, C. Liu, N. Xi, Z. Qiang, S. Zhao, B. Qin, and T. Liu, "Huatuo: Tuning llama model with chinese medical knowledge," 2023.
- [29] O. Byambasuren, Y. Yang, Z. Sui, D. Dai, B. Chang, S. Li, and H. Zan, "Preliminary study on the construction of chinese medical knowledge graph," *J. Chin. Inf. Process.*, vol. 33, no. 10, pp. 1–9, 2019.
- [30] H. Zhang, J. Chen, F. Jiang, F. Yu, Z. Chen, J. Li, G. Chen, X. Wu, Z. Zhang, Q. Xiao *et al.*, "HuatuoGPT, towards taming language model to be a doctor," 2023.
- [31] S. Yang, H. Zhao, S. Zhu, G. Zhou, H. Xu, Y. Jia, and H. Zan, "Zhongjing: Enhancing the chinese medical capabilities of large language model through expert feedback and real-world multi-turn dialogue," 2023.
- [32] W. Y. Wei Zhu and X. Wang, "Shennong-tcm: A traditional chinese medicine large language model," <https://github.com/michael-wzhu/ShenNong-TCM-LLM>, 2023.
- [33] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly, "Parameter-efficient transfer learning for nlp," in *ICML*, 2019, pp. 2790–2799.
- [34] X. Liu, K. Ji, Y. Fu, W. Tam, Z. Du, Z. Yang, and J. Tang, "P-tuning: Prompt tuning can be comparable to fine-tuning across scales and tasks," in *ACL-short*, 2022-05, pp. 61–68.
- [35] X. L. Li and P. Liang, "Prefix-tuning: Optimizing continuous prompts for generation," in *ACL*, 2021, pp. 4582–4597.
- [36] X. Liu, K. Ji, Y. Fu, W. L. Tam, Z. Du, Z. Yang, and J. Tang, "P-tuning v2: Prompt tuning can be comparable to fine-tuning universally across scales and tasks," *arXiv preprint arXiv:2110.07602*, 2021.
- [37] Z. Jiang, X. Zhou, X. Zhang, and S. Chen, "Using link topic model to analyze traditional chinese medicine clinical symptom-herb regularities," in *HealthCom*, 2012, pp. 15–18.
- [38] E. Eroshova, S. Fienberg, and J. Lafferty, "Mixed-membership models of scientific publications," *PNAS*, vol. 101, pp. 5220–5227, 2004.
- [39] R. Nallapati and W. Cohen, "Link-plsa-lda: A new unsupervised model for topics and influence of blogs," in *ICWSM*, vol. 2, no. 1, 2008, pp. 84–92.
- [40] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *JMLR*, vol. 3, pp. 993–1022, 2003.
- [41] T. Hofmann, "Probabilistic latent semantic indexing," in *SIGIR*, 1999, p. 50–57.
- [42] M. Zhang and Z. Zhou, "MI-knn: A lazy learning approach to multi-label learning," *Pattern Recogn.*, vol. 40, no. 7, pp. 2038–2048, 2007.
- [43] X. Dong, Y. Zheng, Z. Shu, K. Chang, D. Yan, J. Xia, Q. Zhu, K. Zhong, X. Wang, K. Yang, and X. Zhou, "Tcmpr: Tcm prescription recommendation based on subnetwork term mapping and deep learning," in *BIBM*, 2021, pp. 3776–3783.
- [44] Y. Yang, Y. Rao, M. Yu, and Y. Kang, "Multi-layer information fusion based on graph convolutional network for knowledge-driven herb recommendation," *Neural Netw.*, vol. 146, pp. 1–10, 2022.
- [45] Z. Zhang, H. Zhang, K. Chen, Y. Guo, J. Hua, Y. Wang, and M. Zhou, "Mengzi: Towards lightweight yet ingenious pre-trained models for chinese," 2021.
- [46] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *JMLR*, vol. 21, no. 140, pp. 1–67, 2020.