

Supplementary Table S1. Results of paired t-tests for the comparison models and MRHAF.

Comparison models	Paired differences			P-value
	Mean	SD	SE	
TCM-33k				
MRHAF vs KGCN	0.11083	0.07611	0.02537	0.0024
MRHAF vs KGAT	0.08839	0.04892	0.01631	0.0006
MRHAF vs RGCN	0.07190	0.03104	0.01035	0.0001
MRHAF vs KDHR	0.03146	0.01178	0.00393	0.0000
MRHAF vs SMGCN	0.04402	0.02136	0.00712	0.0003
MRHAF vs SMRGAT	0.02660	0.01260	0.00420	0.0002
TCM-9k				
MRHAF vs KGCN	0.12529	0.06839	0.02280	0.0006
MRHAF vs KGAT	0.09150	0.02938	0.00979	0.0000
MRHAF vs RGCN	0.09003	0.03713	0.01238	0.0001
MRHAF vs KDHR	0.06196	0.01781	0.00594	0.0000
MRHAF vs SMGCN	0.08231	0.02387	0.00796	0.0000
MRHAF vs SMRGAT	0.05321	0.01827	0.00609	0.0000

Supplementary Table S2. Results of paired t-tests for the ablation study.

Comparison models	Paired differences			T	P
	Mean	SD	SE		
TCM-33k					
MRHAF vs w/o-HACKGAT	0.00841	0.00206	0.00069	12.275	0.0000
MRHAF vs w/o-HSIGAT	0.01099	0.00325	0.00108	10.157	0.0000
MRHAF vs w/o-HESKGAT	0.01317	0.00373	0.00124	10.576	0.0000
TCM-9k					
MRHAF vs w/o-HACKGAT	0.00971	0.00309	0.00103	9.443	0.0000
MRHAF vs w/o-HSIGAT	0.01170	0.00652	0.00217	5.380	0.0007
MRHAF vs w/o-HESKGAT	0.01814	0.00619	0.00206	8.797	0.0000

Supplementary Table S3. Cohen's Kappa analysis of doctors' evaluation results on the TCM_33k dataset.

	Top5		Top10		Top20	
	Kappa	P	Kappa	P	Kappa	P
doctor1 & doctor2	0.065	0.088	0.043	0.211	0.014	0.186
doctor1 & doctor3	0.018	0.167	0.053	0.262	0.172	0.000**
doctor1 & doctor4	0.065	0.022*	0.127	0.004**	0.076	0.086
doctor1 & doctor5	0.179	0.003**	0.102	0.068	0.078	0.051
doctor1 & doctor6	0.030	0.634	0.104	0.049*	0.066	0.277
doctor2 & doctor3	0.099	0.000**	0.088	0.103	0.006	0.342
doctor2 & doctor4	0.127	0.003**	0.107	0.007**	0.038	0.011*
doctor2 & doctor5	0.030	0.462	0.010	0.694	0.007	0.309
doctor2 & doctor6	0.124	0.001**	0.018	0.428	0.031	0.045*
doctor3 & doctor4	0.117	0.001**	0.125	0.010**	0.085	0.034*
doctor3 & doctor5	0.005	0.721	0.026	0.470	0.183	0.000**
doctor3 & doctor6	0.018	0.167	0.076	0.028*	0.030	0.416
doctor4 & doctor5	0.031	0.326	0.027	0.447	0.034	0.408
doctor4 & doctor6	0.100	0.000**	0.161	0.000**	0.130	0.007**

doctor5 & doctor6	0.145	0.013*	0.127	0.013*	0.094	0.016*
-------------------	-------	---------------	-------	---------------	-------	---------------

* $P < 0.05$ ** $P < 0.01$

Supplementary Table S4. Cohen's Kappa analysis of doctors' evaluation results on the TCM_9k dataset.

	Top5		Top10		Top20	
	Kappa	P	Kappa	P	Kappa	P
doctor1 & doctor2	0.084	0.093	0.012	0.710	0.014	0.399
doctor1 & doctor3	0.046	0.080	0.044	0.399	0.174	0.002**
doctor1 & doctor4	0.104	0.045*	0.045	0.351	0.125	0.022*
doctor1 & doctor5	0.048	0.157	0.023	0.433	0.068	0.152
doctor1 & doctor6	0.195	0.008**	0.145	0.001**	0.184	0.004**
doctor2 & doctor3	0.256	0.000**	0.069	0.371	0.021	0.238
doctor2 & doctor4	0.179	0.004**	0.024	0.423	0.034	0.125
doctor2 & doctor5	0.284	0.000**	0.025	0.326	0.071	0.175
doctor2 & doctor6	0.192	0.007**	0.036	0.159	0.016	0.324
doctor3 & doctor4	0.036	0.317	0.002	0.961	0.219	0.007**
doctor3 & doctor5	0.272	0.000**	0.099	0.006**	0.022	0.640
doctor3 & doctor6	0.122	0.002**	0.033	0.399	0.285	0.000**
doctor4 & doctor5	0.123	0.004**	0.073	0.337	0.185	0.001**
doctor4 & doctor6	0.203	0.004**	0.009	0.918	0.137	0.076
doctor5 & doctor6	0.096	0.068	0.340	0.000**	0.110	0.014*

* $P < 0.05$ ** $P < 0.01$

Supplementary Table S5. Results of Kendall's W analysis for doctor ratings on two datasets at different Top-N ranks.

	Kendall's W	P
TCM_33k		
Top5	0.276	0.000**
Top10	0.255	0.000**
Top20	0.247	0.000**
TCM_9k		
Top5	0.326	0.000**
Top10	0.252	0.002**
Top20	0.232	0.012*

* $P < 0.05$ ** $P < 0.01$

Supplementary Table S6. Intraclass Correlation Coefficient (ICC) analysis of inter-rater reliability for doctor evaluations on the two datasets.

Dataset	Recommended herbs	Bidirectional Mixed/Random Consistency	ICC Value	95% CI	P
TCM_33k	Top5	Single Measure ICC(C, 1)	0.120	0.075 ~ 0.173	0.000**
		Average Measure ICC(C, k)	0.450	0.328 ~ 0.557	0.000**
	Top10	Single Measure ICC(C, 1)	0.094	0.052 ~ 0.145	0.000**
		Average Measure ICC(C, k)	0.385	0.249 ~ 0.504	0.000**
	Top20	Single Measure ICC(C, 1)	0.086	0.045 ~ 0.136	0.000**
		Average Measure ICC(C, k)	0.361	0.220 ~ 0.485	0.000**
TCM_9k	Top5	Single Measure ICC(C, 1)	0.177	0.099 ~ 0.275	0.000**
		Average Measure ICC(C, k)	0.563	0.397 ~ 0.695	0.000**
	Top10	Single Measure ICC(C, 1)	0.110	0.042 ~ 0.200	0.000**
		Average Measure ICC(C, k)	0.426	0.209 ~ 0.600	0.000**
	Top20	Single Measure ICC(C, 1)	0.078	0.015 ~ 0.162	0.006**
		Average Measure ICC(C, k)	0.335	0.083 ~ 0.537	0.006**

Note: C stands for Consistency, k represents the number of raters (k=6). Average Measures ICC(C, k) is the primary metric, indicating the reliability of the mean rating across all six experts. * $P < 0.05$ ** $P < 0.01$.