APPENDIX

A. *Prompts for* DIRECT *Paradigm*

You are an exceptionally intelligent assistant that detects anomalies in time series data by listing all the anomalies.
Task Description
For Univariable Tasks: Your task is to determine whether any time steps in the satellite telemetry sequence are anomalous.
For Multivariable Tasks: Your task is to determine whether any time steps in the multivariate satellite telemetry time series are anomalous. The data is represented as an array where each element is an array, corresponding to a variable sequence, with a total of 27 variable sequences collected synchronously over the same time period, reflecting interdependent measurements from multiple sensors.
Requirements
Requirements:
1. Provide the analysis process, starting with "Analysis Process:".
2. Provide the final answer, starting with "Final Answer:", returning the indices of anomalies in the sequence (0-`{WINDOW_SIZE_MAX_INDEX}`). Do not say anything like "the anomalous indices in the sequence are", just return the numbers. If you think there are no anomalies in the sequence, please return None.
3. If reference data or examples are provided, they are intended solely to illustrate normal data patterns and potential anomaly types. Do not directly replicate the answers or anomaly indices from the examples, as they represent specific cases and are not universally applicable. For instance, if an example identifies the latter half of a sequence as anomaly indices, this is merely one scenario, as anomalies may occur anywhere within the entire sequence range. The entire sequence may be entirely anomalous data.
Answer
Input: `{test_sequence}`
Output:

Fig. A1: The zero-shot prompt template used for the DIRECT paradigm.

You are an exceptionally intelligent assistant that detects anomalies in time series data by listing all the anomalies.
Task Description
For Univariable Tasks: Your task is to determine whether any time steps in the satellite telemetry sequence are anomalous.
For Multivariable Tasks: Your task is to determine whether any time steps in the multivariate satellite telemetry time series are anomalous. The data is represented as an array where each element is an array, corresponding to a variable sequence, with a total of 27 variable sequences collected synchronously over the same time period, reflecting interdependent measurements from multiple sensors.
Requirements
Requirements:
1. Provide the analysis process, starting with "Analysis Process:".
2. Provide the final answer, starting with "Final Answer:", returning the indices of anomalies in the sequence (0-`{WINDOW_SIZE_MAX_INDEX}`). Do not say anything like "the anomalous indices in the sequence are", just return the numbers. If you think there are no anomalies in the sequence, please return None.
3. If reference data or examples are provided, they are intended solely to illustrate normal data patterns and potential anomaly types. Do not directly replicate the answers or anomaly indices from the examples, as they represent specific cases and are not universally applicable. For instance, if an example identifies the latter half of a sequence as anomaly indices, this is merely one scenario, as anomalies may occur anywhere within the entire sequence range. The entire sequence may be entirely anomalous data.
Examples
Example 1:
Input: `{positive_samples_str}`
Output: Analysis Process: `{positive_analysis_process}`Final Answer: `{positive_final_answer}`
Example 2:
Input: `{negative_samples_str1}`
Output: Analysis Process: `{negative_analysis_process1}`Final Answer: `{negative_final_answer1}`
Example 3:
Input: `{negative_samples_str2}`
Output: Analysis Process: `{negative_analysis_process2}`Final Answer: `{negative_final_answer2}`
Example 4:
Input:`{negative_samples_str3}`
Output:Analysis Process:`{negative_analysis_process3}`Final Answer:`{negative_final_answer3}`
Answer
Input: `{test_sequence}`
Output:

Fig. A2: The few-shot prompt template for the DIRECT paradigm, which augments the instructions with in-context examples.

You are an exceptionally intelligent assistant that detects anomalies in time series data by listing all the anomalies.
Task Description
For Univariable Tasks: Your task is to determine whether any time steps in the satellite telemetry sequence are anomalous.
For Multivariable Tasks: Your task is to determine whether any time steps in the multivariate satellite telemetry time series are anomalous. The data is represented as an array where each element is an array, corresponding to a variable sequence, with a total of 27 variable sequences collected synchronously over the same time period, reflecting interdependent measurements from multiple sensors.
 # # Requirements
Requirements:
1. Provide the analysis process, starting with "Analysis Process:".
2. Provide the final answer, starting with "Final Answer:", returning the indices of anomalies in the sequence (0-`{WINDOW_SIZE_MAX_INDEX}`). Do not say anything like "the anomalous indices in the sequence are", just return the numbers. If you think there are no anomalies in the sequence, please return None.
3. If reference data or examples are provided, they are intended solely to illustrate normal data patterns and potential anomaly types. Do not directly replicate the answers or anomaly indices from the examples, as they represent specific cases and are not universally applicable. For instance, if an example identifies the latter half of a sequence as anomaly indices, this is merely one scenario, as anomalies may occur anywhere within the entire sequence range. The entire sequence may be entirely anomalous data.
 # # Reference
The following data, retrieved from the satellite telemetry database, is the most similar to the input. Please use it as a reference: `{rag_str}`
Answer
Input: `{test_sequence}`
Output:

Fig. A3: The RAG prompt template for the DIRECT paradigm, which provides a retrieved normal sample as context.

You are an exceptionally intelligent assistant that performs time series forecasting on satellite telemetry data by generating accurate predictions for future values.
Task Description
For Univariable Tasks: Your task is to predict the next `{HORIZON}` time steps of the satellite telemetry time series data.
For Multivariable Tasks: Your task is to predict the next `{HORIZON}` time steps of the multivariate satellite telemetry time series data.The data is represented as a list where each sublist corresponds to a variable sequence, with a total of 27 variable sequences collected synchronously over the same time period, reflecting interdependent measurements from multiple sensors.
Requirements
Requirements:
1. Provide the analysis process, starting with "Analysis Process:".
For Univariable Tasks: 2. Provide the final answer, starting with "Final Answer:", returning the predicted values for the next HORIZON time steps as a list of numbers. Do not include additional explanations in this section, just the predicted values.
For Multivariable Tasks: 2. Provide the final answer, starting with "Final Answer:". Predicted values for the 27 variable sequences must be returned as a list of 27 sublists, each sublist containing `{HORIZON}` predicted values for one sequence. No additional explanations should be included in this section, only the predicted values.
3. If reference data or examples are provided, they are intended to illustrate normal data patterns. You may use them as a reference for normal patterns during prediction, but you must not directly replicate them.
Answer
Input: `{test_sequence}`
Output:

Fig. B1: The zero-shot prompt template used for the PREDICTION-BASED paradigm.

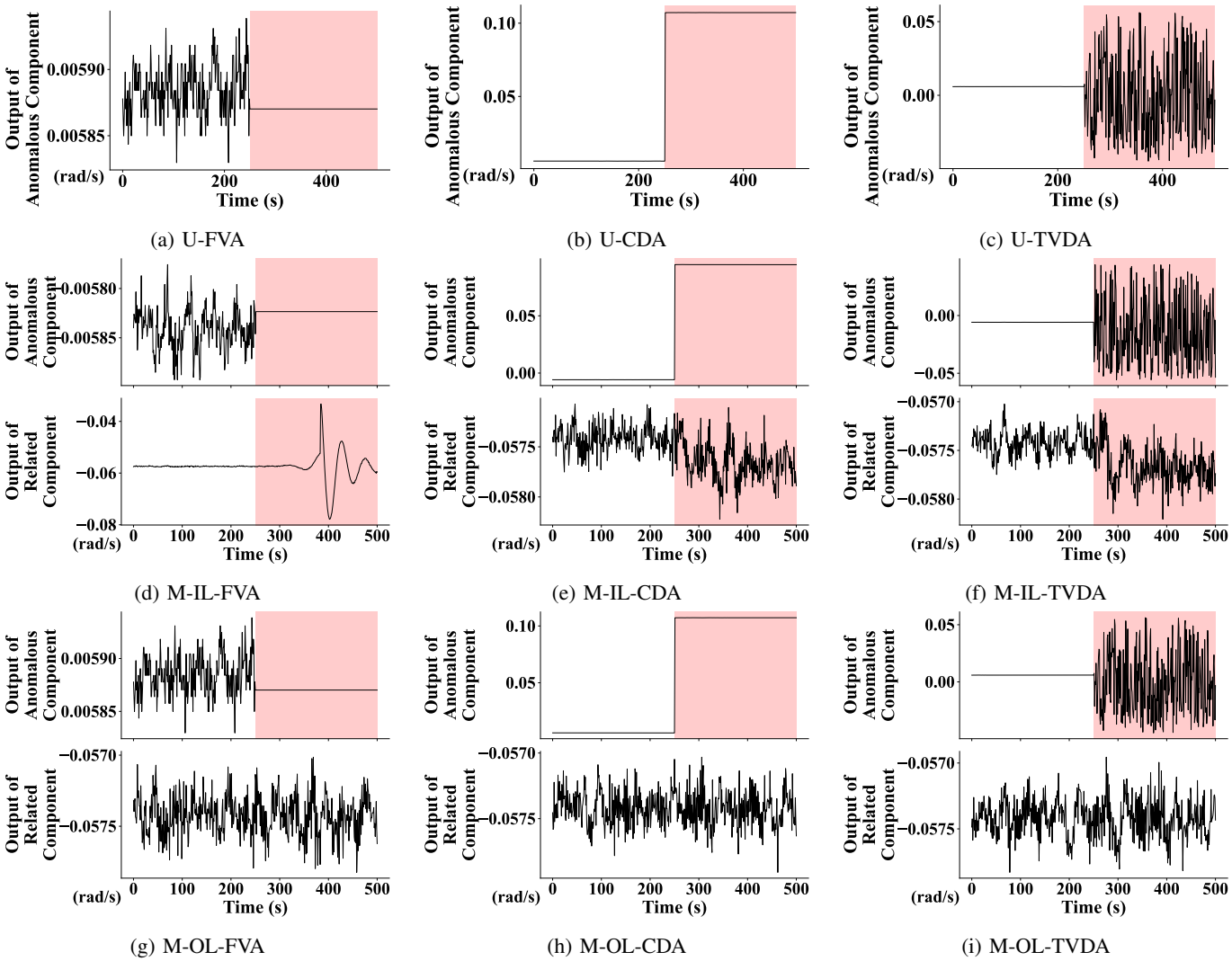
You are an exceptionally intelligent assistant that performs time series forecasting on satellite telemetry data by generating accurate predictions for future values.
Task Description
For Univariable Tasks: Your task is to predict the next `{HORIZON}` time steps of the satellite telemetry time series data.
For Multivariable Tasks: Your task is to predict the next `{HORIZON}` time steps of the multivariate satellite telemetry time series data.The data is represented as a list where each sublist corresponds to a variable sequence, with a total of 27 variable sequences collected synchronously over the same time period, reflecting interdependent measurements from multiple sensors.
Requirements
Requirements:
1. Provide the analysis process, starting with "Analysis Process:".
For Univariable Tasks: 2. Provide the final answer, starting with "Final Answer:", returning the predicted values for the next HORIZON time steps as a list of numbers. Do not include additional explanations in this section, just the predicted values.
For Multivariable Tasks: 2. Provide the final answer, starting with "Final Answer:". Predicted values for the 27 variable sequences must be returned as a list of 27 sublists, each sublist containing `{HORIZON}` predicted values for one sequence. No additional explanations should be included in this section, only the predicted values.
3. If reference data or examples are provided, they are intended to illustrate normal data patterns. You may use them as a reference for normal patterns during prediction, but you must not directly replicate them.
Example
Example:
Input:`{positive_samples_str}`
Output:Analysis Process:`{positive_analysis_process}`Final Answer:`{positive_final_answer_str}`
Answer
Input: `{test_sequence}`
Output:

Fig. B2: The few-shot prompt template for the PREDICTION-BASED paradigm, which augments the instructions with in-context examples.

You are an exceptionally intelligent assistant that performs time series forecasting on satellite telemetry data by generating accurate predictions for future values.
Task Description
For Univariable Tasks: Your task is to predict the next `{HORIZON}` time steps of the satellite telemetry time series data.
For Multivariable Tasks: Your task is to predict the next `{HORIZON}` time steps of the multivariate satellite telemetry time series data.The data is represented as a list where each sublist corresponds to a variable sequence, with a total of 27 variable sequences collected synchronously over the same time period, reflecting interdependent measurements from multiple sensors.
Requirements
Requirements:
1. Provide the analysis process, starting with "Analysis Process:".
For Univariable Tasks: 2. Provide the final answer, starting with "Final Answer:", returning the predicted values for the next HORIZON time steps as a list of numbers. Do not include additional explanations in this section, just the predicted values.
For Multivariable Tasks: 2. Provide the final answer, starting with "Final Answer:". Predicted values for the 27 variable sequences must be returned as a list of 27 sublists, each sublist containing `{HORIZON}` predicted values for one sequence. No additional explanations should be included in this section, only the predicted values.
3. If reference data or examples are provided, they are intended to illustrate normal data patterns. You may use them as a reference for normal patterns during prediction, but you must not directly replicate them.
Reference
The following data, retrieved from the satellite telemetry database, is the most similar to the input. Please use it as a reference:`{rag_str}`
Answer
Input: `{test_sequence}`
Output:

Fig. B3: The RAG prompt template for the PREDICTION-BASED paradigm, which provides a retrieved normal sample as context.

C. *Details of* ATSADBENCH



(a) U-FVA (b) U-CDA (c) U-TVDA

(d) M-IL-FVA (e) M-IL-CDA (f) M-IL-TVDA

(g) M-OL-FVA (h) M-OL-CDA (i) M-OL-TVDA

Fig. C1: Visualization of the nine distinct anomaly detection tasks for evaluation.

D. Hyperparameter Tuning for LLM-based Approaches

TABLE D1: Hyperparameter tuning results for DeepSeek-V3 under the DIRECT paradigm. **Bolded values** indicate the best results for each metric within each task. The highlighted row indicates the hyperparameter configuration ultimately selected.

Task Category	Window Size	Step Size	Precision	Recall	F1 Score	AA	AL	AC
			Multivariate Tasks					
M-IL-FVA	6	6	**0.5829**	**0.4780**	**0.5253**	0.4790	**1.0**	0.0569
	10	10	0.5081	0.4060	0.4514	0.5175	1.5	0.0900
	20	20	0.5095	0.3495	0.4146	**0.5200**	**1.0**	**0.1000**
M-IL-CDA	6	6	**0.5650**	**0.4175**	**0.4802**	**0.4865**	2.0	0.0539
	10	10	0.5031	0.4015	0.4466	0.4800	**1.0**	0.0850
	20	20	0.4174	0.2780	0.3337	0.4550	3.5	**0.1000**
M-IL-TVDA	6	6	0.4993	0.3475	0.4098	0.4865	**1.0**	0.0419
	10	10	**0.5337**	**0.4550**	**0.4912**	0.4725	**1.0**	0.0550
	20	20	0.5195	0.3260	0.4006	**0.4900**	1.5	**0.1300**
M-OL-FVA	6	6	**0.5590**	**0.4240**	**0.4822**	0.4234	**1.5**	0.0569
	10	10	0.5363	0.4695	0.5007	0.4575	2.5	0.0650
	20	20	0.5261	0.3780	0.4399	**0.5250**	**1.5**	**0.1400**
M-OL-CDA	6	6	**0.5414**	**0.4250**	**0.4762**	0.4895	3.0	0.0509
	10	10	0.4788	0.3735	0.4197	**0.5175**	**1.0**	0.0900
	20	20	0.4723	0.3030	0.3692	0.4450	1.5	**0.1100**
M-OL-TVDA	6	6	**0.5137**	0.3655	**0.4271**	0.4414	**1.0**	0.0569
	10	10	0.4997	**0.3860**	0.4355	**0.4825**	2.0	0.0500
	20	20	0.4704	0.3015	0.3675	0.4350	**1.0**	**0.1700**
M-Avg.	6	6	**0.5435**	0.4096	**0.4668**	0.4677	1.5833	0.0529
	10	10	0.5100	**0.4153**	0.4575	**0.4879**	**1.5000**	0.0725
	20	20	0.4859	0.3227	0.3876	0.4783	1.6667	**0.1250**
			Univariate Tasks					
U-FVA	100	20	0.1111	0.0700	0.0859	0.1615	4.0	0.1000
	150	30	0.0893	0.0650	0.0752	**0.7581**	**1.0**	**0.3939**
	200	40	0.2232	0.1600	0.1864	0.1739	2.0	0.1200
	500	100	**0.5081**	**0.3750**	**0.4315**	0.5000	6.5	0.3500
U-CDA	100	20	0.5585	**0.6130**	**0.5845**	0.4948	6.0	0.1400
	150	30	0.5136	0.6050	0.5556	0.5161	3.5	0.1818
	200	40	0.5534	0.5230	0.5378	0.4457	5.0	0.3400
	500	100	0.3707	0.5055	0.4278	**0.6250**	**3.0**	**0.4000**
U-TVDA	100	20	0.4989	0.5465	0.5216	0.4688	**1.0**	0.1100
	150	30	0.5268	0.5350	0.5309	**0.5403**	2.0	0.1667
	200	40	0.4675	0.4705	0.4690	0.3370	1.5	0.1200
	500	100	**0.6125**	**0.6275**	**0.6199**	0.4688	3.0	**0.3500**
U-Avg.	100	20	0.3895	0.4098	0.3973	0.3750	3.6667	0.1167
	150	30	0.3766	0.4017	0.3872	**0.6048**	**2.1667**	0.2475
	200	40	0.4147	0.3845	0.3977	0.3188	2.8333	0.1933
	500	100	**0.4971**	**0.5027**	**0.4931**	0.5313	4.1667	**0.3667**

TABLE D2: Hyperparameter tuning results for DeepSeek-V3 under the PREDICTION-BASED paradigm. **Bolded values** indicate the best results for each metric within each task. The highlighted row indicates the hyperparameter configuration ultimately selected.

Task	Window Size	Step Size	Prediction Horizon	ROC	PRC	Precision	Recall	F1	AA	AL	AC
Multivariate Tasks											
M-IL-FVA	10	5	5	0.4577	0.4891	0.4572	0.4570	0.4571	0.4572	2.0	0.0300
	20	5	5	**0.5348**	**0.5146**	**0.5043**	**0.5040**	**0.5041**	**0.5038**	2.5	**0.0425**
	30	5	5	0.4482	0.4572	0.4042	0.4040	0.4041	0.4025	**1.5**	0.0325
M-IL-CDA	10	5	5	0.6444	**0.6553**	0.6123	0.6120	0.6122	0.6125	2.5	0.0550
	20	5	5	**0.6716**	0.6238	**0.6513**	**0.6510**	**0.6512**	**0.6513**	**2.0**	**0.0650**
	30	5	5	0.6581	0.6484	0.5978	0.5975	0.5976	0.5975	**2.0**	0.0550
M-IL-TVDA	10	5	5	0.6023	0.5739	0.5613	0.5610	0.5611	0.5625	1.0	**0.0525**
	20	5	5	0.6177	**0.5821**	0.5493	0.5490	0.5491	0.5550	1.0	**0.0525**
	30	5	5	**0.6226**	0.5663	**0.6273**	**0.6270**	**0.6272**	**0.6275**	1.0	0.0475
M-OL-FVA	10	5	5	**0.6100**	**0.5963**	**0.6023**	**0.6020**	**0.6022**	**0.6013**	**2.0**	**0.0700**
	20	5	5	0.5112	0.5166	0.4942	0.4940	0.4941	0.4950	3.0	0.0425
	30	5	5	0.4649	0.4861	0.4332	0.4330	0.4331	0.4325	5.0	0.0325
M-OL-CDA	10	5	5	0.5599	0.5445	0.5318	0.5315	0.5316	0.5325	1.5	0.0525
	20	5	5	**0.6156**	**0.5848**	**0.5848**	**0.5845**	**0.5846**	**0.5850**	**1.0**	**0.0575**
	30	5	5	0.5803	0.5528	0.5008	0.5005	0.5006	0.5000	**1.0**	0.0325
M-OL-TVDA	10	5	5	0.5826	0.5604	0.5323	0.5320	0.5321	0.5325	**1.5**	0.0550
	20	5	5	**0.6003**	**0.5668**	0.5958	0.5955	0.5956	0.5975	3.5	**0.0600**
	30	5	5	0.5971	0.5556	**0.6103**	**0.6100**	**0.6102**	**0.6100**	2.5	0.0450
M-Avg.	10	5	5	0.5762	**0.5699**	0.5495	0.5493	0.5494	0.5498	**1.7500**	0.0525
	20	5	5	**0.5918**	0.5648	**0.5633**	**0.5630**	**0.5631**	**0.5646**	2.1667	**0.0533**
	30	5	5	0.5619	0.5444	0.5289	0.5287	0.5288	0.5283	2.1667	0.0408
Univariate Tasks											
U-FVA	100	20	20	0.0723	0.3068	0.0828	0.0825	0.0827	0.0650	**1.0**	0.0700
	250	20	20	0.1365	0.3333	0.1617	0.1615	0.1616	0.1250	1.5	0.1300
	500	20	20	0.2624	0.3704	0.3022	0.3020	0.3021	0.2950	**1.0**	0.2600
	750	20	20	0.4940	0.5262	0.5158	0.5155	0.5156	0.5050	3.0	0.4900
	1000	20	20	**0.5514**	**0.5471**	**0.5368**	**0.5365**	**0.5366**	**0.5200**	**1.0**	**0.5100**
U-CDA	100	20	20	0.5620	0.6390	0.5193	0.5180	0.5186	0.4850	1.0	0.1300
	250	20	20	0.5588	0.5883	0.5213	0.5200	0.5207	0.5100	1.0	0.2100
	500	20	20	0.6242	0.6258	0.5821	0.5815	0.5818	0.5800	1.0	0.3000
	750	20	20	0.6681	0.6406	0.7159	0.7145	0.7152	0.7150	1.0	0.4700
	1000	20	20	**0.7215**	**0.6456**	**0.7799**	**0.7795**	**0.7797**	**0.7800**	1.0	0.5500
U-TVDA	100	20	20	0.8024	0.7495	0.7804	0.7800	0.7802	0.7900	**1.0**	0.5600
	250	20	20	0.8132	0.7497	0.7949	0.7945	0.7947	0.8000	**1.0**	0.6400
	500	20	20	0.7515	0.6316	0.7979	0.7975	0.7977	0.8100	**1.0**	**0.7500**
	750	20	20	**0.8736**	**0.7541**	**0.8354**	**0.8350**	**0.8352**	**0.8550**	1.5	0.5100
	1000	20	20	0.8123	0.6694	0.8194	0.8190	0.8192	0.8300	1.5	0.5800
U-Avg.	100	20	20	0.4789	0.5651	0.4608	0.4602	0.4605	0.4467	**1.0000**	0.2533
	250	20	20	0.5028	0.5571	0.4926	0.4920	0.4923	0.4783	1.1667	0.3267
	500	20	20	0.5460	0.5426	0.5607	0.5603	0.5605	0.5617	**1.0000**	0.4367
	750	20	20	0.6786	**0.6403**	0.6890	0.6883	0.6887	0.6917	1.8333	0.4900
	1000	20	20	**0.6951**	0.6207	**0.7120**	**0.7117**	**0.7118**	**0.7100**	1.1667	**0.5467**

E. Enhancement Strategies

TABLE E1: F1 improvement of Few-Shot and RAG over Zero-Shot setting. Green: positive/zero; Red: negative.

method	U-FVA	U-CDA	U-TVDA	U	MO-FVA	MO-FVA	MO-TVDA	M-OL	MI-FVA	MI-CDA	MI-TVDA	M-IL	Overall
DeepSeek-V3 with DIRECT *pardiagm*													
Zero-Shot	0.4315	0.4278	0.6199	0.4931	0.5007	0.4197	0.4355	0.4520	0.4514	0.4466	0.4912	0.4631	0.4694
Few-Shot	+23.36%	+9.58%	-6.45%	+8.83%	+19.19%	+2.36%	+16.46%	+12.67%	+35.96%	-13.83%	+14.66%	+12.26%	+11.25%
RAG	+40.81%	+11.21%	+4.75%	+18.92%	-9.86%	-11.71%	-15.12%	-12.23%	+0.47%	-12.66%	-22.73%	-11.64%	-1.65%
Qwen3 with DIRECT *pardiagm*													
Zero-Shot	0.7760	0.3341	0.3450	0.4850	0.2507	0.2010	0.2847	0.2455	0.2982	0.2583	0.2597	0.2721	0.3342
Few-Shot	-1.00%	+31.38%	+36.55%	+22.31%	+51.28%	+19.91%	+26.61%	+30.60%	+62.18%	-6.01%	+33.16%	+29.78%	+28.23%
RAG	-3.22%	+6.55%	+37.97%	+13.76%	+18.60%	+14.91%	+9.26%	+14.25%	+32.92%	+11.31%	+11.36%	+18.53%	+15.52%
DeepSeek-V3 with PREDICTION-BASED *pardiagm*													
Zero-Shot	0.5366	0.7797	0.8192	0.7118	0.4941	0.5846	0.5956	0.5581	0.5041	0.6512	0.5491	0.5681	0.6127
Few-Shot	+2.35%	-2.09%	-4.95%	-1.56%	-5.15%	-16.75%	-22.86%	-14.92%	-17.20%	-3.95%	-14.00%	-11.72%	-9.40%
RAG	-1.65%	+0.97%	-1.45%	-0.71%	-9.45%	-0.95%	-13.55%	-7.99%	-12.05%	-10.30%	+5.20%	-5.72%	-4.80%
Qwen3 with PREDICTION-BASED *pardiagm*													
Zero-Shot	0.5066	0.6517	0.9722	0.7102	0.3931	0.4436	0.4226	0.4198	0.4471	0.4386	0.5246	0.4701	0.5333
Few-Shot	+1.35%	+18.85%	-3.05%	+5.72%	-1.80%	-3.30%	+16.40%	+3.77%	+3.60%	-0.75%	+2.05%	+1.63%	+3.71%
RAG	-0.59%	+28.86%	-3.25%	+8.34%	-6.45%	-3.20%	+3.25%	-2.13%	-2.30%	-2.65%	+6.80%	+0.62%	+2.27%

TABLE E2: AA improvement of Few-Shot and RAG over Zero-Shot setting. Green: positive/zero; Red: negative.

method	U-FVA	U-CDA	U-TVDA	U	MO-FVA	MO-FVA	MO-TVDA	M-OL	MI-FVA	MI-CDA	MI-TVDA	M-IL	Overall
DeepSeek-V3 with DIRECT *pardiagm*													
Zero-Shot	0.5000	0.6250	0.4688	0.5313	0.4575	0.5175	0.4825	0.4858	0.5175	0.4800	0.4725	0.4900	0.5024
Few-Shot	+ 0.00%	-12.50%	+15.63%	+1.04%	+4.25%	-1.75%	+2.25%	+1.58%	-18.75%	+2.00%	-4.50%	-7.08%	-1.49%
RAG	-18.75%	+0.00%	-3.13%	-7.29%	-4.50%	+2.75%	+10.00%	+2.75%	-1.50%	+12.50%	+8.00%	+6.33%	+0.60%
Qwen3 with DIRECT *pardiagm*													
Zero-Shot	0.4375	0.5313	0.3750	0.4479	0.5300	0.5100	0.4975	0.5125	0.4900	0.4650	0.4875	0.4808	0.4804
Few-Shot	+3.13%	+9.38%	+6.25%	+6.25%	-3.00%	-1.00%	+0.25%	-1.25%	-21.25%	+3.50%	+1.25%	-5.50%	-0.17%
RAG	+12.50%	-15.63%	+6.25%	+1.04%	-4.00%	+3.00%	-3.00%	-1.33%	-11.00%	+1.50%	+5.00%	-1.50%	-0.60%
DeepSeek-V3 with PREDICTION-BASED *pardiagm*													
Zero-Shot	0.5200	0.7800	0.8300	0.7100	0.4950	0.5850	0.5975	0.5592	0.5038	0.6513	0.5550	0.5700	0.6131
Few-Shot	+4.00%	-2.00%	-2.50%	-0.17%	-5.25%	-16.75%	-22.75%	-14.92%	-17.13%	-3.88%	-14.13%	-11.71%	-8.93%
RAG	+0.00%	+2.00%	+0.00%	+0.67%	-9.50%	-1.00%	-13.75%	-8.08%	-12.13%	-10.75%	+4.62%	-6.08%	-4.50%
Qwen3 with PREDICTION-BASED *pardiagm*													
Zero-Shot	0.5000	0.6100	0.9900	0.7000	0.3725	0.4425	0.4150	0.4100	0.4338	0.4238	0.5413	0.4663	0.5254
Few-Shot	+2.00%	+22.00%	-4.00%	+6.67%	-4.13%	-5.13%	+18.75%	+3.17%	+5.63%	-3.88%	+3.00%	+1.58%	+3.81%
RAG	-1.00%	+33.00%	-5.00%	+9.00%	-10.75%	-5.13%	+1.13%	-4.92%	-2.38%	-4.63%	-1.00%	-2.67%	+0.47%

TABLE E3: AL improvement of Few-Shot and RAG over Zero-Shot setting. Green: positive/zero; Red: negative.

method	U-FVA	U-CDA	U-TVDA	U	MO-FVA	MO-FVA	MO-TVDA	M-OL	MI-FVA	MI-CDA	MI-TVDA	M-IL	overall
ZeroShot	6.50	3.00	3.00	4.17	2.50	1.00	2.00	1.83	1.50	1.00	1.00	1.17	2.39
FewShot	-0.38	+1.00	-0.67	-0.50	-1.50	0.00	-1.00	-0.83	-0.50	0.00	0.00	-0.17	-0.50
RAG	-0.85	0.00	-0.33	-2.17	+2.00	+1.50	0.00	+1.17	-0.50	+0.50	+1.00	+0.33	-0.22
Zero-Shot	1.50	1.50	4.00	2.33	3.00	1.00	1.50	1.83	1.50	1.50	1.00	1.33	1.83
FewShot	0.00	-0.33	+0.50	+0.50	-2.00	0.00	-0.50	-0.83	-0.50	-0.50	0.00	-0.33	-0.22
RAG	-0.33	+1.67	-0.75	-0.33	-2.00	+1.00	0.00	-0.33	-0.50	+1.00	+1.50	+0.67	0.00
Zero-Shot	1.00	1.00	1.50	1.17	3.00	1.00	3.50	2.50	2.50	2.00	1.00	1.83	1.83
FewShot	+4.50	0.00	0.00	+1.50	+1.50	+1.00	-2.00	+0.17	-1.00	-0.50	0.00	-0.50	+0.39
RAG	+1.50	0.00	-0.33	+0.33	+2.50	+1.50	-1.50	+0.83	-1.50	0.00	+2.00	+0.17	+0.44
ZeroShot	1.50	1.00	1.00	1.17	1.00	1.00	1.50	1.17	2.00	1.00	1.00	1.33	1.22
FewShot	+0.67	0.00	0.00	+0.33	+7.00	0.00	-0.50	+2.17	-1.00	0.00	0.00	-0.33	+0.72
RAG	-0.33	0.00	0.00	-0.17	+1.50	0.00	+0.50	+0.67	-0.50	0.00	0.00	-0.17	+0.11

TABLE E4: AC improvement of Few-Shot and RAG over Zero-Shot setting. Green: positive/zero; Red: negative.

method	U-FVA	U-CDA	U-TVDA	U	MO-FVA	MO-FVA	MO-TVDA	M-OL	MI-FVA	MI-CDA	MI-TVDA	M-IL	Overall
DeepSeek-V3 with DIRECT *paradigm*													
Zero-Shot	0.35	0.40	0.35	0.37	0.07	0.09	0.05	0.07	0.09	0.09	0.06	0.08	0.17
Few-Shot	-5.00%	-10.00%	+65.00%	+16.67%	+93.50%	+91.00%	+15.50%	+66.67%	+0.00%	+91.50%	+20.00%	+37.17%	+40.17%
RAG	-15.00%	-10.00%	-15.00%	-13.33%	+1.00%	+3.00%	+6.50%	+3.50%	+2.50%	+3.00%	+6.00%	+3.83%	-2.00%
Qwen3 with DIRECT *paradigm*													
Zero-Shot	0.10	0.25	0.30	0.22	0.09	0.09	0.10	0.09	0.06	0.06	0.08	0.07	0.12
Few-Shot	+15.00%	+75.00%	-10.00%	+26.67%	+91.50%	+91.00%	+90.00%	+90.83%	+3.00%	+94.50%	+92.00%	+63.17%	+60.22%
RAG	+10.00%	-10.00%	-10.00%	-3.33%	-2.00%	-1.50%	-1.50%	-1.67%	-0.50%	-1.50%	-2.50%	-1.50%	-2.17%
DeepSeek-V3 with PREDICTION-BASED *paradigm*													
Zero-Shot	0.51	0.55	0.58	0.55	0.04	0.06	0.06	0.05	0.04	0.07	0.05	0.05	0.22
Few-Shot	+1.00%	+0.00%	-2.00%	-0.33%	-1.75%	-1.50%	-4.00%	-2.42%	-1.75%	-1.00%	-2.25%	-1.67%	-1.47%
RAG	+0.00%	+1.00%	-2.00%	-0.33%	-0.75%	-1.75%	-3.50%	-2.00%	+0.75%	-2.75%	-0.75%	-0.92%	-1.08%
Qwen3 with PREDICTION-BASED *paradigm*													
Zero-Shot	0.51	0.53	1.00	0.68	0.04	0.05	0.03	0.04	0.10	0.04	0.07	0.07	0.26
Few-Shot	+0.00%	+18.00%	-24.00%	-2.00%	-0.25%	-2.25%	+18.75%	+5.42%	-1.75%	+0.75%	+2.25%	+0.42%	+1.28%
RAG	+0.00%	+28.00%	-34.00%	-2.00%	-1.75%	-2.00%	+0.50%	-1.08%	-5.25%	-0.50%	+45.25%	+13.17%	+3.36%