UniTS: A Unified Multi-Task Time Series Model

Shanghua Gao[1], Teddy Koker[2], Owen Queen[1], Thomas Hartvigsen[3], Theodoros Tsiligkaridis[2], Marinka Zitnik[1]

[1]Harvard University, [2]MIT Lincoln Laboratory, [3]University of Virginia

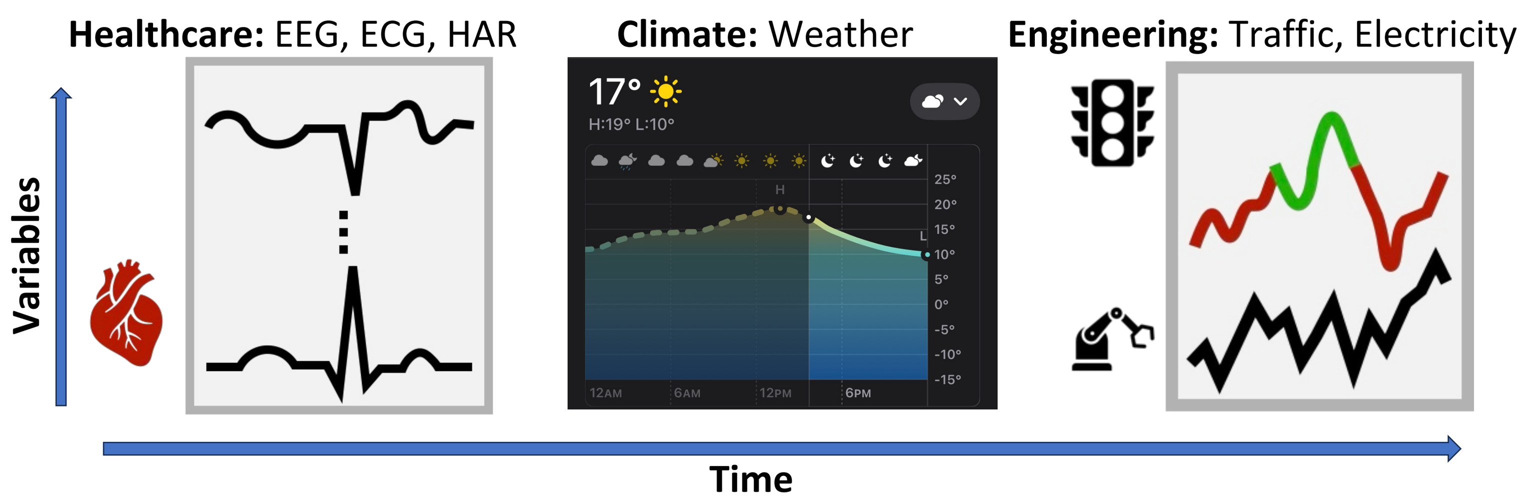
shanghua_gao@hms.harvard.edu ttsili@ll.mit.edu marinka@hms.harvard.edu

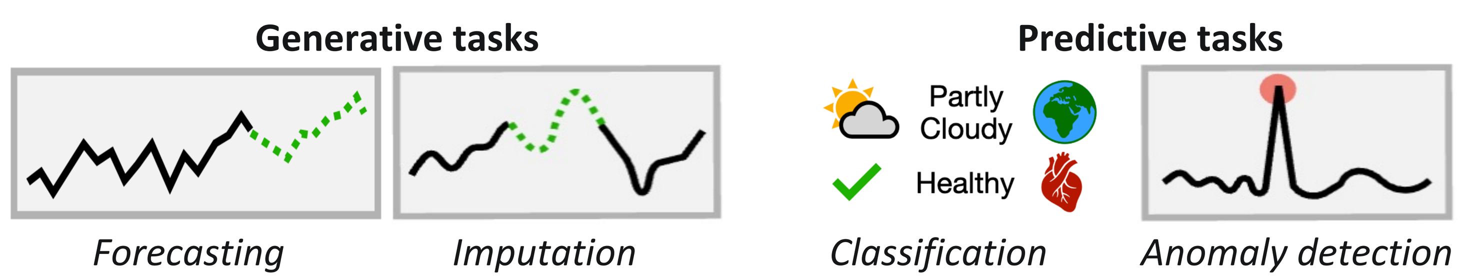
Training code

Model weights

- **Multi-task time series model**
- **Unified predictive & generative tasks**
- **Unified tokenization & model design**

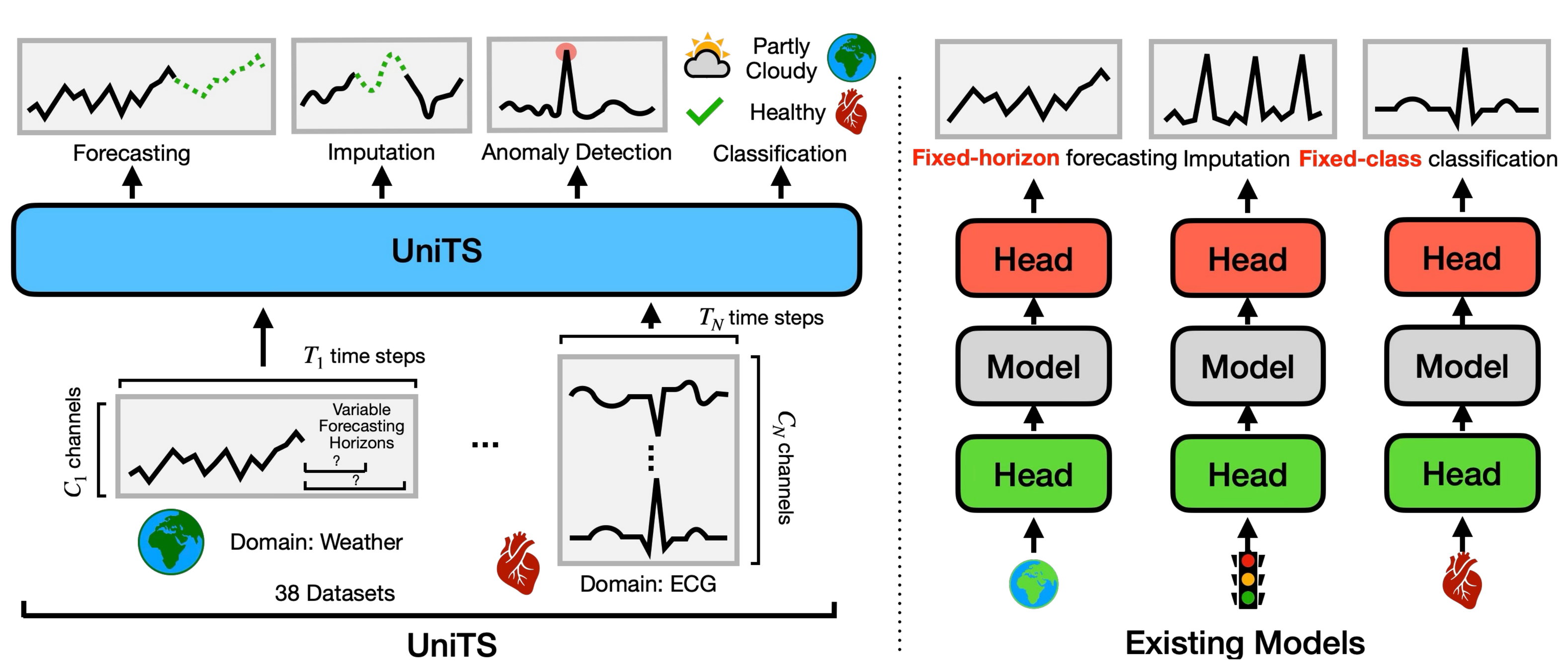
Diverse time series domains and tasks





- There is a growing need for versatile time series models that can accommodate data from diverse domains, such as healthcare, weather, and engineering, while supporting a wide range of tasks, such as forecasting, classification, imputation, and anomaly detection.
- Developing multi-task time series models that unify predictive and generative tasks within a single framework remains an open challenge.

UniTS unifies predictive and generative tasks across domains in a single model

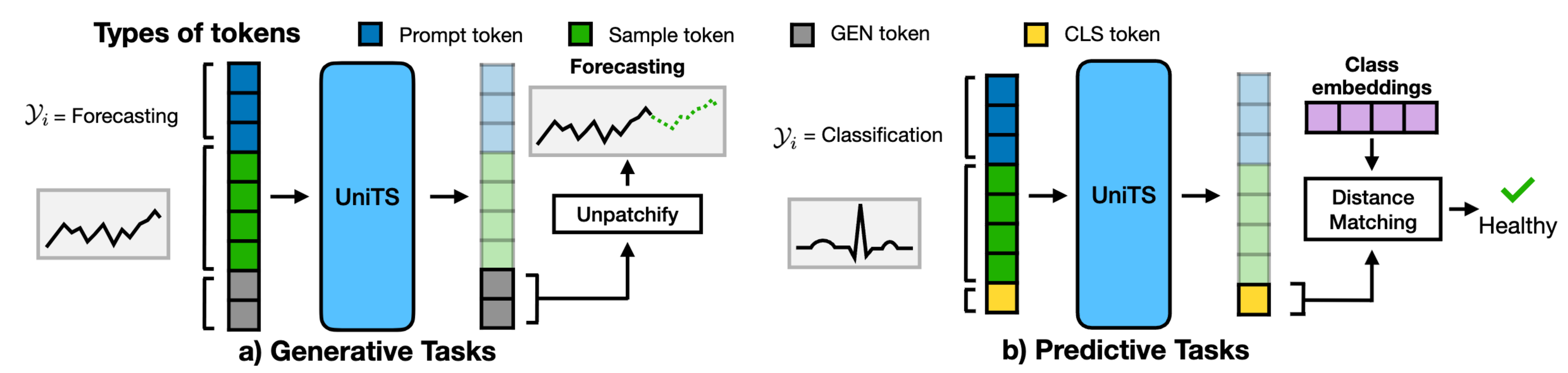


✓ Unified input for all data.
✓ Unified output for all tasks.
✓ One shared model.

✗ Data-specific head.
✗ Task-specific head.
✗ Separate models.

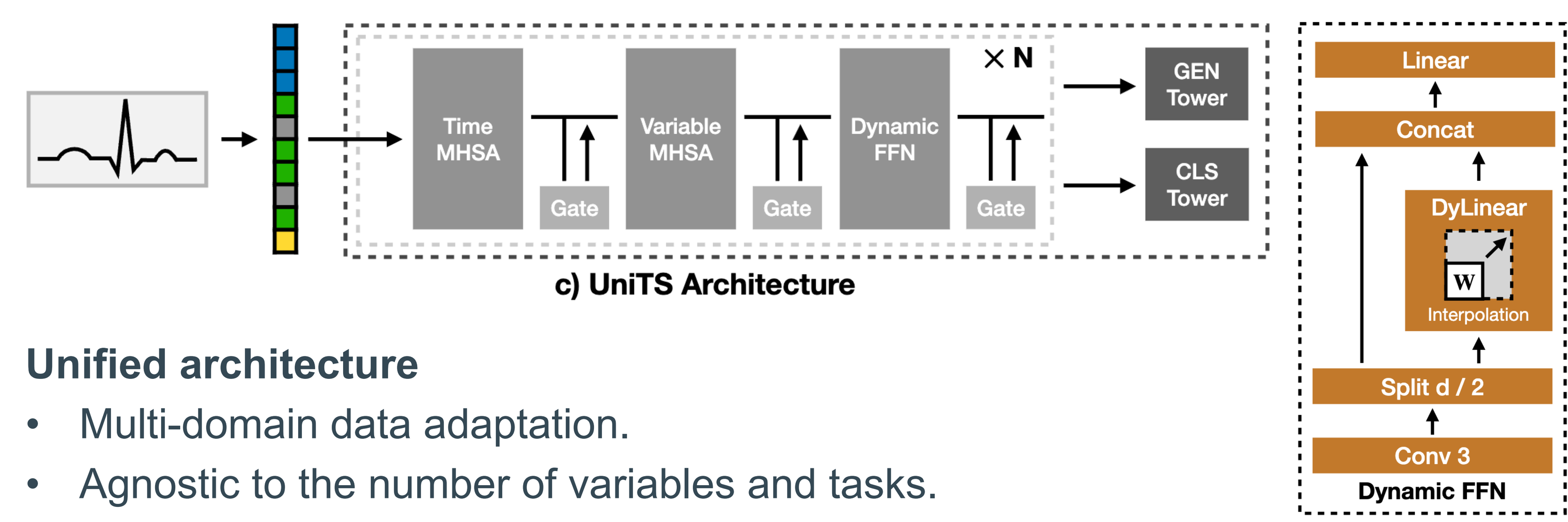
- Prevailing time series models are task-specific, with unique architectures designed and trained for every new task and dataset.
- UniTS is a multi-task time series model that uses task tokenization to unify predictive and generative tasks into a single framework.

UniTS: Unified tokenization, architecture, and self-supervised pre-training



Unified data and task tokenization

- Data tokenization: Diverse time series data → Unified *Sequence Tokens*.
- Task tokenization: Diverse task specifications → Unified *Task Tokens*.
- *Prompt Tokens*: Context tokens as prompts for datasets and tasks.



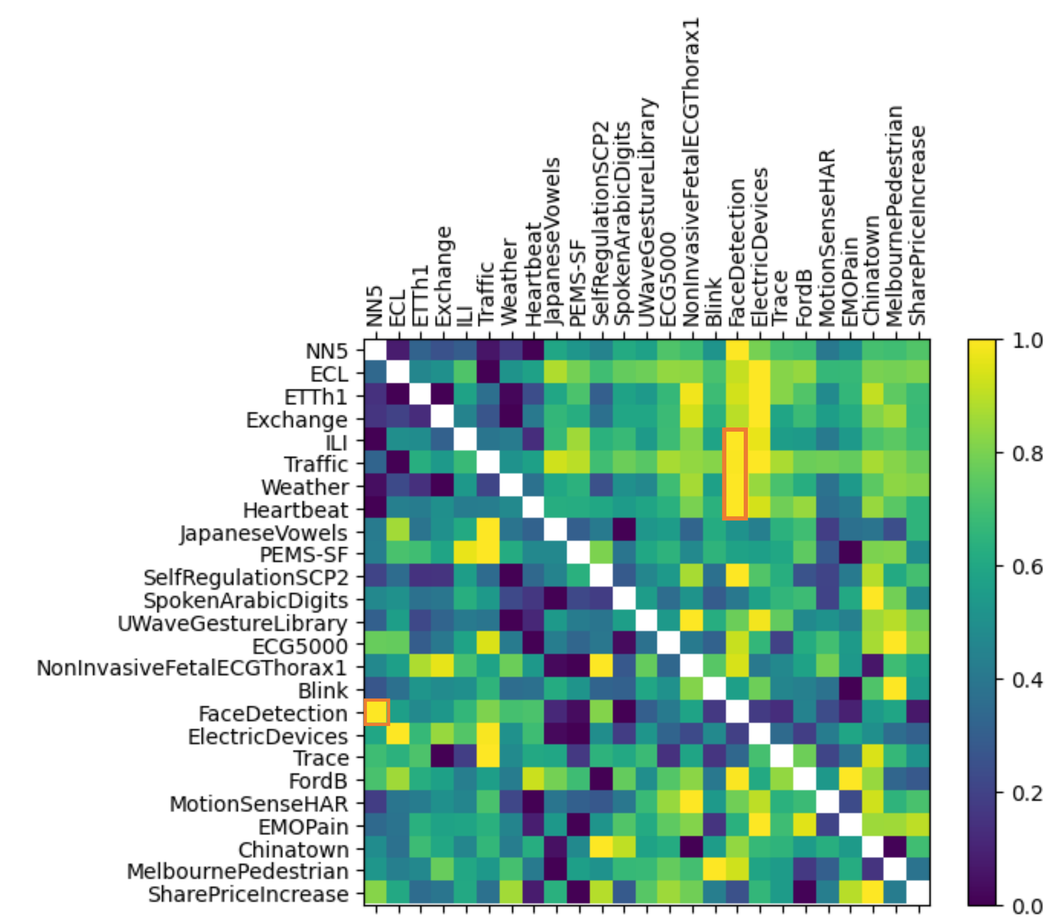
Unified architecture

- Multi-domain data adaptation.
- Agnostic to the number of variables and tasks.

Unified self-supervised pre-training (Dual mask reconstruction modeling)

- Prompt token + GEN token → UniTS → Sequence token.
- Prompt token + CLS token + GEN token → UniTS → Sequence token.

Exploring contextual relationships among datasets

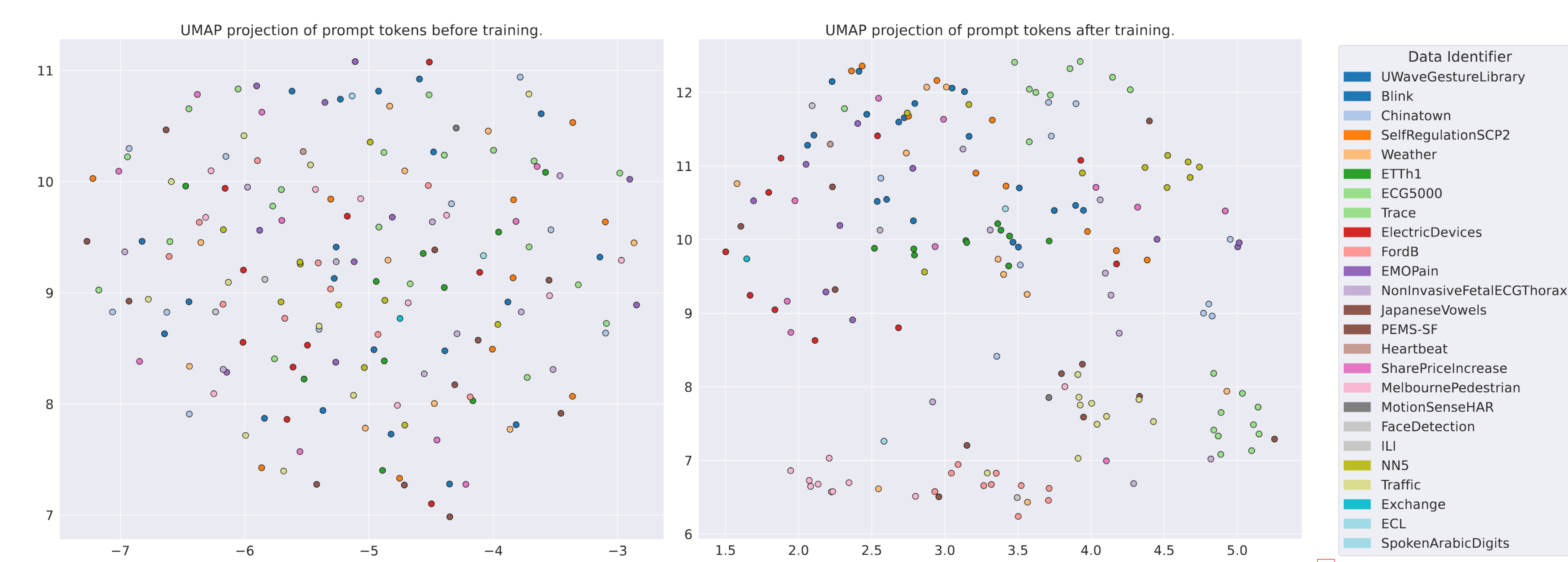


Prompt tokens across datasets

Cross-domain datasets can share similar prompt tokens, suggesting that UniTS representations generalize across time series domains.

UMAP projection of prompt tokens

Clusters within each dataset highlight contextual dataset-dependent features, while clusters across datasets capture task features shared across datasets.



UniTS excels in multi-task learning, few-shot learning, prompt learning, direct new-length forecasting, and single-task learning

Tested on over 38 datasets, UniTS demonstrates better performance than 12 forecasting models, 20 classification models, 18 anomaly detection models, and 16 imputation models, including adapted text-based LLMs. (Refer to the paper for detailed results.)

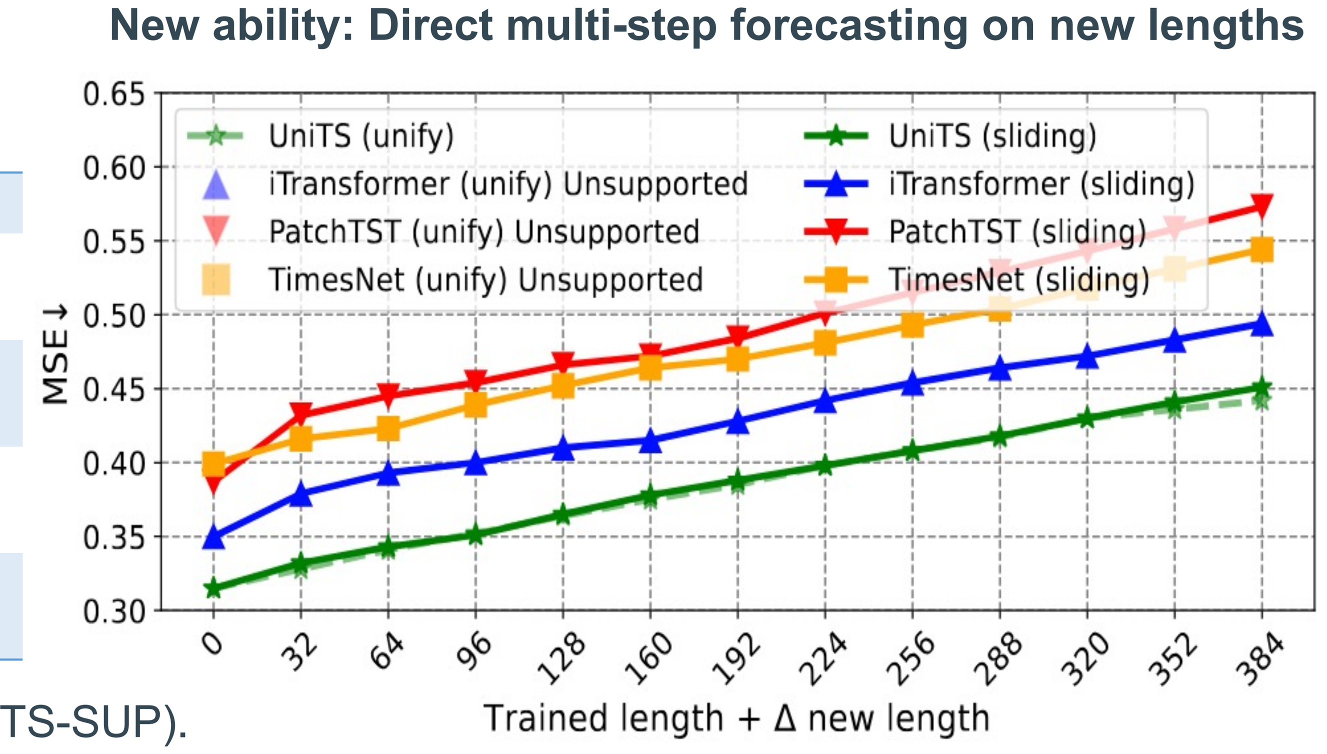
UniTS achieves the best performance in **multi-task learning** across 20 forecasting tasks and 18 classification tasks.

	UNITS-SUP	UNITS-PMT	ITRANS.	TIMESNET	PATCHTST
shared	✓	✗	✗	✗	✗
Average results on 20 forecasting tasks					
MSE/MSE	**.439**/.381	.453/**.376**	.466/.394	.525/.412	.488/.401
Best	**16/20**		3/20	0/20	1/20
Average results on 18 classification tasks					
Acc.	**81.6**	81.2	80.3	80.9	78.1
Best	**10/18**		0/18	4/18	3/18

Multi-task few-shot learning on 6 classification tasks, 9 forecasting tasks, 6 datasets imputation tasks, and 5 anomaly detection tasks.

	UNITS-SUP	UNITS-PMT	ITRANS.
Forecasting (MSE)	**0.481**	0.494	0.510
Classification (Acc)	**65.2**	63.6	59.9
Imputation (MSE)	**0.163**	0.165	0.186
Anomaly det. (F1)	**86.3**	82.3	83.1

New ability: Direct multi-step forecasting on new lengths



Prompt based on SSL pre-trained UniTS (UniTS-PMT) achieves comparable performance to supervised UniTS (UniTS-SUP).

UNITS: A Unified Multi-Task Time Series Model

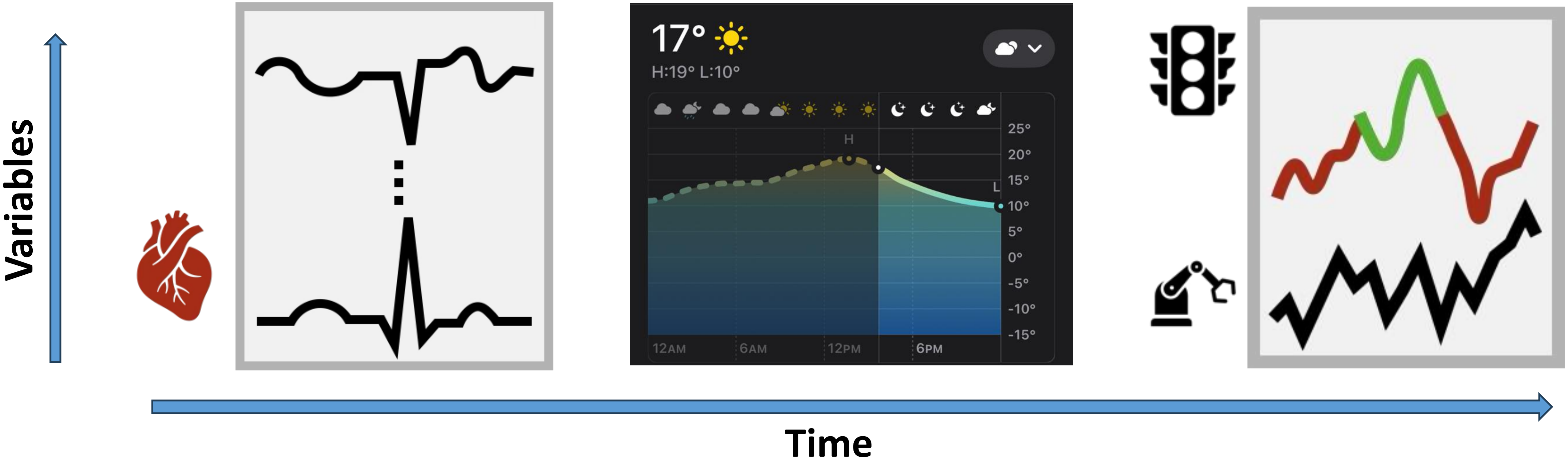
Shanghua Gao[1], Teddy Koker[2], Owen Queen[1],
Thomas Hartvigsen[3], Theodoros Tsiligkaridis[2], Marinka Zitnik[1]

Harvard University[1], MIT Lincoln Laboratory[2], University of Virginia[3]

Time series data comes from diverse domain.

- Time series data varies in domains, duration, and the number of variables.

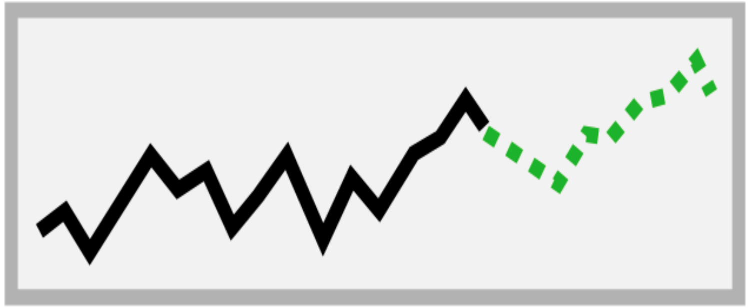
Healthcare: EEG, ECG, HAR **Climate:** Weather **Engineering:** Traffic, Electricity



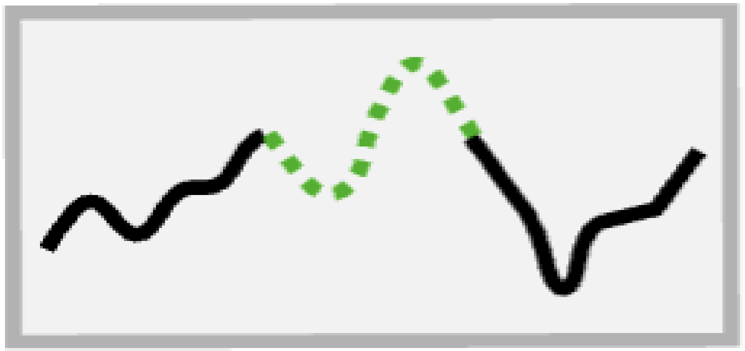
Generative and predictive tasks in time series.

- Time series data encompasses a variety of applications, including both generative and predictive tasks.

Generative tasks

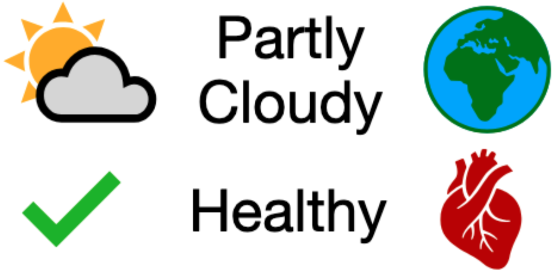


Forecasting

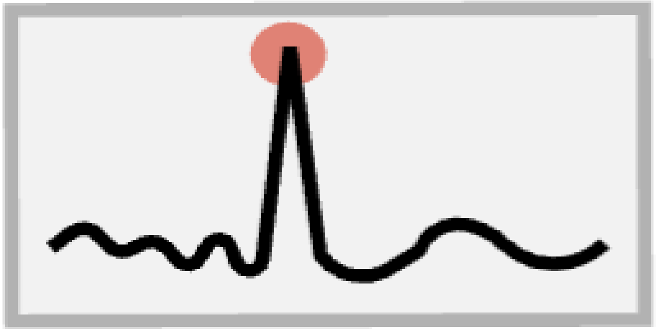


Imputation

Predictive tasks



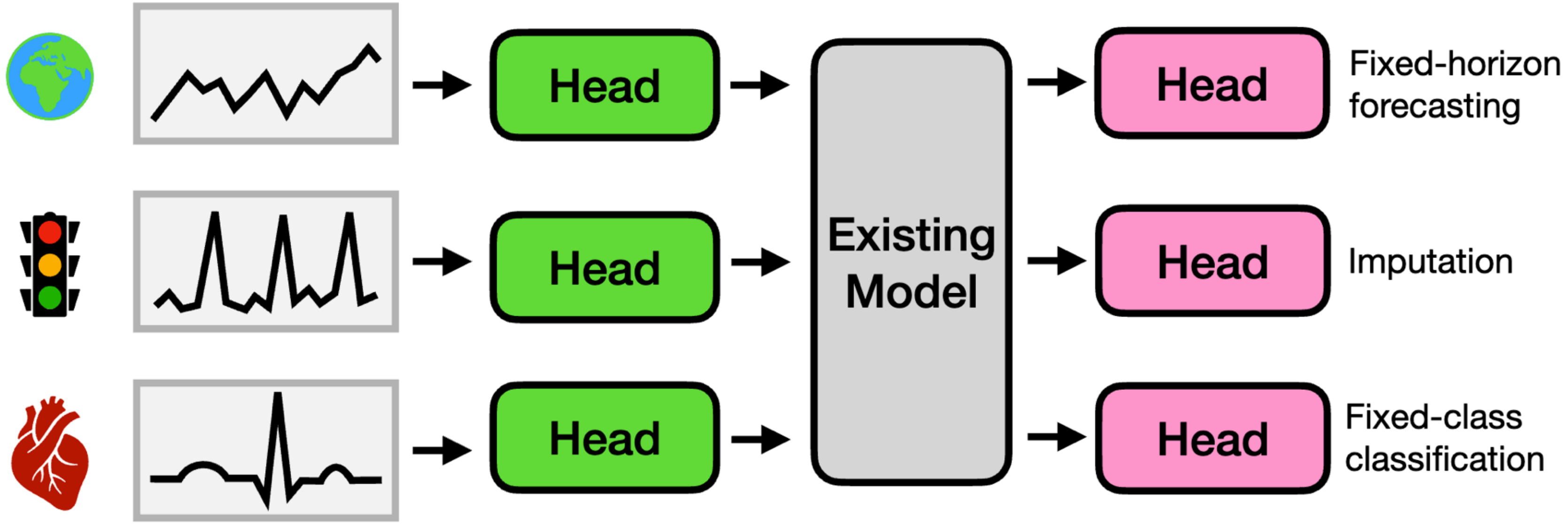
Classification



Anomaly detection

Existing task-specific time series models.

- Current time series models are designed for a single data domain and specific task, limiting adaptability to new applications.



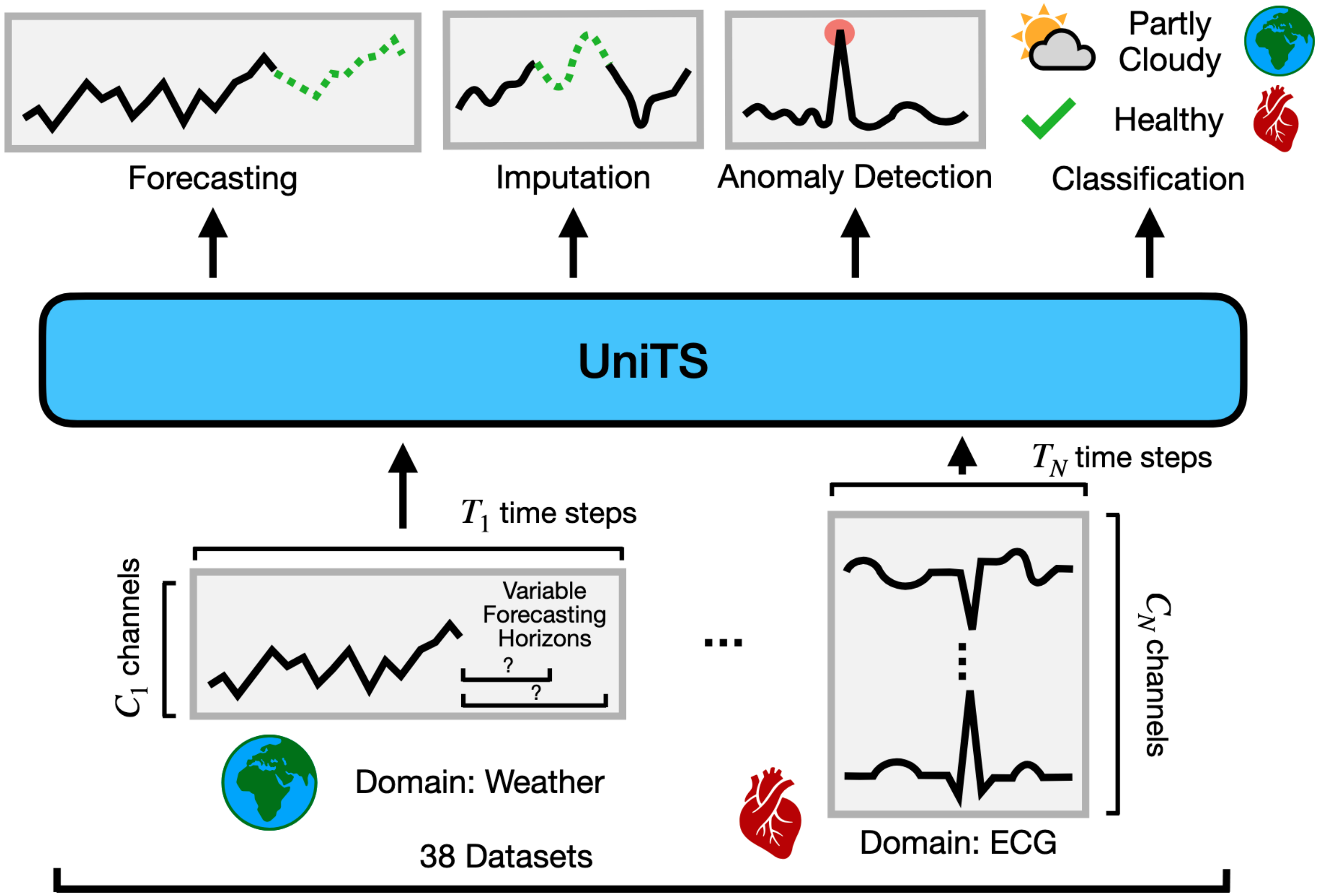
✗ Data-specific head.

✗ Task-specific head.

✗ Separate models.

UniTS: A unified multi-task time series model.

- UniTS is a unified multi-task time series model for predictive and generative tasks.



✓ Unified input for all data. ✓ Unified output for all tasks. ✓ One shared model.

How can UniTS achieve cross-task/data unification?

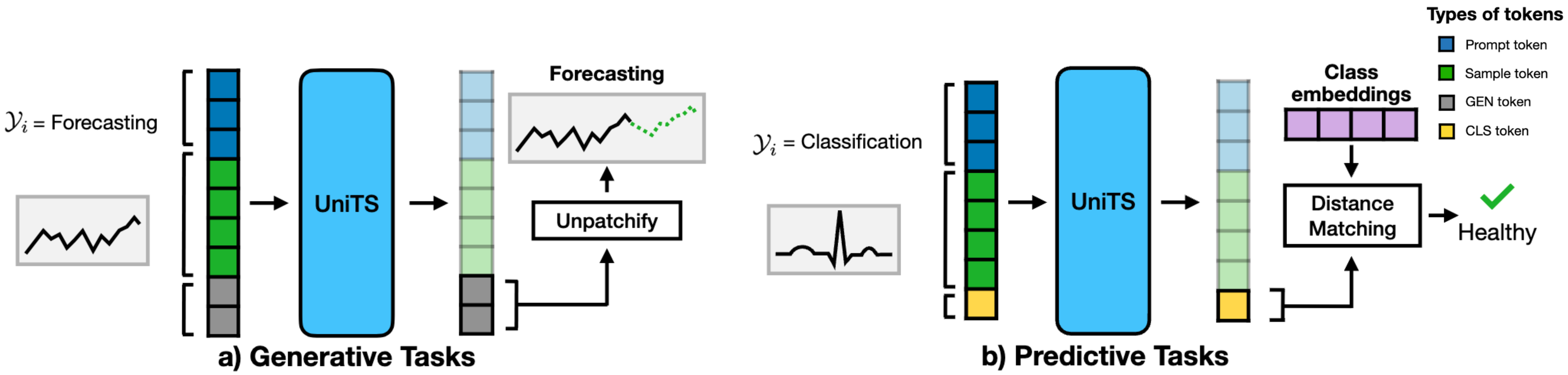
Unified data and task tokenization

Unified network architecture

Unified self-supervised pre-training

Prompting UNITS with unified data/task tokenization.

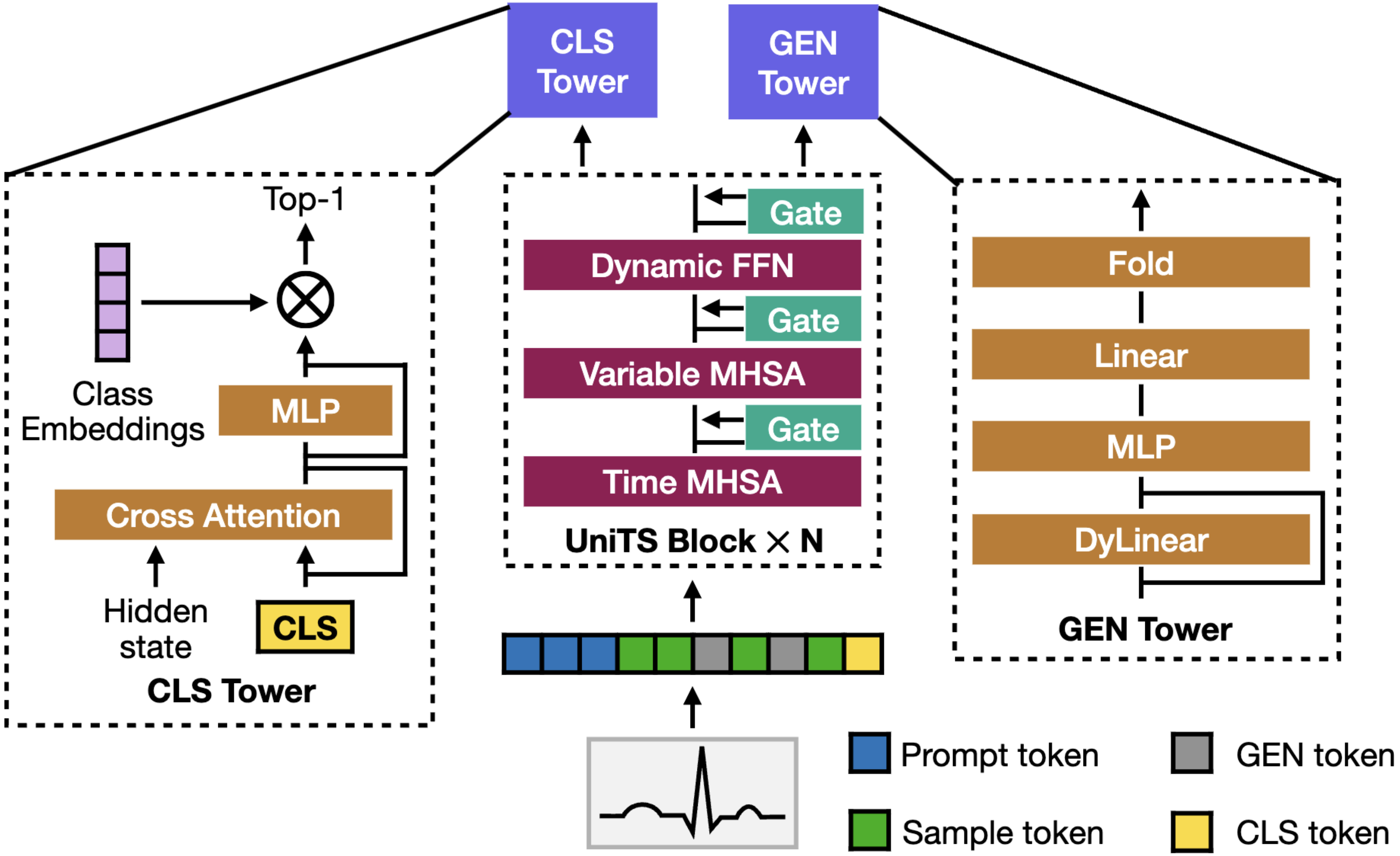
- Data tokenization: Diverse time series data-> Unified ***Sequence Tokens***.
- Task tokenization: Various task specifications-> Unified ***Task Tokens***.
- ***Prompt Tokens***: Context tokens as prompts for datasets and tasks.



UniTS use unified tokens to unify different task types and data.

Unified network architecture.

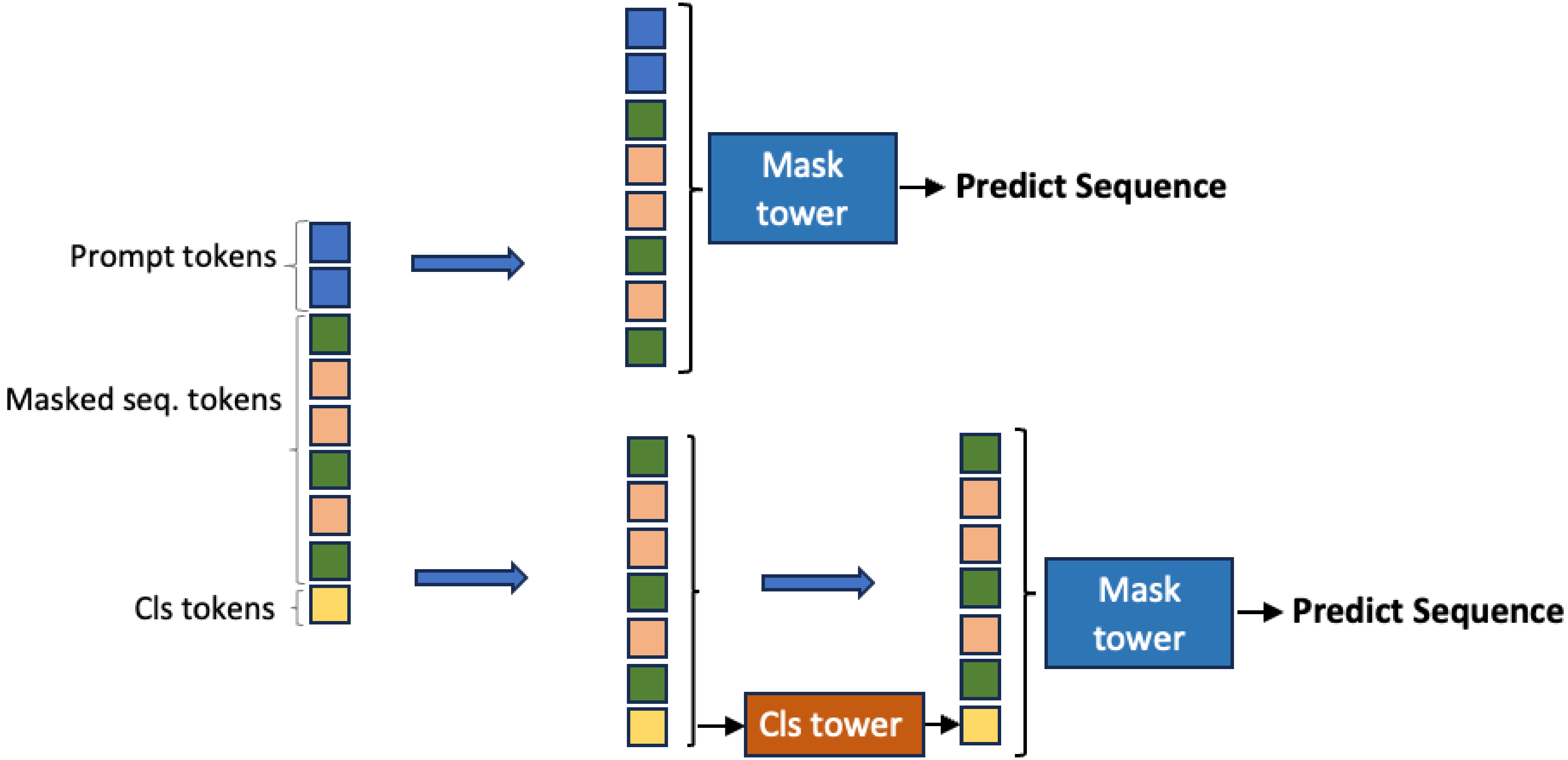
GEN/CLS Tower: transform task tokens into generation and prediction results.



UniTS blocks: handle multi-domain data with varying dynamics and the number of variables.

Unified mask reconstruction SSL pre-training.

- Learning both generative and recognition representation.



$$L_u = |H_m(\mathbf{z}_p, \mathbf{z}_s) - x|^2 + |H_m(\hat{\mathbf{z}}_c, \mathbf{z}_s) - x|^2.$$

Left side: **Prompt token** + Masked seq. tokens -> masked token prediction

Right side: **CLS token** + Masked seq. tokens -> masked token prediction

UniTS achieves multi-task learning with one model.

UniTS (supervised/Prompt)

Forecasting

Best count 16/20 (MSE). 14/20 (MAE).
Best Average score

Classification

Best count 10/18 (Acc).
Best Average score

Forecasting Data-Task	UniTS Sup. MSE↓	UniTS Sup. MAE↓	Prompt UniTS$_{SSL}$ MSE↓	Prompt UniTS$_{SSL}$ MAE↓	iTransformer MSE↓	iTransformer MAE↓	TimesNet MSE↓	TimesNet MAE↓	PatchTST MSE↓	PatchTST MAE↓	Pyraformer MSE↓	Pyraformer MAE↓	Autoformer MSE↓	Autoformer MAE↓
NN5$_{p112}$	**0.611**	0.549	0.622	0.546	0.623	0.554	0.629	**0.541**	0.634	0.568	1.069	0.791	1.232	0.903
ECL$_{p96}$	0.167	0.271	**0.157**	**0.258**	0.204	0.288	0.184	0.289	0.212	0.299	0.39	0.456	0.262	0.364
ECL$_{p192}$	0.181	0.282	**0.173**	**0.272**	0.208	0.294	0.204	0.307	0.213	0.303	0.403	0.463	0.34	0.421
ECL$_{p336}$	0.197	0.296	**0.185**	**0.284**	0.224	0.31	0.217	0.32	0.228	0.317	0.417	0.466	0.624	0.608
ECL$_{p720}$	0.231	0.324	**0.219**	**0.314**	0.265	0.341	0.284	0.363	0.27	0.348	0.439	0.483	0.758	0.687
ETTh1$_{p96}$	0.386	0.409	0.39	0.411	**0.382**	**0.399**	0.478	0.448	0.389	0.4	0.867	0.702	0.505	0.479
ETTh1$_{p192}$	**0.429**	0.436	0.432	0.438	0.431	**0.426**	0.561	0.504	0.44	0.43	0.931	0.751	0.823	0.601
ETTh1$_{p336}$	**0.466**	0.457	0.48	0.46	0.476	**0.449**	0.612	0.537	0.482	0.453	0.96	0.763	0.731	0.58
ETTh1$_{p720}$	0.494	0.483	0.542	0.508	0.495	0.487	0.601	0.541	**0.486**	**0.479**	0.994	0.782	0.699	0.59
Exchange$_{p192}$	0.243	0.351	0.2	0.32	**0.175**	**0.297**	0.259	0.37	0.178	0.301	1.221	0.916	0.306	0.409
Exchange$_{p336}$	0.431	0.476	0.346	0.425	**0.322**	**0.409**	0.478	0.501	0.328	0.415	1.215	0.917	0.462	0.508
ILI$_{p60}$	**1.986**	**0.878**	2.372	0.945	1.989	0.905	2.367	0.966	2.307	0.97	4.791	1.459	3.812	1.33
Traffic$_{p96}$	0.47	0.318	**0.465**	**0.298**	0.606	0.389	0.611	0.336	0.643	0.405	0.845	0.465	0.744	0.452
Traffic$_{p192}$	0.485	0.323	**0.484**	**0.306**	0.592	0.382	0.643	0.352	0.603	0.387	0.883	0.477	1.086	0.638
Traffic$_{p336}$	0.497	0.325	**0.494**	**0.312**	0.6	0.384	0.662	0.363	0.612	0.389	0.907	0.488	1.185	0.692
Traffic$_{p720}$	**0.53**	0.34	0.534	**0.335**	0.633	0.401	0.678	0.365	0.652	0.406	0.974	0.522	1.344	0.761
Weather$_{p96}$	0.158	0.208	**0.157**	**0.206**	0.193	0.232	0.169	0.22	0.194	0.233	0.239	0.323	0.251	0.315
Weather$_{p192}$	**0.207**	0.253	0.208	**0.251**	0.238	0.269	0.223	0.264	0.238	0.268	0.323	0.399	0.289	0.335
Weather$_{p336}$	**0.264**	0.294	**0.264**	**0.291**	0.291	0.306	0.279	0.302	0.29	0.304	0.333	0.386	0.329	0.356
Weather$_{p720}$	**0.341**	**0.344**	0.344	**0.344**	0.365	0.354	0.359	0.355	0.363	0.35	0.424	0.447	0.39	0.387
Best Count	8/20	2/20	**9/20**	**12/20**	3/20	5/20	0/20	1/20	1/20	1/20	0/20	0/20	0/20	0/20
Average Score	**0.439**	0.381	0.453	**0.376**	0.466	0.394	0.525	0.412	0.488	0.401	0.931	0.623	0.809	0.571
Fully Shared Model	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×

Classification Data-Task heads	UniTS Sup. Accuracy↑	Prompt UniTS$_{SSL}$ Accuracy↑	iTransformer Accuracy↑	TimesNet Accuracy↑	PatchTST Accuracy↑	Pyraformer Accuracy↑	Autoformer Accuracy↑
2 Category$_{7\text{datasets}}$	**0.731**	**0.731**	0.724	0.73	0.708	0.615	0.662
3 Category$_{1\text{datasets}}$	0.797	**0.814**	0.794	0.78	0.792	**0.814**	0.699
4 Category$_{1\text{datasets}}$	0.96	**0.99**	0.79	0.91	0.77	0.74	0.6
5 Category$_{1\text{datasets}}$	0.928	0.924	0.933	0.926	**0.943**	0.914	0.919
6 Category$_{1\text{datasets}}$	0.951	**0.958**	0.936	0.906	0.758	0.887	0.302
7 Category$_{2\text{datasets}}$	0.727	0.726	0.702	0.635	0.716	**0.743**	0.677
8 Category$_{1\text{datasets}}$	0.822	**0.853**	0.822	0.844	0.819	0.722	0.422
9 Category$_{1\text{datasets}}$	0.922	0.903	0.959	**0.976**	0.941	0.854	0.941
10 Category$_{2\text{datasets}}$	0.922	0.897	0.935	**0.972**	0.889	0.722	0.861
52 Category$_{1\text{datasets}}$	**0.896**	0.808	0.882	0.889	0.865	0.214	0.217
Best Count	3/18	**7/18**	0/18	4/18	3/18	4/18	0/18
Average Score	**0.816**	0.812	0.803	0.809	0.781	0.688	0.656
Fully Shared Model	✓	✓	×	×	×	×	×

UniTS with **fully shared model** outperforms baseline models that utilize **separate input/output heads**.

SSL UniTS + Prompting achieve strong performance.

UniTS (supervised/Prompt)

Forecasting

Best count **17/20 (MSE). 14/20 (MAE).**
Best Average score

Classification

Best count **10/18 (Acc).**
Best Average score

Forecasting Data-Task	UniTS Sup. MSE↓	UniTS Sup. MAE↓	Prompt UniTS$_{SSL}$ MSE↓	Prompt UniTS$_{SSL}$ MAE↓	iTransformer MSE↓	iTransformer MAE↓	TimesNet MSE↓	TimesNet MAE↓	PatchTST MSE↓	PatchTST MAE↓	Pyraformer MSE↓	Pyraformer MAE↓	Autoformer MSE↓	Autoformer MAE↓
NN5$_{p112}$	**0.611**	0.549	0.622	0.546	0.623	0.554	0.629	**0.541**	0.634	0.568	1.069	0.791	1.232	0.903
ECL$_{p96}$	0.167	0.271	**0.157**	**0.258**	0.204	0.288	0.184	0.289	0.212	0.299	0.39	0.456	0.262	0.364
ECL$_{p192}$	0.181	0.282	**0.173**	**0.272**	0.208	0.294	0.204	0.307	0.213	0.303	0.403	0.463	0.34	0.421
ECL$_{p336}$	0.197	0.296	**0.185**	**0.284**	0.224	0.31	0.217	0.32	0.228	0.317	0.417	0.466	0.624	0.608
ECL$_{p720}$	0.231	0.324	**0.219**	**0.314**	0.265	0.341	0.284	0.363	0.27	0.348	0.439	0.483	0.758	0.687
ETTh1$_{p96}$	0.386	0.409	0.39	0.411	**0.382**	**0.399**	0.478	0.448	0.389	0.4	0.867	0.702	0.505	0.479
ETTh1$_{p192}$	**0.429**	0.436	0.432	0.438	0.431	**0.426**	0.561	0.504	0.44	0.43	0.931	0.751	0.823	0.601
ETTh1$_{p336}$	**0.466**	0.457	0.48	0.46	0.476	**0.449**	0.612	0.537	0.482	0.453	0.96	0.763	0.731	0.58
ETTh1$_{p720}$	0.494	0.483	0.542	0.508	0.495	0.487	0.601	0.541	**0.486**	**0.479**	0.994	0.782	0.699	0.59
Exchange$_{p192}$	0.243	0.351	0.2	0.32	**0.175**	**0.297**	0.259	0.37	0.178	0.301	1.221	0.916	0.306	0.409
Exchange$_{p336}$	0.431	0.476	0.346	0.425	**0.322**	**0.409**	0.478	0.501	0.328	0.415	1.215	0.917	0.462	0.508
ILI$_{p60}$	**1.986**	**0.878**	2.372	0.945	1.989	0.905	2.367	0.966	2.307	0.97	4.791	1.459	3.812	1.33
Traffic$_{p96}$	0.47	0.318	**0.465**	**0.298**	0.606	0.389	0.611	0.336	0.643	0.405	0.845	0.465	0.744	0.452
Traffic$_{p192}$	0.485	0.323	**0.484**	**0.306**	0.592	0.382	0.643	0.352	0.603	0.387	0.883	0.477	1.086	0.638
Traffic$_{p336}$	0.497	0.325	**0.494**	**0.312**	0.6	0.384	0.662	0.363	0.612	0.389	0.907	0.488	1.185	0.692
Traffic$_{p720}$	**0.53**	0.34	0.534	**0.335**	0.633	0.401	0.678	0.365	0.652	0.406	0.974	0.522	1.344	0.761
Weather$_{p96}$	0.158	0.208	**0.157**	**0.206**	0.193	0.232	0.169	0.22	0.194	0.233	0.239	0.323	0.251	0.315
Weather$_{p192}$	**0.207**	0.253	0.208	**0.251**	0.238	0.269	0.223	0.264	0.238	0.268	0.323	0.399	0.289	0.335
Weather$_{p336}$	**0.264**	0.294	**0.264**	**0.291**	0.291	0.306	0.279	0.302	0.29	0.304	0.333	0.386	0.329	0.356
Weather$_{p720}$	**0.341**	**0.344**	0.344	**0.344**	0.365	0.354	0.359	0.355	0.363	0.35	0.424	0.447	0.39	0.387
Best Count	8/20	2/20	**9/20**	**12/20**	3/20	5/20	0/20	1/20	1/20	1/20	0/20	0/20	0/20	0/20
Average Score	**0.439**	0.381	0.453	**0.376**	0.466	0.394	0.525	0.412	0.488	0.401	0.931	0.623	0.809	0.571
Fully Shared Model	✓	✓	✓	✓	×	×	×	×	×	×	×	×	×	×

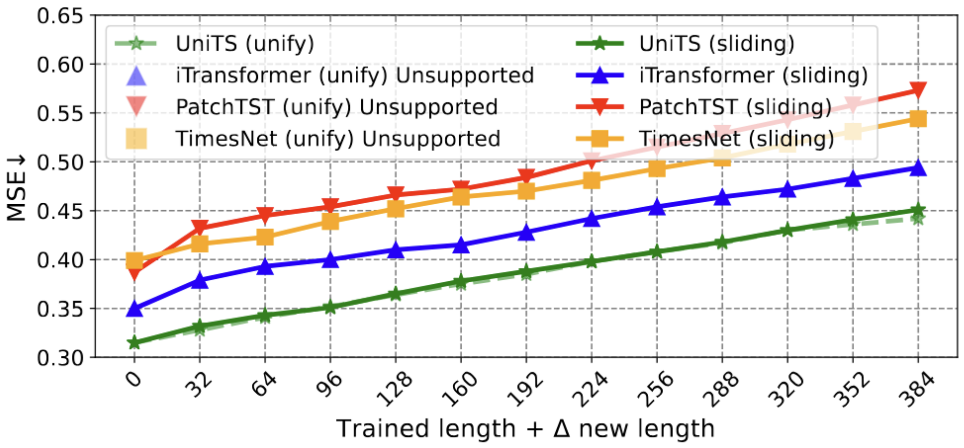
Classification Data-Task Heads	UniTS Sup. Accuracy↑	Prompt UniTS$_{SSL}$ Accuracy↑	iTransformer Accuracy↑	TimesNet Accuracy↑	PatchTST Accuracy↑	Pyraformer Accuracy↑	Autoformer Accuracy↑
2 Category$_{7\text{datasets}}$	**0.731**	**0.731**	0.724	0.73	0.708	0.615	0.662
3 Category$_{1\text{datasets}}$	0.797	**0.814**	0.794	0.78	0.792	**0.814**	0.699
4 Category$_{1\text{datasets}}$	0.96	**0.99**	0.79	0.91	0.77	0.74	0.6
5 Category$_{1\text{datasets}}$	0.928	0.924	0.933	0.926	**0.943**	0.914	0.919
6 Category$_{1\text{datasets}}$	0.951	**0.958**	0.936	0.906	0.758	0.887	0.302
7 Category$_{2\text{datasets}}$	0.727	0.726	0.702	0.635	0.716	**0.743**	0.677
8 Category$_{1\text{datasets}}$	0.822	**0.853**	0.822	0.844	0.819	0.722	0.422
9 Category$_{1\text{datasets}}$	0.922	0.903	0.959	**0.976**	0.941	0.854	0.941
10 Category$_{2\text{datasets}}$	0.922	0.897	0.935	**0.972**	0.889	0.722	0.861
52 Category$_{1\text{datasets}}$	**0.896**	0.808	0.882	0.889	0.865	0.214	0.217
Best Count	3/18	**7/18**	0/18	4/18	3/18	4/18	0/18
Average Score	**0.816**	0.812	0.803	0.809	0.781	0.688	0.656
Fully Shared Model	✓	✓	×	×	×	×	×

Prompting the SSL pretrained model has the comparable performance to supervised learning!

UniTS achieves few-shot learning on new tasks/data.

Model	Data Ratio	Acc↑	MSE↓	MAE↓	Shared
iTransformer (Finetune)	5%	56.4	0.598	0.487	×
UNITS (Prompt)	5%	55.7	**0.508**	**0.440**	✓
UNITS (Finetune)	5%	**57.4**	0.530	0.448	✓
iTransformer (Finetune)	15%	56.5	0.524	0.447	×
UNITS (Prompt)	15%	59.5	0.496	0.435	✓
UNITS (Finetune)	15%	**61.8**	**0.487**	**0.428**	✓
iTransformer (Finetune)	20%	59.9	0.510	0.438	×
UNITS (Prompt)	20%	63.6	0.494	0.435	✓
UNITS (Finetune)	20%	**65.2**	**0.481**	**0.425**	✓

Forecasting & Classification



Direct multi-step forecasting on new lengths

Impu. (MSE)	Ratio	ECL	ETTh1	ETTh2	ETTm1	ETTm2	Weather	Avg	Best	Shared
TimesNet-*FT*	25%	0.245	0.369	0.193	0.442	0.119	0.106	0.246	0/6	×
	50%	0.258	0.412	0.211	0.607	0.140	0.125	0.292	0/6	×
PatchTST-*FT*	25%	0.195	0.315	0.147	0.309	0.092	0.089	0.191	0/6	×
	50%	0.230	0.353	0.175	0.442	0.111	0.105	0.236	0/6	×
iTrans-*FT*	25%	0.174	0.301	0.185	0.254	0.113	0.087	0.186	0/6	×
	50%	0.203	0.332	0.205	0.372	0.136	0.106	0.226	0/6	×
UNITS-*PMT*	25%	**0.117**	0.281	**0.177**	0.247	0.095	0.075	0.165	2/6	✓
	50%	**0.135**	0.323	**0.246**	0.343	0.131	**0.093**	0.212	3/6	✓
UNITS-*FT*	25%	0.143	**0.277**	0.194	**0.204**	**0.088**	**0.074**	**0.163**	**4/6**	✓
	50%	0.161	**0.313**	0.252	**0.295**	**0.119**	0.096	**0.206**	**3/6**	✓

Imputation

Anomaly (F1↑)	MSL	PSM	SMAP	SMD	SWAT	Avg	Best	Shared
Anomaly Trans.	78.0	90.2	68.3	77.8	81.5	79.2	0/5	×
TimesNet-*FT*	33.9	91.0	68.5	84.0	**93.4**	74.2	1/5	×
iTransfomer-*FT*	80.4	96.5	67.2	82.4	89.0	83.1	0/5	×
PatchTST-*FT*	79.9	96.6	68.7	83.8	92.6	84.3	0/5	×
UNITS-*PMT*	75.4	95.5	65.8	82.3	92.5	82.3	0/5	✓
UNITS-*FT*	**81.2**	**97.3**	**76.0**	**84.7**	92.5	**86.3**	**4/5**	✓

Anomaly detection

Thank you!

- **Code:**
  - https://github.com/mims-harvard/UniTS