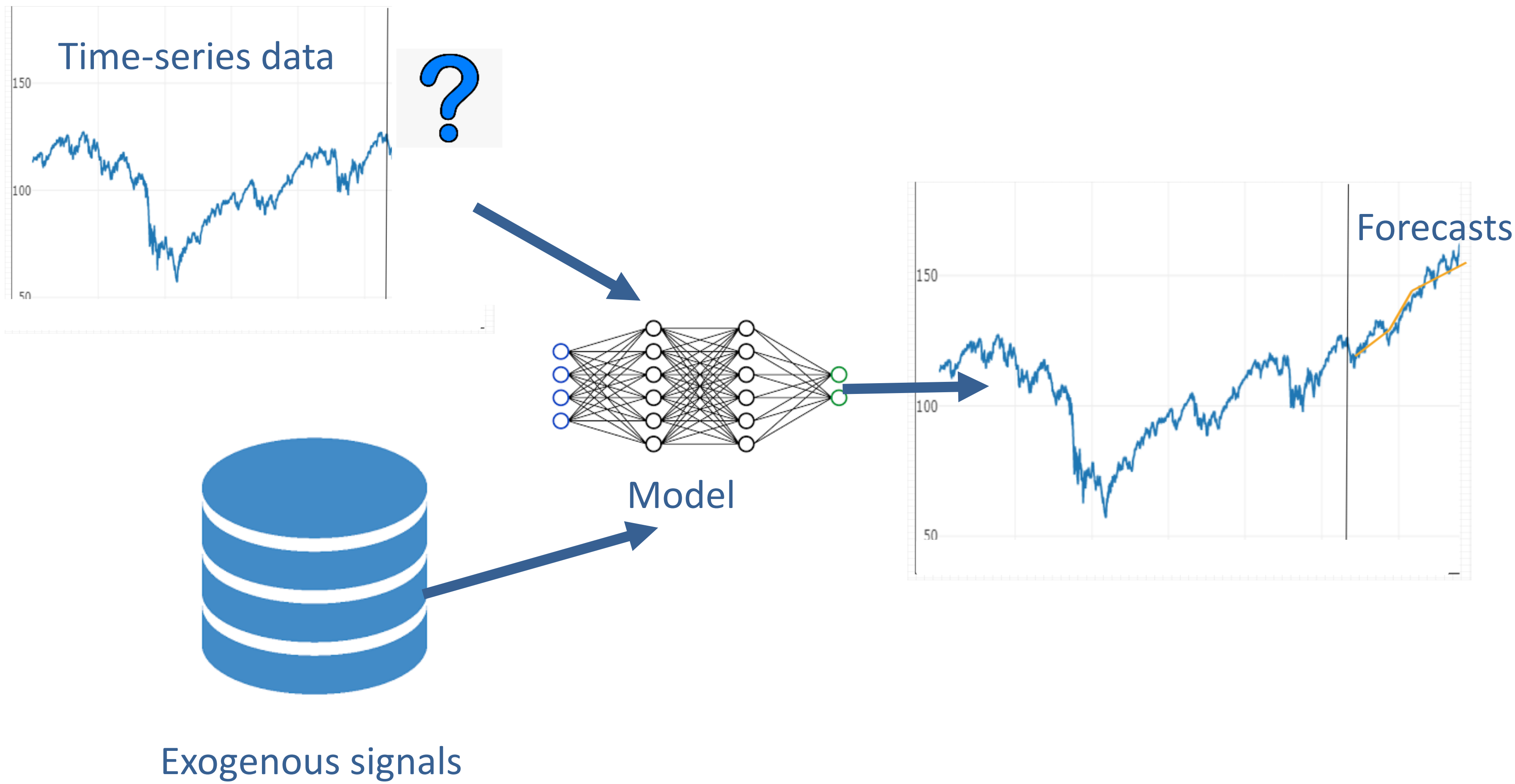
Large Pre-trained Time-series models for Cross-domain Time-series Analysis Tasks

Harsha Kamarthi, B Aditya Prakash

NeurIPS 2024

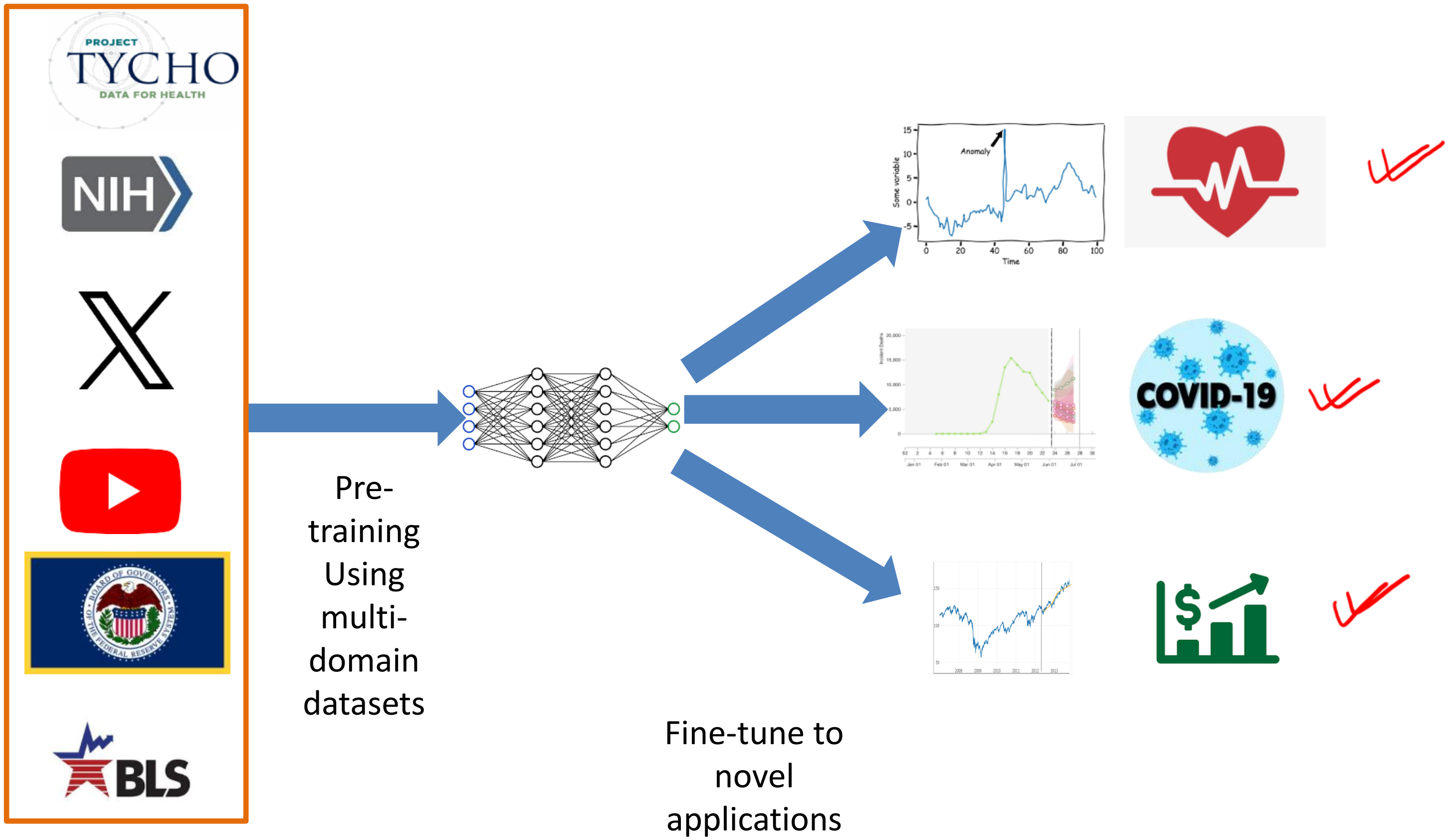
Traditional Forecasting Setup



Challenges with traditional setup

- Takes human and compute time to design the model and tune hyperparameters
- Hard to deploy in applications with low data availability
- Eg: Forecasting Pandemic during initial stages
- Generalize knowledge across multiple tasks (forecasting, classification, anomaly detection, etc.)

Time-series Foundational model

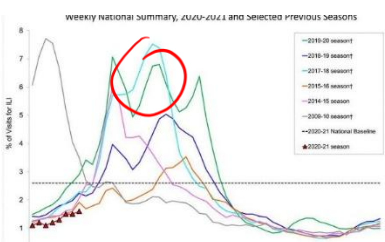


Advantages: Foundation Models

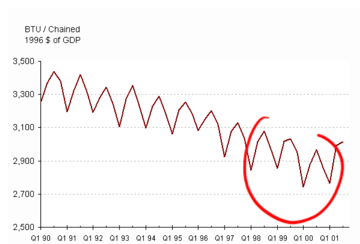
- Better generalized performance across domains
- Requires significantly less/no training time for adaptation to domain
- Much smaller data requirement for adaptation to SOTA performance

Challenge

- Adding time-series data to foundation models is non-trivial. Why?
- Complex and Heterogenous patterns across domains



Seasonal Epidemics



Power Consumption

Stock Prices

Pre-trained time-series models across multiple domains

- Time-series from multiple domains: Power, Economics, Epidemiology, etc.
- Need to effectively capture patterns across different domains

Tokenization for Time-series

- Single time-steps not meaningful

Language Tokens

Natural Language Processing

['Natural', 'Language', 'Processing']

Semanitically Meaningful

Time-series Tokens

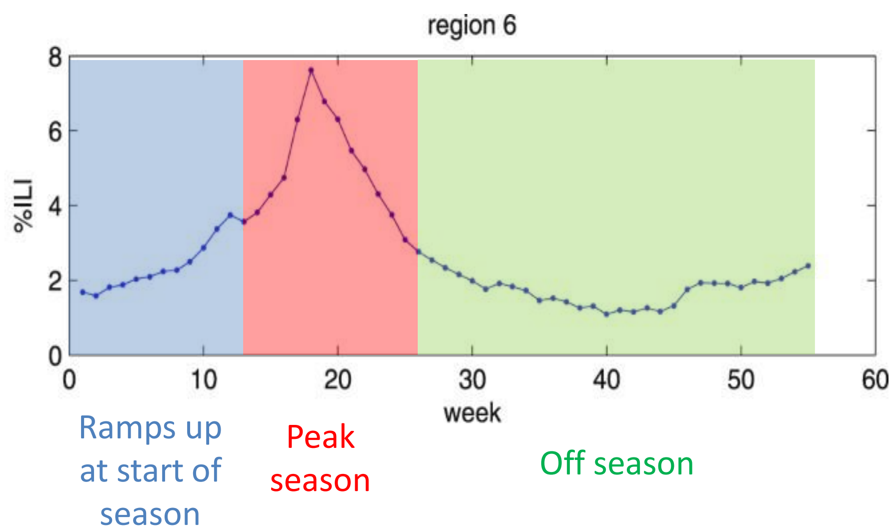
10, 33.4, 55.1

10 | 33.4 | 55.1

Not Meaningful

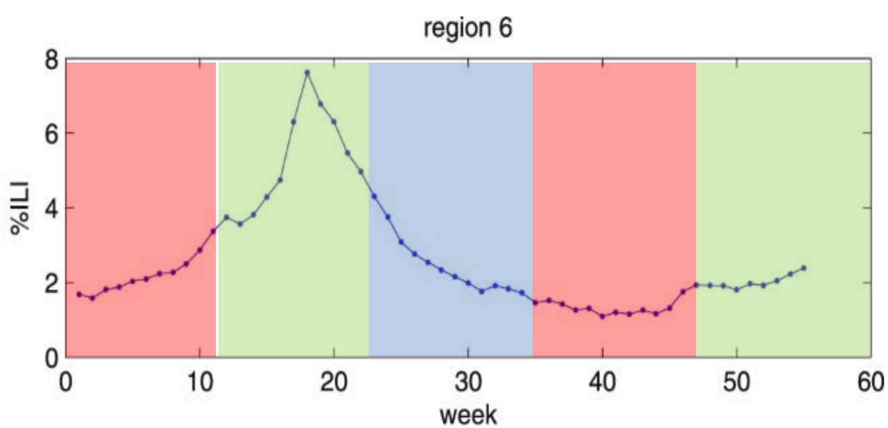
Segment Tokens

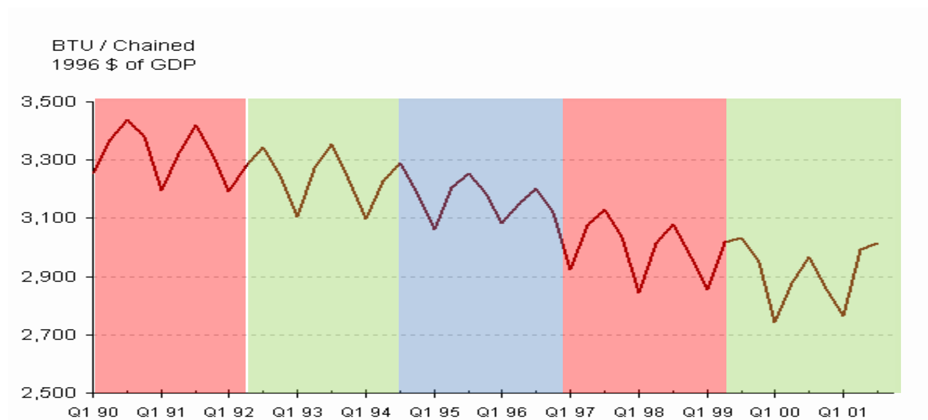
- Segments of time-series can provide semantic meaning:



Segmentation Strategy

- Option 1: Choose a uniform strategy
- But different domains have different dynamics and require different strategies
- Different time-periods in same time-series may require different strategies





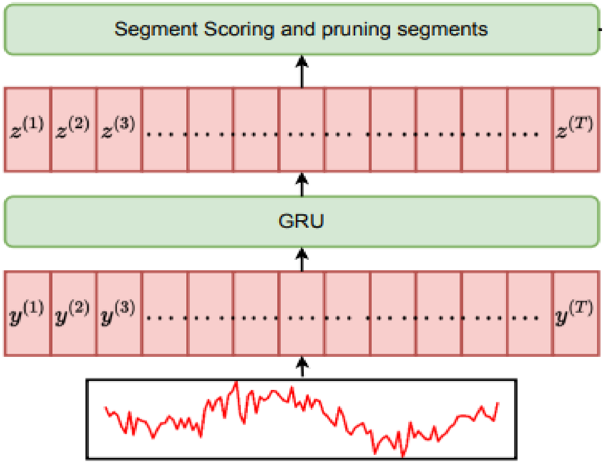
Soln: Adaptive Segmentation

- Model figures optimal segmentation strategy via scoring mechanism during pre-training
- Step 1: Scoring model to score the importance of each segment
- Step 2: Prune optimal set of segments based on the scores of all segments

Scoring Model

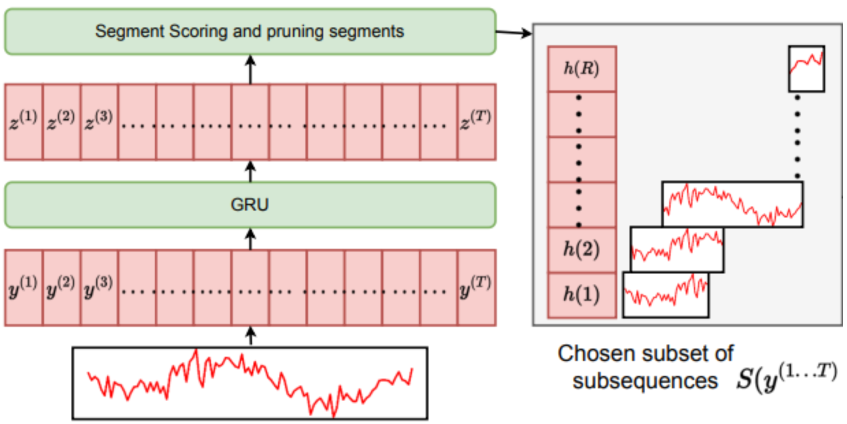
- Step 1: Scoring model to score the importance of each segment

$$s(i, j) = Atten(GRU(y^{(1\dots,i)}, GRU(y^{(1\dots,j)})$$



Pruning Segments (Contd.)

- Select segments with highest scores that cover the entire time-series



Learning Optimal Segmentation

- Segmentation scoring model trained on the SSL losses:
- Choose segmentation strategy that minimizes the SSL losses

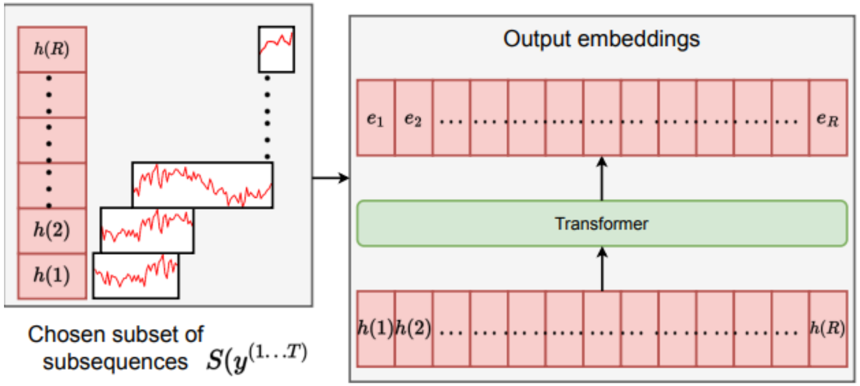
$$\mathcal{L}_g = \left(\sum_{(i,j) \in S(y^{(1 \ldots t)})} s(i,j) + \log(\mathcal{L}_{SSL}) \right)^2$$

Set of chosen segments

SSL Loss

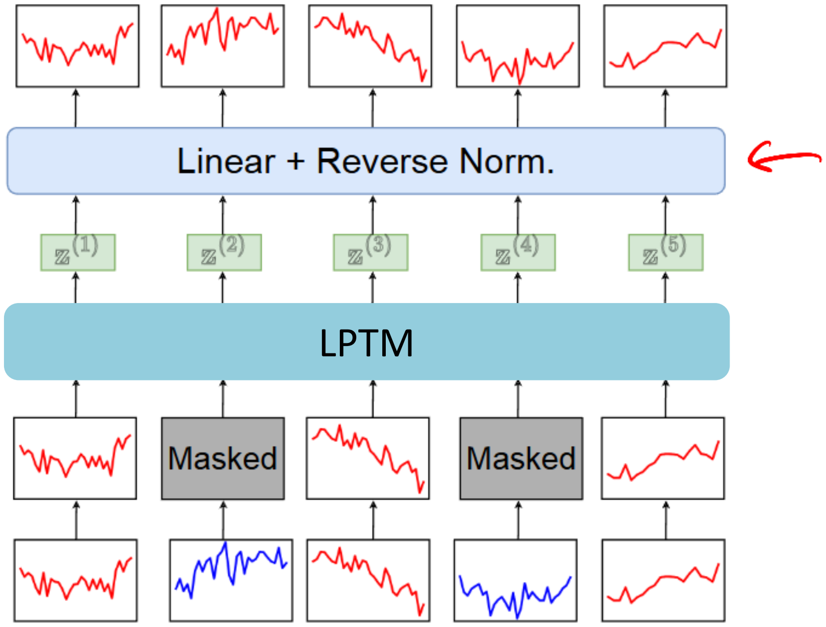
Adaptive Segmentation (Contd.)

- Chosen segments are used by the model



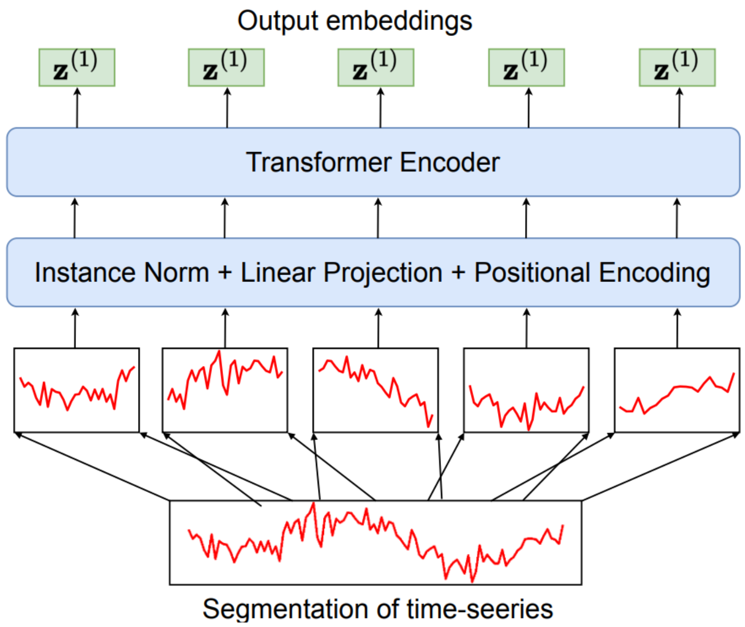
Self-supervised learning task

- Random masking: Reconstruct segments of time-series
- Enables model to learn temporal dynamics across the time-series w/o supervision



Segmented Time-series as tokens

- Each segment is encoded as fixed length token embedding
- Bonus: Faster and memory efficient vs time-steps as tokens



Experiments

- Pre-trained on datsets from epidemiology, Power, Traffic, Demand/Sales, Stocks, Behavioral data
- Trained on unseen applications & datasets from these domains
- Tasks: Forecasting & Classification

Forecasting

- Best ranked model across tasks from diverse domains: Electricity, Epidemiology, Traffic, Demand, Retail
- Beats domain-specific models in most cases

Model	Flu-US	Flu-Japan	ETT1	ETT2	PEM-B	NY-B	NY-T	Nasdaq	M4	Rank
AutoARIMA	2.14	1344	0.73	0.64	4.1	4.13	16.43	0.62	1.89	25.06
Informer	1.62	1139	0.57	0.71	3.1	2.89	12.33	0.83	1.055	15.89
Autoformer	1.41	1227	0.72	0.82	2.7	2.73	12.71	0.19	0.887	13.67
PatchTST	0.96	1113	0.52	0.63	**2.5**	2.64	11.95	0.15	0.877	7.5
N-HITS	1.42	1211	0.53	0.62	2.9	2.74	11.87	0.57	0.968	13.0
TiDE	1.21	1186	**0.49**	0.49	3.5	3.86	11.95	0.57	1.078	13.44
MICN	0.95	1145	**0.49**	0.57	3.6	2.61	11.56	0.13	0.931	7.77
TimesNet	1.04	1194	0.56	0.62	3.9	2.83	11.82	0.19	1.055	12.11
TFT	1.21	1876	0.52	0.51	4.6	2.95	12.55	0.24	1.18	16.11
iTransformer	1.14	1256	0.57	0.59	4.3	2.83	13.16	0.29	1.125	17.5
LLM-Time	1.21	1319	0.52	0.49	3.9	3.7	12.11	0.21	1.064	14.05
TimesFM	1.32	1214	0.58	0.49	3.7	2.8	12.19	0.22	1.07	13.44
Lag-LLAMA	1.46	1416	0.61	0.57	3.9	2.9	13.43	0.28	1.33	20.16
Chronos	1.21	1228	0.59	0.52	3.7	3.1	12.82	0.27	1.04	15.55
MOIRAI	1.31	1336	0.62	0.55	3.9	3.5	13.71	0.24	1.21	19.22
STEP	1.17	983	0.54	0.93	2.7	2.52	**10.37**	**0.11**	1.331	10.33
EpiFNP	**0.52**	872	0.81	1.25	4.1	2.98	12.11	0.28	1.281	16.77
ColaGNN	1.65	694	0.72	1.19	3.9	3.19	14.97	0.25	1.185	19.22
TS2Vec	1.85	905.9	0.99	1.74	3.5	3.11	13.48	0.94	1.344	21.94
SimMTM	1.31	1289	0.61	0.55	3.4	3.1	12.79	0.28	1.284	17.94
TS-TCC	1.94	1134.6	0.75	1.29	3.3	2.97	15.55	0.76	1.274	21
LPTM	0.79	704	**0.49**	**0.46**	**2.5**	**2.37**	11.84	0.17	**0.872**	**2.55**

Kamarthi and Prakash 2024

Classification

- Adapt to classification by simply adding a final fine-tuned classification layer.
- Outperforms SOTA classification models across over 35 tasks

	Informer	Autoformer	TimesNet	TARNet	TS2Vec	TS-TCC	TST	SimMTM	CRT	LPTM
BasicMotions	0.95	0.93	0.92	**1.00**	0.99	**1.00**	0.92	0.86	0.88	**1.00**
FaceDetection	0.51	0.49	0.59	0.63	0.51	0.54	0.62	0.73	0.78	**0.79**
FingerMovements	0.58	0.54	0.58	0.62	0.46	0.47	0.59	0.68	0.72	**0.78**
PEMS-SF	0.67	0.71	0.84	**0.94**	0.75	0.73	0.93	0.86	0.89	0.93
RacketSports	0.83	0.86	0.91	**0.98**	0.77	0.85	0.79	0.84	0.87	0.93
EigenWorms	0.49	0.62	0.73	0.89	0.84	0.77	0.72	0.82	0.79	**0.94**
ArticularyWordRecognition	0.83	0.82	0.79	0.97	0.89	0.97	0.92	0.92	0.88	**0.98**
AtrialFibrillation	0.57	0.55	0.68	**1.00**	0.44	0.37	0.72	0.85	0.89	0.93
CharacterTrajectories	0.57	0.55	0.68	0.97	0.98	0.96	**0.99**	0.97	0.93	0.98
Cricket	0.94	0.87	0.88	**1.00**	0.98	0.97	0.84	0.96	0.94	0.99
DuckGeese	0.54	0.44	0.56	0.75	0.39	0.57	0.74	0.58	0.55	**0.79**

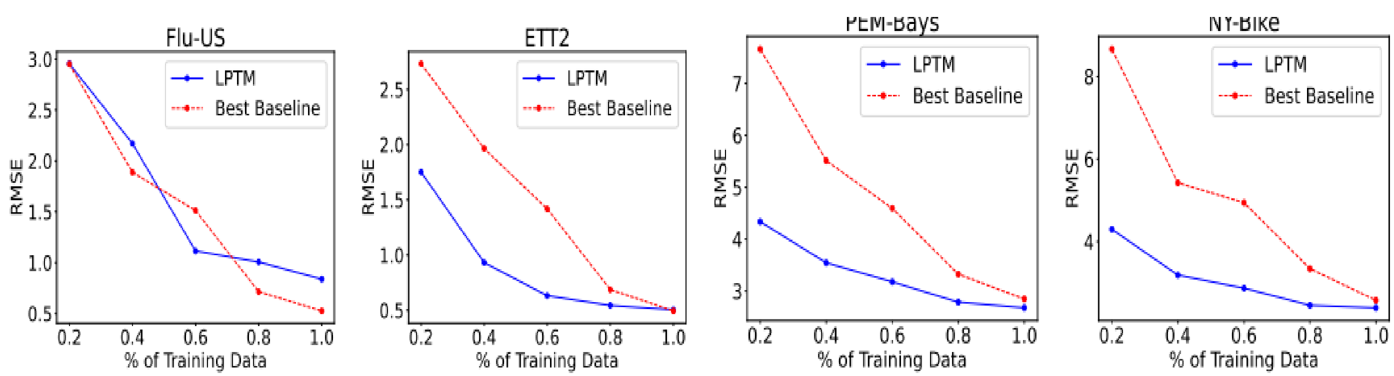
Zero-shot forecasting

- LPTM outperforms SOTA pre-trained models in zero-shot forecasting
- Includes foundational models with 2x-10x more parameters (TimesFM, MOIRAI, Chronos)

Model	Flu-US	Flu-japan	ETT1	ETT2	PEM-B	NY-B	NY-T	Nasdaq	M4
LLM-Time	1.38	1411	0.57	0.54	4.3	4.5	13.53	0.29	1.189
TimesFM	1.35	1259	0.61	0.59	3.9	3.9	13.11	0.29	1.211
Lag-LLAMA	1.52	1488	0.83	1.06	5.3	3.8	12.84	0.24	1.311
Chronos	1.29	1274	0.62	0.56	4.2	3.6	13.74	0.29	1.125
MOIRAI	1.39	1411	0.69	0.52	4.2	4.4	13.82	0.27	1.192
TS2Vec	1.94	1233.1	1.33	1.82	3.7	4.1	14.39	0.87	1.616
TS-TCC	2.17	1356.15	1.14	1.57	4.1	3.8	15.72	0.92	1.492
SimMTM	2.17	1356.15	1.14	1.57	4.1	3.8	15.72	0.92	1.492
LPTM	**1.14**	**996**	**0.53**	**0.49**	**3.4**	**3.2**	**13.12**	**0.22**	**0.972**

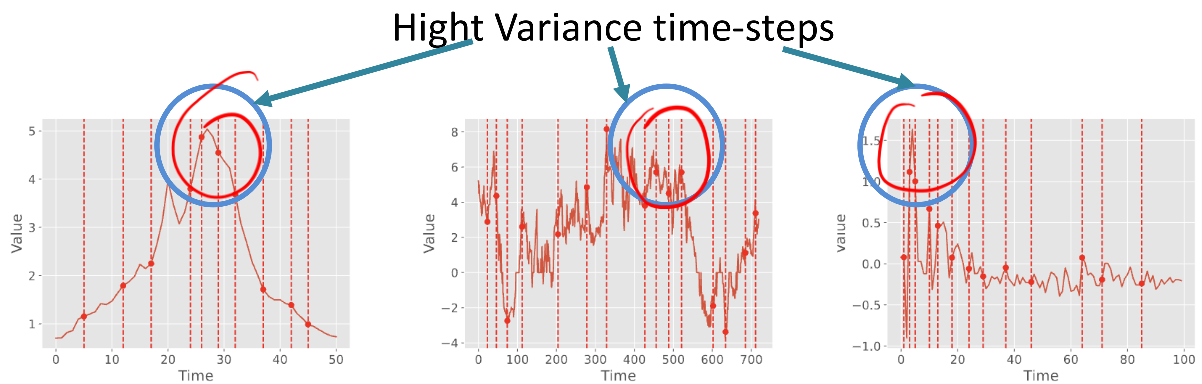
Fine-tuned performance

- LPTM outperforms SOTA pre-trained models in zero-shot forecasting
- Requires much less data and 20-50% lesser compute to converge to best fine-tuned performance



Other Results

- Quicker training time than most competitive baselines in all benchmarks
- Segments selected capture high variance and important time-series regions (peak of epidemic, changepoints)



(a) Flu-US (b) ETT1 (c) BasicMotions

Thank you!

Harsha
Kamarthi

B Aditya
Prakash

- Code: [AdityaLab/LPTM]
- Acknowledgements: NSF (Expeditions CCF-1918770, CAREER IIS-2028586,Medium IIS-1955883, Medium IIS-2403240, Medium IIS-2106961, PIPP CCF-2200269), CDC MInD program, Meta faculty gifts, and funds/computing resources from Georgia Tech

Code

Paper