From Self-supervised learning to LLMs for Timeseries: Adopting "GPT" paradigm for modelling behaviours at scale

Flora Salim

Cisco Chair of Digital Transport & AI
Professor, School of Computer Science and Engineering

Deputy Director (Engagement), UNSW AI Institute
UNSW Sydney

Co-Lead; Machines Program; Mobilities Focus Area,
ARC Centre of Excellence for Automated Decision Making and Society (ADM+S)

NIIN National Industry Innovation Network

CISCO Cisco Research

ADM +S

UNSW AI Institute

UNSW SYDNEY

Collective and Robust Ubiquitous Intelligence (CRUISE) lab

CISCO

Innovation Central

ADM+S ARC Centre of Excellence for Automated Decision-Making and Society

Prof Flora Salim
Cisco Chair of Digital Transport
Deputy Director, *UNSW AI Institute*
Co-Lead, Mobility and Transport, Machines Program, *ADM+S Centre*

Members:

- 3 Faculty members
- 3+ postdocs
- 9+ PhD students
- 7+ Masters and Honours students
- 1 U.S. Fulbright Professor
- 4+ Visiting Researchers

Research Capabilities

- Machine Learning; Unsupervised Learning; Deep Learning; Adversarial Learning
- Time-series, Spatio-temporal Data
- Analytics and Forecasting
- Prediction + Optimization
- Natural Language Processing
- Mobility data science
- On-device AI; Edge and federated learning
- Behaviour modelling
- Mobile and wearable sensing and AI
- Personalization and profiling
- Activity recognition, emotion recognition
- Recommender systems
- Responsible, Ethical, Equitable AI (Fairness, Debiasing, Transparency, Explainability)

Application Domains

- Transport and Mobility
- Energy, Climate, and Sustainability
- Health and wellbeing AI
- Pandemic, Emergency Preparedness

Current Projects

- *Understanding Bias in AI Models for the Prediction of Infectious Disease Spread*, CSIRO and NSF.
- *Mobility Question Answering (Q&A) for Natural-Language-based Spatio-Temporal Forecasting*, Cisco Research USA.
- *NSW Clean Technology Research Development & Commercialisation Infrastructure Grant*, CSIRO, NSW Gov
- *Rail Passenger Ride Comfort Modelling using In-vehicle IoT Sensor Data: A Feasibility Assessment*, TfNSW, Sydney Trains, Cisco
- *IoT Data Security and Assurance Framework for Intelligent Transport*, CyberSecurity CRC & Cisco
- *Towards AI on the edge: Developing data-efficient machine learning models for multimodal sensing devices and IoT*, Data61 NextGen Program
- *ARC Centre of Excellence for Automated Decision Making and Society (ADM+S)*, 2020-2027, Co-Lead of Transport & Mobility Focus Area, Co-Lead of Machines Program.

UNSW AI Institute

NIIN

National Industry Innovation Network

Partnerships

Innovation Central

Specialised Centres

Research Chairs

Skill & Talent Development

Future Investments

UNSW Engineering Faculty Profile

Mechanical Manufacturing

Electrical & Telecommunications

Australian Graduate School of Engineering

Chemical Engineering

Graduate School of Biomedical Engineering

Civil & Environmental

Computer Science & Engineering

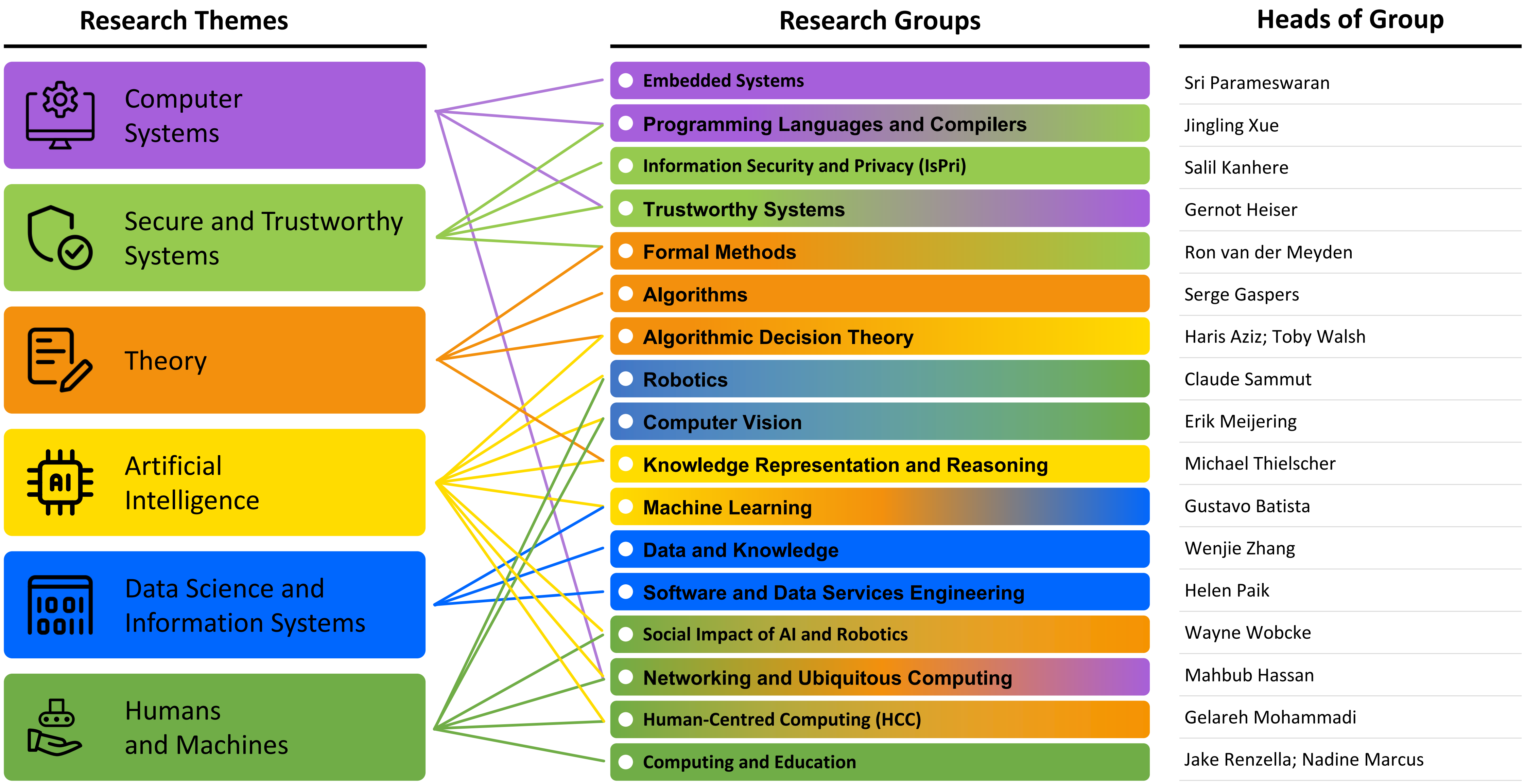
Minerals & Energy Resources

Photovoltaics & Renewable Energy

Innovation Central Sydney (ICS) is a part of UNSW Faculty of Engineering

Research in the School of Computer Science and Engineering

UNSW
SYDNEY



Research Themes

- Computer Systems
- Secure and Trustworthy Systems
- Theory
- Artificial Intelligence
- Data Science and Information Systems
- Humans and Machines

Research Groups	Heads of Group
Embedded Systems	Sri Parameswaran
Programming Languages and Compilers	Jingling Xue
Information Security and Privacy (IsPri)	Salil Kanhere
Trustworthy Systems	Gernot Heiser
Formal Methods	Ron van der Meyden
Algorithms	Serge Gaspers
Algorithmic Decision Theory	Haris Aziz; Toby Walsh
Robotics	Claude Sammut
Computer Vision	Erik Meijering
Knowledge Representation and Reasoning	Michael Thielscher
Machine Learning	Gustavo Batista
Data and Knowledge	Wenjie Zhang
Software and Data Services Engineering	Helen Paik
Social Impact of AI and Robotics	Wayne Wobcke
Networking and Ubiquitous Computing	Mahbub Hassan
Human-Centred Computing (HCC)	Gelareh Mohammadi
Computing and Education	Jake Renzella; Nadine Marcus

ARC Centre of Excellence in Automated Decision Making and Society (ADM+S)

AUTOMATED DECISION-MAKING THAT BENEFITS **ALL AUSTRALIANS**

Goal

The ARC Centre of Excellence for Automated Decision-Making and Society brings together universities, industry, government and the community to support the development of responsible, ethical and inclusive automated decision-making.

At a Glance

- 9 Australian Universities
- 21 Partner Organisations
- 79 Collaborating Organisations
- 247 Centre Members:
    - Social Scientists
    - Historians
    - Humanities
    - Data Scientists
    - Mathematicians
    - Computer Scientists
    - Law and Regulation
    - Economists

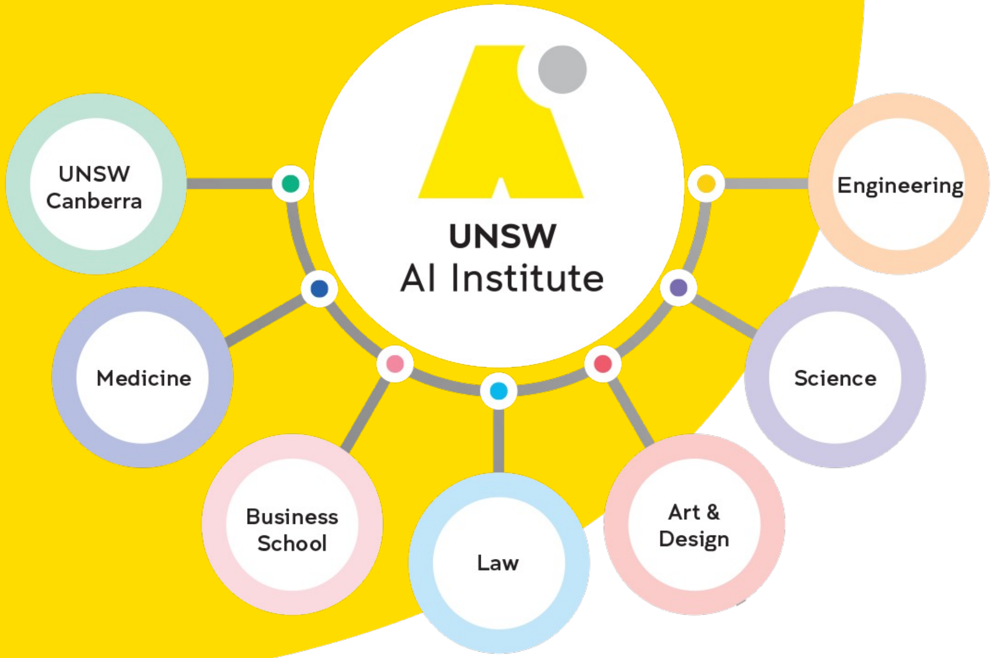
2022 report: https://cloudstor.aarnet.edu.au/plus/s/OxWXdb8RAjnu9kl

UNSW AI Institute

UNSW
AI Institute

UNSW AI Institute is the flagship research institute of UNSW (which is one of the World's Top 100 Universities and 23rd globally for Engineering and Technology*). Its researchers' extensive track record in AI research and development capabilities across several faculties is well recognised globally.

* QS World University Rankings by Subject 2022



UNSW
AI Institute

UNSW Canberra

Engineering

Medicine

Science

Business School

Law

Art & Design

SUPPORTING OVER

300 UNSW ACADEMICS

45 RESEARCH GROUPS, LABS AND CENTRES

Website:
UNSW.ai

Twitter:
@unsw_ai

LinkedIn:
linkedin.com/company/unsw-ai-institute

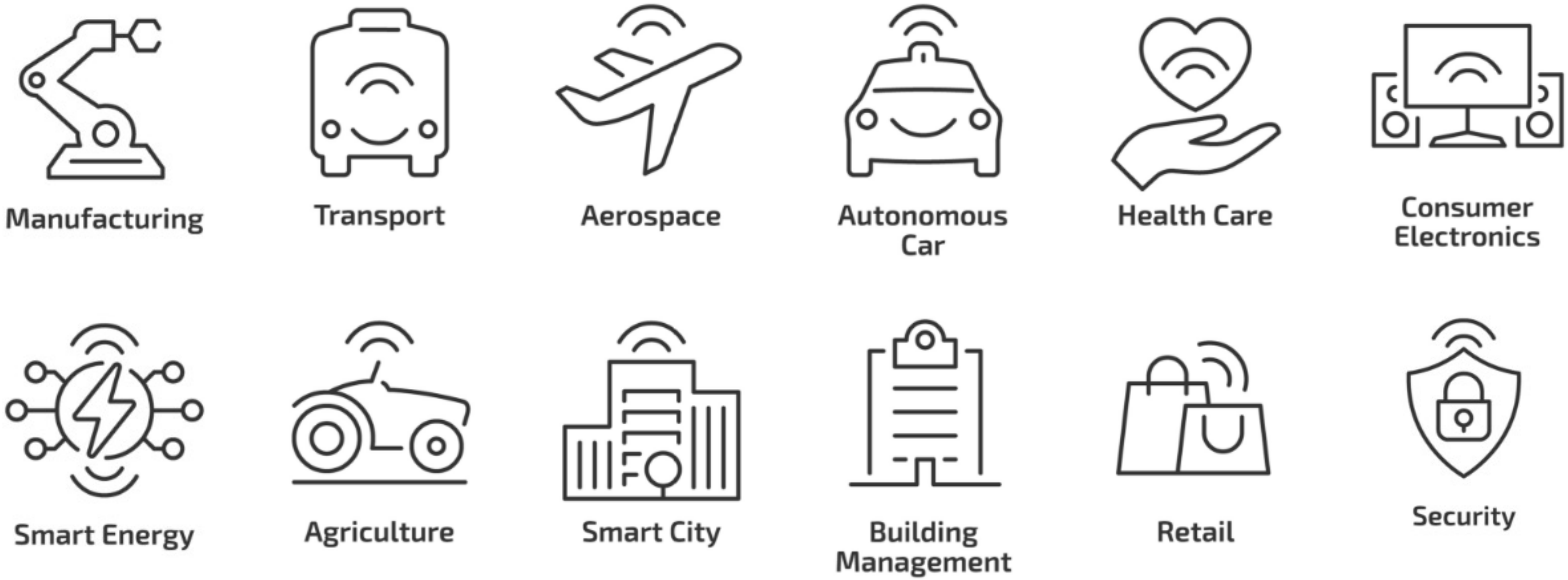
Youtube:
unsw.to/youtube_ai

UNSW
SYDNEY

My research

Learning **Complex Behaviours** using Heterogenous Sensor Data *in the Wild*

Proliferation of IoT in many domains



- Manufacturing
- Transport
- Aerospace
- Autonomous Car
- Health Care
- Consumer Electronics
- Smart Energy
- Agriculture
- Smart City
- Building Management
- Retail
- Security

Mobility Data & AI at the Core

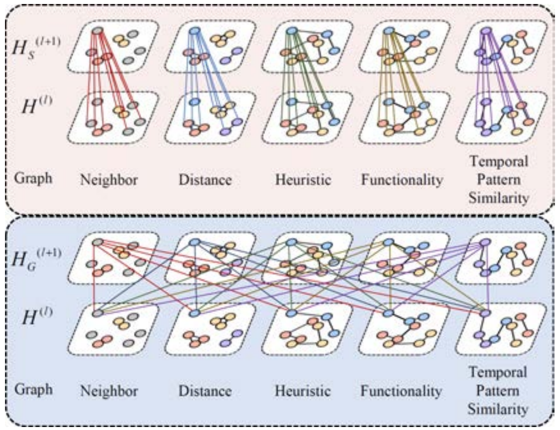
Mokbel et al, 2023, Towards Mobility Data Science (Vision paper) https://arxiv.org/abs/2307.05717

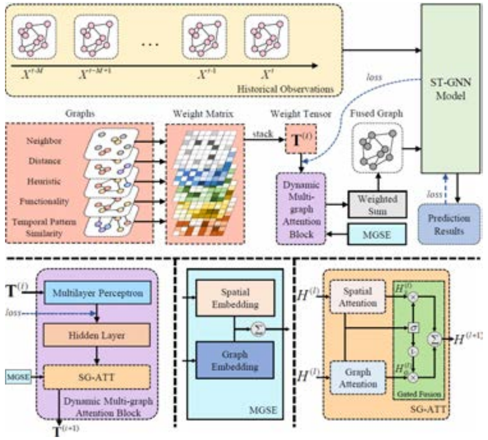
Multi-Scale Behaviour intelligence

Situational Awareness in Dynamic and Challenging Contexts

Representation learning for situational awareness

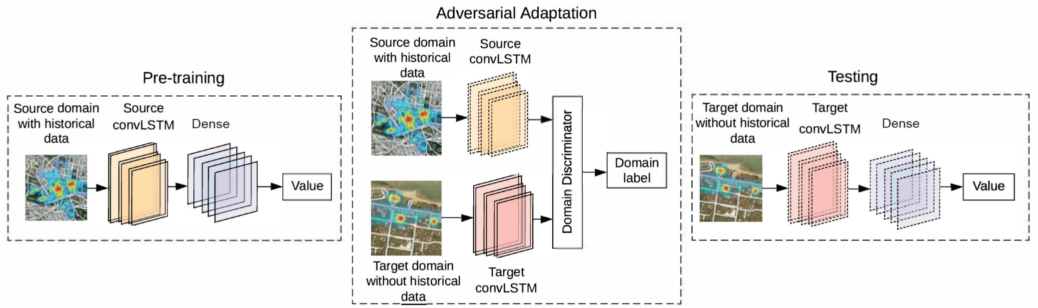
- Handling multi-source time-series, irregularity, unequal-length sequences, inconsistency
- Adaptive to unprecedented volatility, capability to handle zero shot, unseen, adversarial events
- Aware of the dynamic interactions among heterogeneous spatio-temporal data
- Multi-resolution: individual-, group- or urban-level





Applications:

- Traffic and mobility forecasting
- Smart City and Sustainability: air pollution, energy, urban health
- Defence and Intelligence



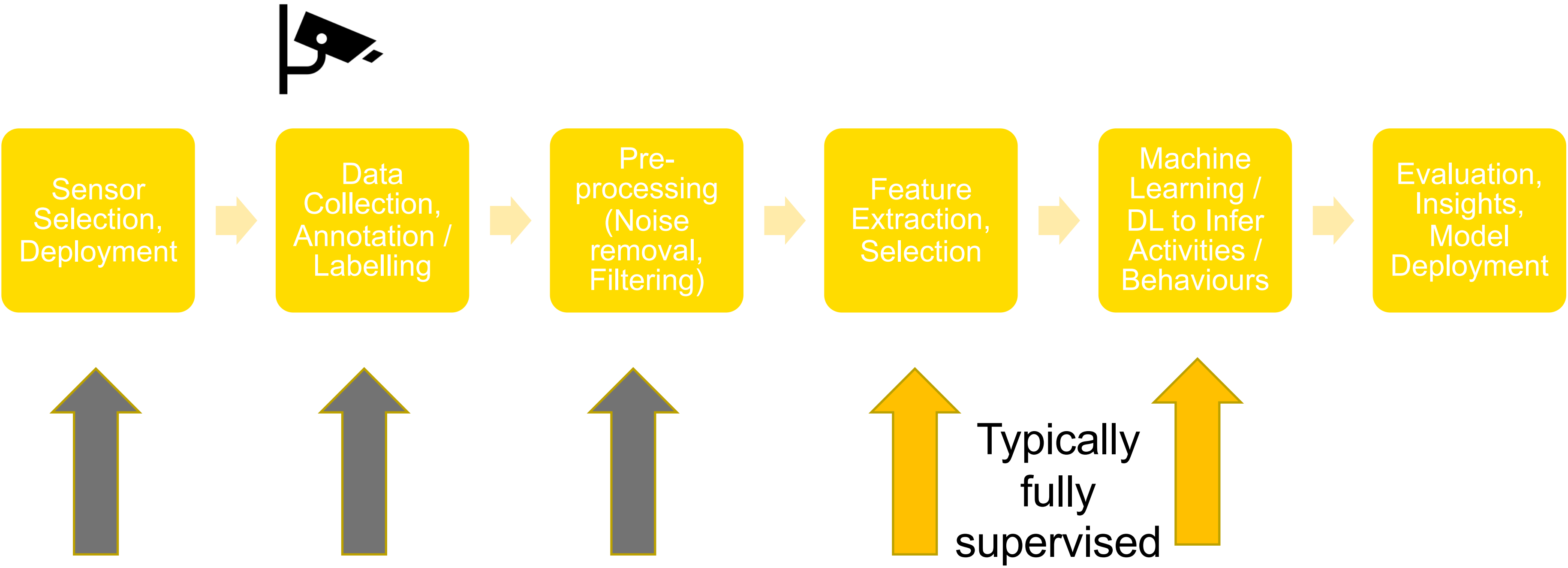
Supported by ARC Discovery Project DP190101485, Qatar National Research Foundation (QNRF), C4NET, CSIRO, Cisco

Patterns of Dynamic Behaviours

- Physiology, health, and mental health
- Mobile Information Needs
- Mobility Behaviour (individual, group, and city-scale)
- Urban-Scale Consumption Behaviour (buildings, retail, energy)

Typical pipeline for behaviour recognition

Daily Environment



Sensor Selection, Deployment

Data Collection, Annotation / Labelling

Pre-processing (Noise removal, Filtering)

Feature Extraction, Selection

Machine Learning / DL to Infer Activities / Behaviours

Evaluation, Insights, Model Deployment

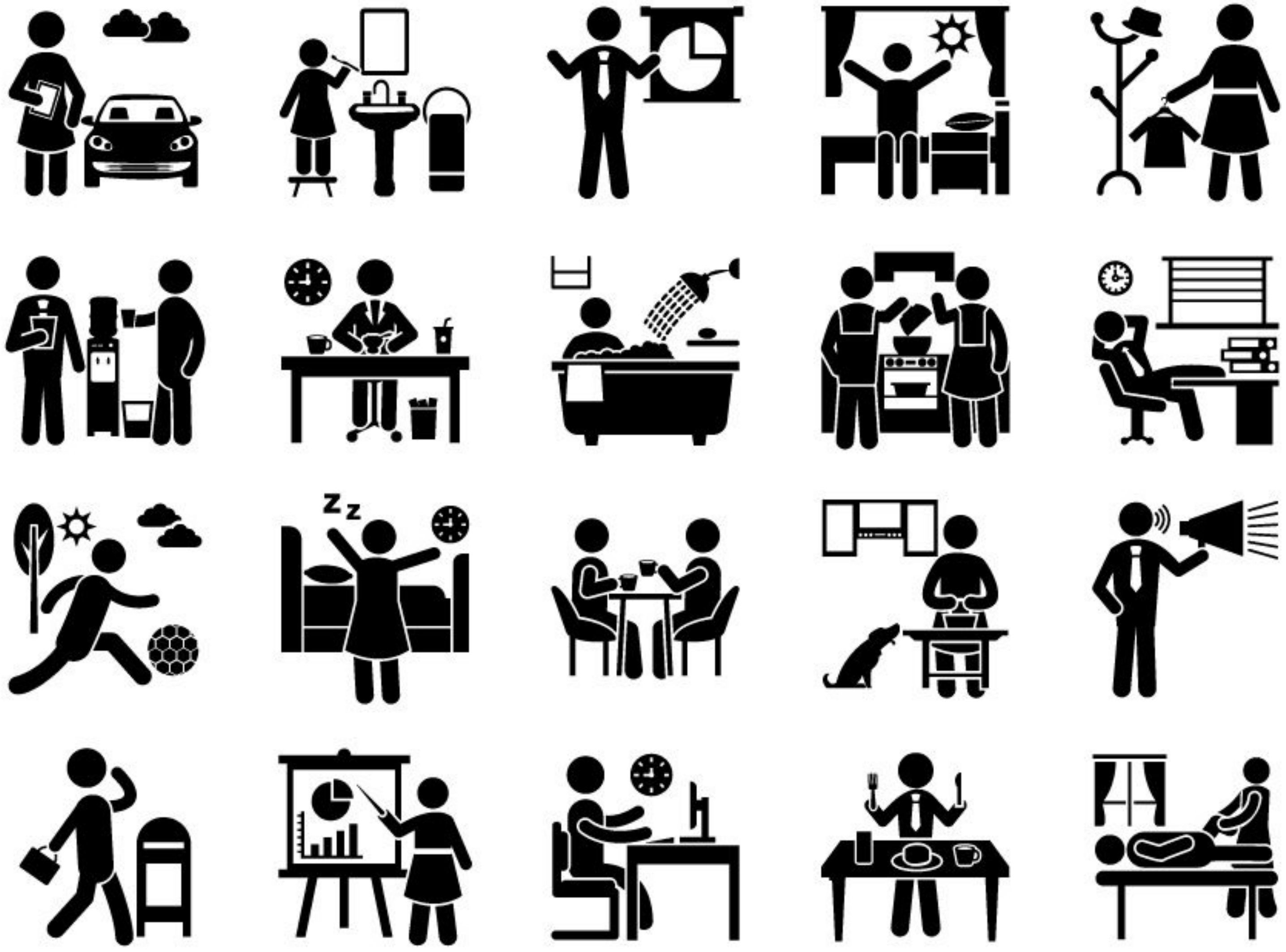
Typically fully supervised

Major bottlenecks

Expensive
Require domain input
Too narrow / specific

A wide variety of inference tasks

- Physical activity recognition
- Step counting
- Mobility / transport usage prediction
- Emotion / stress recognition
- Sleep analytics
- Concentration inference
- Productivity and task inference
- Anomaly detection
- Criminal, safe/unsafe behaviour
- Brain activity inference
- Sport/ muscle activity inference
- Information needs inference
- App usage prediction
- Health condition inference
- Psychometric inference
- Social distancing, contact tracing

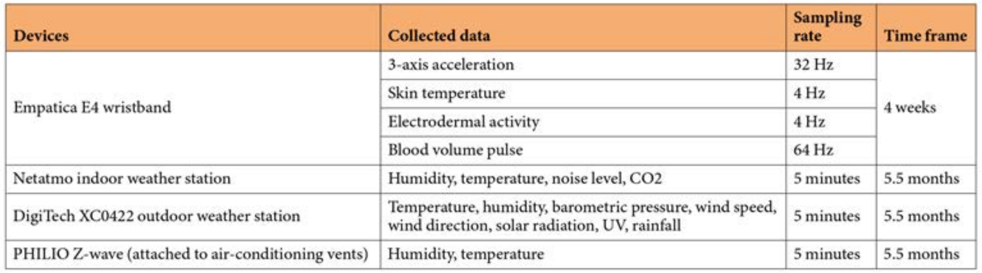


Digital phenotyping

Multi-dimensional data collection and user/ group profiling:

- Behaviour engagement
- Emotion and mental health
- Concentration and stress
- Indoor and thermal comfort
- Cyber / online and physical activities
- Mobility behaviour (individual, groups, and at precinct and city scale)

(a) Empatica E4 wristbands (b) Netatmo indoor weather station (c) Classroom for Year 10 students



Devices	Collected data	Sampling rate	Time frame
Empatica E4 wristband	3-axis acceleration	32 Hz	4 weeks
	Skin temperature	4 Hz	
	Electrodermal activity	4 Hz	
	Blood volume pulse	64 Hz	
Netatmo indoor weather station	Humidity, temperature, noise level, CO2	5 minutes	5.5 months
DigiTech XC0422 outdoor weather station	Temperature, humidity, barometric pressure, wind speed, wind direction, solar radiation, UV, rainfall	5 minutes	5.5 months
PHILIO Z-wave (attached to air-conditioning vents)	Humidity, temperature	5 minutes	5.5 months

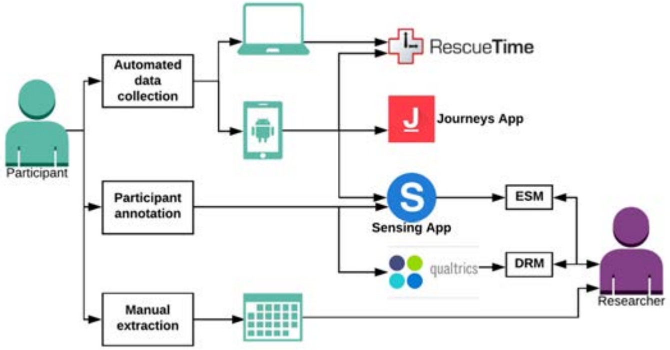


Figure 1: Overview of logging apps and their instruments.

[1] Gao, N., Marschall, M., Burry, J., Watkins, S. and Salim, F.D., 2022. Understanding occupants' behaviour, engagement, emotion, and comfort indoors with heterogeneous sensors and wearables. *Scientific Data*, *9*(1).

[2] Liono, J., Trippas, J.R., Spina, D., Rahaman, M.S., Ren, Y., Salim, F.D., Sanderson, M., Scholer, F. and White, R.W., 2019. Building a benchmark for task progress in digital assistants. *Proceedings of WSDM*, *19*.

[3] Rahaman, M.S., Liono, J., Ren, Y., Chan, J., Kudo, S., Rawling, T. and Salim, F.D., 2020. An ambient–physical system to infer concentration in open-plan workplace. *IEEE Internet of Things Journal*, *7*(12), pp.11576-11586.

A multimodal sensor dataset: occupants' behaviour, engagement, emotion, and comfort indoors

Gao N, Marshall M, Burry J, Watkins S, Salim FD. Understanding occupants' behaviour, engagement, emotion, and comfort indoors with heterogeneous sensors and wearables.

(*Nature Scientific Data*)

The research is funded by ARC linkage program (No. LP150100246) 'Swarming: micro-flight data capture and analysis in architectural design' by Prof J Burry, Prof, S Watkins, A/Prof F Salim, RMIT University. Industry Partner of this research is Aurecon.

Dataset Introduction

In-Gauge Dataset

a 5-month longitudinal field study across 17 classrooms

En-Gage Dataset

4-week cross-sectional study with 29 participants

Devices	Collected data	Sampling rate	Time frame
Empatica E4 wristband	3-axis acceleration Skin temperature Electrodermal activity Blood volume pulse	32 Hz 4 Hz 4 Hz 64 Hz	4 weeks
Netamo indoor weather station	Humidity, temperature, noise level, CO_2	5 minutes	5.5 months
DigiTech XC0422 outdoor weather station	Temperature, humidity, barometric pressure, wind speed, wind direction, solar radiation, UV, rainfall	5 minutes	5.5 months
PHILIO Z-wave (attached to air-conditioning vents)	Humidity, temperature	5 minutes	5.5 months

Indoor station

Weather station

Window/door sensor

AC sensor

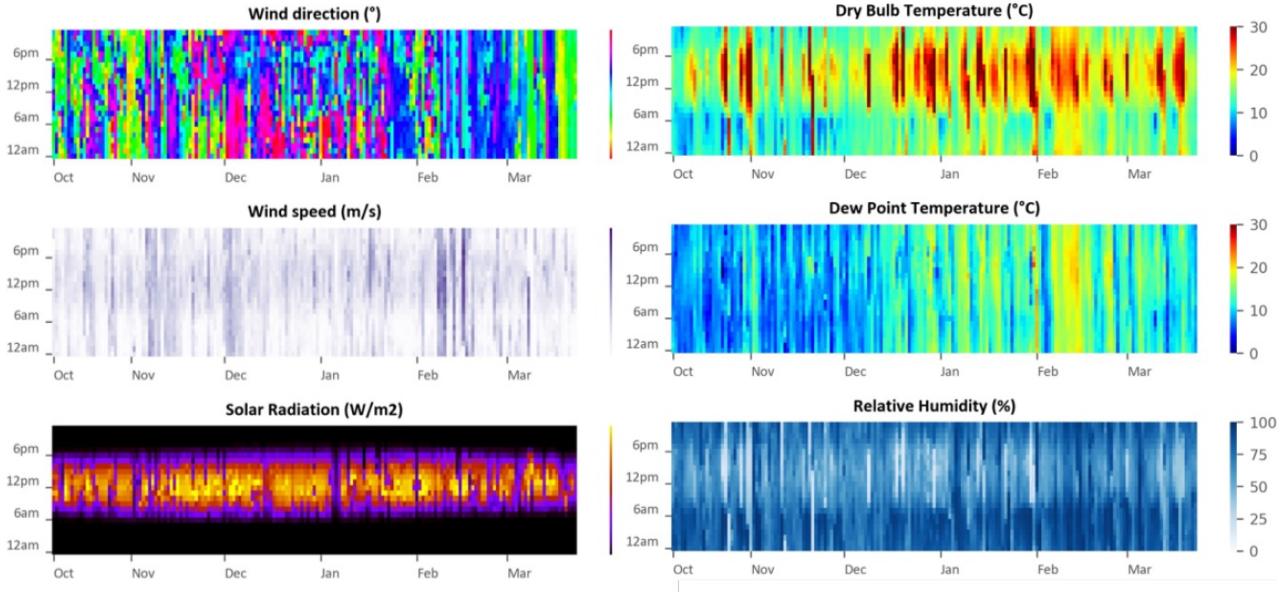
Wearables

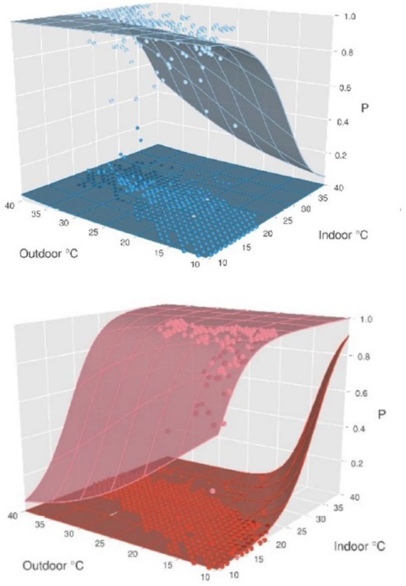
Touch the photo that best captures how you feel right now: (optional)

Please rate your confidence level for your answers in this survey: (optional)

Daily Survey

In-Gauge Dataset





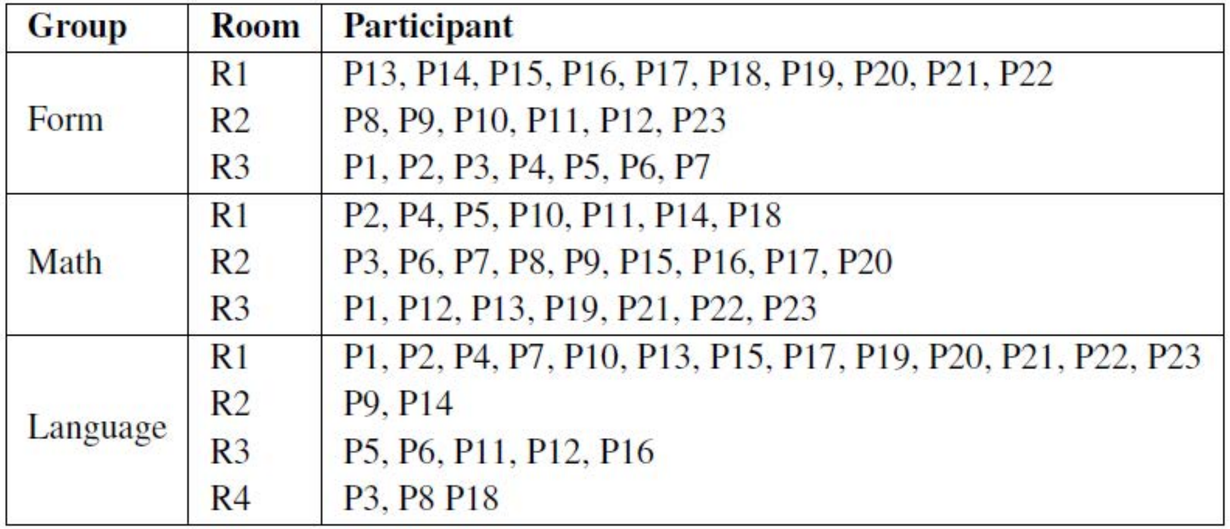
	Units	Range	Accuracy	Resolution
Dry Bulb Temperature	°C	-40 °C – 60 °C	±1 %	0.1 °C
Dew Point Temperature	°C	-40 °C – 60 °C	±1 %	0.1 °C
Relative Humidity	%	1 % - 99 %	±5 %	1%
Wind Speed	m/s	0 m/s - 50 m/s	±1 m/s (<5 m/s) ±10 % (>= 5 m/s)	0.1 m/s
Gust Speed	m/s	0 m/s - 50 m/s	±1 m/s (<5 m/s) ±10 % (>= 5 m/s)	0.1 m/s
Wind Direction	°	0 ° - 360 °	±22.5 °	22.5 °
Rainfall	mm	0 mm - 9999 mm	±10 %	0.3 mm (<1000 mm) 1 mm (>= 1000 mm)
Light	Lux	0k Lux - 400k Lux	±15 %	0.1 Lux
Solar Radiation	W/m^2	-	-	-

	Units	Range	Accuracy	Resolution
Dry Bulb Temperature	°C	0 °C to 50 °C	± 0,3 °C	0.1 °C
Relative Humidity	%	0 to 100 %	± 3 %	1 %
CO$_2$	ppm	0 to 5,000 ppm	±50 ppm (<1,000 ppm) ±5 % (>= 1,000 ppm)	1 ppm
Noise	dB	35 dB to 120 dB	-	1 dB

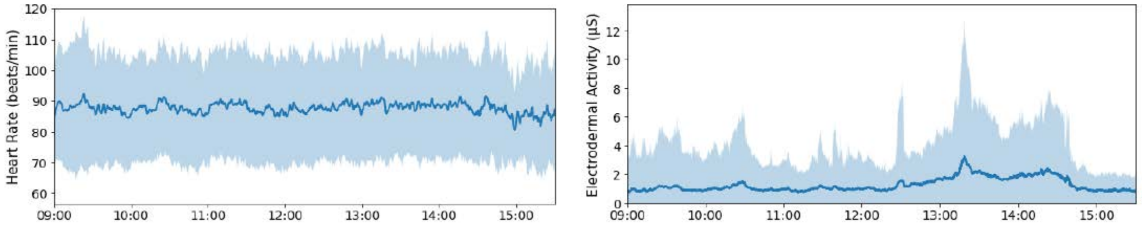
Netamo Home Coach Logging Specifications

DigiTech XC0422 logging specifications

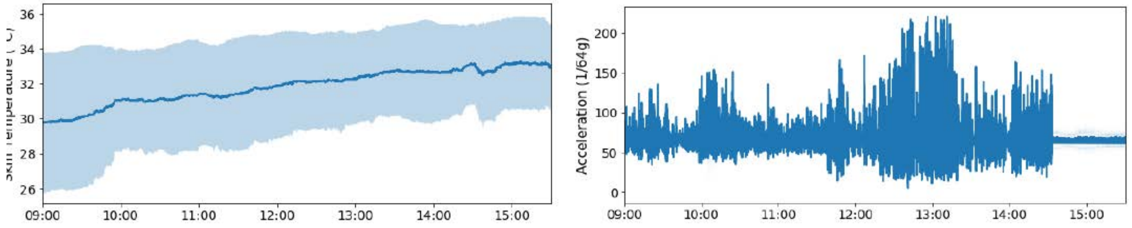
En-Gage Dataset



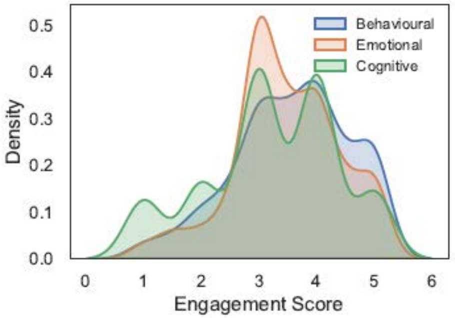
Group	Room	Participant
Form	R1	P13, P14, P15, P16, P17, P18, P19, P20, P21, P22
	R2	P8, P9, P10, P11, P12, P23
	R3	P1, P2, P3, P4, P5, P6, P7
Math	R1	P2, P4, P5, P10, P11, P14, P18
	R2	P3, P6, P7, P8, P9, P15, P16, P17, P20
	R3	P1, P12, P13, P19, P21, P22, P23
Language	R1	P1, P2, P4, P7, P10, P13, P15, P17, P19, P20, P21, P22, P23
	R2	P9, P14
	R3	P5, P6, P11, P12, P16
	R4	P3, P8 P18



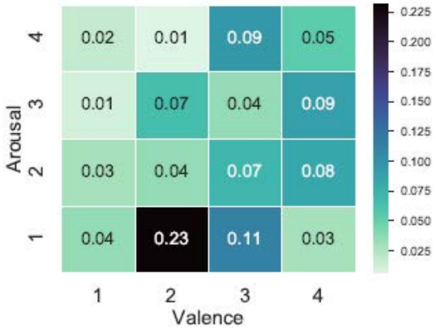
(a) Heart Rate (b) Electrodermal Activity



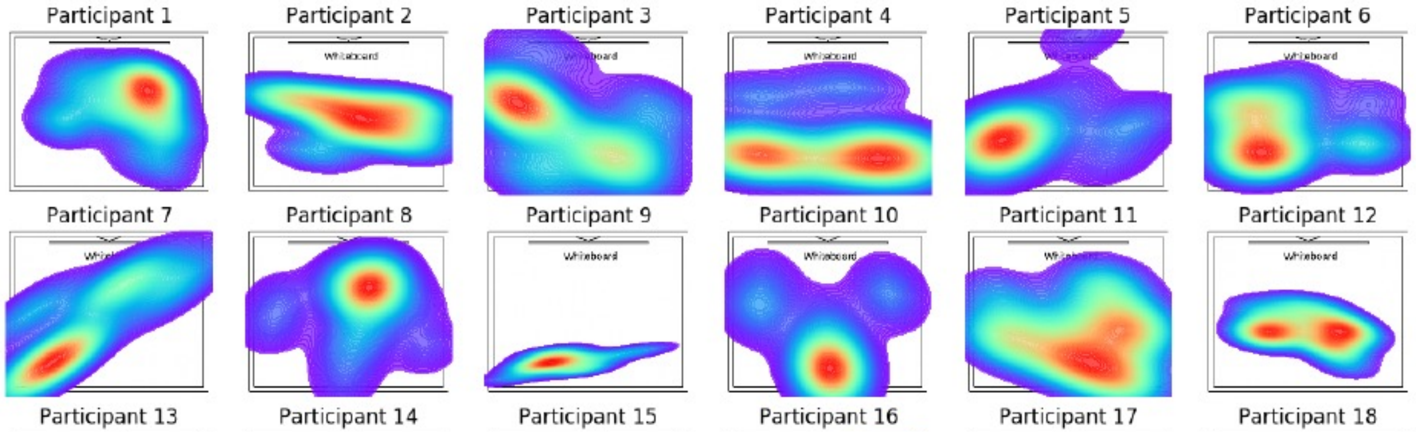
(c) Skin Temperature (d) 3-axis Acceleration (Magnitude)



(a) Multi-dimensional Engagement



(b) Valence and Arousal



Microsoft Cortana Intelligence Institute

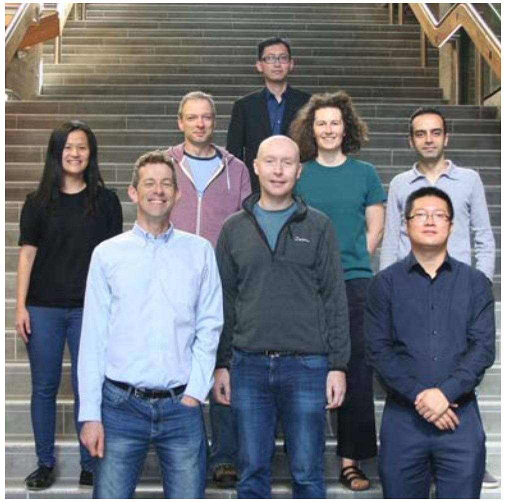
Overview:

Cortana Intelligence Institute is driving the next-generation of capabilities for Microsoft’s digital assistant, Cortana. Focused on researching work-related tasks and using sensors in mobile phones, the CII team builds a complex multidimensional data set, used to model and predict user’s work-related tasks.

Impact:

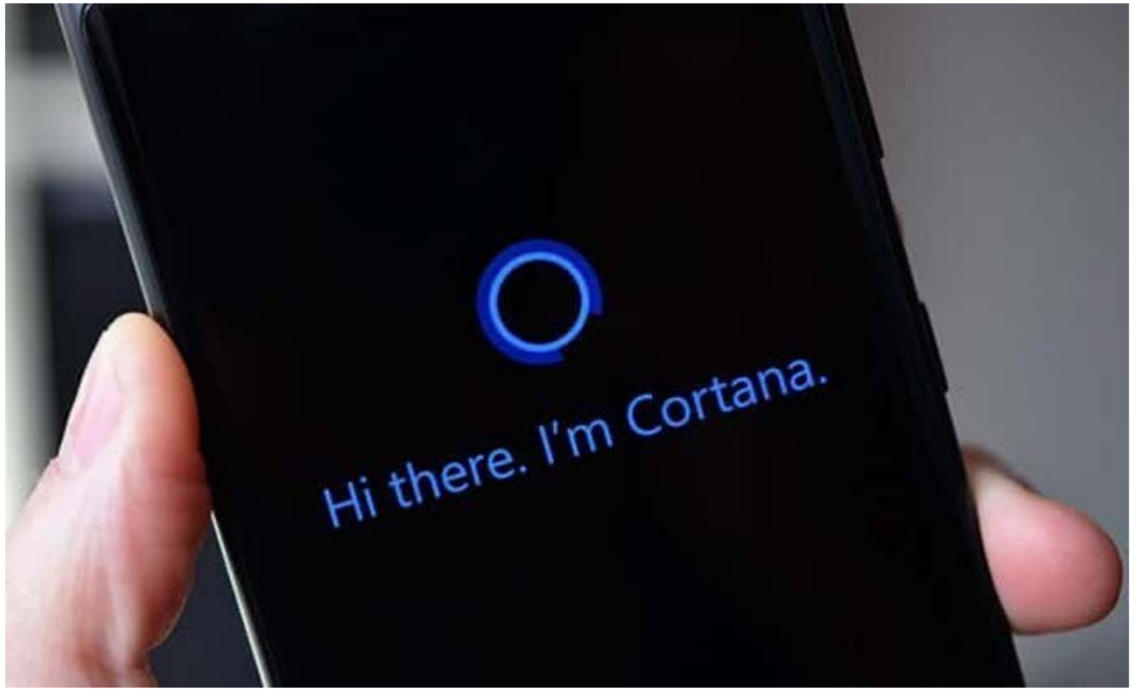
- Task intelligence, to support complex tasks such as tracking a person's progress on a task, reminders, or assisting with completion of a task.
- Create a virtual assistant that can manage a calendar, understand the user, be aware of context, and support multi-turn dialogues.

Hey Cortana



Understanding Mobile Information Needs

Microsoft Cortana Intelligence Institute, RMIT University



Khaokaew, Y., Holcombe-James, I., Rahaman, M.S., Liono, J., Trippas, J.R., Spina, D., Bailey, P., Belkin, N.J., Bennett, P.N., Ren, Y. and Sanderson, M., 2022. Imagining future digital assistants at work: A study of task management needs. *International Journal of Human-Computer Studies*, *168*, p.102905.

Research Gaps (User’s need on Digital assistants)

Contextual information from sensors

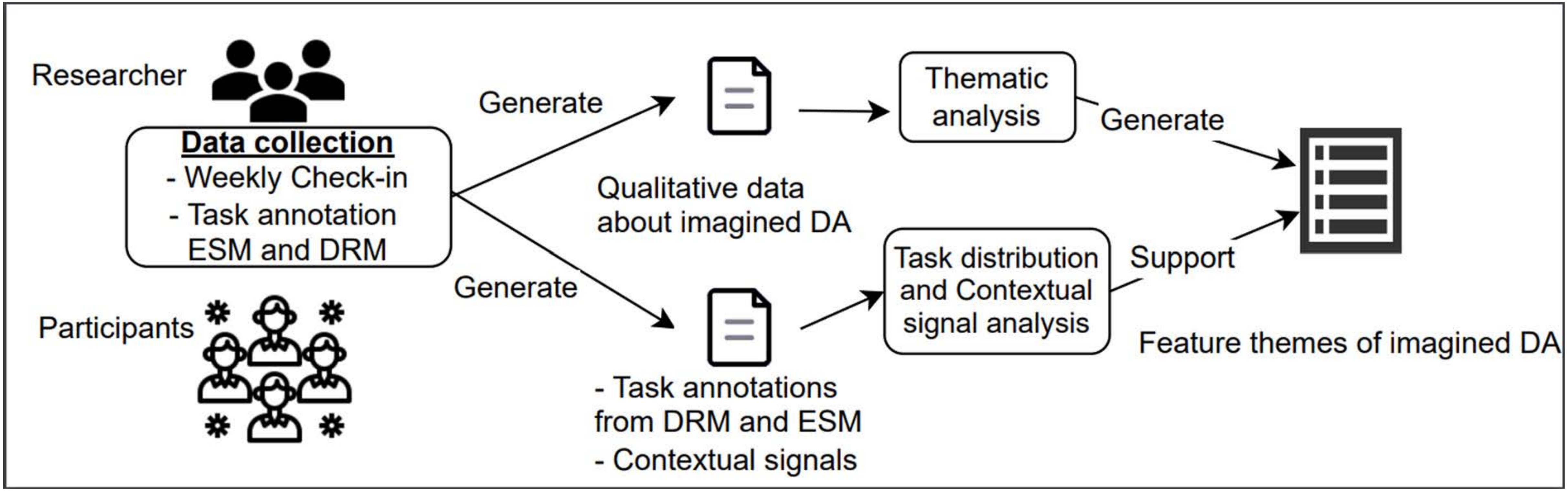
?

Use Survey-based approach

?

The performed tasks based on occupations

Methodology (Mixed method)



Khaokaew, Y., Holcombe-James, I., Rahaman, M.S., Liono, J., Trippas, J.R., Spina, D., Bailey, P., Belkin, N.J., Bennett, P.N., Ren, Y. and Sanderson, M., 2022. Imagining future digital assistants at work: A study of task management needs. *International Journal of Human-Computer Studies*, *168*, p.102905.

Task annotation

Task [XYZ]

Categorized into

Task taxonomy [*]

1. Travel
2. Physical
3. Education
4. Meals breaks
5. Communication
6. Planning
7. Project
8. Documentation
9. Low level
10. Admin + management
11. Finance
12. IT
13. Customer care
14. Problem solving

---

* Trippas, J.R., Spina, D., Scholer, F., Awadallah, A.H., Bailey, P., Bennett, P.N., White, R.W., Liono, J., Ren, Y., Salim, F.D. and Sanderson, M., 2019, March. **Learning about work tasks to inform intelligent assistant design.** In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (pp. 5-14).

Methodology (Contextual signals)

Category	Signals	Description
Device usage (Cyber)	Social	associated with user distraction
	Messaging	indication of personal and direct communication
	Travel	indication of journey planning
	Email	indicator of professional communication
	Calendar	indication of daily scheduling
Journey (Movement)	Number of regular places	average number of regular places
	Number irregular places	average number irregular places
	Number of long-range journeys	average number of long-range journeys
	Number of short-range journeys	average number of short-range journeys
Social (Social interact)	Meeting per day	average number of meeting per day
	Meeting attendees	average number of Meeting attendees

Khaokaew, Y., Holcombe-James, I., Rahaman, M.S., Liono, J., Trippas, J.R., Spina, D., Bailey, P., Belkin, N.J., Bennett, P.N., Ren, Y. and Sanderson, M., 2022. Imagining future digital assistants at work: A study of task management needs. *International Journal of Human-Computer Studies*, *168*, p.102905.

Result (relationship between Tasks and Sub-themes)

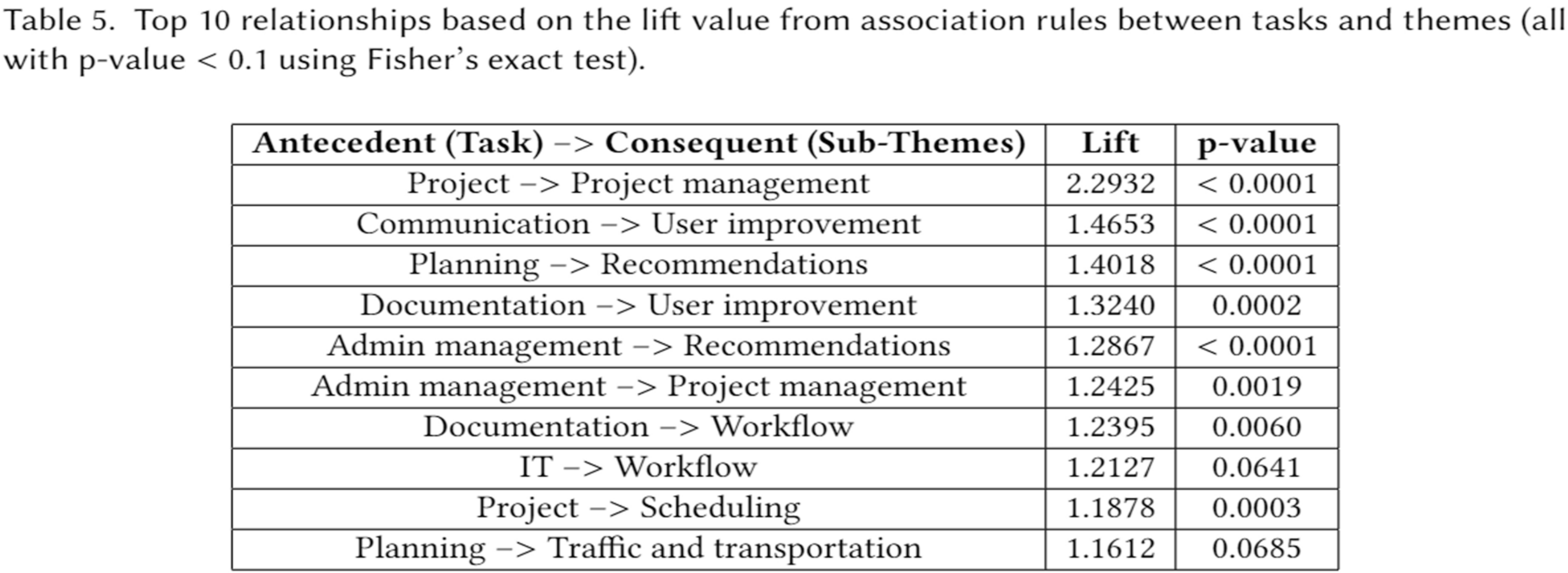
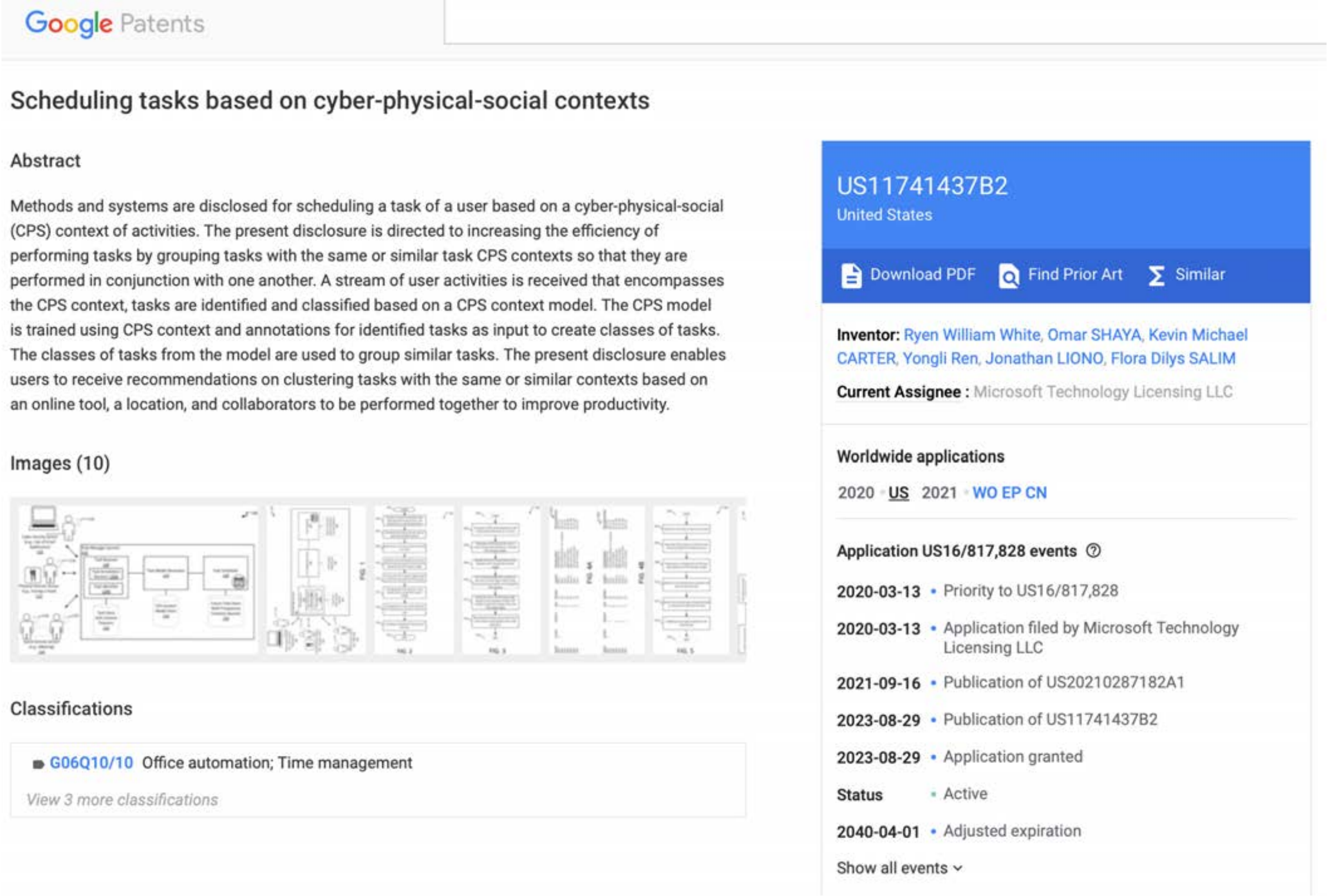


Table 5. Top 10 relationships based on the lift value from association rules between tasks and themes (all with p-value < 0.1 using Fisher's exact test).

Antecedent (Task) –> Consequent (Sub-Themes)	**Lift**	**p-value**
Project –> Project management	2.2932	< 0.0001
Communication –> User improvement	1.4653	< 0.0001
Planning –> Recommendations	1.4018	< 0.0001
Documentation –> User improvement	1.3240	0.0002
Admin management –> Recommendations	1.2867	< 0.0001
Admin management –> Project management	1.2425	0.0019
Documentation –> Workflow	1.2395	0.0060
IT –> Workflow	1.2127	0.0641
Project –> Scheduling	1.1878	0.0003
Planning –> Traffic and transportation	1.1612	0.0685

Khaokaew, Y., Holcombe-James, I., Rahaman, M.S., Liono, J., Trippas, J.R., Spina, D., Bailey, P., Belkin, N.J., Bennett, P.N., Ren, Y. and Sanderson, M., 2022. Imagining future digital assistants at work: A study of task management needs. *International Journal of Human-Computer Studies*, *168*, p.102905.

A non-typical output: a US patent with Microsoft Research



Scheduling tasks based on cyber-physical-social contexts

Abstract

Methods and systems are disclosed for scheduling a task of a user based on a cyber-physical-social (CPS) context of activities. The present disclosure is directed to increasing the efficiency of performing tasks by grouping tasks with the same or similar task CPS contexts so that they are performed in conjunction with one another. A stream of user activities is received that encompasses the CPS context, tasks are identified and classified based on a CPS context model. The CPS model is trained using CPS context and annotations for identified tasks as input to create classes of tasks. The classes of tasks from the model are used to group similar tasks. The present disclosure enables users to receive recommendations on clustering tasks with the same or similar contexts based on an online tool, a location, and collaborators to be performed together to improve productivity.

Images (10)

Classifications

G06Q10/10 Office automation; Time management

View 3 more classifications

US11741437B2
United States

Inventor: Ryen William White, Omar SHAYA, Kevin Michael CARTER, Yongli Ren, Jonathan LIONO, Flora Dilys SALIM

Current Assignee : Microsoft Technology Licensing LLC

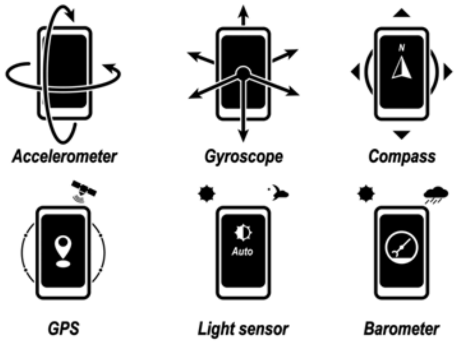
Worldwide applications

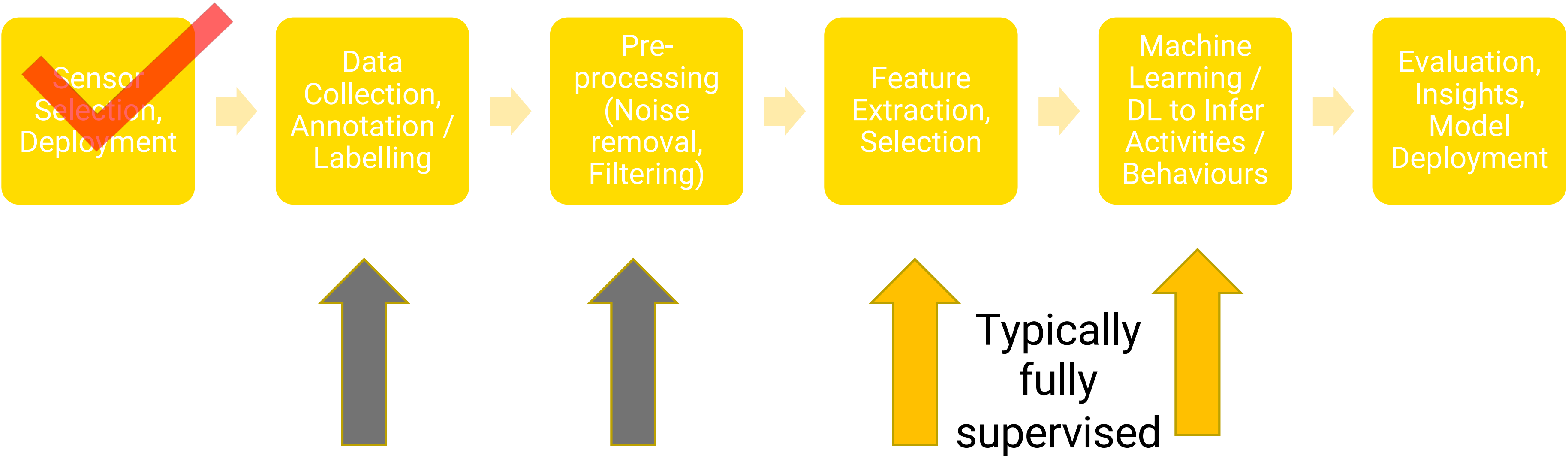
2020 US 2021 WO EP CN

Application US16/817,828 events

2020-03-13	Priority to US16/817,828
2020-03-13	Application filed by Microsoft Technology Licensing LLC
2021-09-16	Publication of US20210287182A1
2023-08-29	Publication of US11741437B2
2023-08-29	Application granted
Status	Active
2040-04-01	Adjusted expiration

Typical pipeline for human activity recognition





Sensor Selection, Deployment

Data Collection, Annotation / Labelling

Pre-processing (Noise removal, Filtering)

Feature Extraction, Selection

Machine Learning / DL to Infer Activities / Behaviours

Evaluation, Insights, Model Deployment

Typically fully supervised

Major bottlenecks

Expensive
Require domain input
Too narrow / specific

Learning from the Open World with Data in the Wild: Different types of spatio-temporal data from urban sensors

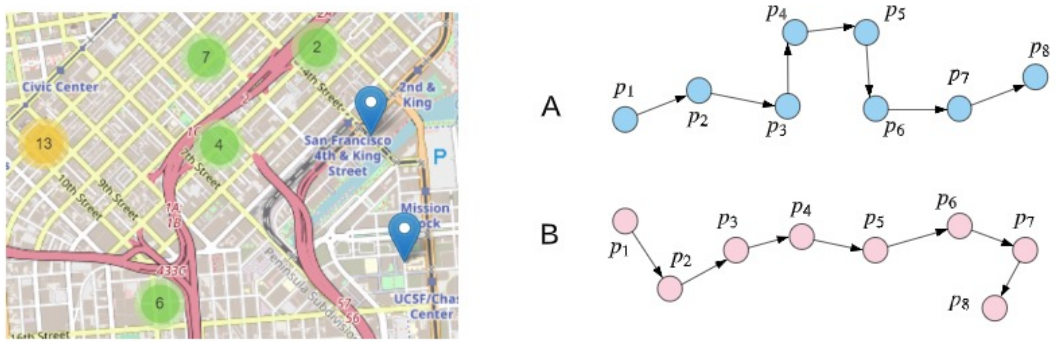
ST Event Data

ST Trajectory Data

ST Raster Data

ST Graph Data

Time-Series with (sporadic) spatial reference



(a) Spatio-temporal events

(b) Two trajectories

Fig. 1. Examples of spatio-temporal events and trajectories

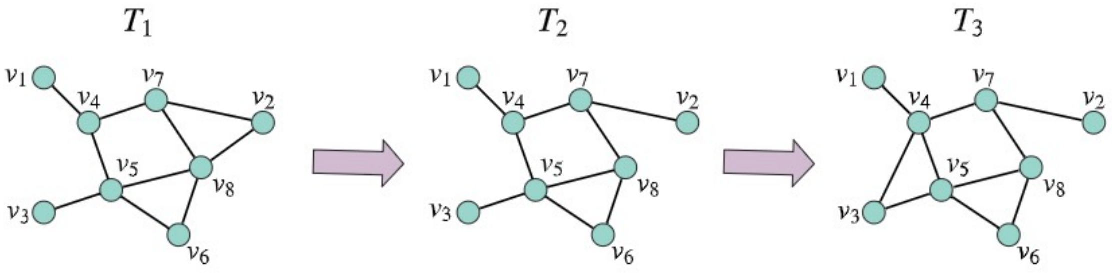


Fig. 2. Example of Spatio-temporal Graph Data

Gao, N., Xue, H., Shao, W., Zhao, S., Qin, K.K., Prabowo, A., Rahaman, M.S. and Salim, F.D., 2022. Generative adversarial networks for spatio-temporal data: A survey. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *13*(2), pp.1-25.

Commercial Applications

Medical Application

- Noise
- Data distribution shift
- Behaviour shift

Biological Applications

Geological & Climate Application

ChatGPT:

Chat

Generative

Pre-trained

Transformer

The evolution of Generative AIs on vision, text, and multimodal data

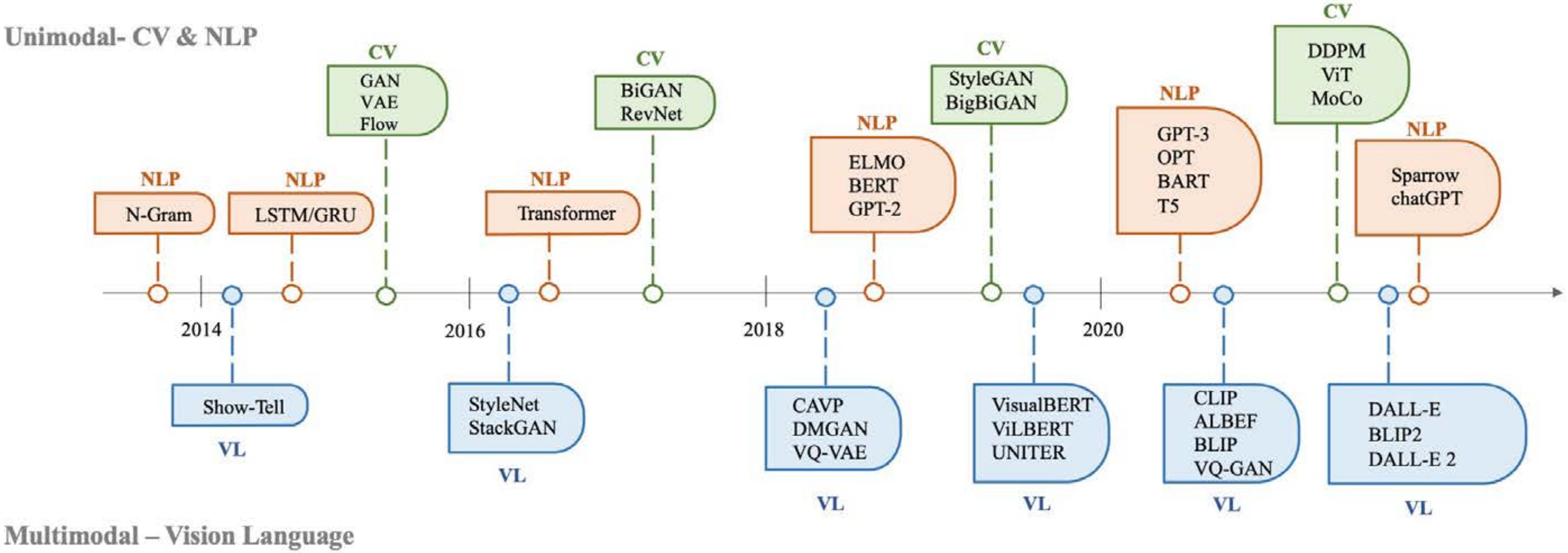
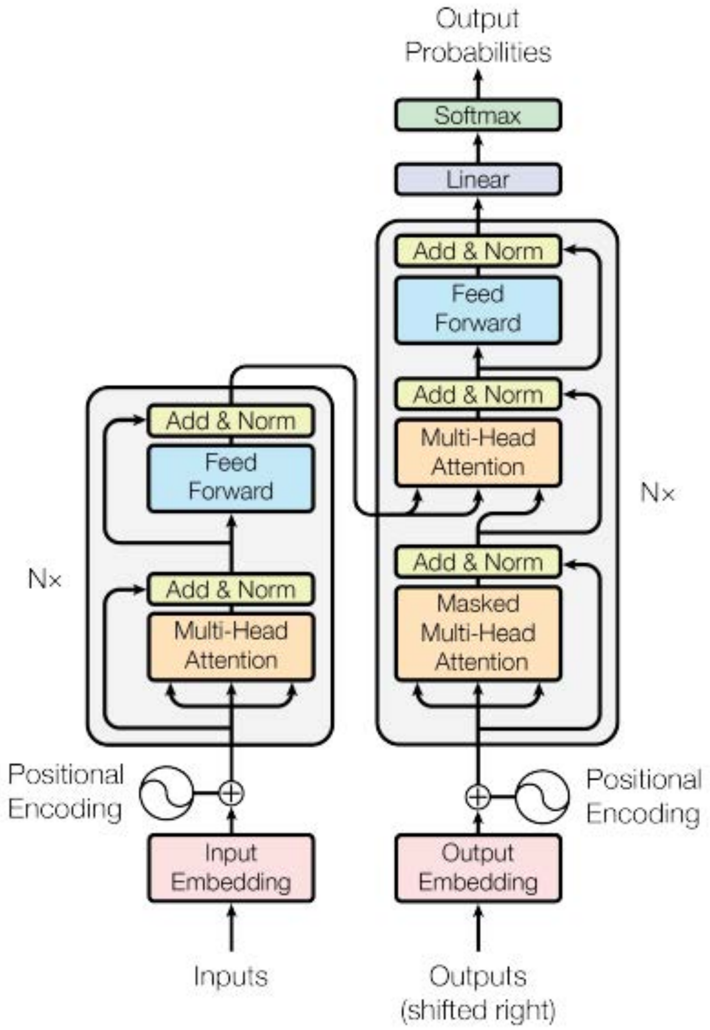


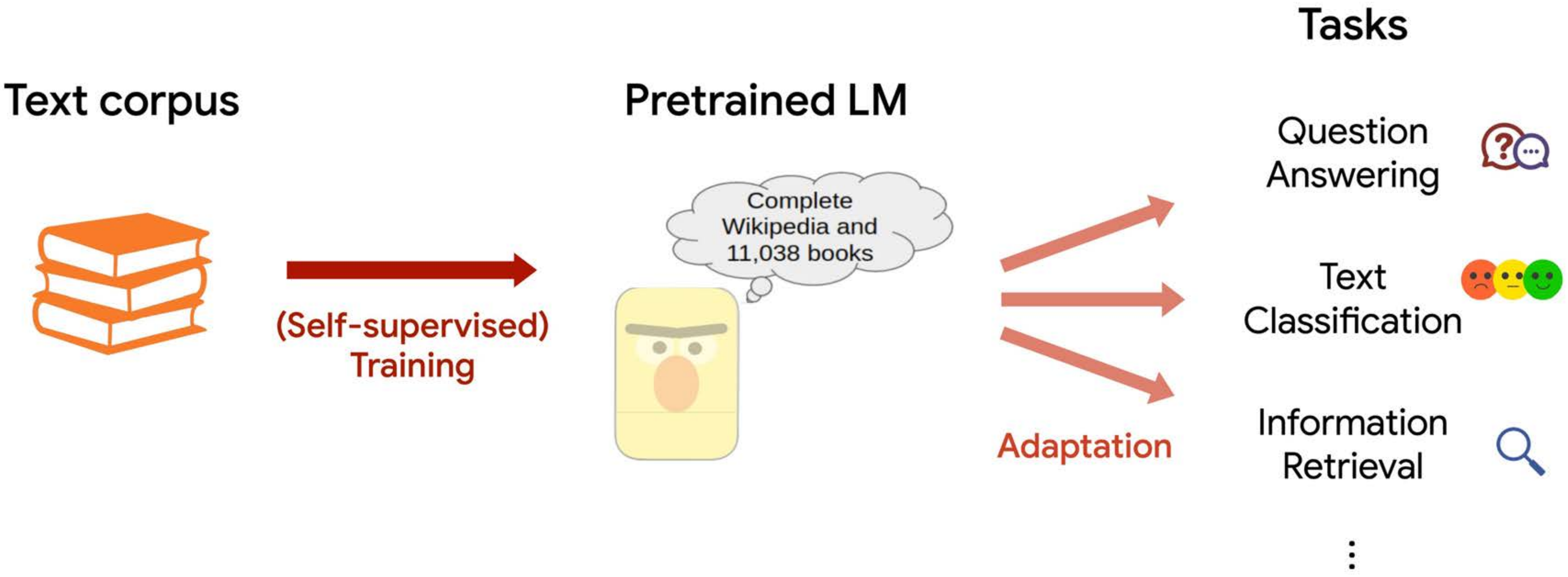
Image credit: Cao et al., 2023, A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT

Transformer marked the birth of Large Language Models (LLMs)



From “Attention is all you need” paper by Vaswani, et al., 2017

Pre-trained Language Model



http://ai.stanford.edu/blog/linkbert/

How can I harness GPT paradigm for multimodal and mobility sensor data?

Self-supervised Pretraining for Multimodal and Mobility Sensor Data

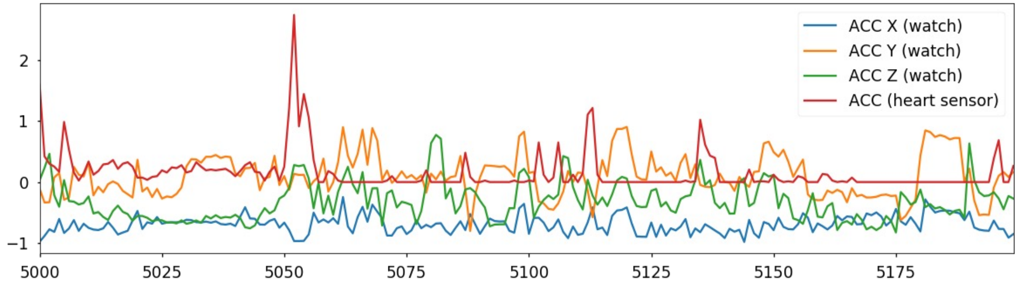
UNSW
AI Institute

Typical machine learning algorithms require a high volume of **training** data

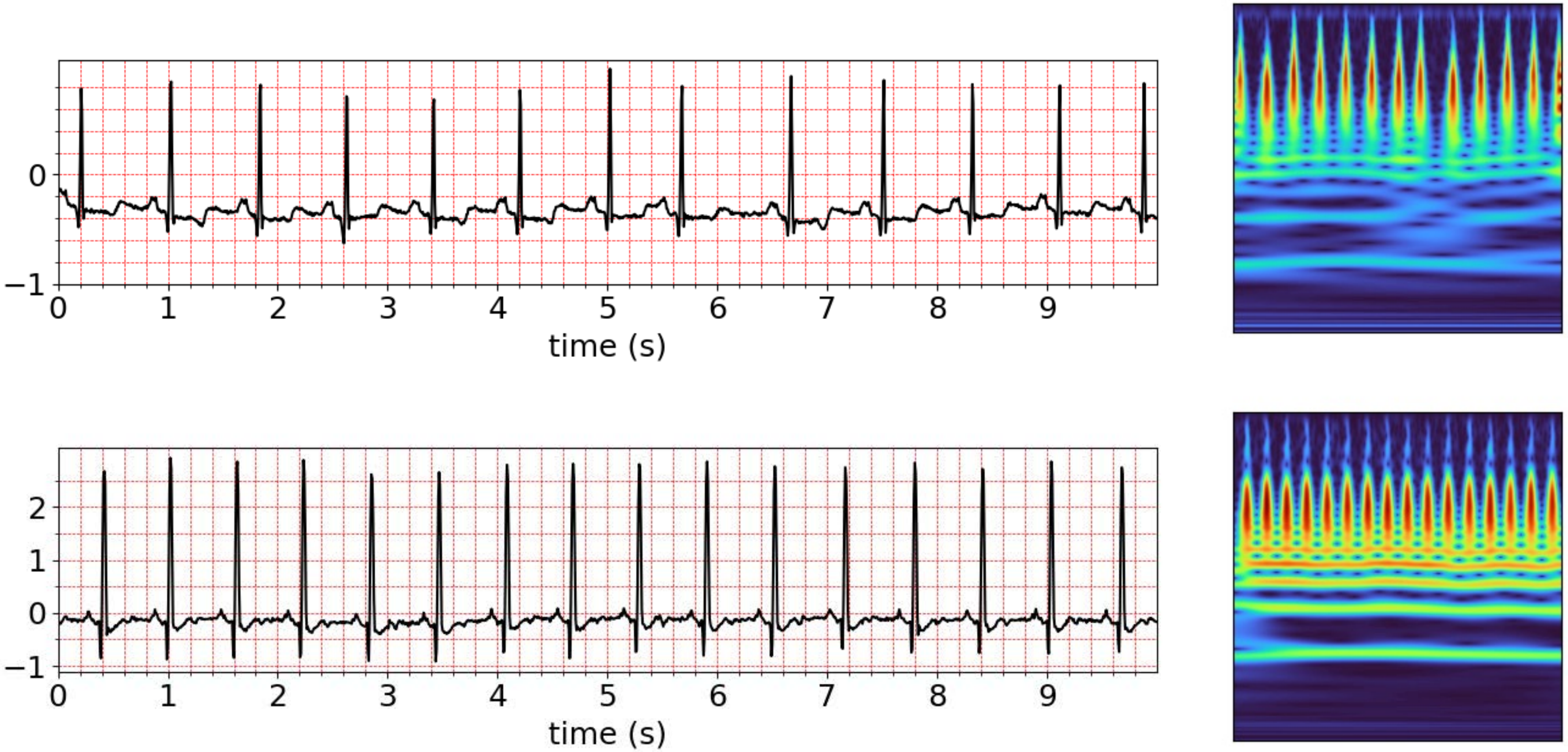
Limited resources of labeled *sensor* data

Annotating

vs.



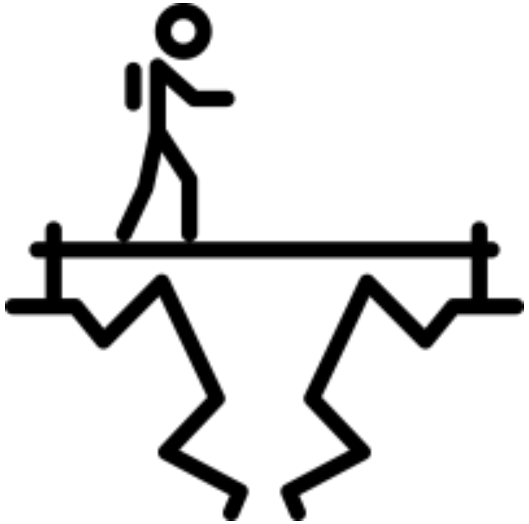
Machine learning traditionally requires handcrafted feature engineering



Which of the two ECG sensor sample representation shows someone with cardiac arrythmia?

Deldari, S., Xue, H., Saeed, A., He, J., Smith, D. V., & Salim, F. D. (2022). Beyond Just Vision: A Review on Self-Supervised Representation Learning on Multimodal and Temporal Data. *arXiv preprint arXiv:2206.02353*.

Challenges with real-world timeseries and spatiotemporal data



Irregularity

Multi-resolution

Missing Values

Multi-sources

Unlabelled

Annotation of big sensor data is often infeasible

Annotation of big sensor data is ***infeasible***:

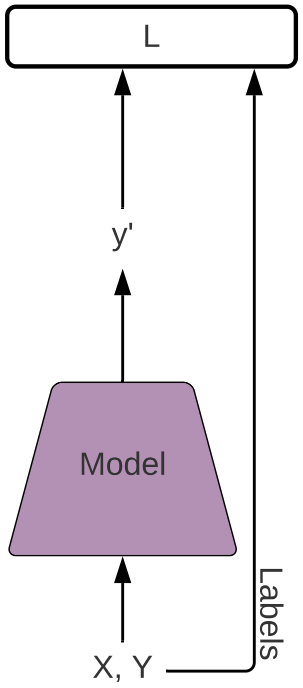
- Challenging
- Time-consuming
- Expensive
- Inaccurate in some cases

Supervised vs self-supervised

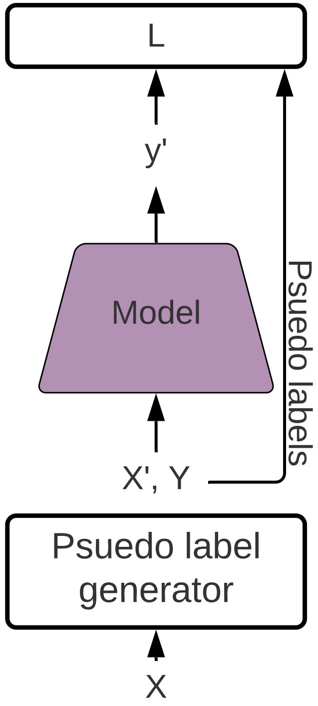
Self-supervised learning doesn’t require labels.

It obtains supervisory signals from the data itself by learning a pretext task

Supervised Model

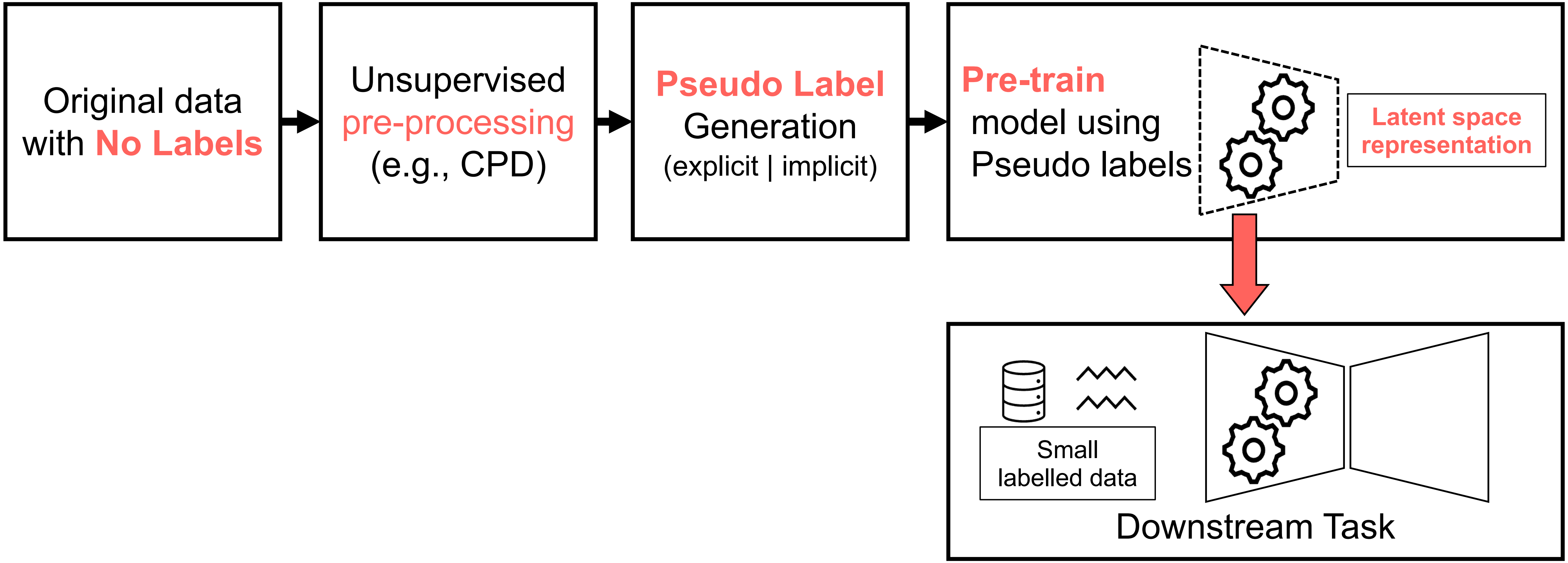


Self-Supervised Model



Deldari, S., Xue, H., Saeed, A., He, J., Smith, D. V., & Salim, F. D. (2022). Beyond Just Vision: A Review on Self-Supervised Representation Learning on Multimodal and Temporal Data. *arXiv preprint arXiv:2206.02353*.

Self-supervised Learning Framework



Original data with **No Labels**

Unsupervised pre-processing (e.g., CPD)

Pseudo Label Generation (explicit | implicit)

Pre-train model using Pseudo labels

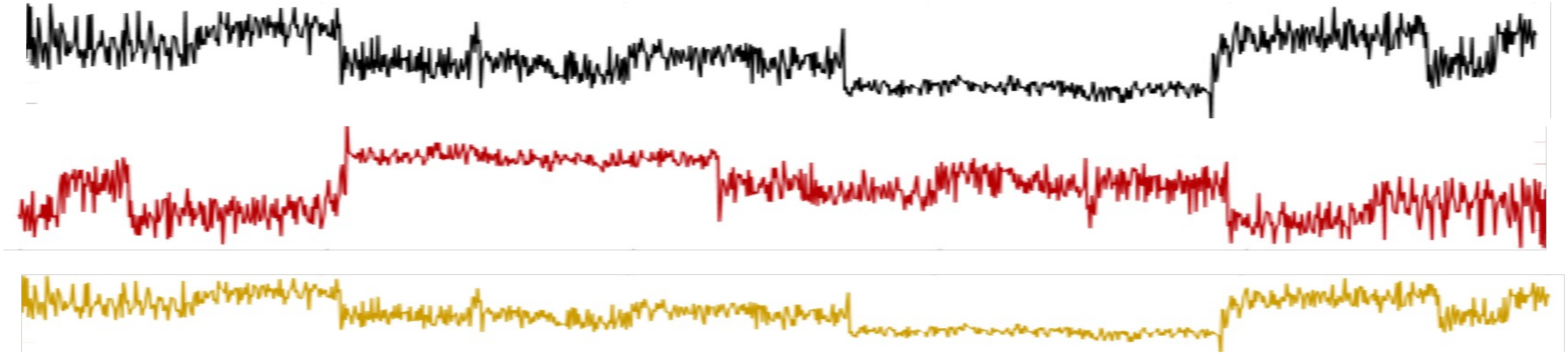
Latent space representation

Small labelled data

Downstream Task

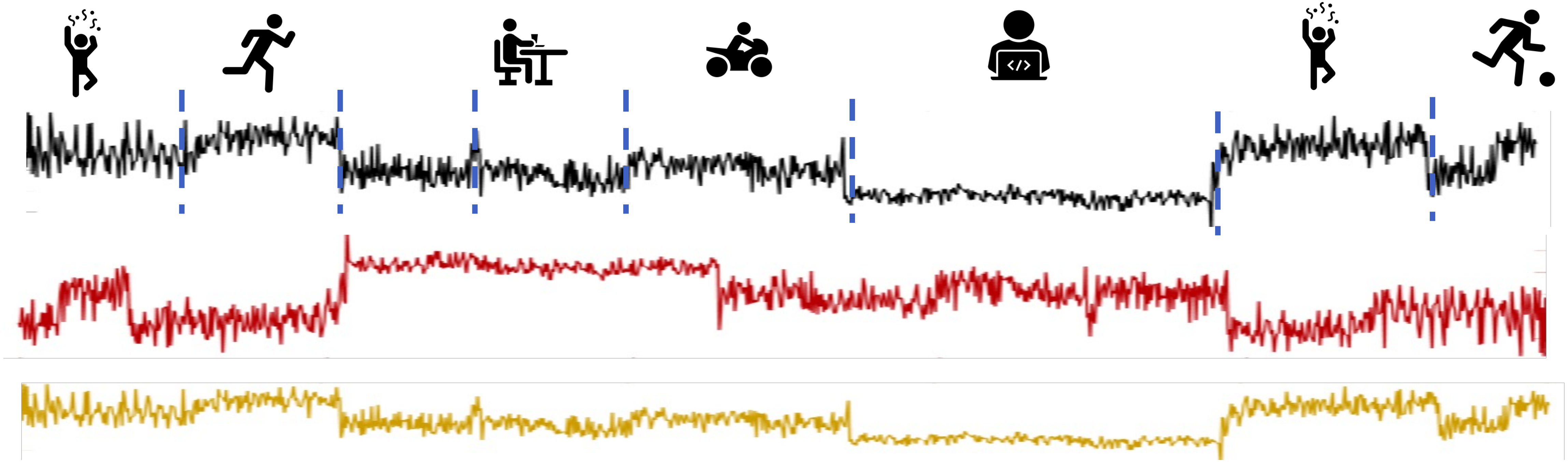
No label, No boundary

Detects constituent segments and change points within time series.



No label, No boundary

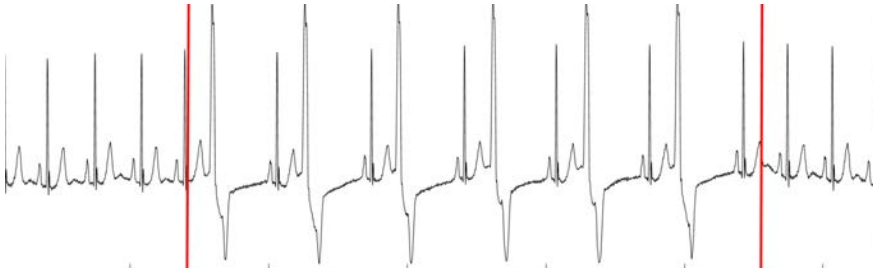
Detects constituent segments and change points within time series.



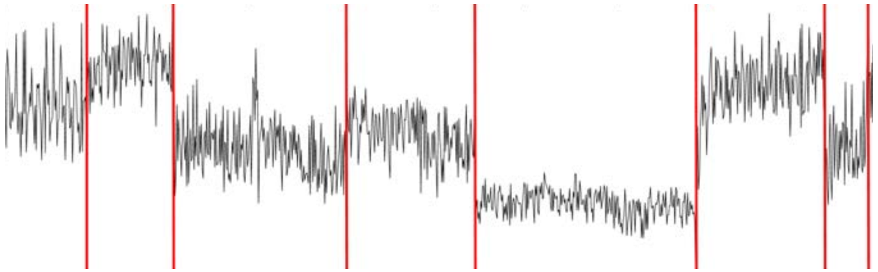
Various types of changes in sensor data

We categorize the input data based on whether they show:

Repetitive / non-repetitive temporal shape



Gradual / sudden changes



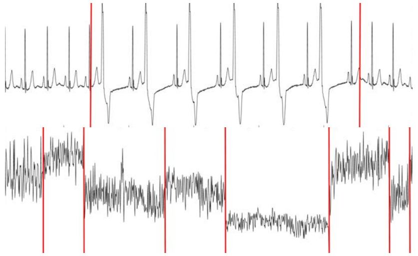
- Combination of those

Categories	Gradual changes	Sudden changes
Repetitive patterns	1. Hand Gesture	2. PAMAP, 3. RFID, 4. USC-HAD
Non-Repetitive patterns	5. EYE, 6. Emotion	7. WESAD

ESPRESSO:
Entropy and **S**ha**P**e awa**R**e tim**E**-**S**eries **S**egmentati**O**n

ESPRESSO can handle:

1. Repetitive/non-repetitive temporal changes
2. Gradual/sudden changes
3. Combination of both



How can we brew the ESPRESSO?

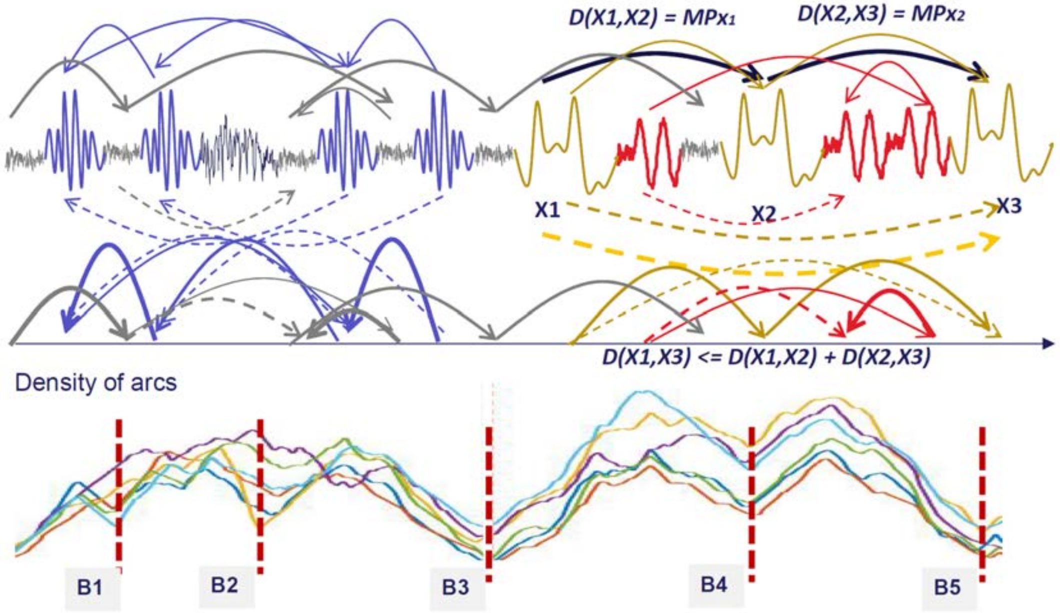
1. Extract the most similar patterns

2. Extract chain of similar patterns

3. Assign weight of arcs based on similarity and temporal distance

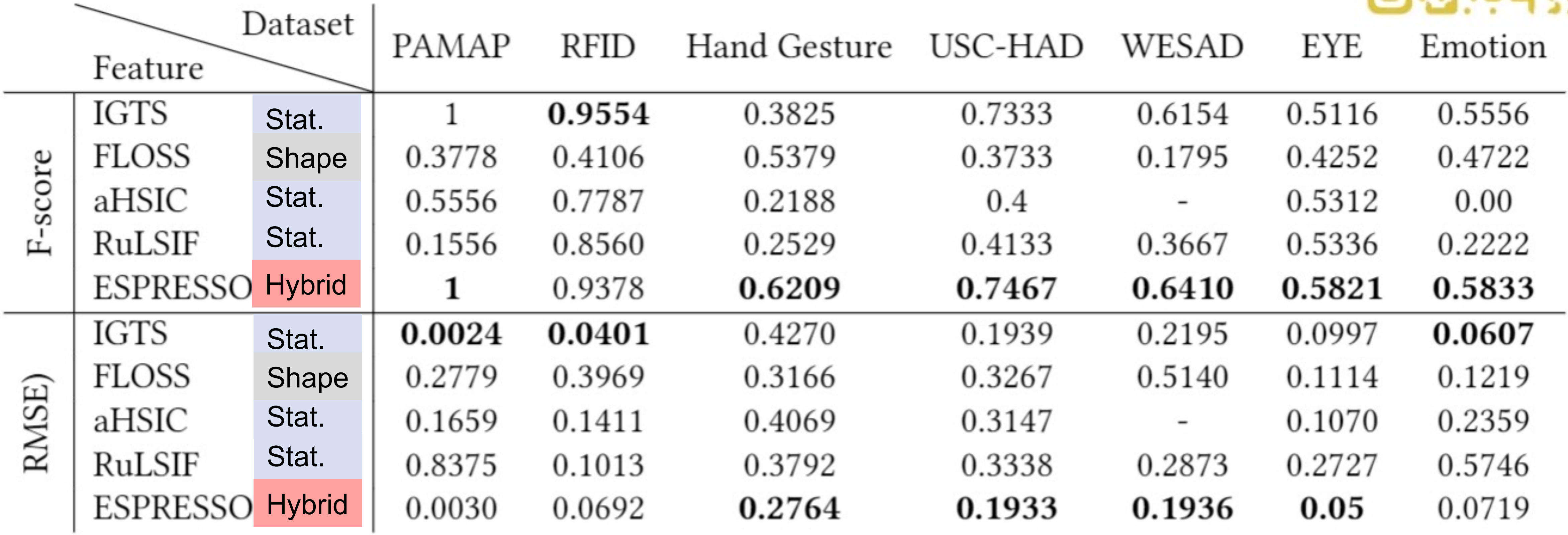
4. Extract Segment Boundary Candidates

5. Greedy search over candidates to find combination of segments with the least Entropy



ESPRESSO: Evaluation

SOTA comparison

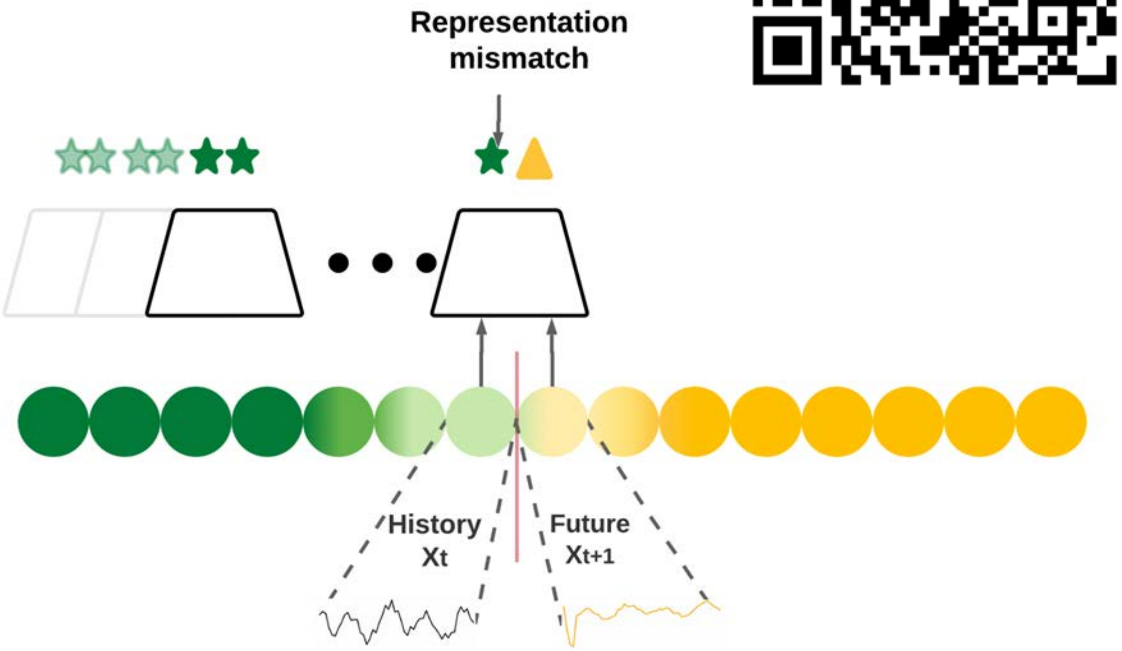


	Dataset Feature		PAMAP	RFID	Hand Gesture	USC-HAD	WESAD	EYE	Emotion
F-score	IGTS	Stat.	1	**0.9554**	0.3825	0.7333	0.6154	0.5116	0.5556
	FLOSS	Shape	0.3778	0.4106	0.5379	0.3733	0.1795	0.4252	0.4722
	aHSIC	Stat.	0.5556	0.7787	0.2188	0.4	-	0.5312	0.00
	RuLSIF	Stat.	0.1556	0.8560	0.2529	0.4133	0.3667	0.5336	0.2222
	ESPRESSO	Hybrid	**1**	0.9378	**0.6209**	**0.7467**	**0.6410**	**0.5821**	**0.5833**
RMSE)	IGTS	Stat.	**0.0024**	**0.0401**	0.4270	0.1939	0.2195	0.0997	**0.0607**
	FLOSS	Shape	0.2779	0.3969	0.3166	0.3267	0.5140	0.1114	0.1219
	aHSIC	Stat.	0.1659	0.1411	0.4069	0.3147	-	0.1070	0.2359
	RuLSIF	Stat.	0.8375	0.1013	0.3792	0.3338	0.2873	0.2727	0.5746
	ESPRESSO	Hybrid	0.0030	0.0692	**0.2764**	**0.1933**	**0.1936**	**0.05**	0.0719

TS-CP^2 : Time series Change Point Detection using Contrastive Learning

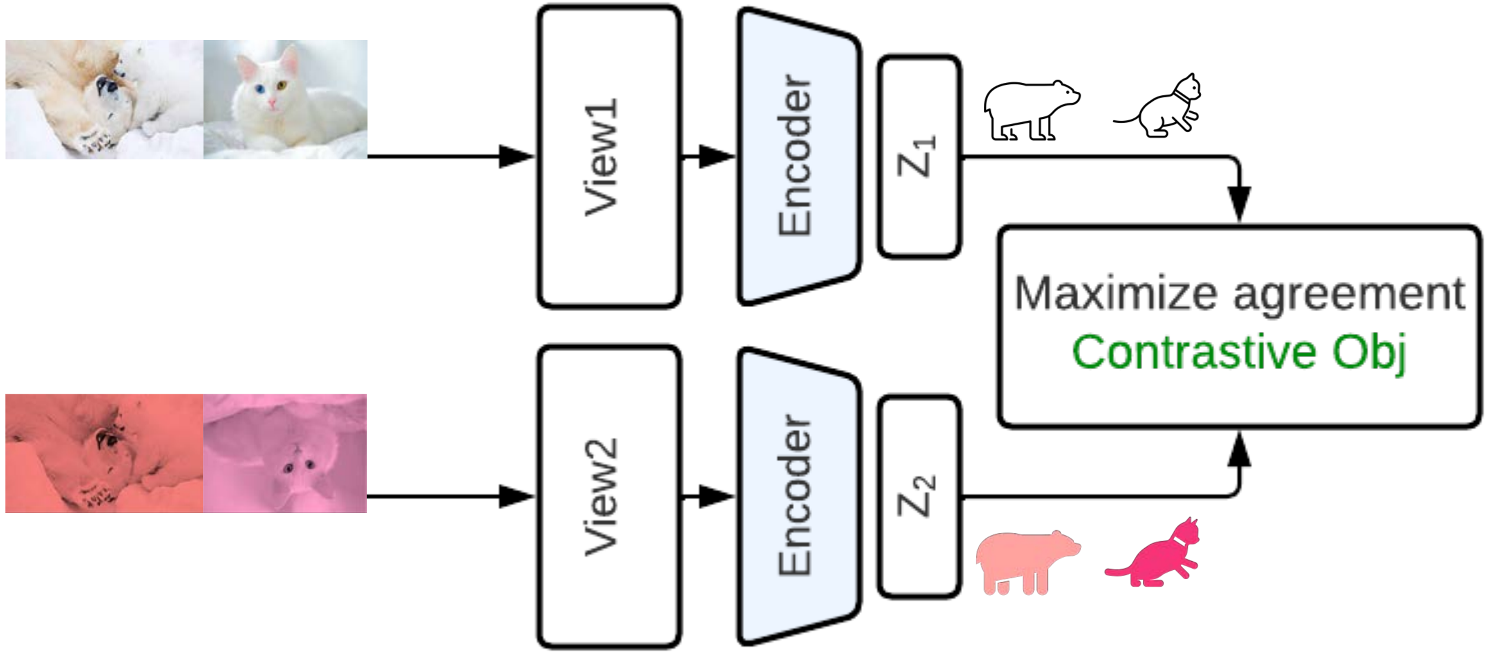
Source code on GitHub:

- We propose a multi-task model to
  1. **Learn representation** using Contrastive learning
  2. Predict future based on Contrastive Predictive Coding
- We use the representation learning task as an auxiliary task to do **Change Point Detection**.
- Also applicable for **Anomaly detection with no labelled data**



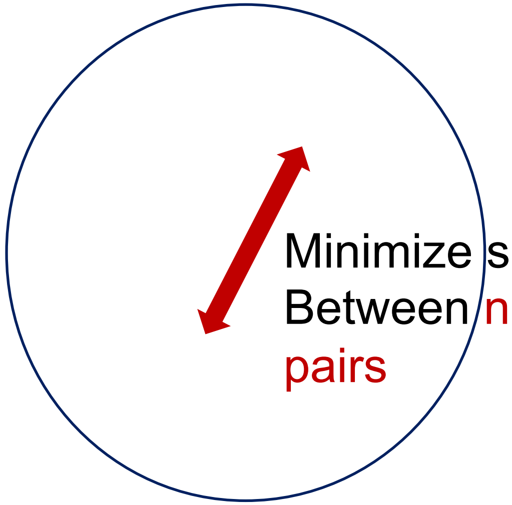
Deldari, S., Smith, D. V., Xue, H., & Salim, F. D. (2021, April). Time Series Change Point Detection with Self-Supervised Contrastive Predictive Coding. In Proceedings of the Web Conference 2021 (WWW 2021) pp. 3124-3135 .

Self-Supervised Contrastive Learning



Maximize agreement
Contrastive Obj

Maximize similarity
Between Positive pairs

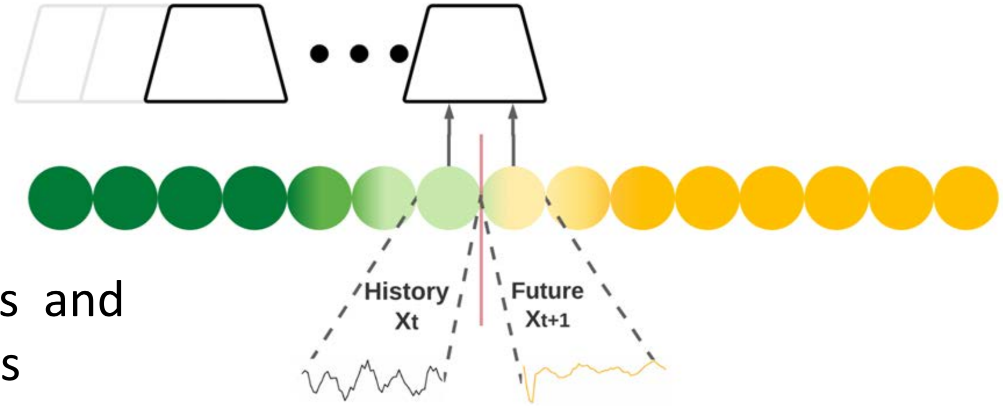


Minimize similarity
Between negative pairs

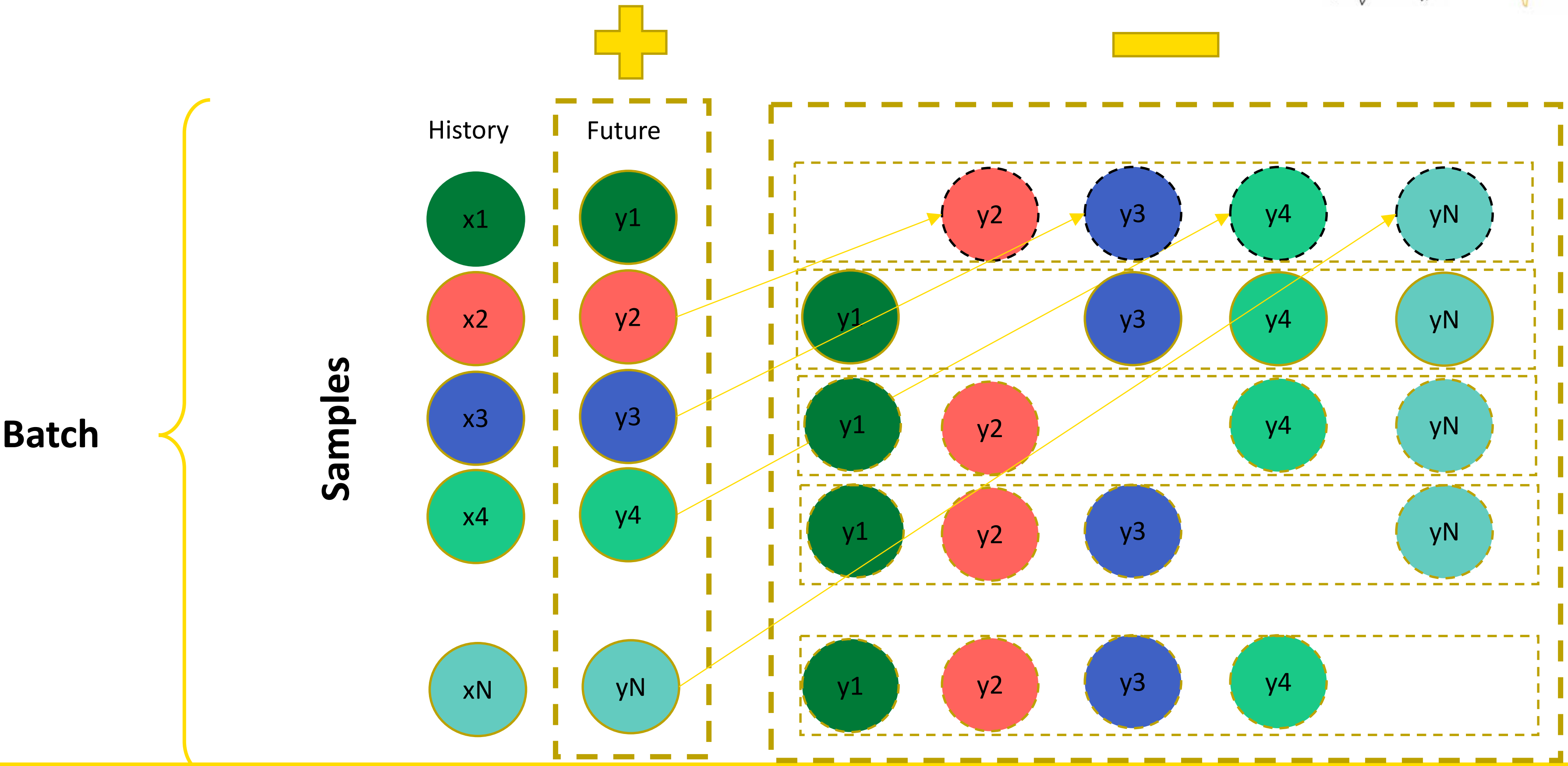
Representation Space

Chen, T., Kornblith, S., Norouzi, M. and Hinton, G., 2020, November. A simple framework for contrastive learning of visual representations. In *International conference on machine learning.*

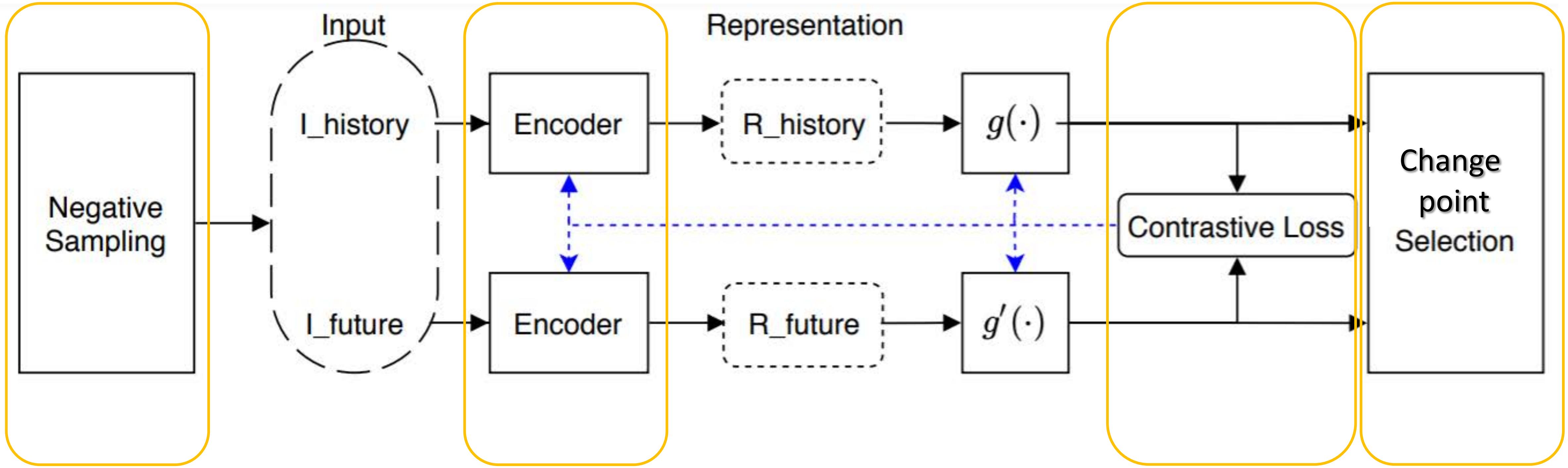
Negative Sampling



Contrastive learning: Maximize agreement between positive pairs and
Minimize agreement between negative pairs



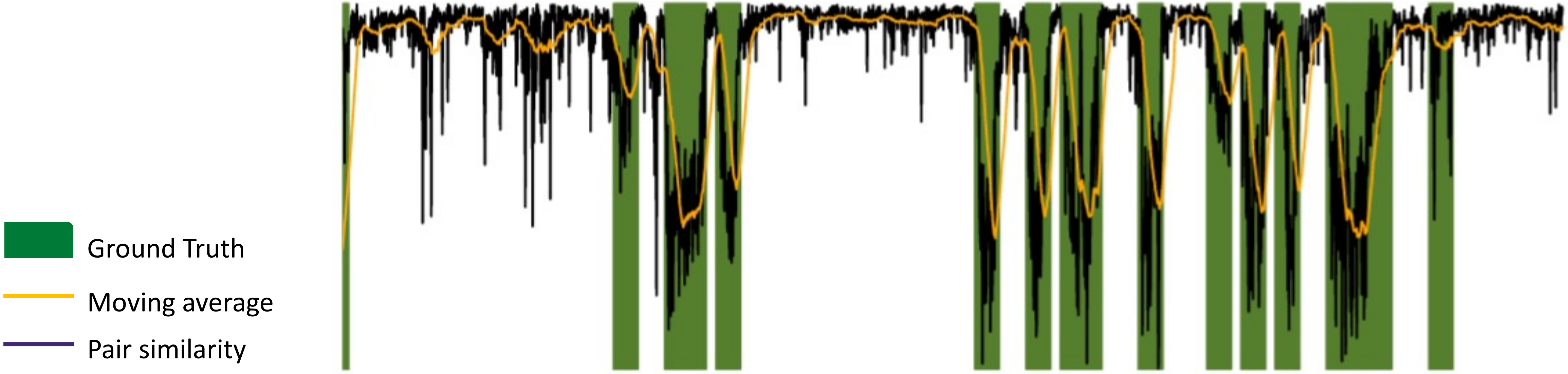
TS-CP^2 : Framework



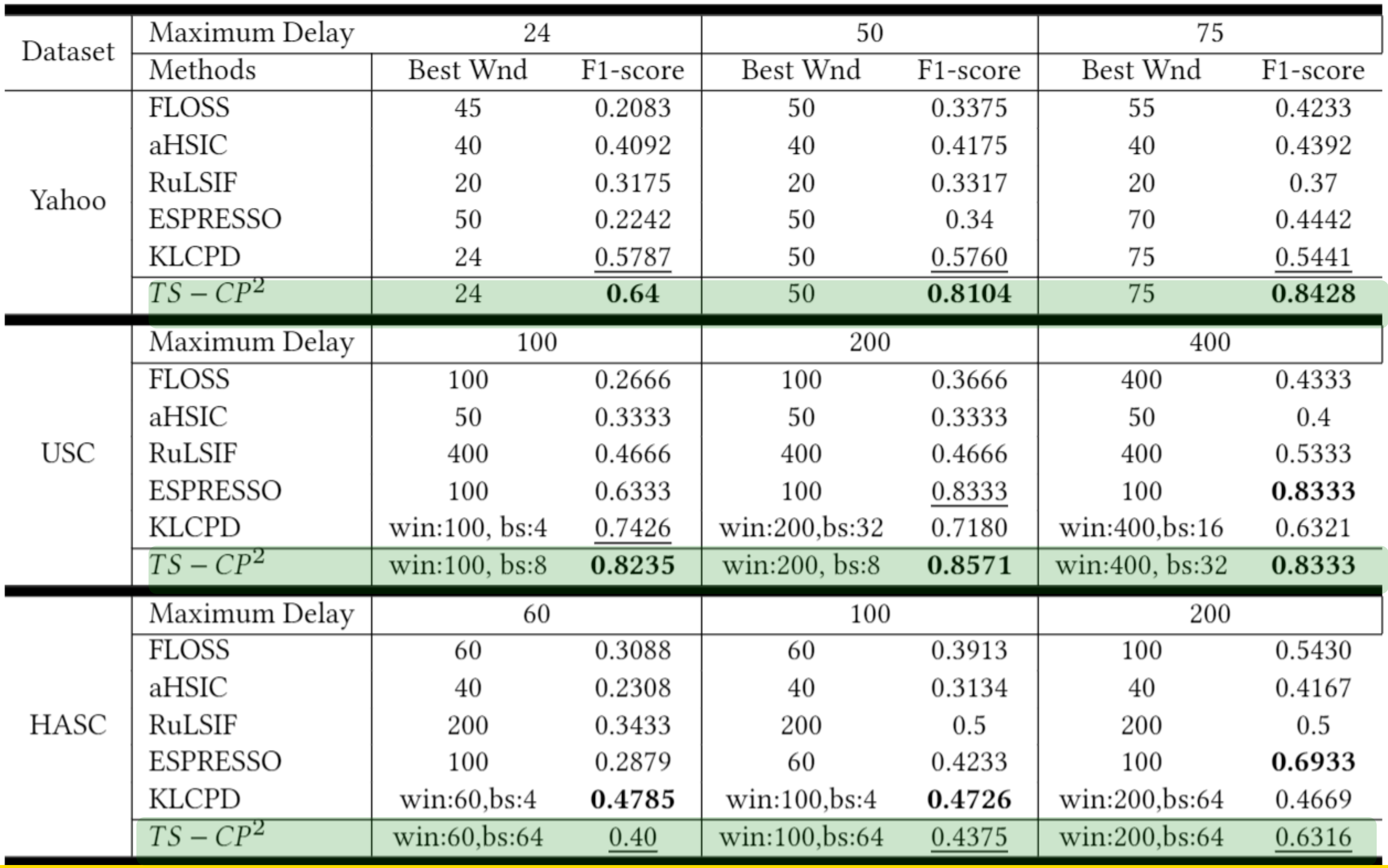
Evaluate the representations: Change Point Detection

- We hypothesize that when a change point intersects the embeddings of the history and future intervals, they will be distributed differently.

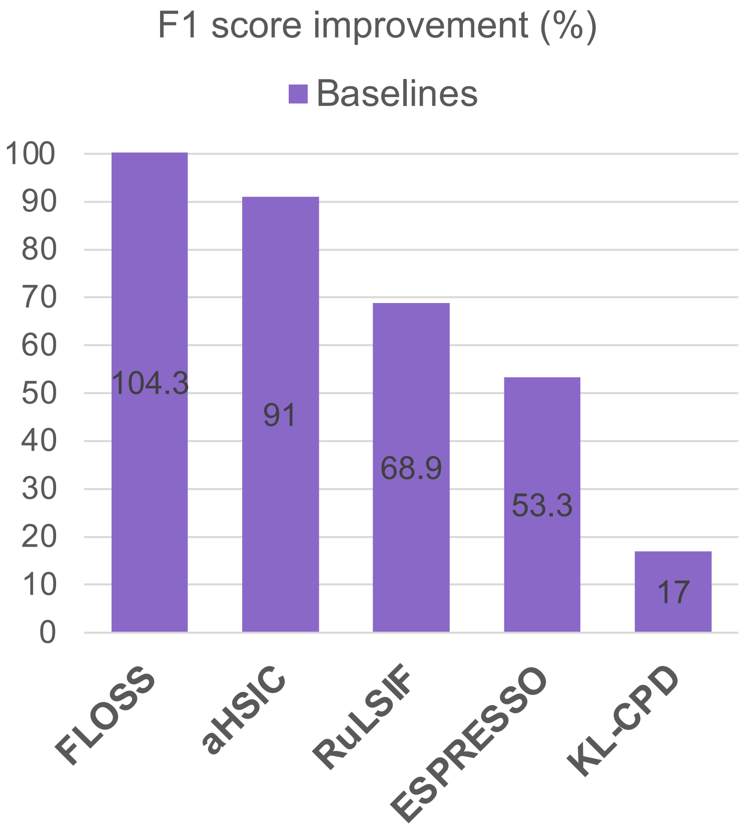
- To detect change points we utilise the cosine similarity between the latent embeddings of the history and future intervals



Evaluation



Dataset	Maximum Delay	24		50		75	
	Methods	Best Wnd	F1-score	Best Wnd	F1-score	Best Wnd	F1-score
Yahoo	FLOSS	45	0.2083	50	0.3375	55	0.4233
	aHSIC	40	0.4092	40	0.4175	40	0.4392
	RuLSIF	20	0.3175	20	0.3317	20	0.37
	ESPRESSO	50	0.2242	50	0.34	70	0.4442
	KLCPD	24	0.5787	50	0.5760	75	0.5441
	$TS - CP^2$	24	**0.64**	50	**0.8104**	75	**0.8428**
USC	Maximum Delay	100		200		400	
	FLOSS	100	0.2666	100	0.3666	400	0.4333
	aHSIC	50	0.3333	50	0.3333	50	0.4
	RuLSIF	400	0.4666	400	0.4666	400	0.5333
	ESPRESSO	100	0.6333	100	0.8333	100	**0.8333**
	KLCPD	win:100, bs:4	0.7426	win:200,bs:32	0.7180	win:400,bs:16	0.6321
	$TS - CP^2$	win:100, bs:8	**0.8235**	win:200, bs:8	**0.8571**	win:400, bs:32	**0.8333**
HASC	Maximum Delay	60		100		200	
	FLOSS	60	0.3088	60	0.3913	100	0.5430
	aHSIC	40	0.2308	40	0.3134	40	0.4167
	RuLSIF	200	0.3433	200	0.5	200	0.5
	ESPRESSO	100	0.2879	60	0.4233	100	**0.6933**
	KLCPD	win:60,bs:4	**0.4785**	win:100,bs:4	**0.4726**	win:200,bs:64	0.4669
	$TS - CP^2$	win:60,bs:64	0.40	win:100,bs:64	0.4375	win:200,bs:64	0.6316



Source code on GitHub:

$TS - CP^2$

Deldari, S., Smith, D. V., Xue, H., & Salim, F. D. (2021, April). Time Series Change Point Detection with Self-Supervised Contrastive Predictive Coding. In Proceedings of the Web Conference 2021 (WWW 2021) pp. 3124-3135 .

Exploring Self-Supervised Representation Ensembles for COVID-19 Cough Classification

Fully-supervised based classification methods inevitably need to rely on well-annotated cough sounds data

Not easy to get annotated labels (especially for COVID-19 sounds)

- *Manually annotation*: require experts with medical expertise, hard to label large-scale datasets
- *Survey*: asking participants to report their health status -> privacy concern about sensitive medical data -> limited access of datasets

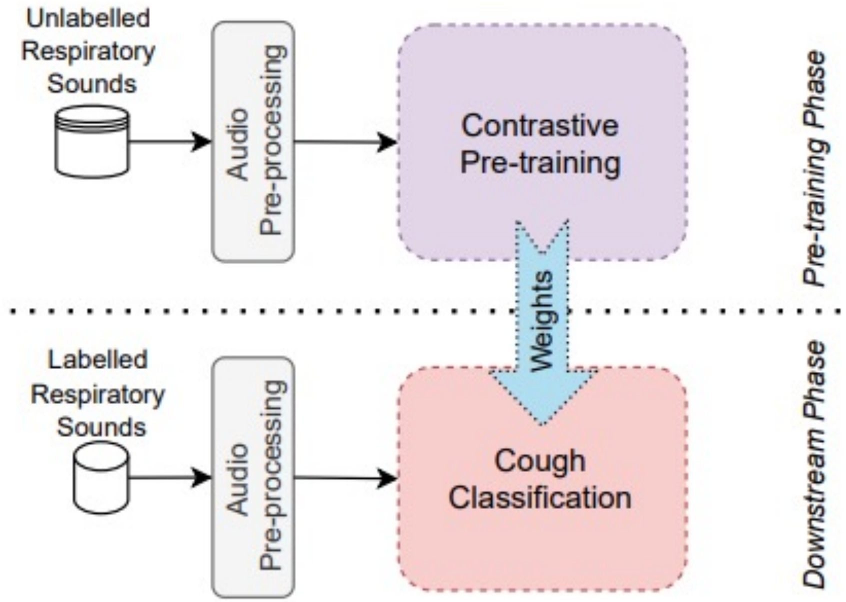
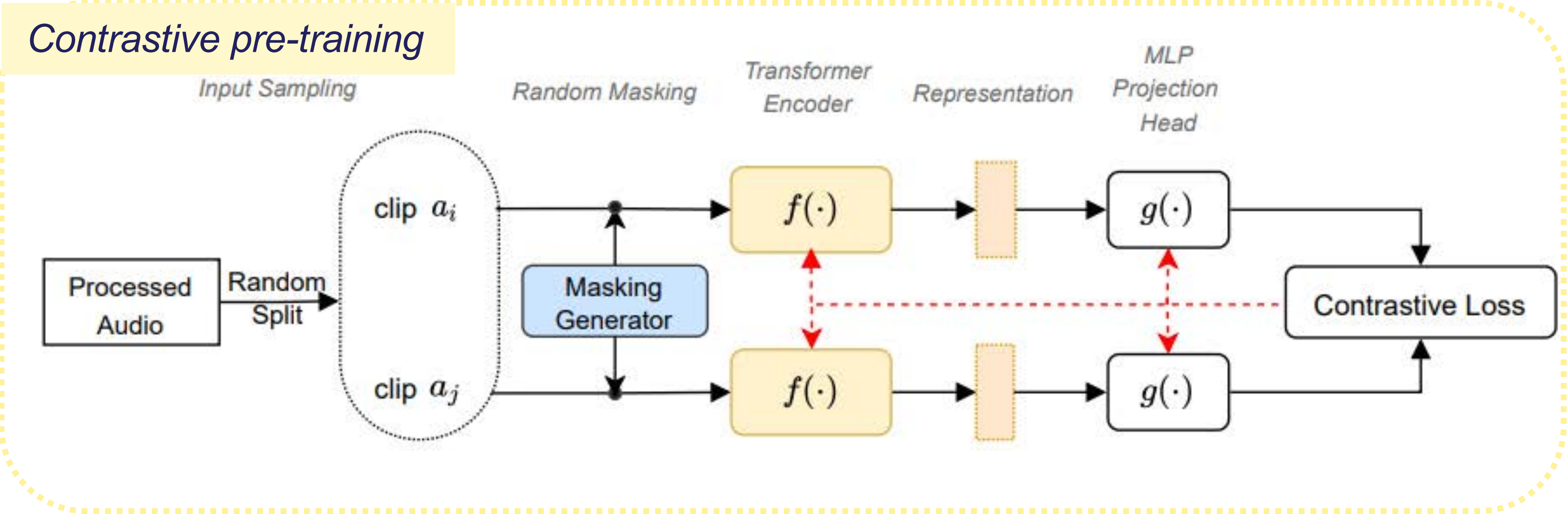
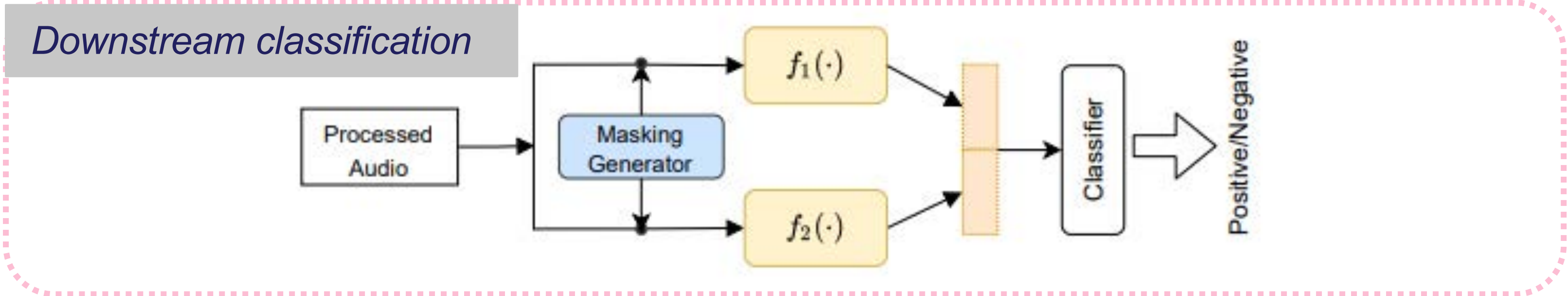


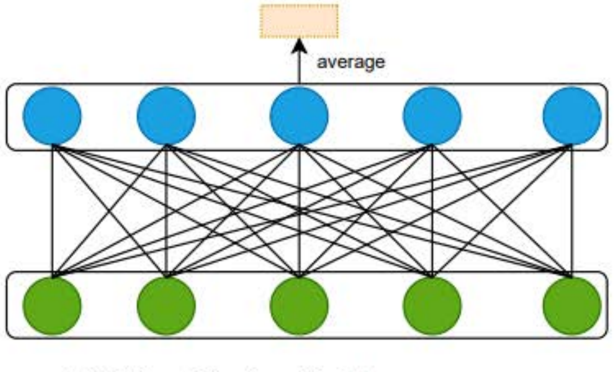
Figure 1: Concept illustration of the proposed framework.

Patent: Hao Xue, Flora D. Salim. Method and systems for respiratory sound classification. (Australian Patent 2022224848, Worldwide Patent WO2023015361A1)

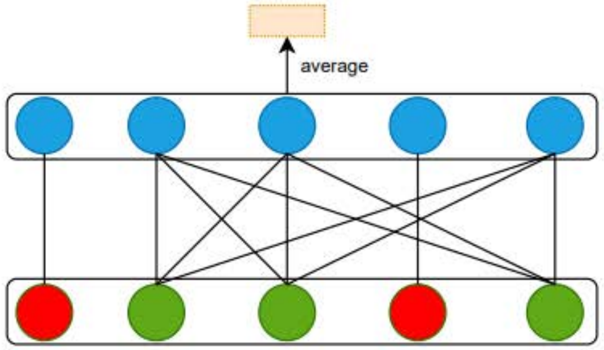
Publication: Xue, H. and Salim, F.D., 2021, August. Exploring self-supervised representation ensembles for COVID-19 cough classification. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining.







(a) Without Random Masking



(b) With Random Masking

Contrastive learning-based pre-training phase with unlabeled data

Ensembled architecture for downstream classification fine-tuning with labelled data

Introduce random masking generator in both phases to further improve the performance

X. Hao, F. D. Salim, Exploring Self-Supervised Representation Ensembles for COVID-19 Cough Classification, ACM KDD 2021

Pretrained on IIS-Coswara dataset, Tested on Cambridge COVID-sounds dataset

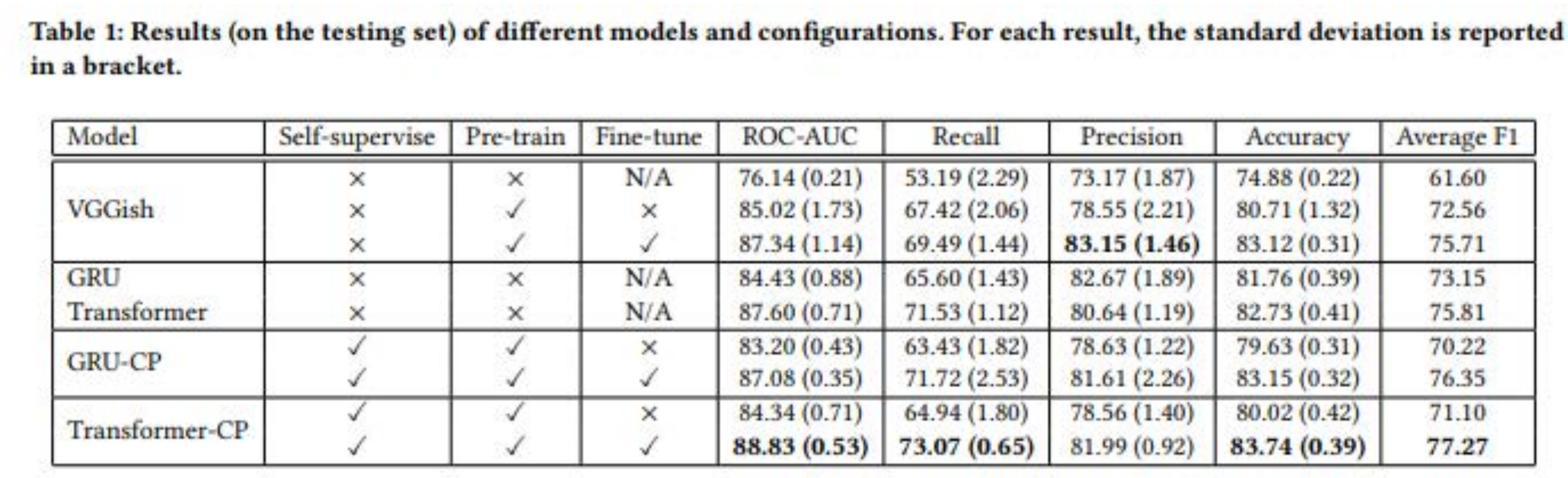


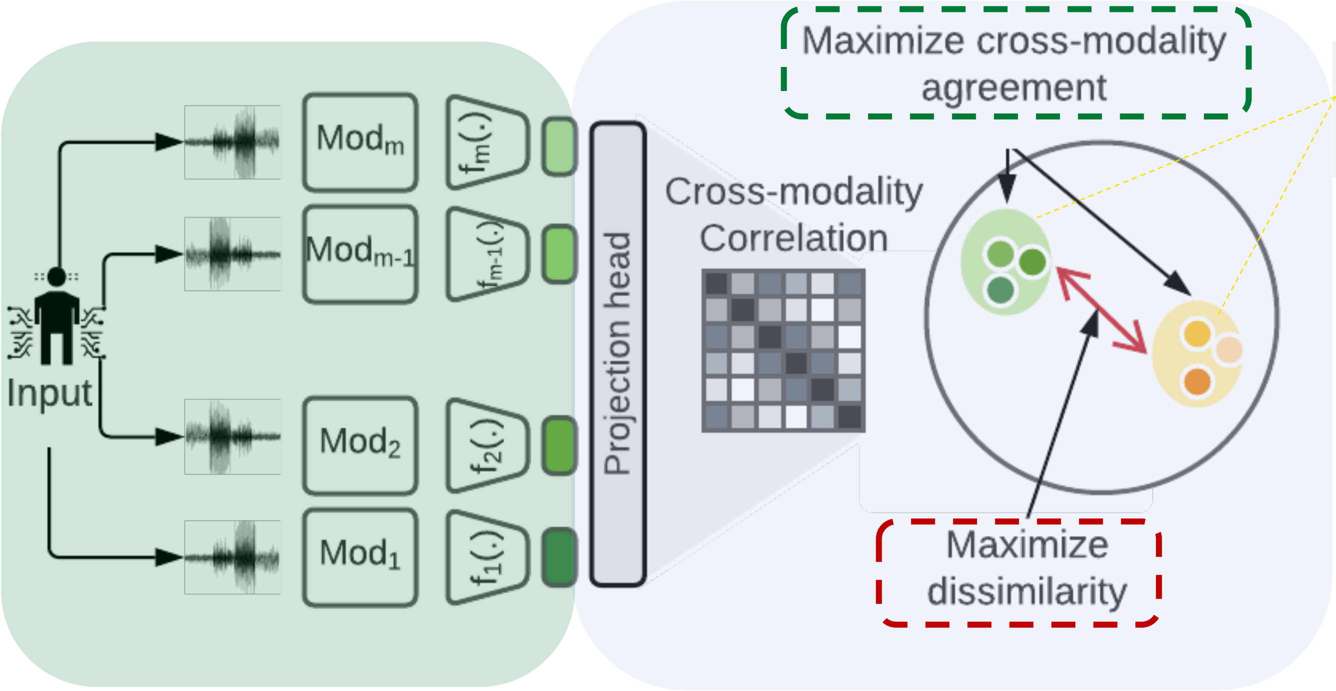
Table 1: Results (on the testing set) of different models and configurations. For each result, the standard deviation is reported in a bracket.

Model	Self-supervise	Pre-train	Fine-tune	ROC-AUC	Recall	Precision	Accuracy	Average F1
VGGish	×	×	N/A	76.14 (0.21)	53.19 (2.29)	73.17 (1.87)	74.88 (0.22)	61.60
	×	✓	×	85.02 (1.73)	67.42 (2.06)	78.55 (2.21)	80.71 (1.32)	72.56
	×	✓	✓	87.34 (1.14)	69.49 (1.44)	**83.15 (1.46)**	83.12 (0.31)	75.71
GRU	×	×	N/A	84.43 (0.88)	65.60 (1.43)	82.67 (1.89)	81.76 (0.39)	73.15
Transformer	×	×	N/A	87.60 (0.71)	71.53 (1.12)	80.64 (1.19)	82.73 (0.41)	75.81
GRU-CP	✓	✓	×	83.20 (0.43)	63.43 (1.82)	78.63 (1.22)	79.63 (0.31)	70.22
	✓	✓	✓	87.08 (0.35)	71.72 (2.53)	81.61 (2.26)	83.15 (0.32)	76.35
Transformer-CP	✓	✓	×	84.34 (0.71)	64.94 (1.80)	78.56 (1.40)	80.02 (0.42)	71.10
	✓	✓	✓	**88.83 (0.53)**	**73.07 (0.65)**	81.99 (0.92)	**83.74 (0.39)**	**77.27**

X. Hao, F. D. Salim, Exploring Self-Supervised Representation Ensembles for COVID-19 Cough Classification, ACM KDD 2021

COCOA

Cross mOdality COntrastive leArning for Sensor Data



Representations of Multiple modalities from a single sample

$$\mathcal{L}_{positive} = \sum_{v \neq w} 1 - Sim\,(z_v, z_w)$$

$$\mathcal{L}_{negative} = \sum_{t \neq t'} Sim\,(z_t, z_{t'})$$

$$\mathcal{L}_{\mathrm{COCOA}} = \sum_{t} \mathcal{L}^{t}_{pos} + \sum_{v} \mathcal{L}^{v}_{neg}$$

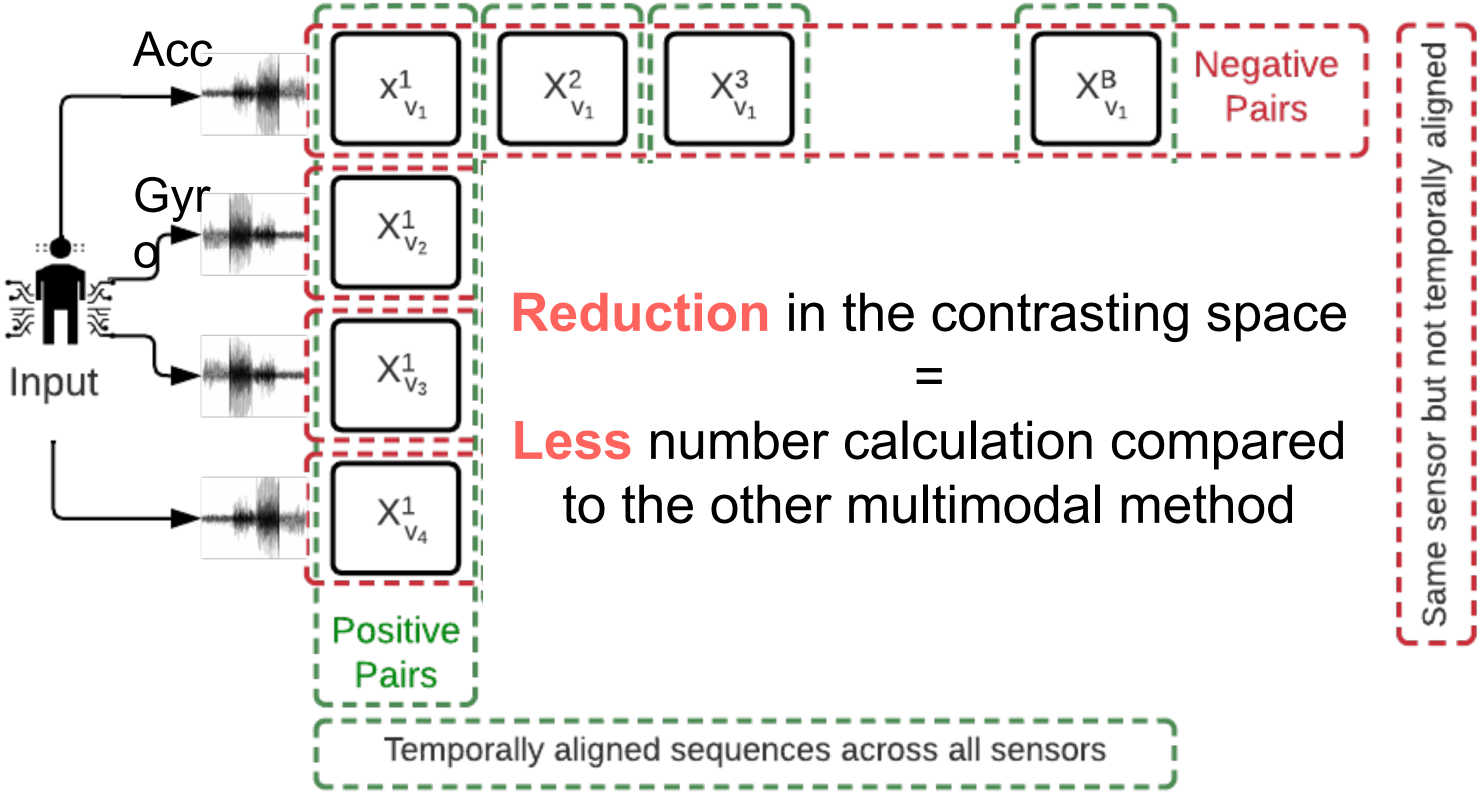
Deldari, Shohreh, Hao Xue, Aaqib Saeed, Daniel V. Smith, and Flora D. Salim. "COCOA: Cross Modality Contrastive Learning for Sensor Data." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, no. 3 (2022): 1-28.

COCOA

Cross M**O**dality **CO**ntrastive Le**A**rning for Sensor Data

- ✓ COCOA is the **first** to apply contrastive learning for **multiple sensors** of different modalities.
- ✓ COCOA has linear complexity in terms of the **number of sensors**.
- ✓ COCOA is highly **label efficient**.
- ✓ COCOA performs well with **smaller batch sizes.**

Positive and Negative Pair Sampling



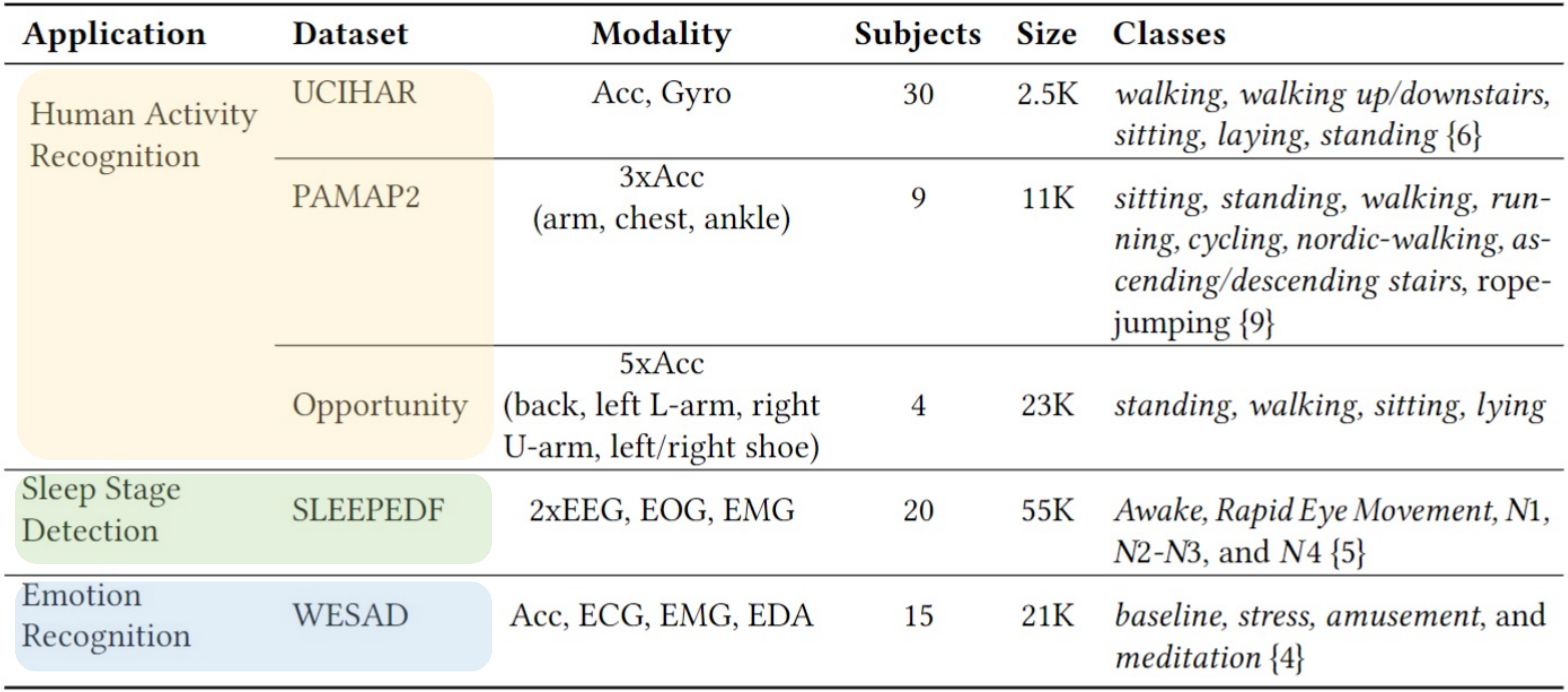
Reduction in the contrasting space
=
Less number calculation compared to the other multimodal method

Negative Pairs are selected from **the same** source. Higher similar characteristics results in **harder negatives**.

Positive Pairs are selected from **different** sources which results in **harder positives**.

COCOA

Evaluation

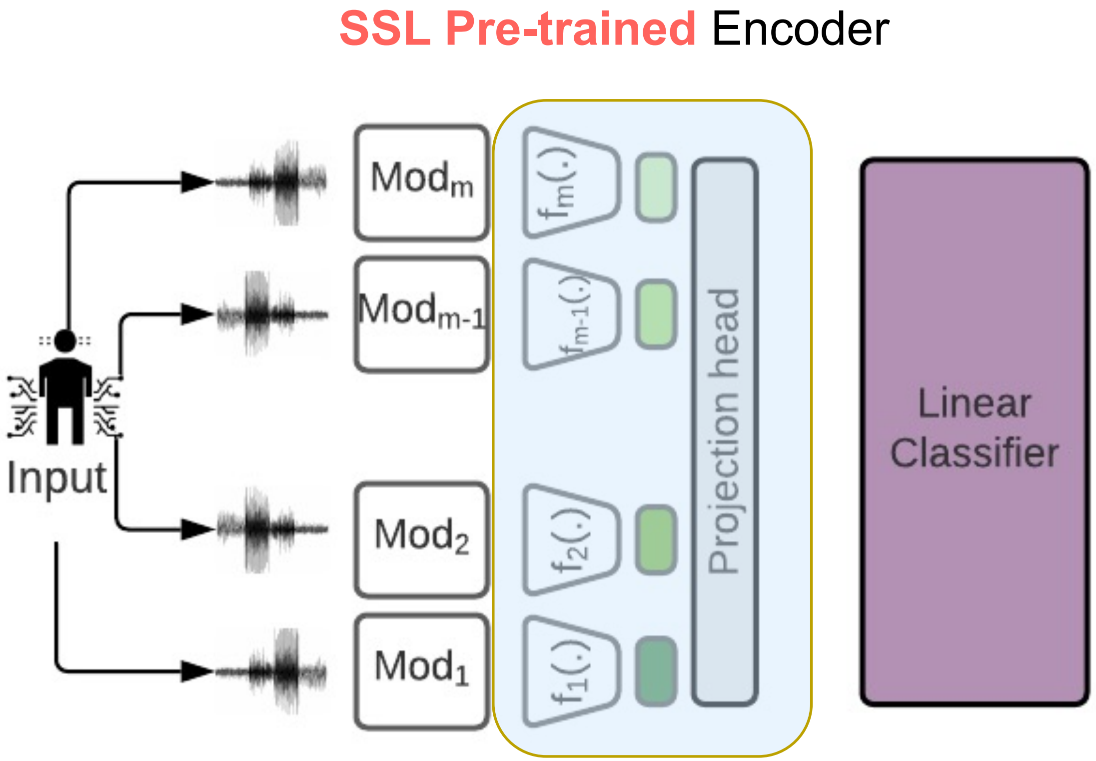


Application	Dataset	Modality	Subjects	Size	Classes
Human Activity Recognition	UCIHAR	Acc, Gyro	30	2.5K	*walking, walking up/downstairs, sitting, laying, standing* {6}
	PAMAP2	3xAcc (arm, chest, ankle)	9	11K	*sitting, standing, walking, running, cycling, nordic-walking, ascending/descending stairs*, rope-jumping {9}
	Opportunity	5xAcc (back, left L-arm, right U-arm, left/right shoe)	4	23K	*standing, walking, sitting, lying*
Sleep Stage Detection	SLEEPEDF	2xEEG, EOG, EMG	20	55K	*Awake, Rapid Eye Movement, N1, N2-N3*, and *N4* {5}
Emotion Recognition	WESAD	Acc, ECG, EMG, EDA	15	21K	*baseline, stress, amusement*, and *meditation* {4}

COCOA

Evaluation setup

We used **the same encoder** for all **10 SOTA baselines**.



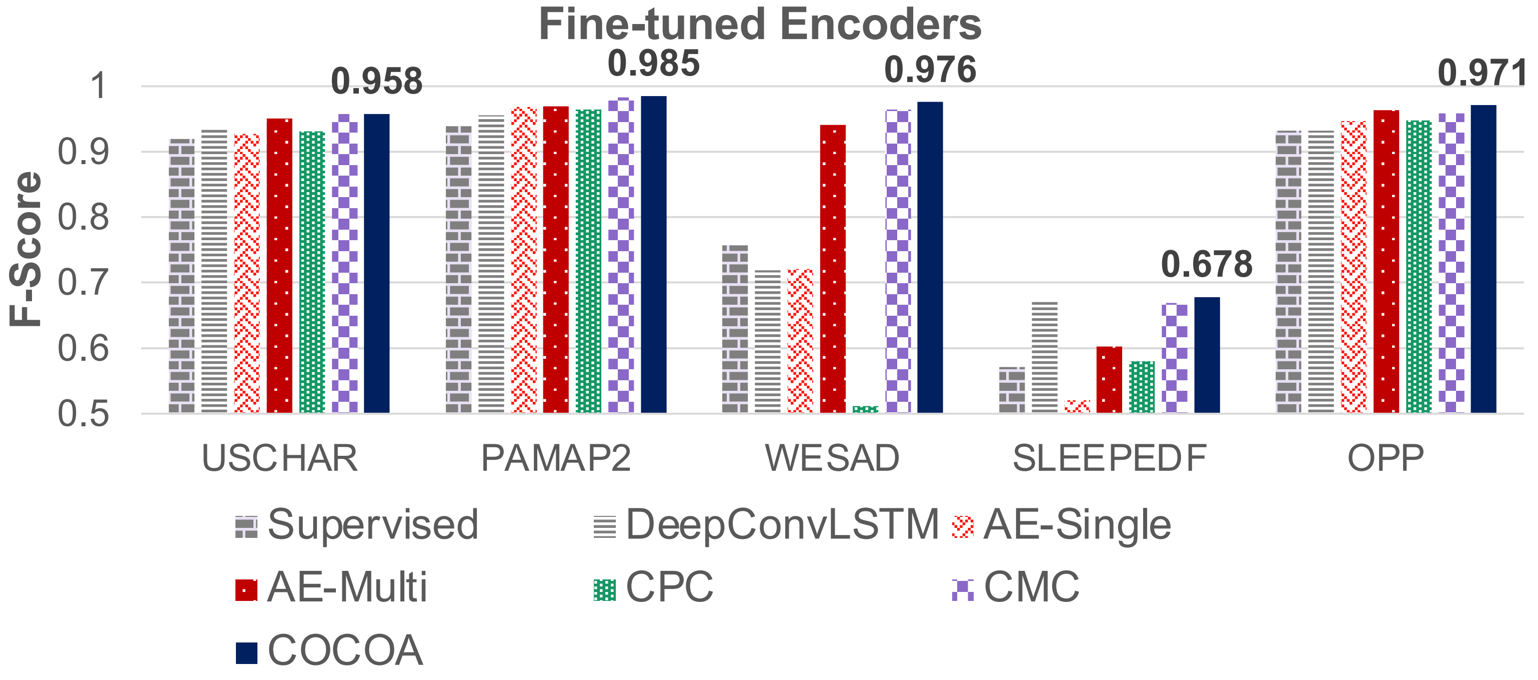
- **Fixed:** Frozen SSL pre-trained encoder + Linear classifier
- **Fine-tuned:** SSL pre-trained encoder + Linear classifier
- **Fully supervised** encoder + Linear classifier

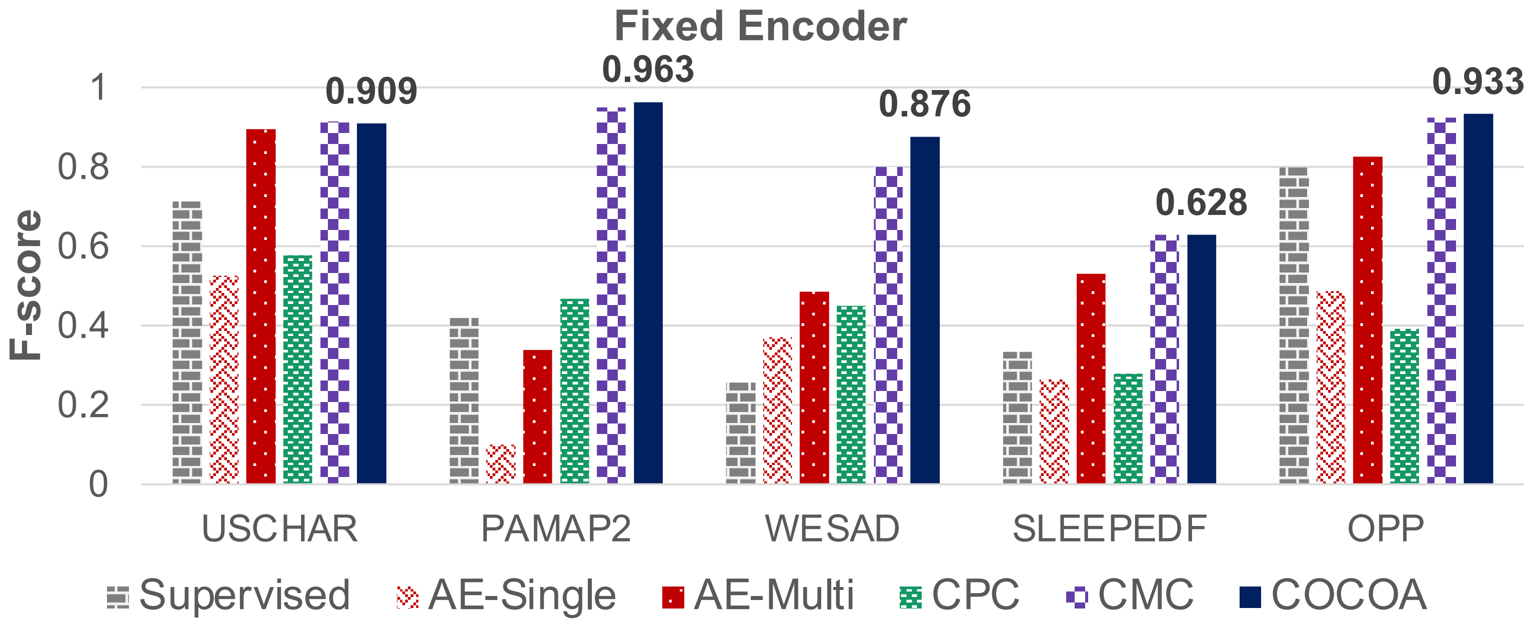
COCOA

Baseline Evaluation & Ablation

Outperform the fully supervised baseline

Less computation in COCOA compared to CMC

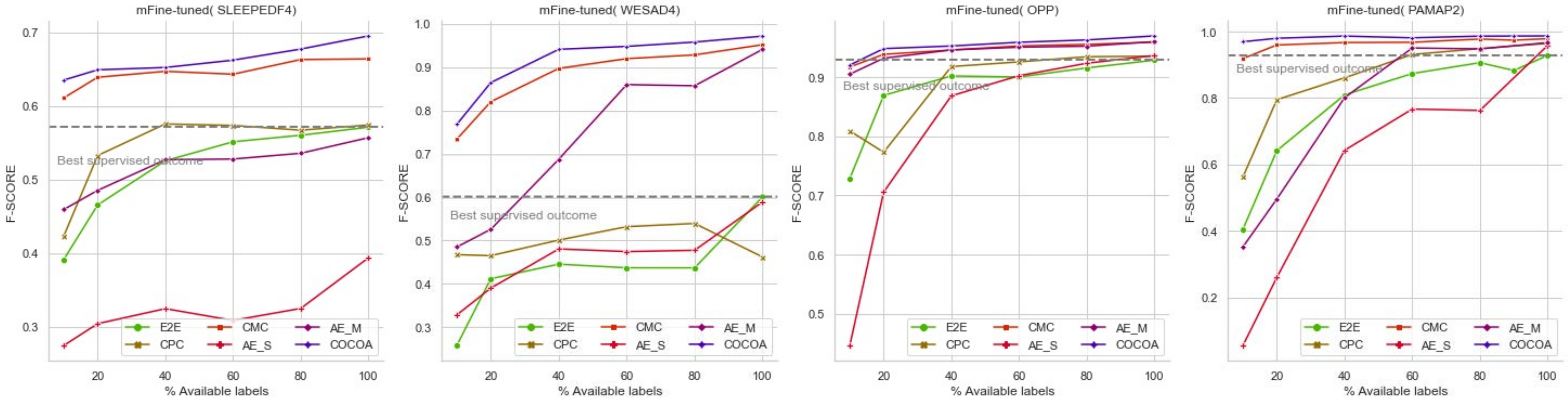


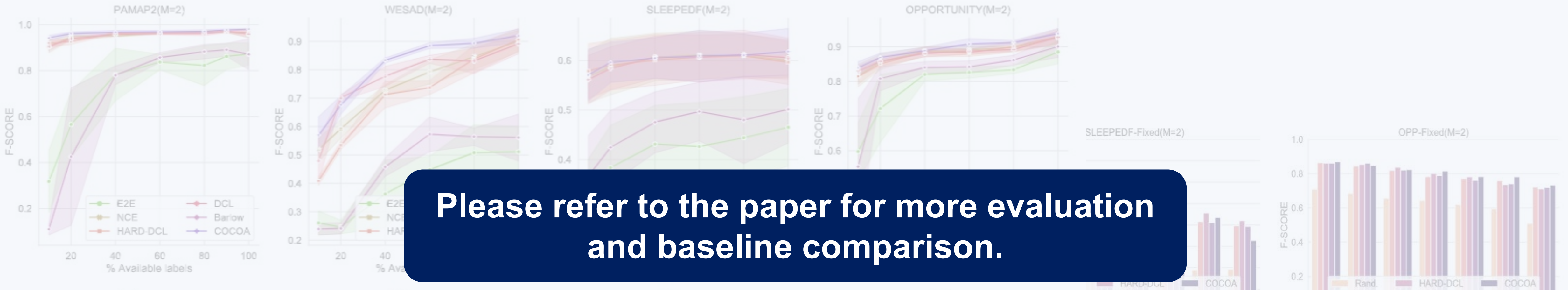


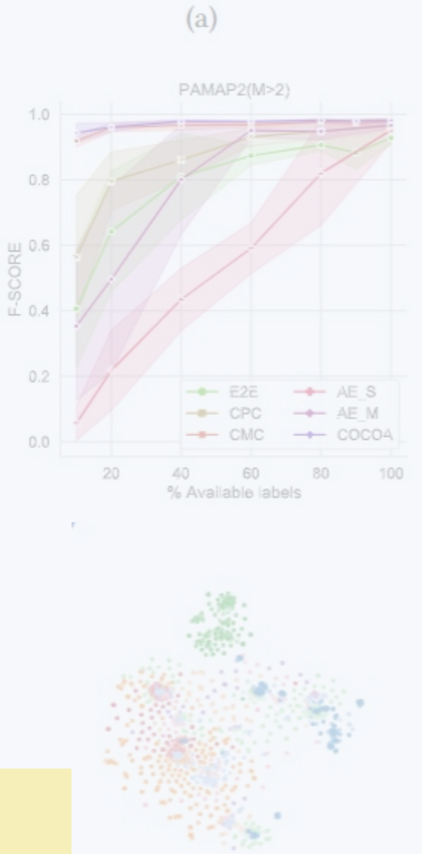
COCOA

Data Efficiency

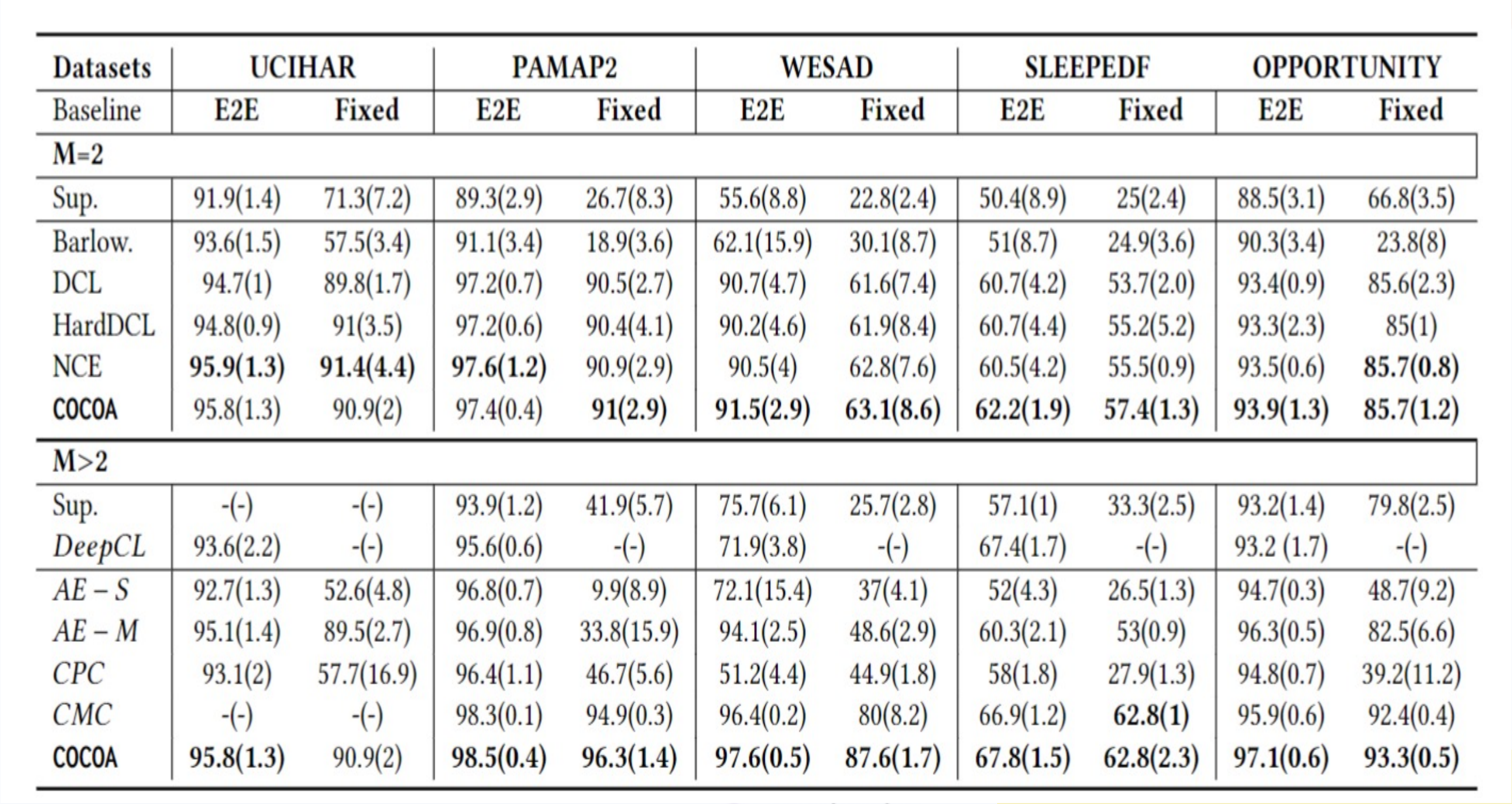
Even with 10% of labels we can outperform the fully supervised baseline with 100% labels.







Please refer to the paper for more evaluation and baseline comparison.



Datasets	UCIHAR		PAMAP2		WESAD		SLEEPEDF		OPPORTUNITY	
Baseline	E2E	Fixed	E2E	Fixed	E2E	Fixed	E2E	Fixed	E2E	Fixed
M=2										
Sup.	91.9(1.4)	71.3(7.2)	89.3(2.9)	26.7(8.3)	55.6(8.8)	22.8(2.4)	50.4(8.9)	25(2.4)	88.5(3.1)	66.8(3.5)
Barlow.	93.6(1.5)	57.5(3.4)	91.1(3.4)	18.9(3.6)	62.1(15.9)	30.1(8.7)	51(8.7)	24.9(3.6)	90.3(3.4)	23.8(8)
DCL	94.7(1)	89.8(1.7)	97.2(0.7)	90.5(2.7)	90.7(4.7)	61.6(7.4)	60.7(4.2)	53.7(2.0)	93.4(0.9)	85.6(2.3)
HardDCL	94.8(0.9)	91(3.5)	97.2(0.6)	90.4(4.1)	90.2(4.6)	61.9(8.4)	60.7(4.4)	55.2(5.2)	93.3(2.3)	85(1)
NCE	**95.9(1.3)**	**91.4(4.4)**	**97.6(1.2)**	90.9(2.9)	90.5(4)	62.8(7.6)	60.5(4.2)	55.5(0.9)	93.5(0.6)	**85.7(0.8)**
COCOA	95.8(1.3)	90.9(2)	97.4(0.4)	**91(2.9)**	**91.5(2.9)**	**63.1(8.6)**	**62.2(1.9)**	**57.4(1.3)**	**93.9(1.3)**	**85.7(1.2)**
M>2										
Sup.	-(-)	-(-)	93.9(1.2)	41.9(5.7)	75.7(6.1)	25.7(2.8)	57.1(1)	33.3(2.5)	93.2(1.4)	79.8(2.5)
DeepCL	93.6(2.2)	-(-)	95.6(0.6)	-(-)	71.9(3.8)	-(-)	67.4(1.7)	-(-)	93.2 (1.7)	-(-)
AE – S	92.7(1.3)	52.6(4.8)	96.8(0.7)	9.9(8.9)	72.1(15.4)	37(4.1)	52(4.3)	26.5(1.3)	94.7(0.3)	48.7(9.2)
AE – M	95.1(1.4)	89.5(2.7)	96.9(0.8)	33.8(15.9)	94.1(2.5)	48.6(2.9)	60.3(2.1)	53(0.9)	96.3(0.5)	82.5(6.6)
CPC	93.1(2)	57.7(16.9)	96.4(1.1)	46.7(5.6)	51.2(4.4)	44.9(1.8)	58(1.8)	27.9(1.3)	94.8(0.7)	39.2(11.2)
CMC	-(-)	-(-)	98.3(0.1)	94.9(0.3)	96.4(0.2)	80(8.2)	66.9(1.2)	**62.8(1)**	95.9(0.6)	92.4(0.4)
COCOA	**95.8(1.3)**	90.9(2)	**98.5(0.4)**	**96.3(1.4)**	**97.6(0.5)**	**87.6(1.7)**	**67.8(1.5)**	**62.8(2.3)**	**97.1(0.6)**	**93.3(0.5)**

Federated Self-Supervised Learning of Multi-Sensor Representations for Embedded Intelligence

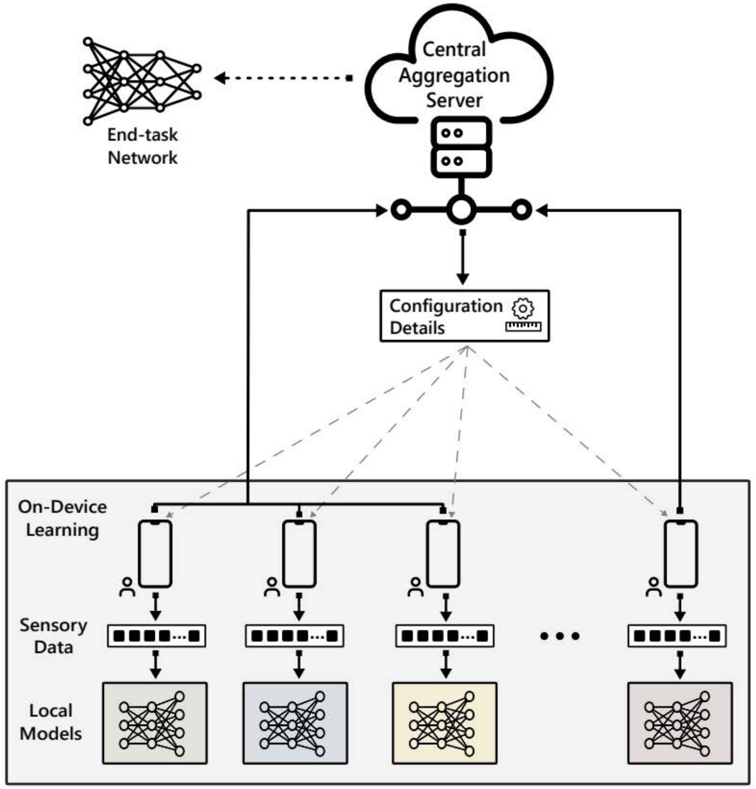
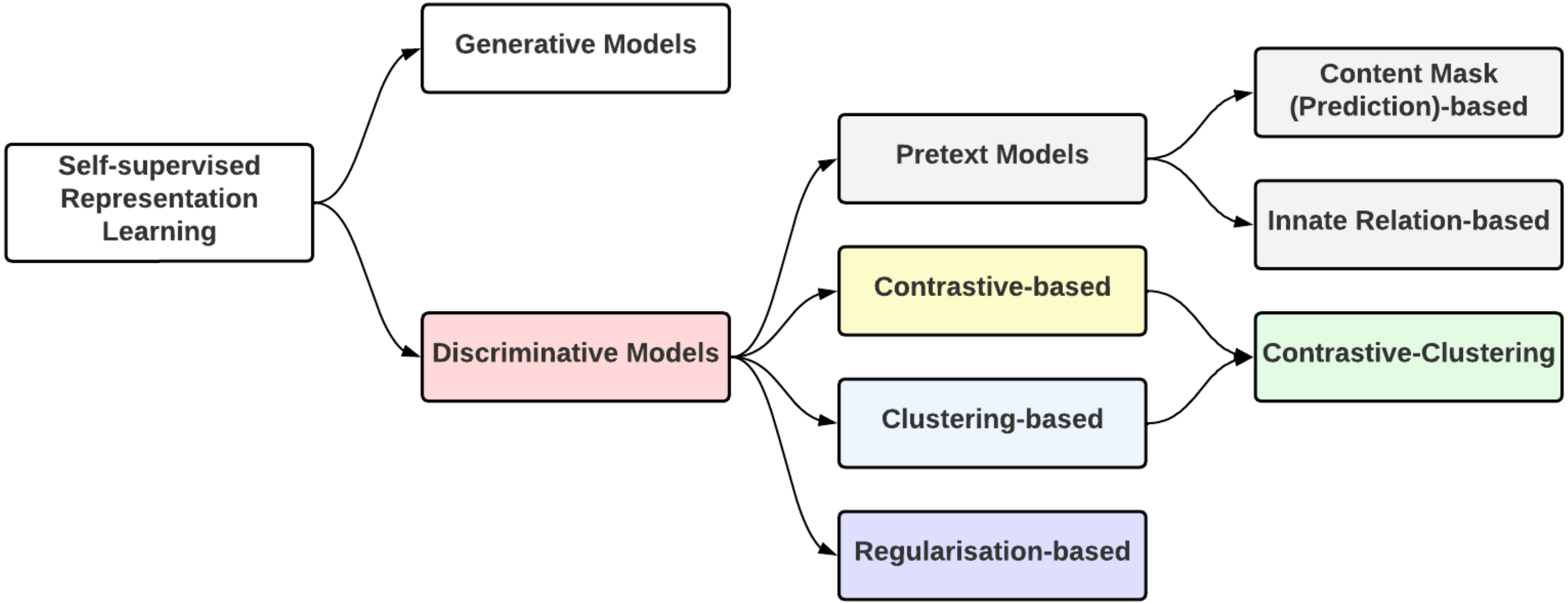


Fig. 3: Overview of federated learning framework. A central server dispatches a randomly initialized model and other training configuration details to the selected clients' devices, as depicted by dashed gray lines. The clients train local models on their private data and send the models back to the server illustrated with solid black lines. The models are aggregated to produce a unified model that is used for the end-task.

Saeed, A., Salim, F.D., Ozcelebi, T. and Lukkien, J., 2020. Federated Self-Supervised Learning of Multisensor Representations for Embedded Intelligence. IEEE Internet of Things Journal, 8(2), pp.1030-1040.

Beyond Just Vision:
A Review on Self-Supervised Representation Learning on Multimodal and Temporal Data



Self-supervised Representation Learning

Generative Models

Discriminative Models

Pretext Models

Contrastive-based

Clustering-based

Regularisation-based

Content Mask (Prediction)-based

Innate Relation-based

Contrastive-Clustering

Deldari, S., Xue, H., Saeed, A., He, J., Smith, D.V. and Salim, F.D., 2022. Beyond Just Vision: A Review on Self-Supervised Representation Learning on Multimodal and Temporal Data.

Self-supervised Activity Representation Learning with Incremental Data: An Empirical Study

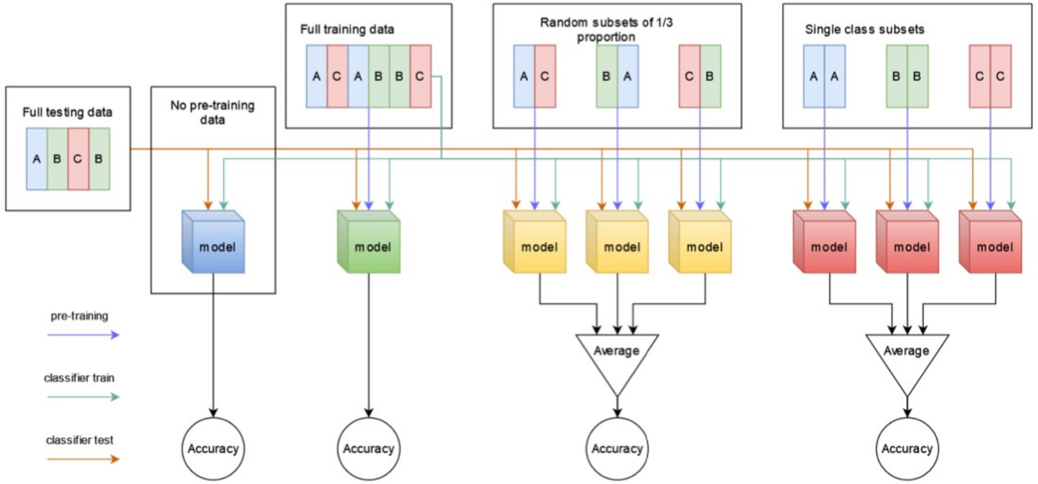


Fig. 3. Workflow of the UEA dataset investigation. Full training data were split into random subsets with size according to the number of classes in a dataset, and also single class subsets. These were used to pre-train encoder models, and compared with model trained based on no and full pre-training data.

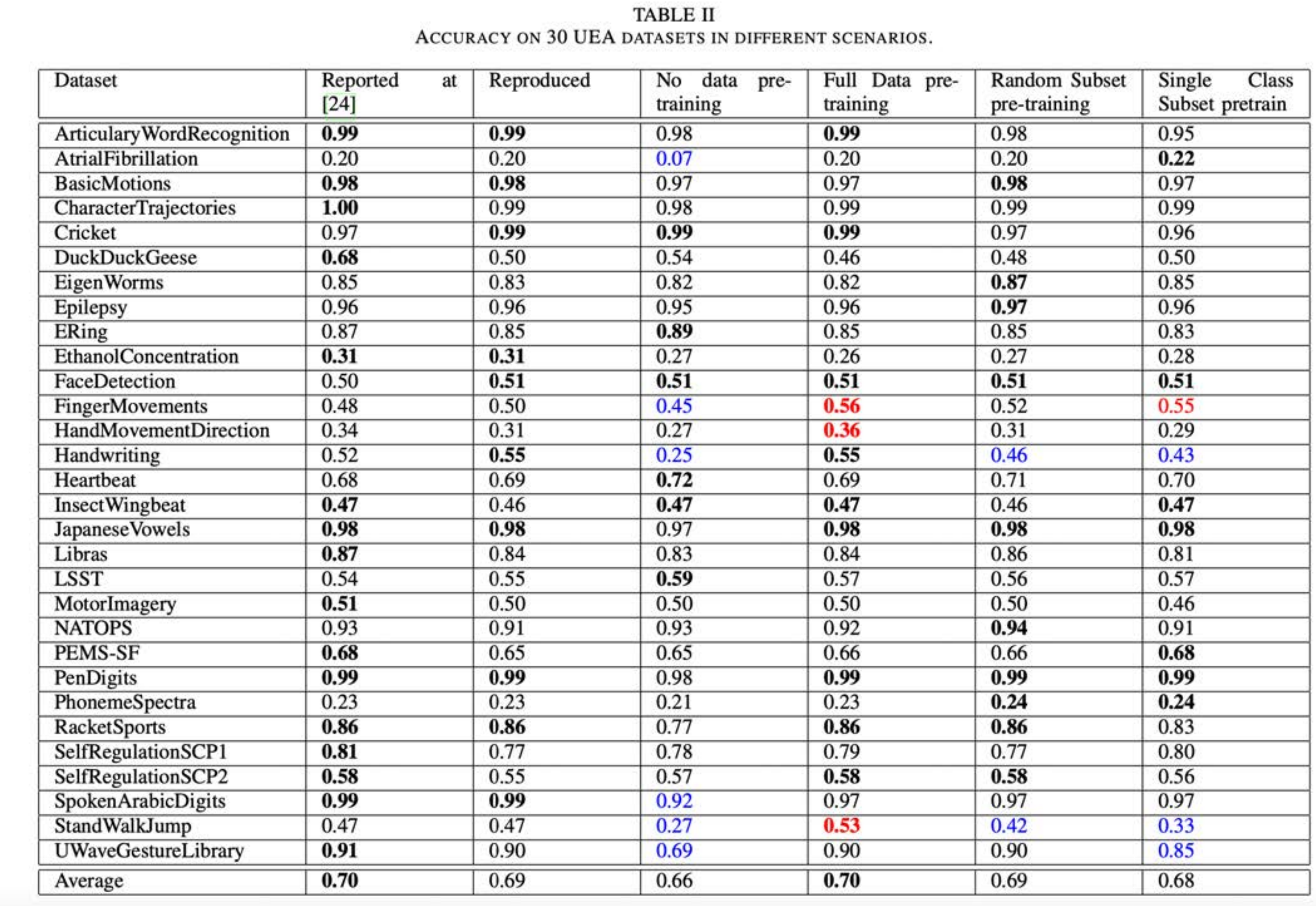


TABLE II
ACCURACY ON 30 UEA DATASETS IN DIFFERENT SCENARIOS.

Dataset	Reported at [24]	Reproduced	No data pre-training	Full Data pre-training	Random Subset pre-training	Single Class Subset pretrain
ArticularyWordRecognition	**0.99**	**0.99**	0.98	**0.99**	0.98	0.95
AtrialFibrillation	0.20	0.20	0.07	0.20	0.20	**0.22**
BasicMotions	**0.98**	**0.98**	0.97	0.97	**0.98**	0.97
CharacterTrajectories	**1.00**	0.99	0.98	0.99	0.99	0.99
Cricket	0.97	**0.99**	**0.99**	**0.99**	0.97	0.96
DuckDuckGeese	**0.68**	0.50	0.54	0.46	0.48	0.50
EigenWorms	0.85	0.83	0.82	0.82	**0.87**	0.85
Epilepsy	0.96	0.96	0.95	0.96	**0.97**	0.96
ERing	0.87	0.85	**0.89**	0.85	0.85	0.83
EthanolConcentration	**0.31**	**0.31**	0.27	0.26	0.27	0.28
FaceDetection	0.50	**0.51**	**0.51**	**0.51**	**0.51**	**0.51**
FingerMovements	0.48	0.50	0.45	**0.56**	0.52	0.55
HandMovementDirection	0.34	0.31	0.27	**0.36**	0.31	0.29
Handwriting	0.52	**0.55**	0.25	**0.55**	0.46	0.43
Heartbeat	0.68	0.69	**0.72**	0.69	0.71	0.70
InsectWingbeat	**0.47**	0.46	**0.47**	**0.47**	0.46	**0.47**
JapaneseVowels	**0.98**	**0.98**	0.97	**0.98**	**0.98**	**0.98**
Libras	**0.87**	0.84	0.83	0.84	0.86	0.81
LSST	0.54	0.55	**0.59**	0.57	0.56	0.57
MotorImagery	**0.51**	0.50	0.50	0.50	0.50	0.46
NATOPS	0.93	0.91	0.93	0.92	**0.94**	0.91
PEMS-SF	**0.68**	0.65	0.65	0.66	0.66	**0.68**
PenDigits	**0.99**	**0.99**	0.98	**0.99**	**0.99**	**0.99**
PhonemeSpectra	0.23	0.23	0.21	0.23	**0.24**	**0.24**
RacketSports	**0.86**	**0.86**	0.77	**0.86**	**0.86**	0.83
SelfRegulationSCP1	**0.81**	0.77	0.78	0.79	0.77	0.80
SelfRegulationSCP2	**0.58**	0.55	0.57	**0.58**	**0.58**	0.56
SpokenArabicDigits	**0.99**	**0.99**	0.92	0.97	0.97	0.97
StandWalkJump	0.47	0.47	0.27	**0.53**	0.42	0.33
UWaveGestureLibrary	**0.91**	0.90	0.69	0.90	0.90	0.85
Average	**0.70**	0.69	0.66	**0.70**	0.69	0.68

Liu, J., Deldari, S., Xue, H., Nguyen, V. and Salim, F.D., 2023. Self-supervised Activity Representation Learning with Incremental Data: An Empirical Study. *arXiv preprint arXiv:2305.00619*, IEEE Mobile Data Management (MDM) 2023

Contrastive Learning-based Imputation-Prediction Networks for In-hospital Mortality Risk Modeling using EHRs

Lorenzo Yuxi Liu[1], Zhenhao Zhang[2], Shaowen Qin[3] , **Flora D. Salim**[4], Antonio Jimeno Yepes[5]

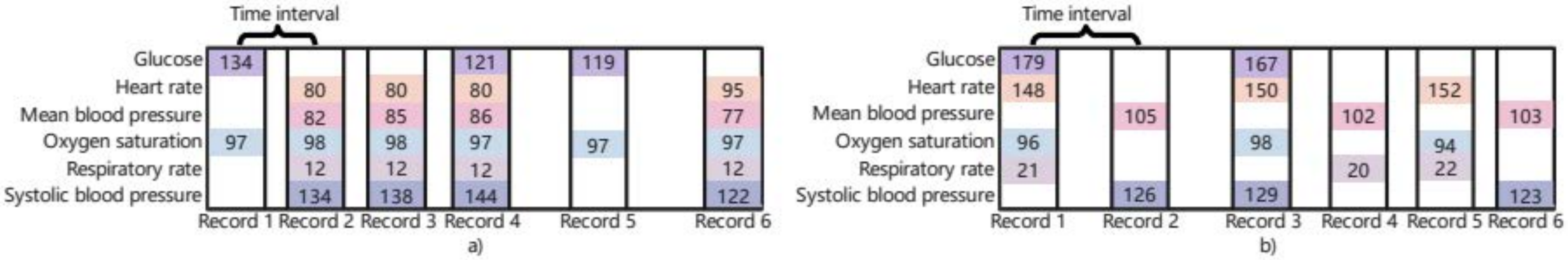
[1,3]**Flinders University**, [2]Northwest A&F University, [4]**University of New South Wales (UNSW)**, [5]RMIT University

EHR-based risk prediction tasks

- predict the mortality risks based on historical EHR data
- used to provide early warnings when a patient's health condition deteriorates

Open Challenge

- The data is highly irregular with many missing values
- The interval between medical records vary across patients



	Record 1	Record 2	Record 3	Record 4	Record 5	Record 6
Glucose	134			121	119	
Heart rate		80	80	80		95
Mean blood pressure		82	85	86		77
Oxygen saturation	97	98	98	97	97	97
Respiratory rate		12	12	12		12
Systolic blood pressure		134	138	144		122

	Record 1	Record 2	Record 3	Record 4	Record 5	Record 6
Glucose	179		167			
Heart rate	148		150		152	
Mean blood pressure		105		102		103
Oxygen saturation	96		98		94	
Respiratory rate	21			20	22	
Systolic blood pressure		126	129			123

Fig. 1. Illustration of medical records of patients A and B.

The MIMIC-III Database

Existing Approaches

Imputation first then prediction

Imputation as a separate preprocessing task

Imputation Step

Exploit correlations of variables in patient medical records to impute missing values

Establish time-decay mechanisms to consider the effect of varying time intervals between records

Prediction Step

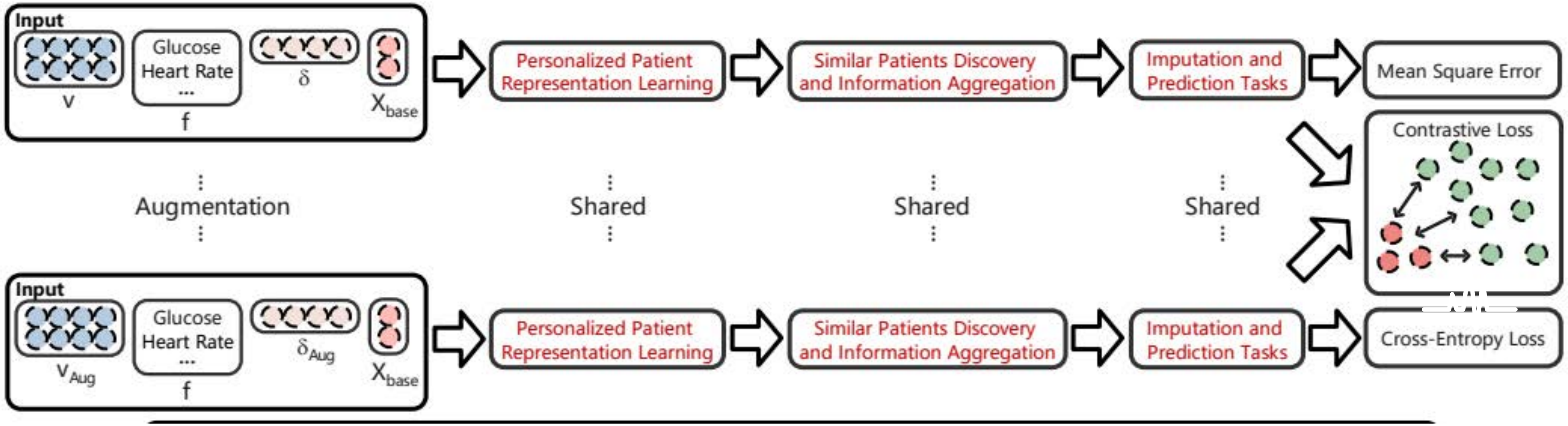
The complete data matrices are then used as input for downstream prediction tasks.

Our intution:
Combine Imputation with Prediction with Patient Stratification & Contrastive Learning

Patient stratification refers to the method of dividing a patient population into subgroups based on specific disease characteristics and symptom severity.

Patients in the same subgroup generally had more similar health trajectories.

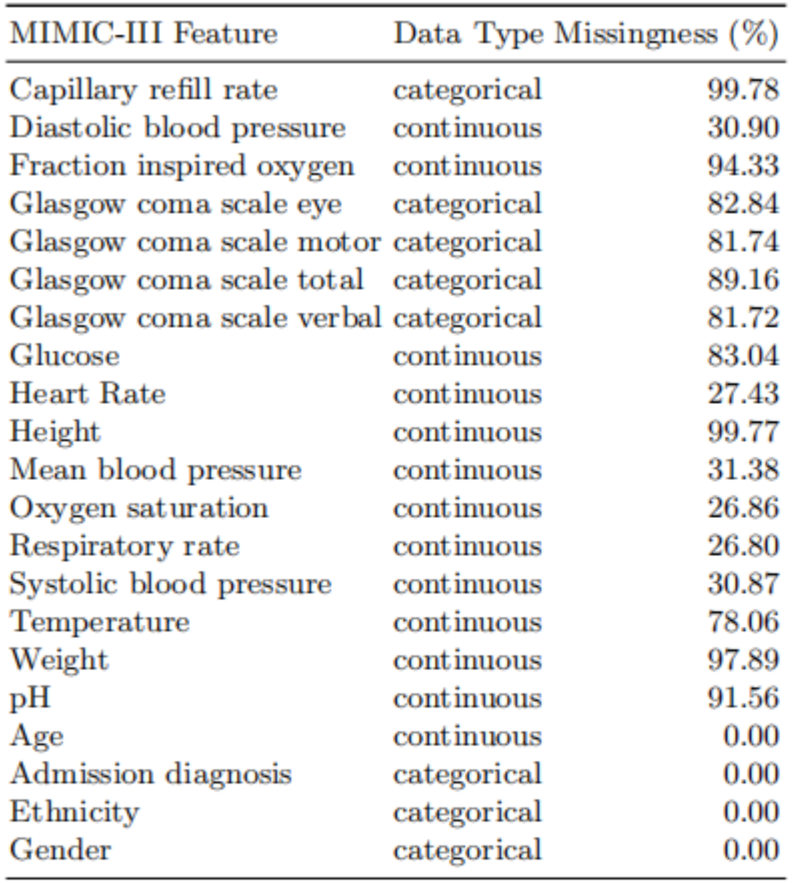
Proposed Method:
Contrastive Learning-based Imputation-Prediction Networks for In-hospital Mortality Risk Modeling



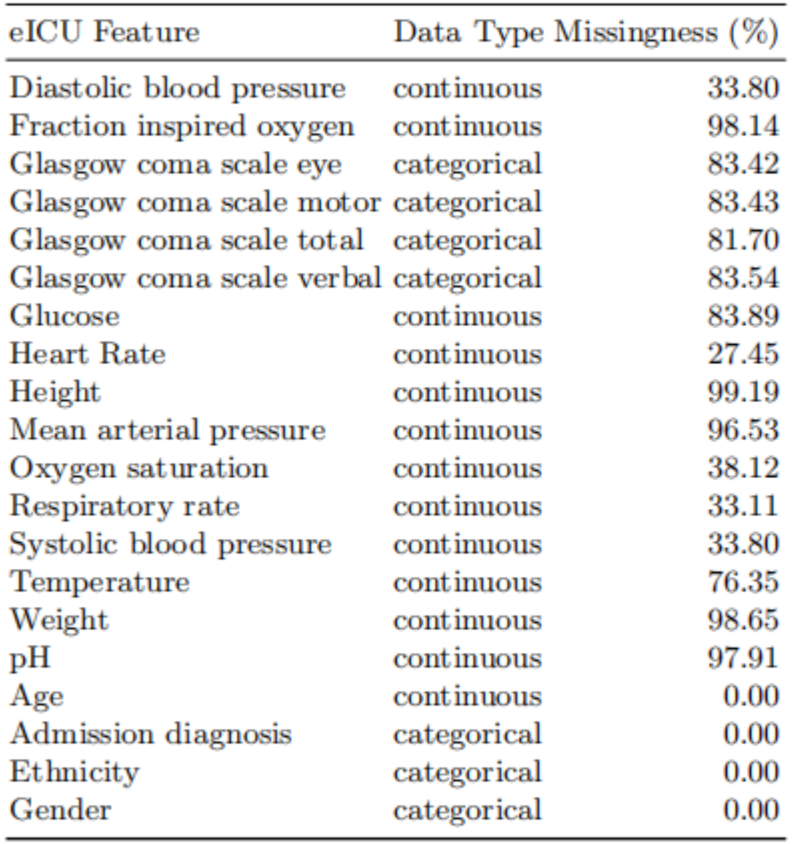
Experiments: Datasets

MIMIC-III [1] and eICU [2] Databases

Inputs: clinical times series (e.g., heart rate) and demographics (i.e., age, sex, and ethnicity)

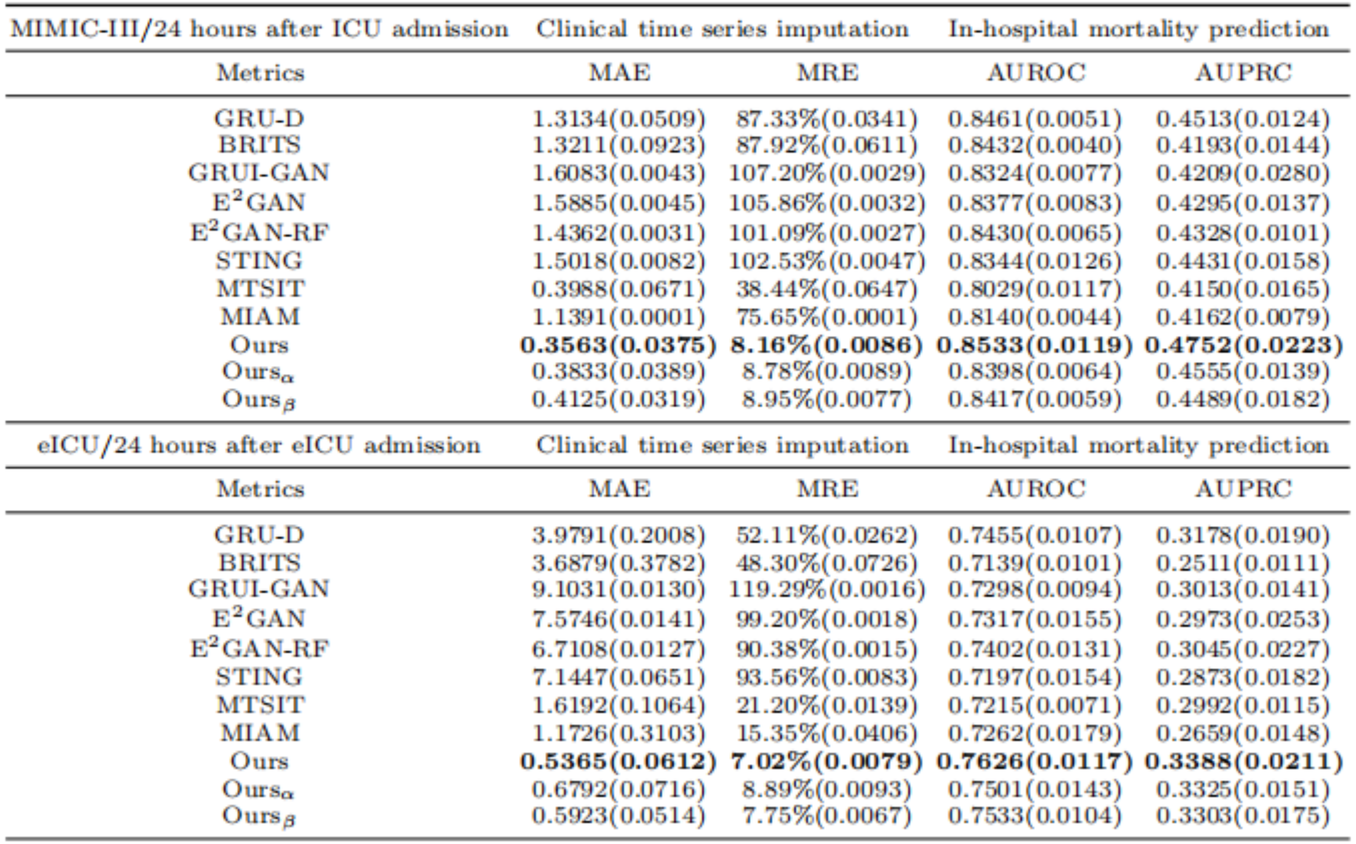


MIMIC-III Feature	Data Type	Missingness (%)
Capillary refill rate	categorical	99.78
Diastolic blood pressure	continuous	30.90
Fraction inspired oxygen	continuous	94.33
Glasgow coma scale eye	categorical	82.84
Glasgow coma scale motor	categorical	81.74
Glasgow coma scale total	categorical	89.16
Glasgow coma scale verbal	categorical	81.72
Glucose	continuous	83.04
Heart Rate	continuous	27.43
Height	continuous	99.77
Mean blood pressure	continuous	31.38
Oxygen saturation	continuous	26.86
Respiratory rate	continuous	26.80
Systolic blood pressure	continuous	30.87
Temperature	continuous	78.06
Weight	continuous	97.89
pH	continuous	91.56
Age	continuous	0.00
Admission diagnosis	categorical	0.00
Ethnicity	categorical	0.00
Gender	categorical	0.00



eICU Feature	Data Type	Missingness (%)
Diastolic blood pressure	continuous	33.80
Fraction inspired oxygen	continuous	98.14
Glasgow coma scale eye	categorical	83.42
Glasgow coma scale motor	categorical	83.43
Glasgow coma scale total	categorical	81.70
Glasgow coma scale verbal	categorical	83.54
Glucose	continuous	83.89
Heart Rate	continuous	27.45
Height	continuous	99.19
Mean arterial pressure	continuous	96.53
Oxygen saturation	continuous	38.12
Respiratory rate	continuous	33.11
Systolic blood pressure	continuous	33.80
Temperature	continuous	76.35
Weight	continuous	98.65
pH	continuous	97.91
Age	continuous	0.00
Admission diagnosis	categorical	0.00
Ethnicity	categorical	0.00
Gender	categorical	0.00

Experiment Results

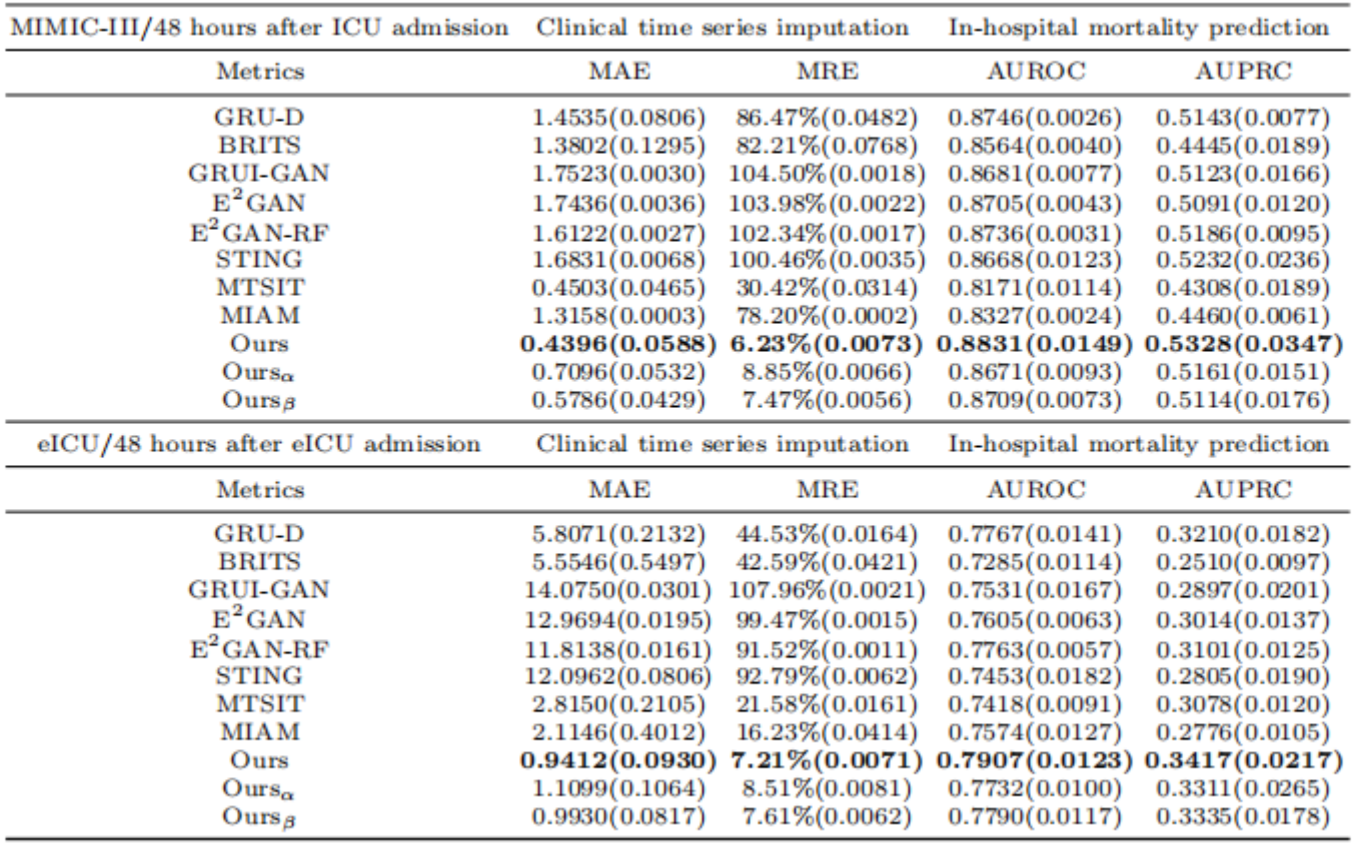


MIMIC-III/24 hours after ICU admission	Clinical time series imputation		In-hospital mortality prediction	
Metrics	MAE	MRE	AUROC	AUPRC
GRU-D	1.3134(0.0509)	87.33%(0.0341)	0.8461(0.0051)	0.4513(0.0124)
BRITS	1.3211(0.0923)	87.92%(0.0611)	0.8432(0.0040)	0.4193(0.0144)
GRUI-GAN	1.6083(0.0043)	107.20%(0.0029)	0.8324(0.0077)	0.4209(0.0280)
E^2GAN	1.5885(0.0045)	105.86%(0.0032)	0.8377(0.0083)	0.4295(0.0137)
E^2GAN-RF	1.4362(0.0031)	101.09%(0.0027)	0.8430(0.0065)	0.4328(0.0101)
STING	1.5018(0.0082)	102.53%(0.0047)	0.8344(0.0126)	0.4431(0.0158)
MTSIT	0.3988(0.0671)	38.44%(0.0647)	0.8029(0.0117)	0.4150(0.0165)
MIAM	1.1391(0.0001)	75.65%(0.0001)	0.8140(0.0044)	0.4162(0.0079)
Ours	**0.3563(0.0375)**	**8.16%(0.0086)**	**0.8533(0.0119)**	**0.4752(0.0223)**
Ours_α	0.3833(0.0389)	8.78%(0.0089)	0.8398(0.0064)	0.4555(0.0139)
Ours_β	0.4125(0.0319)	8.95%(0.0077)	0.8417(0.0059)	0.4489(0.0182)
eICU/24 hours after eICU admission	Clinical time series imputation		In-hospital mortality prediction	
Metrics	MAE	MRE	AUROC	AUPRC
GRU-D	3.9791(0.2008)	52.11%(0.0262)	0.7455(0.0107)	0.3178(0.0190)
BRITS	3.6879(0.3782)	48.30%(0.0726)	0.7139(0.0101)	0.2511(0.0111)
GRUI-GAN	9.1031(0.0130)	119.29%(0.0016)	0.7298(0.0094)	0.3013(0.0141)
E^2GAN	7.5746(0.0141)	99.20%(0.0018)	0.7317(0.0155)	0.2973(0.0253)
E^2GAN-RF	6.7108(0.0127)	90.38%(0.0015)	0.7402(0.0131)	0.3045(0.0227)
STING	7.1447(0.0651)	93.56%(0.0083)	0.7197(0.0154)	0.2873(0.0182)
MTSIT	1.6192(0.1064)	21.20%(0.0139)	0.7215(0.0071)	0.2992(0.0115)
MIAM	1.1726(0.3103)	15.35%(0.0406)	0.7262(0.0179)	0.2659(0.0148)
Ours	**0.5365(0.0612)**	**7.02%(0.0079)**	**0.7626(0.0117)**	**0.3388(0.0211)**
Ours_α	0.6792(0.0716)	8.89%(0.0093)	0.7501(0.0143)	0.3325(0.0151)
Ours_β	0.5923(0.0514)	7.75%(0.0067)	0.7533(0.0104)	0.3303(0.0175)

Imputation and prediction results: 24 hours after ICU admission

CODE:

Experiment Results



MIMIC-III/48 hours after ICU admission	Clinical time series imputation		In-hospital mortality prediction	
Metrics	MAE	MRE	AUROC	AUPRC
GRU-D	1.4535(0.0806)	86.47%(0.0482)	0.8746(0.0026)	0.5143(0.0077)
BRITS	1.3802(0.1295)	82.21%(0.0768)	0.8564(0.0040)	0.4445(0.0189)
GRUI-GAN	1.7523(0.0030)	104.50%(0.0018)	0.8681(0.0077)	0.5123(0.0166)
E^2GAN	1.7436(0.0036)	103.98%(0.0022)	0.8705(0.0043)	0.5091(0.0120)
E^2GAN-RF	1.6122(0.0027)	102.34%(0.0017)	0.8736(0.0031)	0.5186(0.0095)
STING	1.6831(0.0068)	100.46%(0.0035)	0.8668(0.0123)	0.5232(0.0236)
MTSIT	0.4503(0.0465)	30.42%(0.0314)	0.8171(0.0114)	0.4308(0.0189)
MIAM	1.3158(0.0003)	78.20%(0.0002)	0.8327(0.0024)	0.4460(0.0061)
Ours	**0.4396(0.0588)**	**6.23%(0.0073)**	**0.8831(0.0149)**	**0.5328(0.0347)**
Ours_α	0.7096(0.0532)	8.85%(0.0066)	0.8671(0.0093)	0.5161(0.0151)
Ours_β	0.5786(0.0429)	7.47%(0.0056)	0.8709(0.0073)	0.5114(0.0176)
eICU/48 hours after eICU admission	Clinical time series imputation		In-hospital mortality prediction	
Metrics	MAE	MRE	AUROC	AUPRC
GRU-D	5.8071(0.2132)	44.53%(0.0164)	0.7767(0.0141)	0.3210(0.0182)
BRITS	5.5546(0.5497)	42.59%(0.0421)	0.7285(0.0114)	0.2510(0.0097)
GRUI-GAN	14.0750(0.0301)	107.96%(0.0021)	0.7531(0.0167)	0.2897(0.0201)
E^2GAN	12.9694(0.0195)	99.47%(0.0015)	0.7605(0.0063)	0.3014(0.0137)
E^2GAN-RF	11.8138(0.0161)	91.52%(0.0011)	0.7763(0.0057)	0.3101(0.0125)
STING	12.0962(0.0806)	92.79%(0.0062)	0.7453(0.0182)	0.2805(0.0190)
MTSIT	2.8150(0.2105)	21.58%(0.0161)	0.7418(0.0091)	0.3078(0.0120)
MIAM	2.1146(0.4012)	16.23%(0.0414)	0.7574(0.0127)	0.2776(0.0105)
Ours	**0.9412(0.0930)**	**7.21%(0.0071)**	**0.7907(0.0123)**	**0.3417(0.0217)**
Ours_α	1.1099(0.1064)	8.51%(0.0081)	0.7732(0.0100)	0.3311(0.0265)
Ours_β	0.9930(0.0817)	7.61%(0.0062)	0.7790(0.0117)	0.3335(0.0178)

Imputation and prediction results: 48 hours after ICU admission

CODE:

Embedding Contextual Dynamics using Transformers

UNSW
AI Institute

Positional Encoding in Transformers - Suitable for ST data

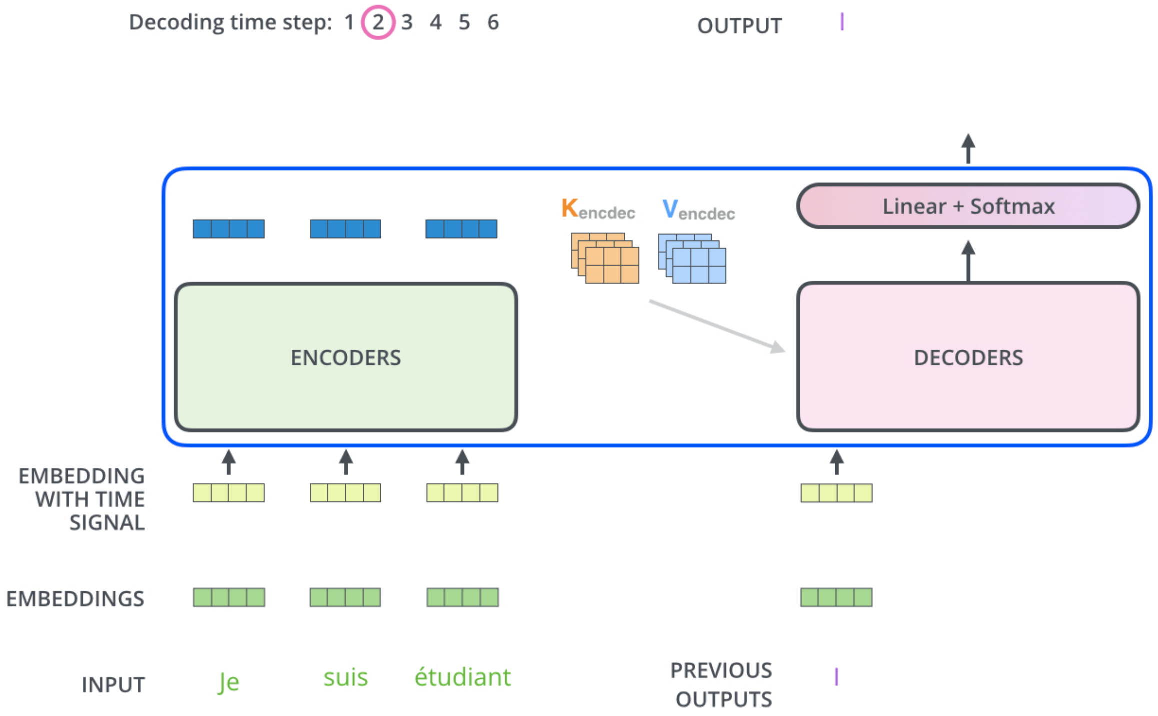
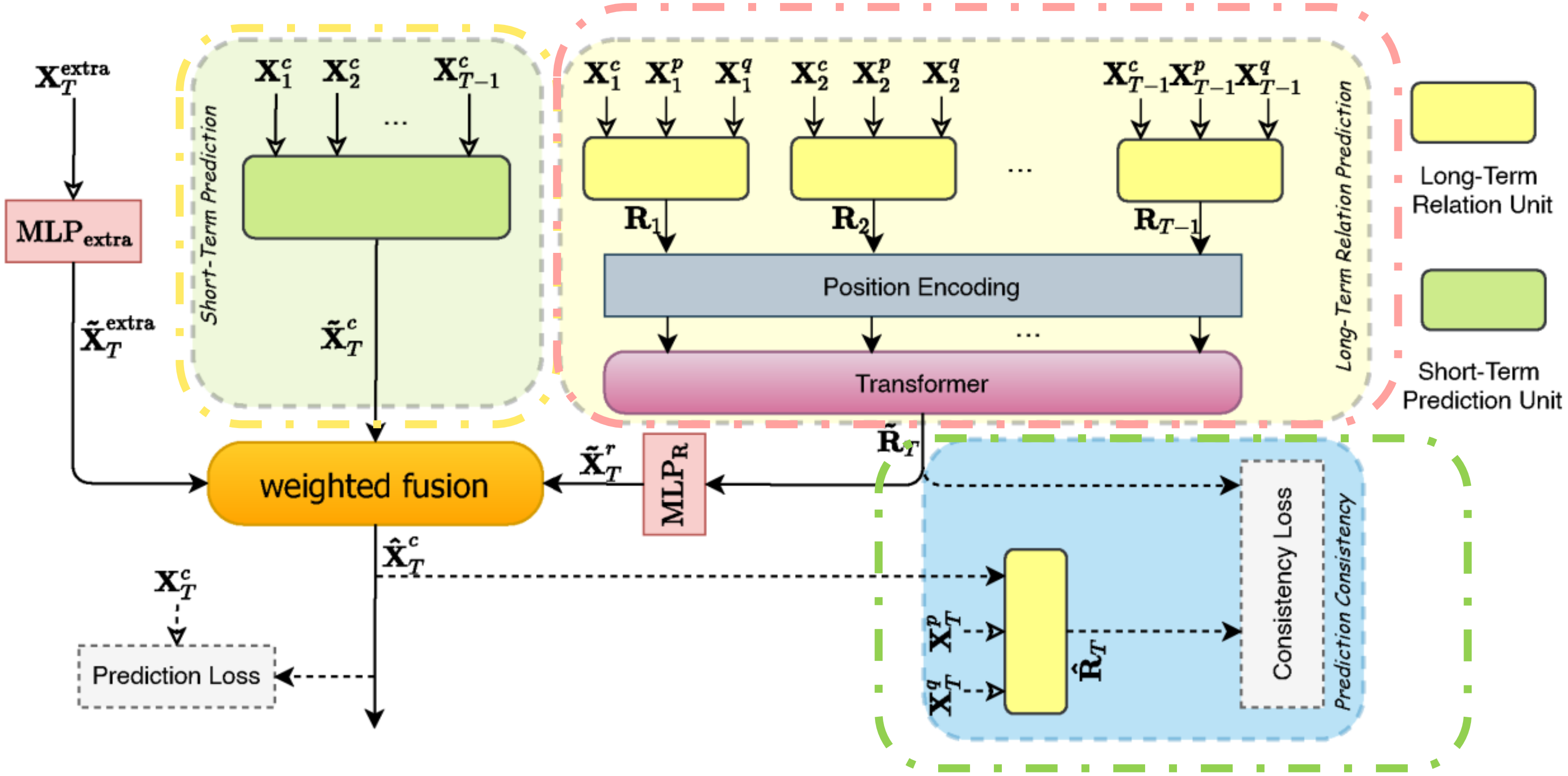


Image Credit: https://jalammar.github.io/illustrated-transformer/

Effective Urban Flow Forecasting (Traffic Flow, People Movement Flow, Check-ins)

- Short-Term Prediction module
- Long-Term Relation Prediction module
- Prediction Consistency module



Xue, H. and Salim, F.D., 2020. TERMCast: Temporal Relation Modeling for Effective Urban Flow Forecasting. *PAKDD 2020*

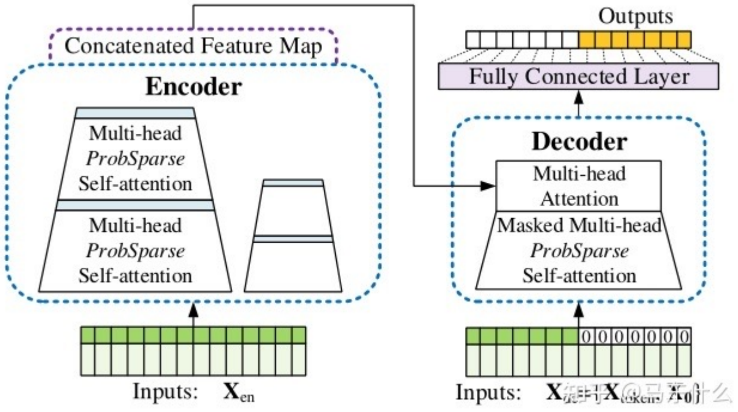
Other Transformer-based Models

- Transformer starts to replace RNNs in sequential forecasting

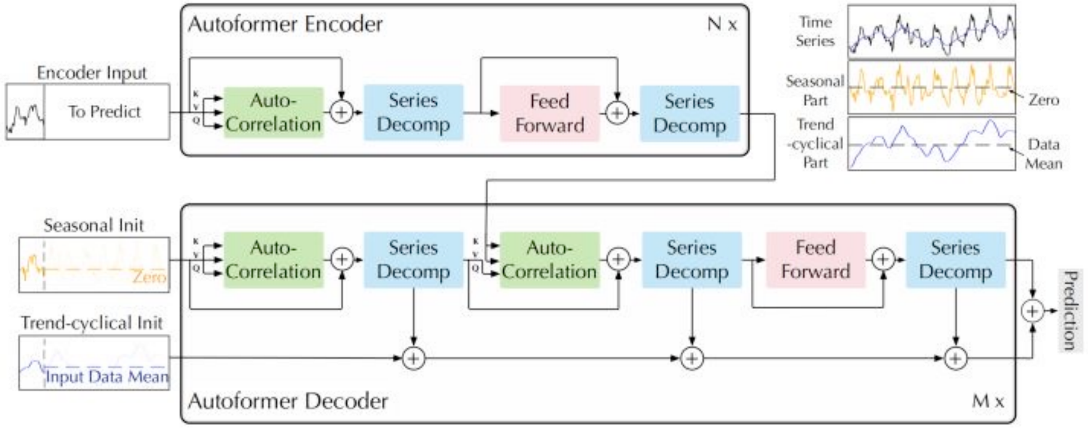
- Informer:
  - enhance the prediction capacity especially for long sequence forecasting
  - ProbSparse self-attention mechanism to efficiently replace the canonical self-attention

- Autoformer:
  - Also powerful in long sequence prediction
  - Auto-Correlation mechanism with dependencies discovery and information aggregation at the series level

Informer



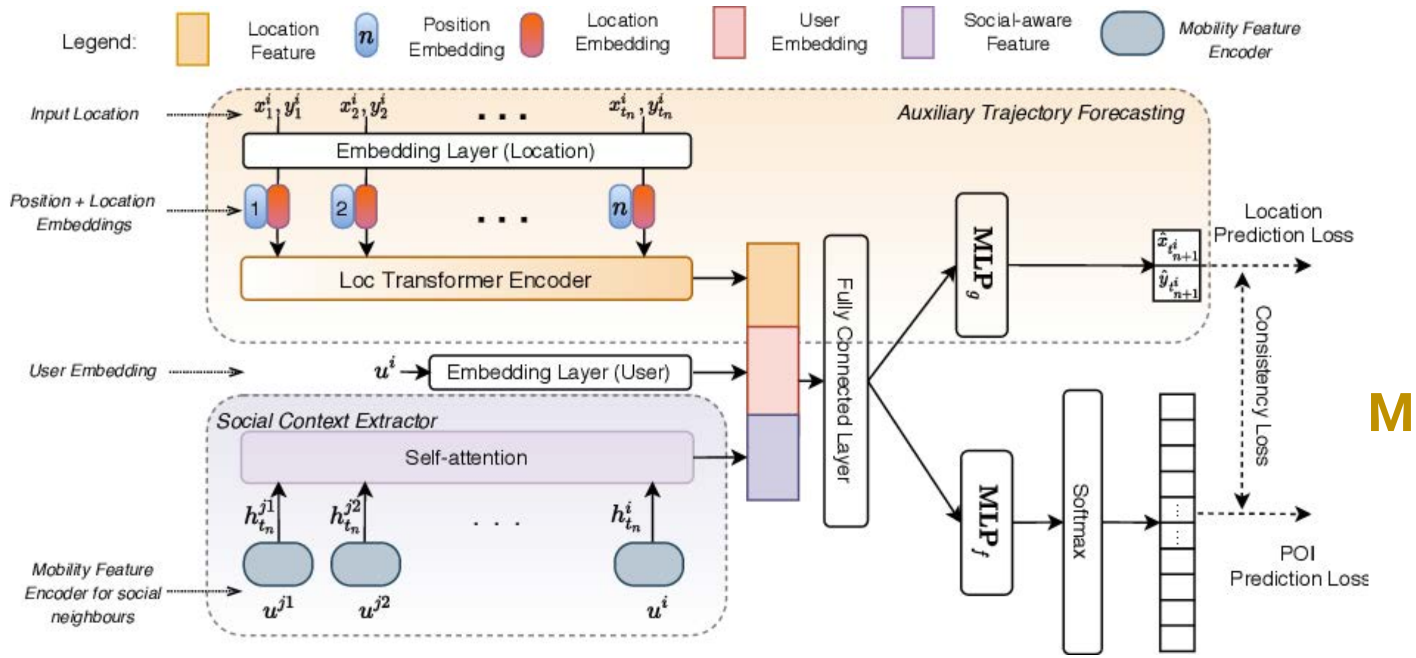
Autoformer



Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of AAAI

Jiehui Xu, Jianmin Wang, Mingsheng Long, et al. 2021. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in Neural Information Processing Systems 34

Transformers-based architecture for Mobility Prediction



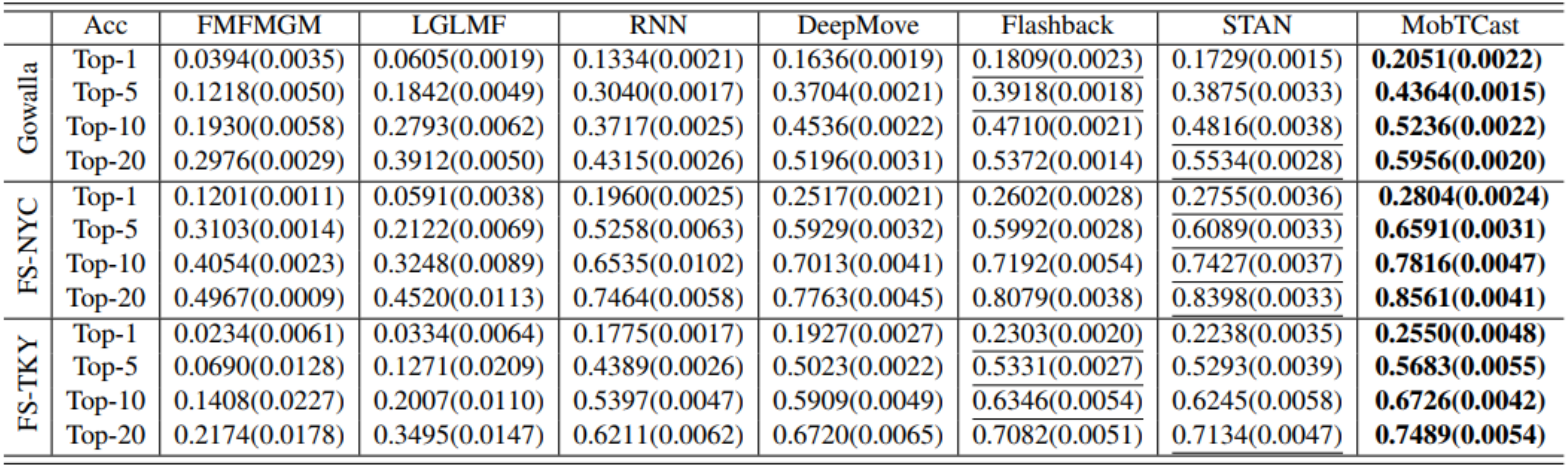
MobTCast

- Embed semantics and contexts e.g., groups, category etc on Transformers, the architecture would be complicated
- Need to design extra layers/modules for auxiliary information

Hao Xue, Flora Salim, Yongli Ren, and Nuria Oliver. 2021. MobTCast: Leveraging Auxiliary Trajectory Forecasting for Human Mobility Prediction. Advances in Neural Information Processing Systems (NeurIPS) 34 (2021).

Experiments: Results

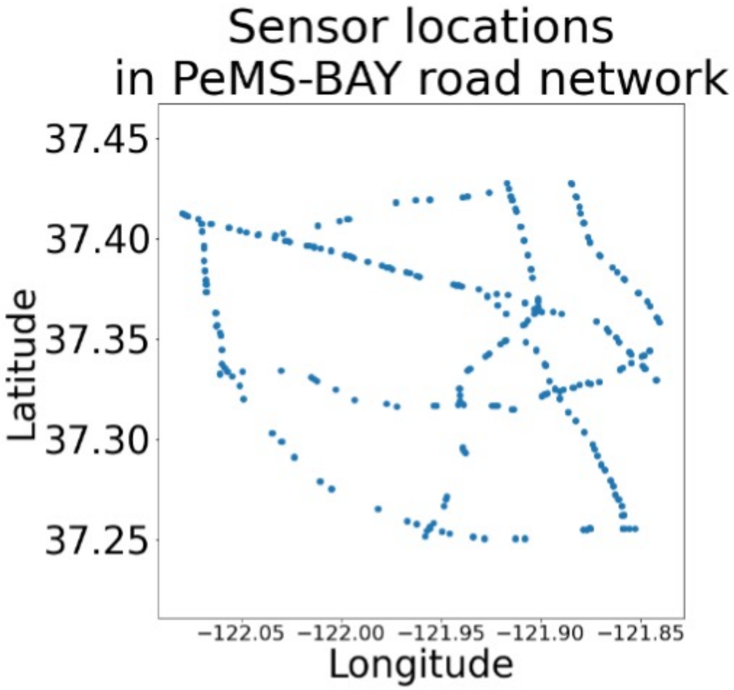
- Dataset: Gowalla, FourSquare-NYC, FourSquare-TKY
- Metric: Top-k (k=1, 5, 10, 20)



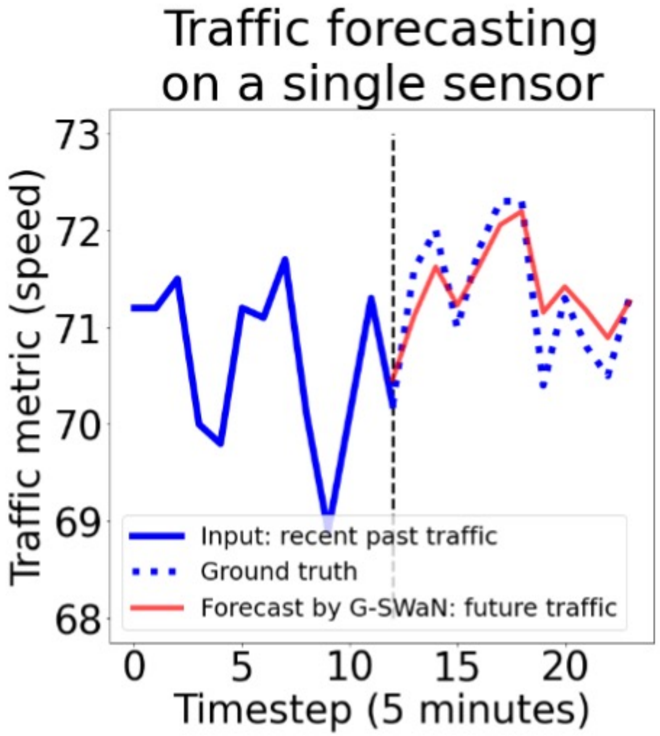
	Acc	FMFMGM	LGLMF	RNN	DeepMove	Flashback	STAN	MobTCast
Gowalla	Top-1	0.0394(0.0035)	0.0605(0.0019)	0.1334(0.0021)	0.1636(0.0019)	0.1809(0.0023)	0.1729(0.0015)	**0.2051(0.0022)**
	Top-5	0.1218(0.0050)	0.1842(0.0049)	0.3040(0.0017)	0.3704(0.0021)	0.3918(0.0018)	0.3875(0.0033)	**0.4364(0.0015)**
	Top-10	0.1930(0.0058)	0.2793(0.0062)	0.3717(0.0025)	0.4536(0.0022)	0.4710(0.0021)	0.4816(0.0038)	**0.5236(0.0022)**
	Top-20	0.2976(0.0029)	0.3912(0.0050)	0.4315(0.0026)	0.5196(0.0031)	0.5372(0.0014)	0.5534(0.0028)	**0.5956(0.0020)**
FS-NYC	Top-1	0.1201(0.0011)	0.0591(0.0038)	0.1960(0.0025)	0.2517(0.0021)	0.2602(0.0028)	0.2755(0.0036)	**0.2804(0.0024)**
	Top-5	0.3103(0.0014)	0.2122(0.0069)	0.5258(0.0063)	0.5929(0.0032)	0.5992(0.0028)	0.6089(0.0033)	**0.6591(0.0031)**
	Top-10	0.4054(0.0023)	0.3248(0.0089)	0.6535(0.0102)	0.7013(0.0041)	0.7192(0.0054)	0.7427(0.0037)	**0.7816(0.0047)**
	Top-20	0.4967(0.0009)	0.4520(0.0113)	0.7464(0.0058)	0.7763(0.0045)	0.8079(0.0038)	0.8398(0.0033)	**0.8561(0.0041)**
FS-TKY	Top-1	0.0234(0.0061)	0.0334(0.0064)	0.1775(0.0017)	0.1927(0.0027)	0.2303(0.0020)	0.2238(0.0035)	**0.2550(0.0048)**
	Top-5	0.0690(0.0128)	0.1271(0.0209)	0.4389(0.0026)	0.5023(0.0022)	0.5331(0.0027)	0.5293(0.0039)	**0.5683(0.0055)**
	Top-10	0.1408(0.0227)	0.2007(0.0110)	0.5397(0.0047)	0.5909(0.0049)	0.6346(0.0054)	0.6245(0.0058)	**0.6726(0.0042)**
	Top-20	0.2174(0.0178)	0.3495(0.0147)	0.6211(0.0062)	0.6720(0.0065)	0.7082(0.0051)	0.7134(0.0047)	**0.7489(0.0054)**

Mobility data captures dynamics of human behaviour at scale

Because Every Sensor Is Unique, so Is Every Pair: Handling Dynamicity in Traffic Forecasting



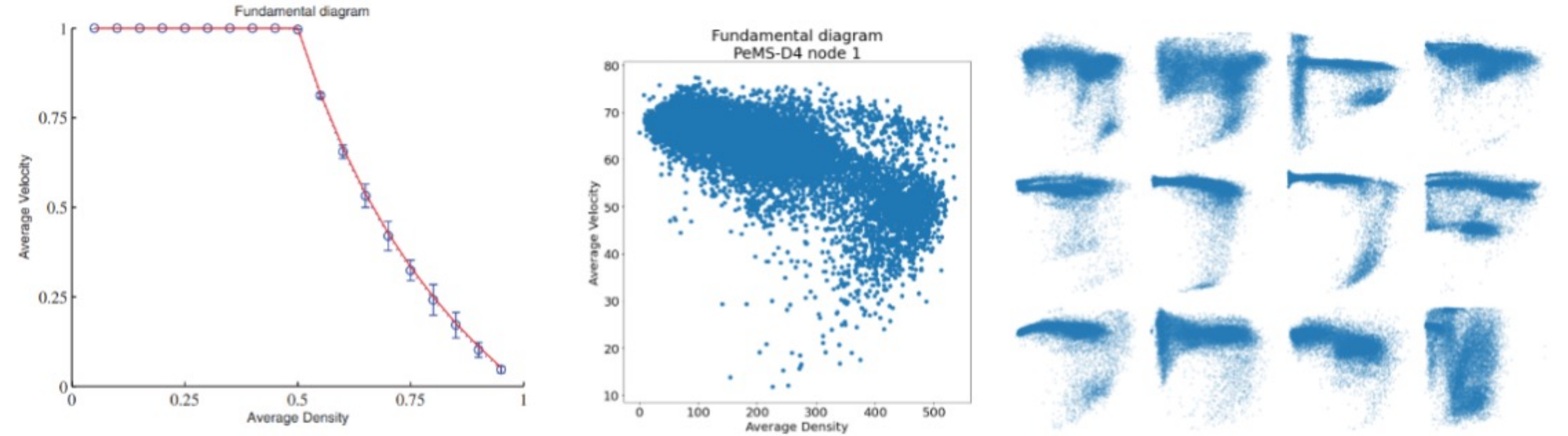
(a) Locations of the sensors on the Californian highway network surrounding the bay area. Installing a network of sensors on a road infrastructure enables traffic forecasting and smarter cities.



(b) At each sensor, traffic forecasting uses the recent sensor readings (solid blue line) to predict the future traffic (red line). This forecast is made by our proposed architecture Graph Self-attenion WaveNet (G-SWaN). Our forecasts accurately predict the future traffic (dotted blue line).

Arian Prabowo, Wei Shao, Hao Xue, Piotr Koniusz, and Flora D. Salim. 2023. Because Every Sensor Is Unique, so Is Every Pair: Handling Dynamicity in Traffic Forecasting. In International Conference on Internet-of-Things Design and Implementation (ACM IoTDI ’23).

Every sensor is unique



(a) Idealized fundamental diagram. Figure is copied from [5].

(b) Fundamental diagram of a sensor in PeMS-D4 dataset.

(c) Fundamental diagrams of 12 selected sensors in PeMS-D4 dataset showing great diversity.

Fig. 2. Fundamental diagrams showing the dynamics between flow (density) and speed (velocity).

so is every pair

(a) Association plots of 8 random sensor pairs.

(b) Association plots of a sensor pair over three weeks. Blue lines are weekdays, red lines are weekends.

Fig. 3. Association plots of different pairs of sensor readings in PeMS-D4. The x-value of a point is the flow at one sensor, while the y-value is the flow at the other sensor. Consecutive data points are connected by a line.

G-SWaN architecture

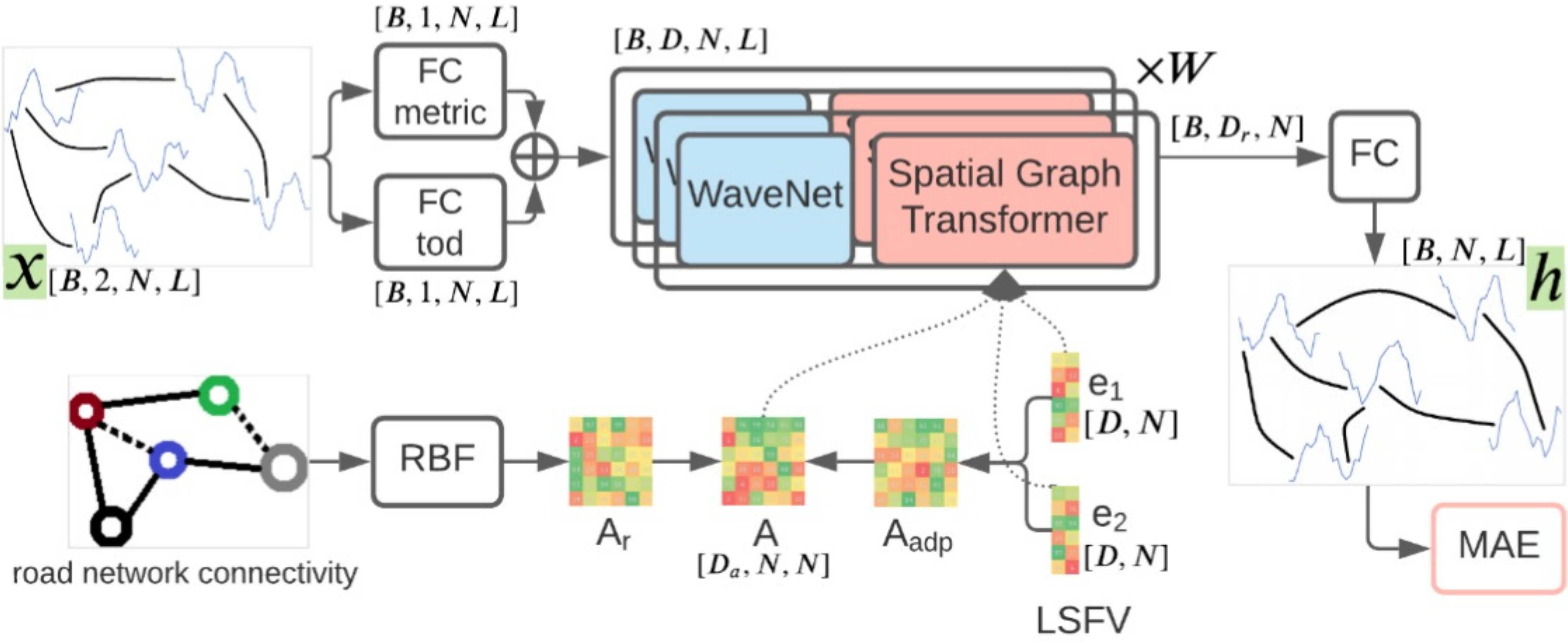


Fig. 4. System architecture of Graph Self-attention WaveNet (G-SWaN). Spatial Graph Transformers (SGT) is the novel module proposed that uses the node embeddings e_1 and e_2 to capture the unique sensor dynamics in the self-attention mechanisms. The notations are described in Table 1.

Results

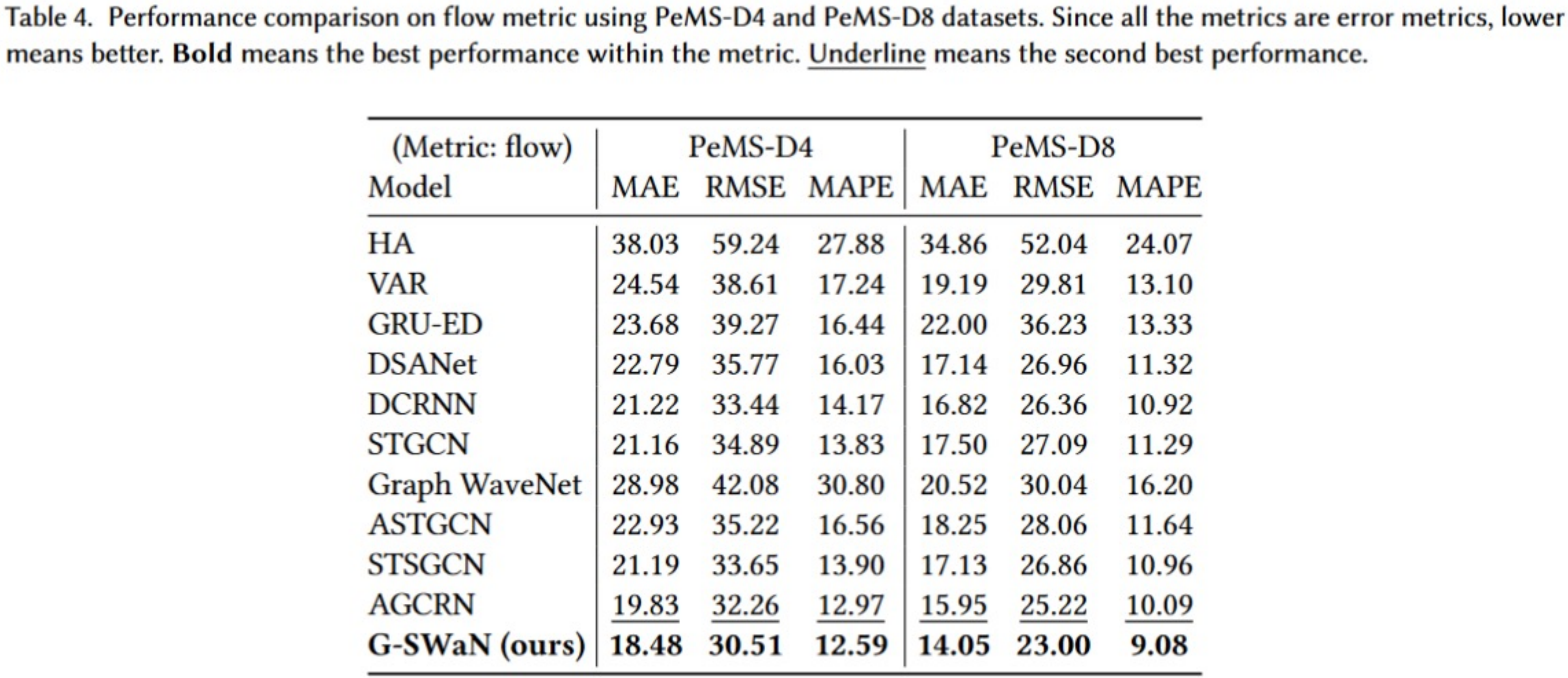
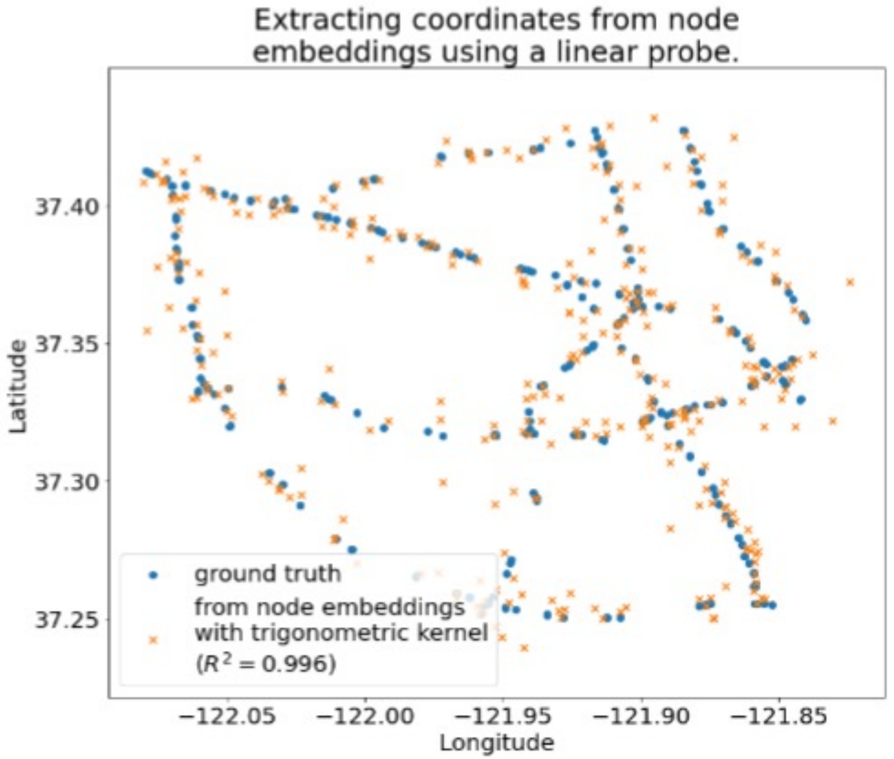


Table 4. Performance comparison on flow metric using PeMS-D4 and PeMS-D8 datasets. Since all the metrics are error metrics, lower means better. **Bold** means the best performance within the metric. Underline means the second best performance.

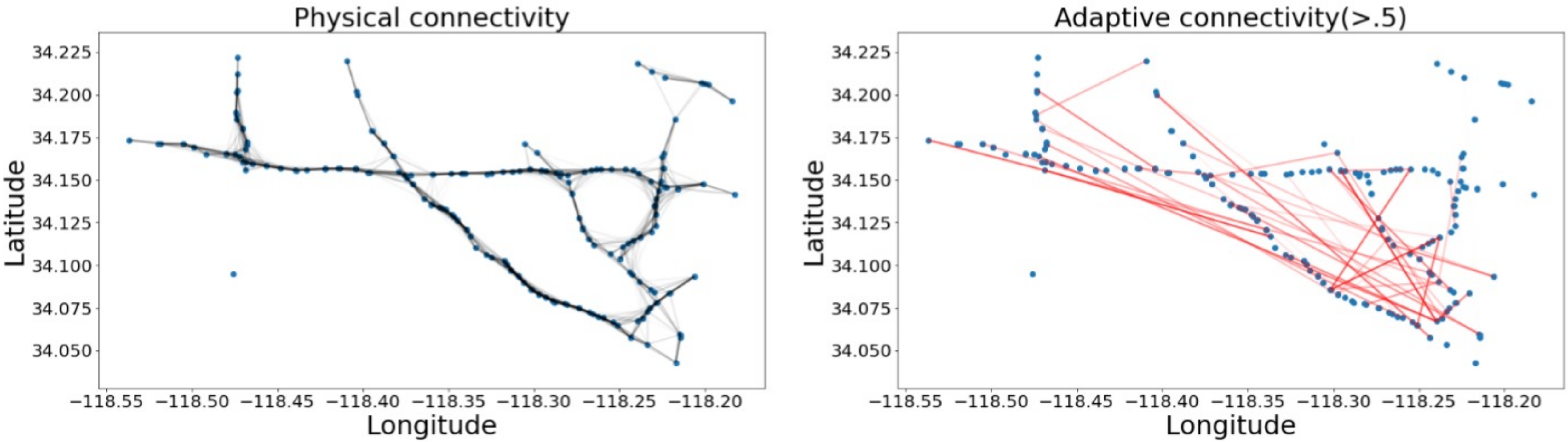
(Metric: flow)	PeMS-D4			PeMS-D8		
Model	MAE	RMSE	MAPE	MAE	RMSE	MAPE
HA	38.03	59.24	27.88	34.86	52.04	24.07
VAR	24.54	38.61	17.24	19.19	29.81	13.10
GRU-ED	23.68	39.27	16.44	22.00	36.23	13.33
DSANet	22.79	35.77	16.03	17.14	26.96	11.32
DCRNN	21.22	33.44	14.17	16.82	26.36	10.92
STGCN	21.16	34.89	13.83	17.50	27.09	11.29
Graph WaveNet	28.98	42.08	30.80	20.52	30.04	16.20
ASTGCN	22.93	35.22	16.56	18.25	28.06	11.64
STSGCN	21.19	33.65	13.90	17.13	26.86	10.96
AGCRN	19.83	32.26	12.97	15.95	25.22	10.09
G-SWaN (ours)	**18.48**	**30.51**	**12.59**	**14.05**	**23.00**	**9.08**

Node embeddings encode the coordinate from the traffic dynamics alone.



(a) Recovering sensor coordinates from node embeddings using a linear probe with trigonometric kernels.

Adaptive connectivity connect distant nodes.



(b) Qualitative comparison between physical and adaptive adjacency matrix in METR-LA dataset. The line transparency is proportional to the edge weight.

What if we have limited traffic data?

E.g.: New Roads, Unseen roads in training

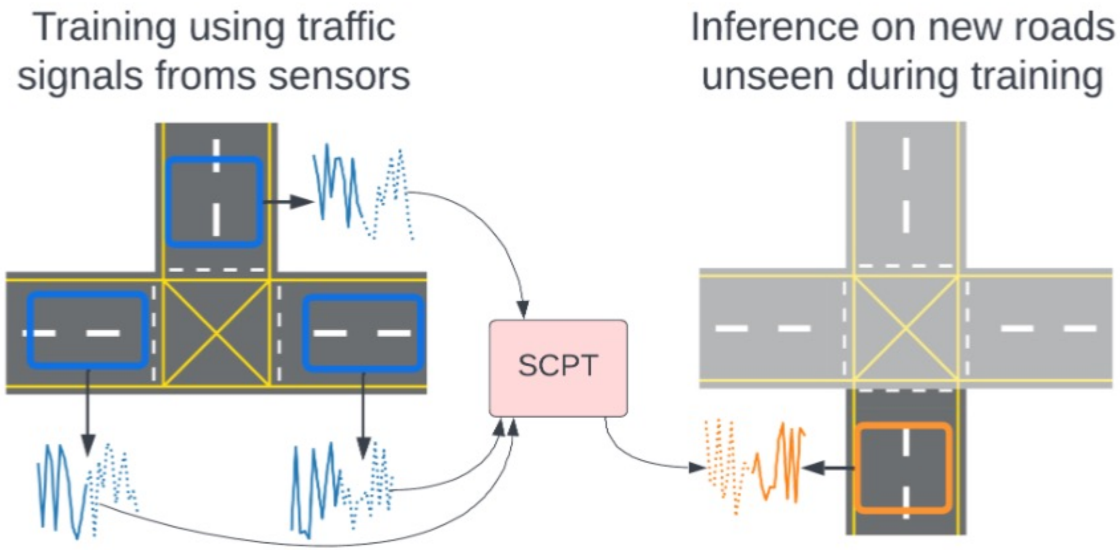


Fig. 1: A novel setup for traffic forecasting. Traffic signals (blue timeseries) are generated from sensors (blue boxes) at many different locations. Models are trained using these historical traffic data. During inference, forecasting is performed on traffic signals (orange timeseries) from new sensors (orange box) on roads which are not previously seen during training. We also introduce a novel framework called Spatial Contrastive Pre-Training (SCPT) that is effective for this new setup.

We pre-trained a spatial encoder using SCPT.

During inference time, it infers the spatial embedding of new roads from minimal data.

Prabowo, A., Xue, H., Shao, W., Koniusz, P., Salim, F. Traffic forecasting on new roads using spatial contrastive pre-training (SCPT). Data Mining and Knowledge Discovery (2023). https://doi.org/10.1007/s10618-023-00982-0

Spatio-temporal split

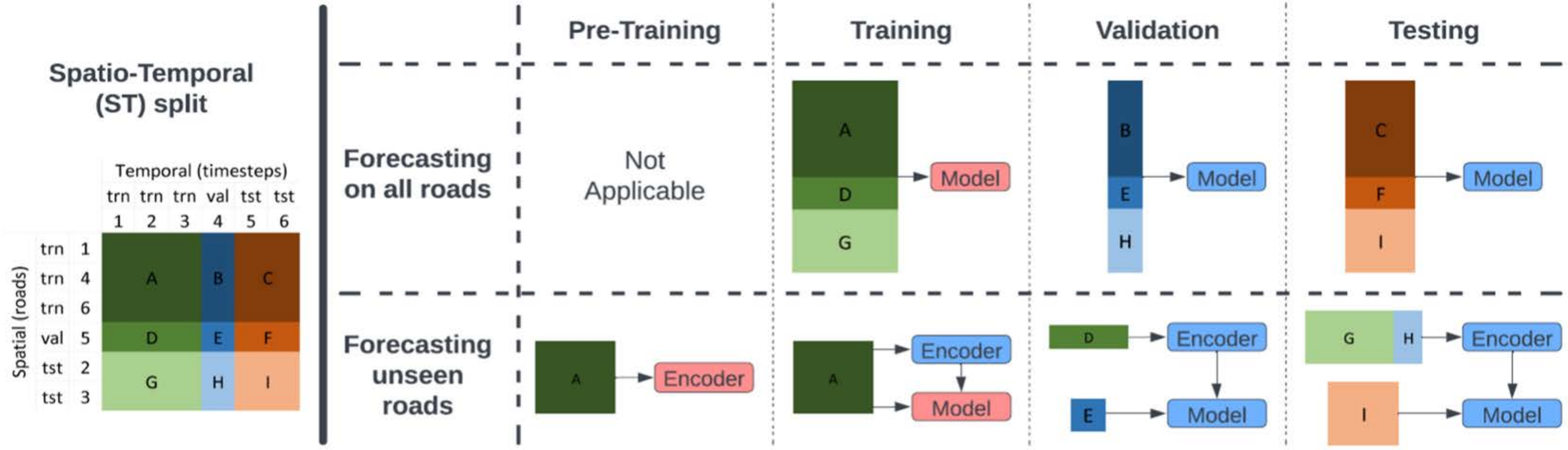


Fig. 2: The ST splitting strategy divides the dataset into nine subsets (left side), while the right side illustrates the usage of different subset combinations at different stages.

This is SimCLR like, very popular in CV.

However, the encoder is stochastic

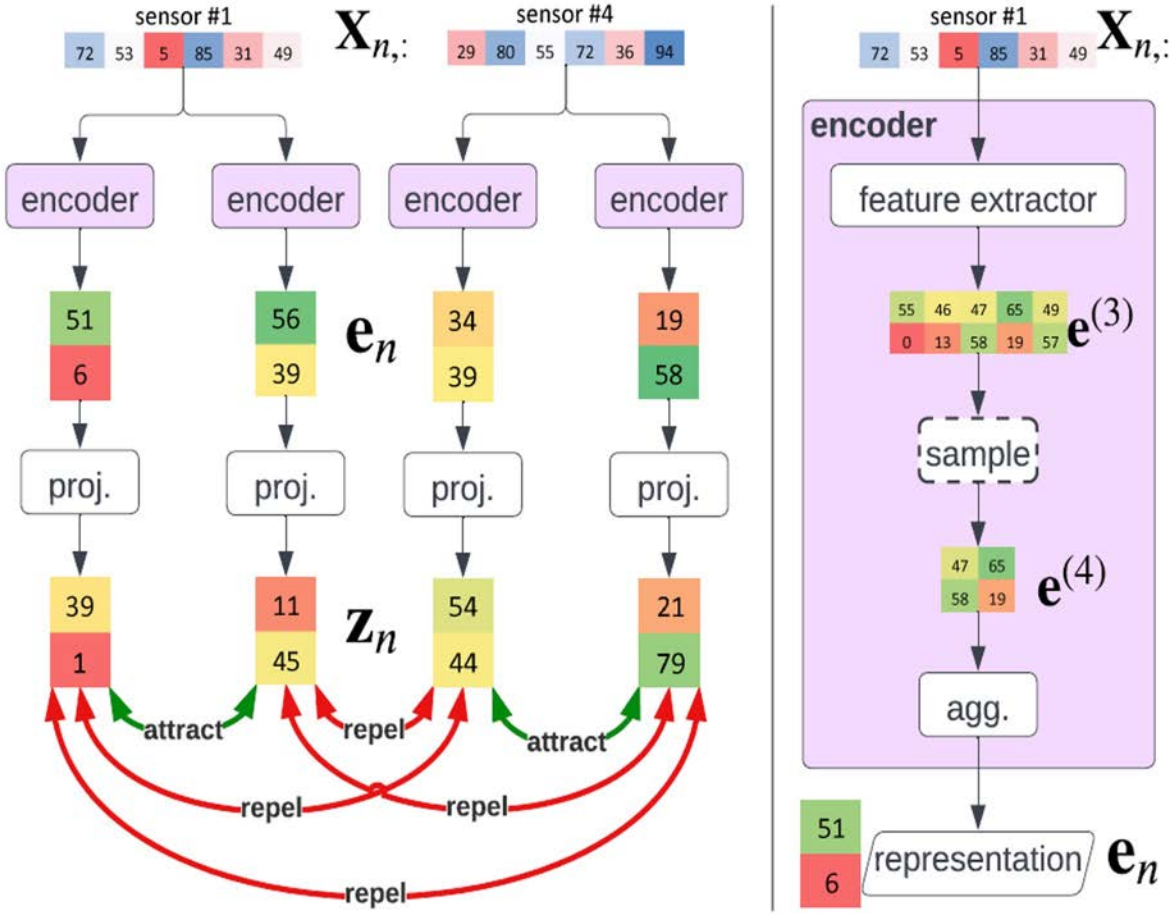


Fig. 3: On the left, the use of contrastive loss to pre-train the spatial encoder is depicted, while on the right, the framework of the (spatial) encoder is illustrated.

We used Graph WaveNet as the backbone

But this is architecture agnostic

Only that such architecture uses node embeddings

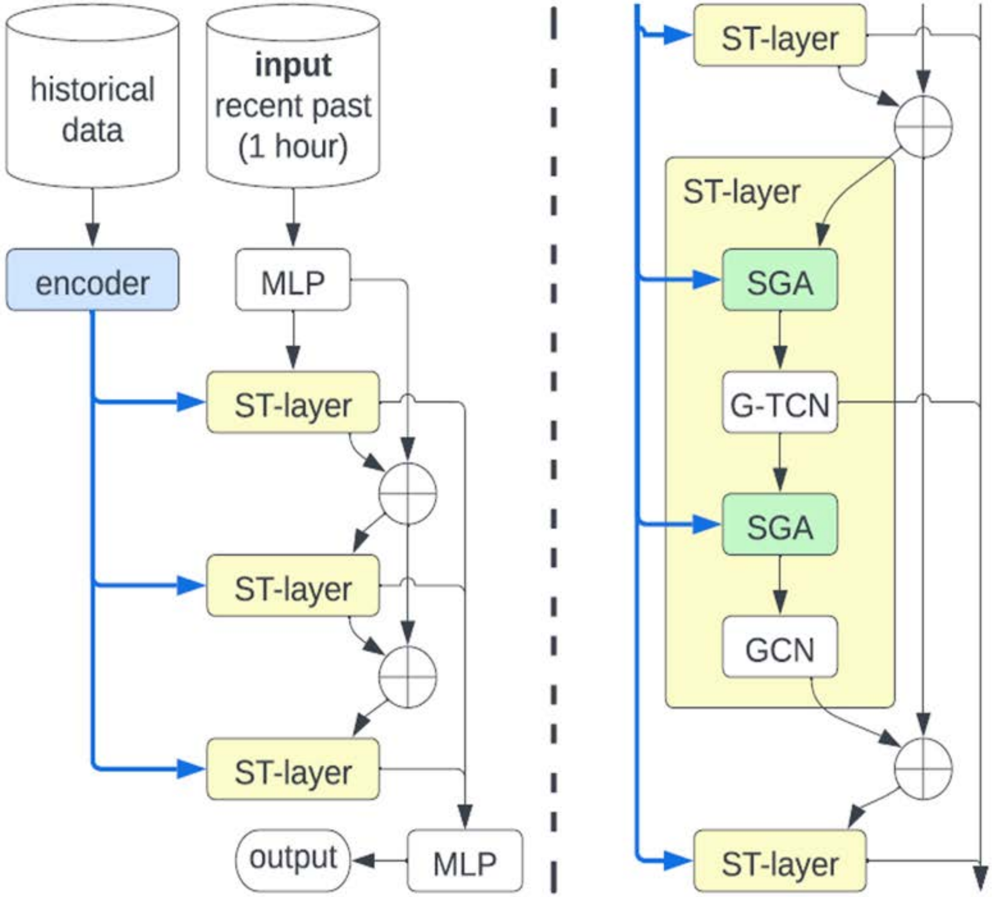


Fig. 4: The left side of the figure illustrates the flow of outputs from the spatial encoder (blue) into the spatio-temporal (ST) layers (yellow). On the right side, the usage of Spatially Gated Addition (SGA) to integrate spatial information from the spatial encoder into the input of the G-TCN and GCN layers within the ST-layer is depicted.

Results

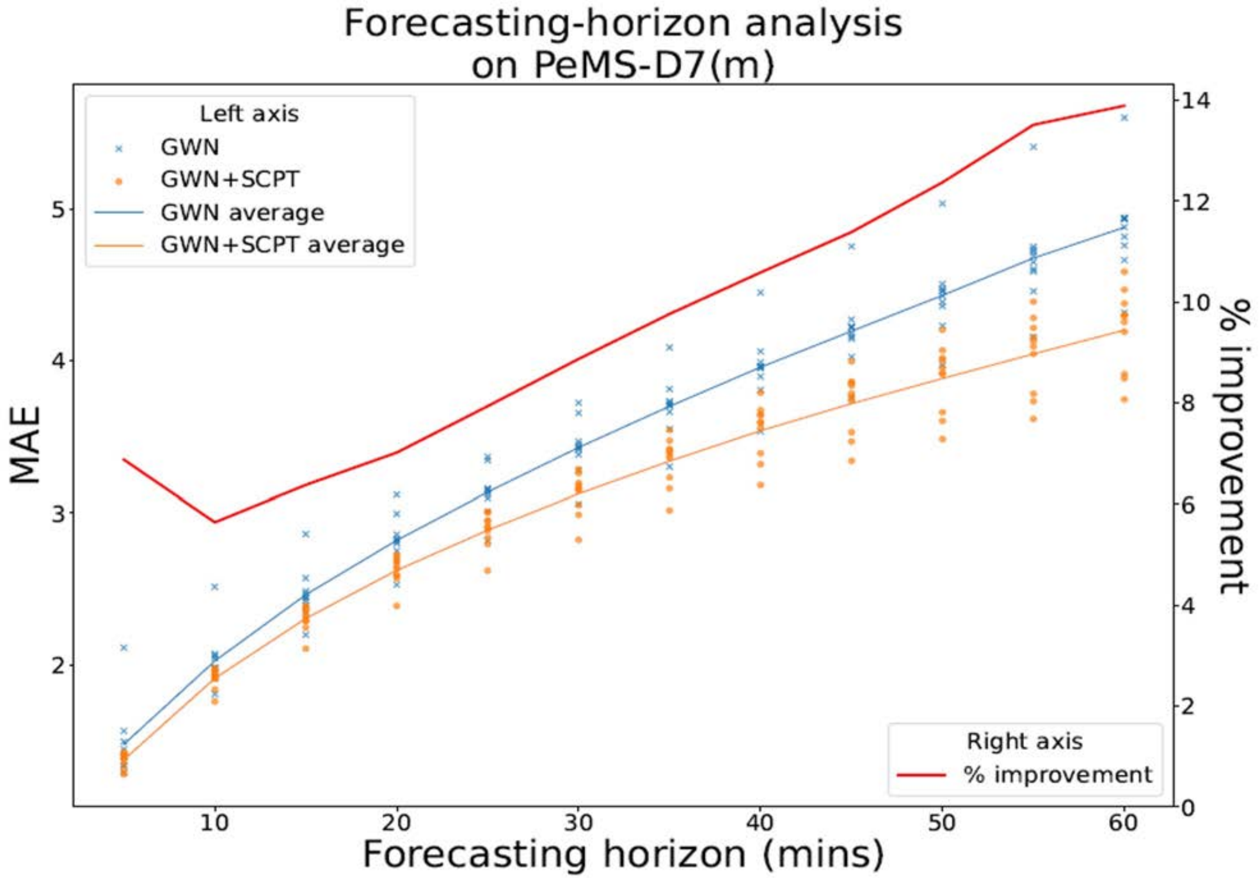


Fig. 5: Performance across forecasting horizons.

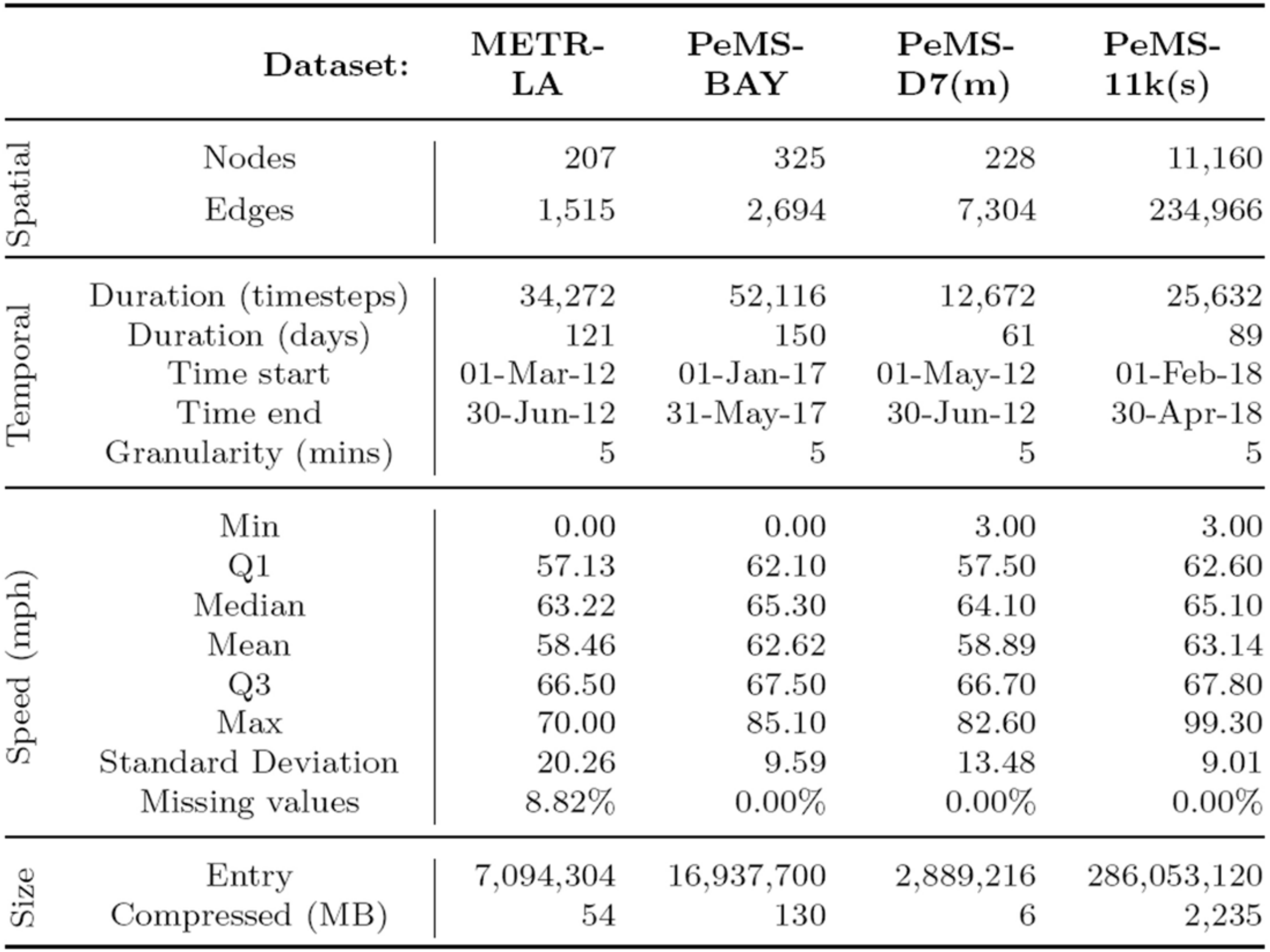


Table 4: Detailed statistics on the real world datasets.

	Dataset:	**METR-LA**	**PeMS-BAY**	**PeMS-D7(m)**	**PeMS-11k(s)**
Spatial	Nodes	207	325	228	11,160
	Edges	1,515	2,694	7,304	234,966
Temporal	Duration (timesteps)	34,272	52,116	12,672	25,632
	Duration (days)	121	150	61	89
	Time start	01-Mar-12	01-Jan-17	01-May-12	01-Feb-18
	Time end	30-Jun-12	31-May-17	30-Jun-12	30-Apr-18
	Granularity (mins)	5	5	5	5
Speed (mph)	Min	0.00	0.00	3.00	3.00
	Q1	57.13	62.10	57.50	62.60
	Median	63.22	65.30	64.10	65.10
	Mean	58.46	62.62	58.89	63.14
	Q3	66.50	67.50	66.70	67.80
	Max	70.00	85.10	82.60	99.30
	Standard Deviation	20.26	9.59	13.48	9.01
	Missing values	8.82%	0.00%	0.00%	0.00%
Size	Entry	7,094,304	16,937,700	2,889,216	286,053,120
	Compressed (MB)	54	130	6	2,235

Most traffic dataset is artificially small.

A sub network is selected for research only.

Favorable trade-off between error and speed.

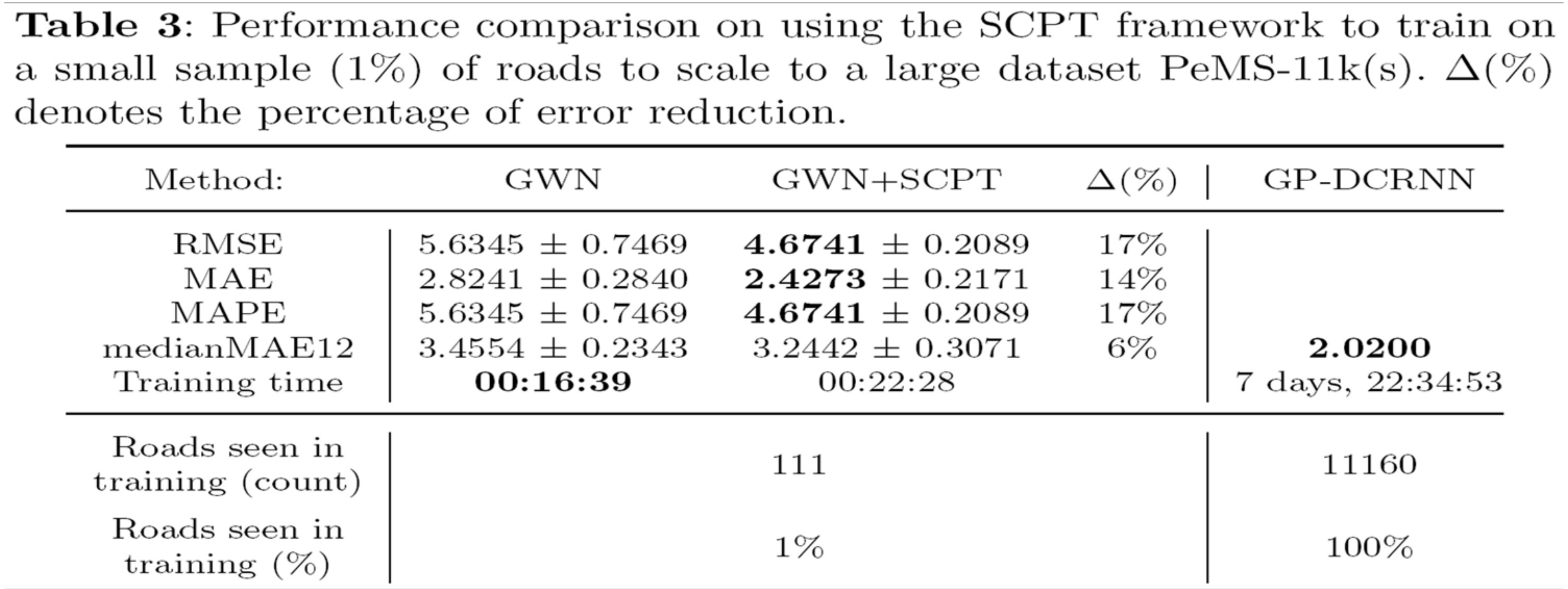
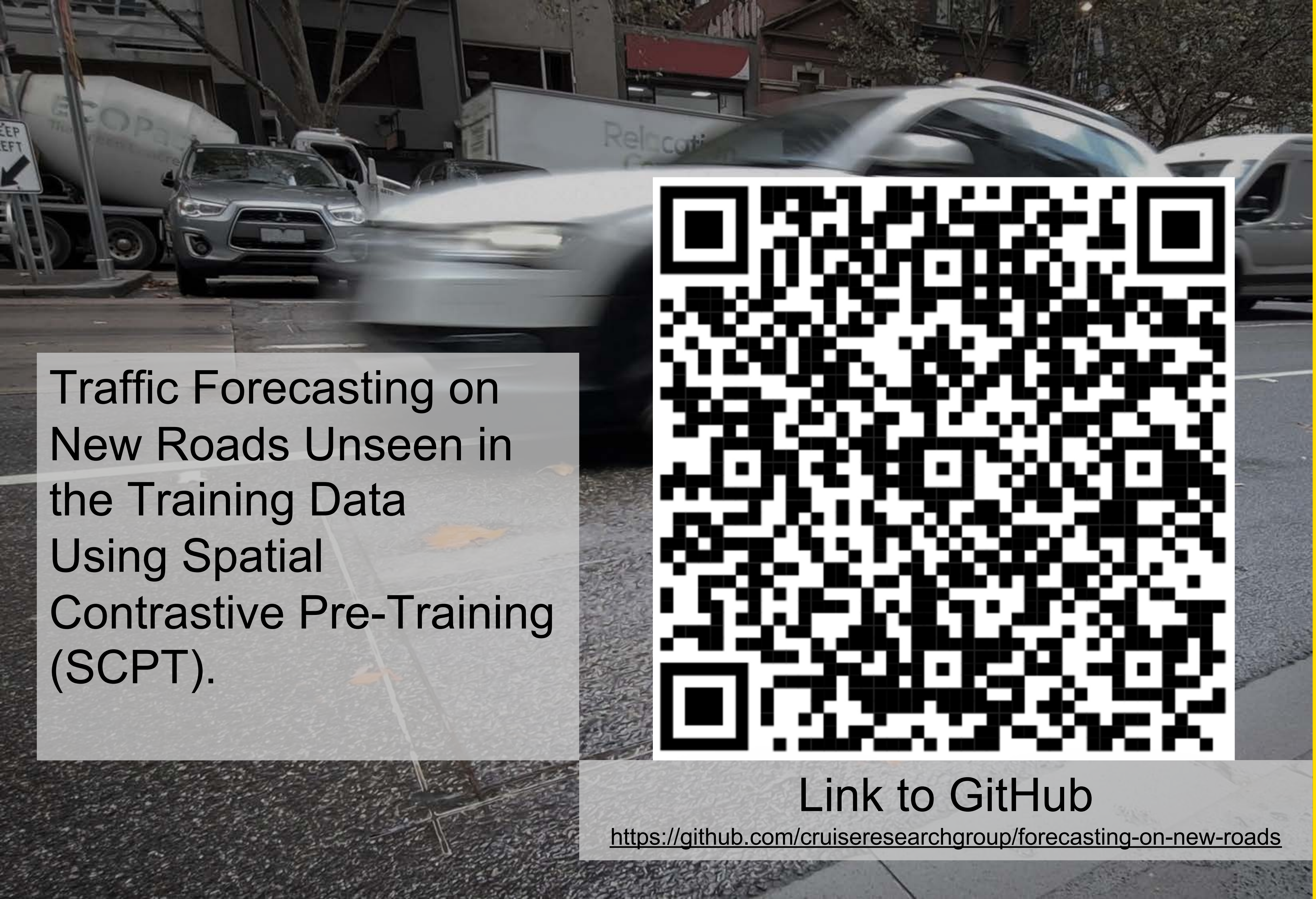


Table 3: Performance comparison on using the SCPT framework to train on a small sample (1%) of roads to scale to a large dataset PeMS-11k(s). Δ(%) denotes the percentage of error reduction.

Method:	GWN	GWN+SCPT	Δ(%)	GP-DCRNN
RMSE	5.6345 ± 0.7469	**4.6741** ± 0.2089	17%	
MAE	2.8241 ± 0.2840	**2.4273** ± 0.2171	14%	
MAPE	5.6345 ± 0.7469	**4.6741** ± 0.2089	17%	
medianMAE12	3.4554 ± 0.2343	3.2442 ± 0.3071	6%	**2.0200**
Training time	**00:16:39**	00:22:28		7 days, 22:34:53
Roads seen in training (count)	111			11160
Roads seen in training (%)	1%			100%

Prabowo, A., Xue, H., Shao, W., Koniusz, P., Salim, F. Traffic forecasting on new roads using spatial contrastive pre-training (SCPT). Data Mining and Knowledge Discovery (2023). https://doi.org/10.1007/s10618-023-00982-0



Traffic Forecasting on New Roads Unseen in the Training Data Using Spatial Contrastive Pre-Training (SCPT).

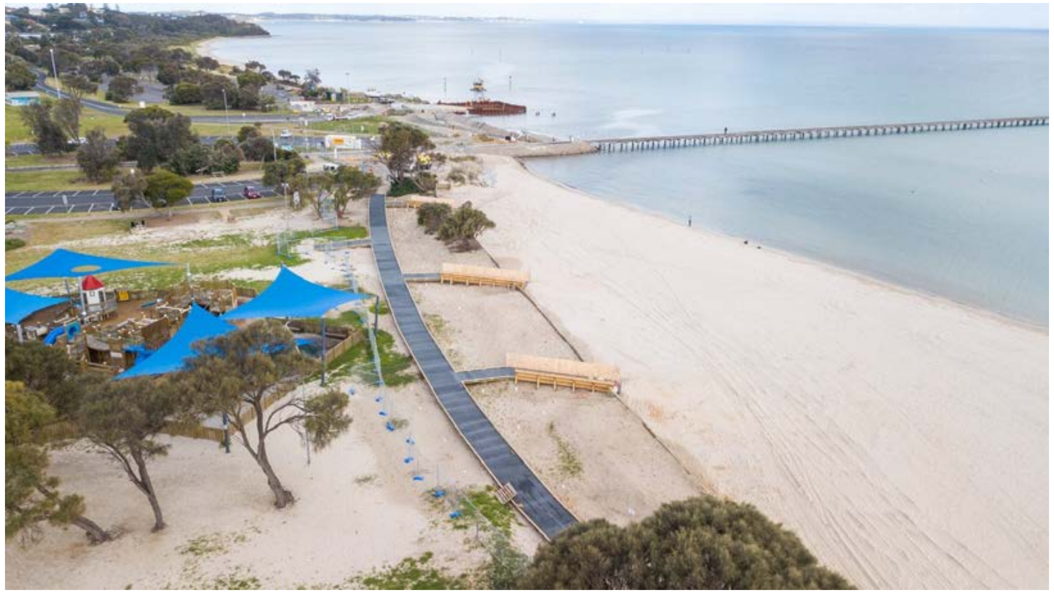
Link to GitHub

https://github.com/cruiseresearchgroup/forecasting-on-new-roads

Arian Prabowo,
Hao Xue,
Wei Shao,
Piotr Koniusz, and
Flora D. Salim.

Parking Availability Prediction

$1M Smarter Cities & Suburbs Project funded by Fed Govt & Mornington Peninsula Shire



Rye Township: a few days (<30 days) data in early 2020, then COVID hit

Rye Data:

- Collected by the Mornington Peninsula Shire;
- Includes **179,288** records across **527** devices in Rye.

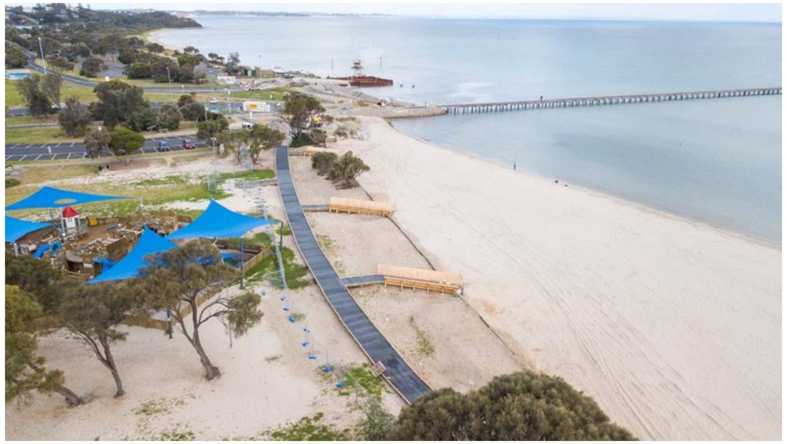
Melbourne CBD: >5 years of data

Melbourne On-street Parking Data:

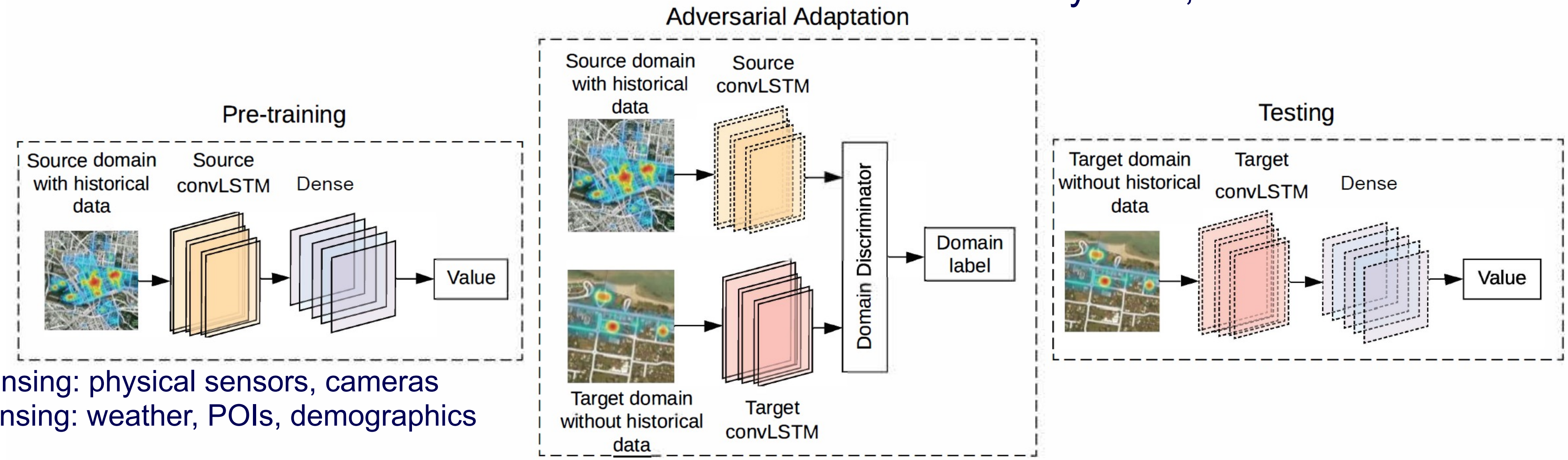
- From the City of Melbourne Open Data;
- Has **35.9** million records of on-street car parking in Melbourne, containing **5044** sensor devices and **4695** parking slots.

Few-Shot Parking Availability Prediction

Melbourne CBD: >5 years of data



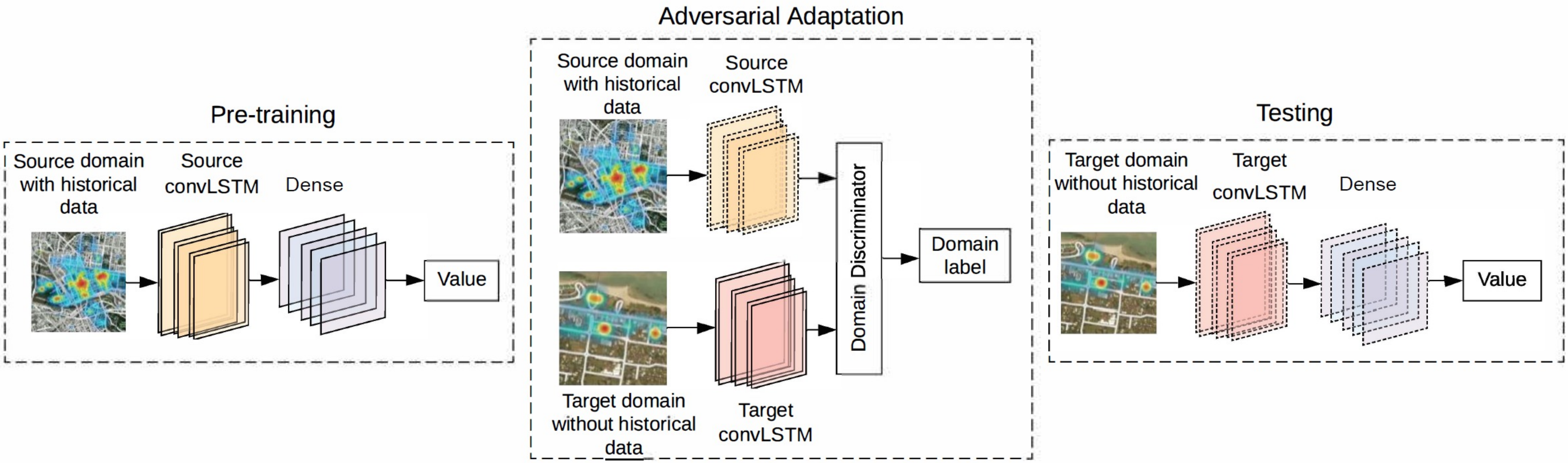
Rye Township: a few days data in early 2020, then COVID hit



Explicit Sensing: physical sensors, cameras
Implicit Sensing: weather, POIs, demographics

Architecture

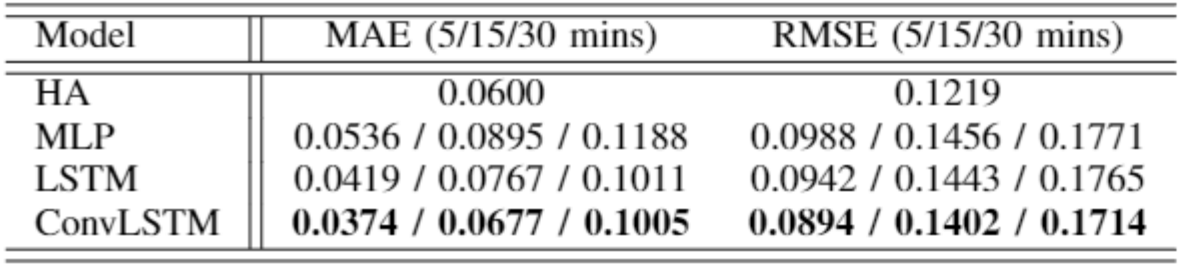
Refined ADDA (Adversarial Discriminative Domain Adaption)



Shao, W., Zhao, S., …. and Salim, F.D., 2021, March. FADACS: A Few-Shot Adversarial Domain Adaptation Architecture for Context-Aware Parking Availability Sensing. In *2021 IEEE International Conference on Pervasive Computing and Communications (PerCom).* *https://arxiv.org/pdf/2007.08551.pdf ;* Codes: https://github.com/cruiseresearchgroup/FADACS_Parking_Prediction

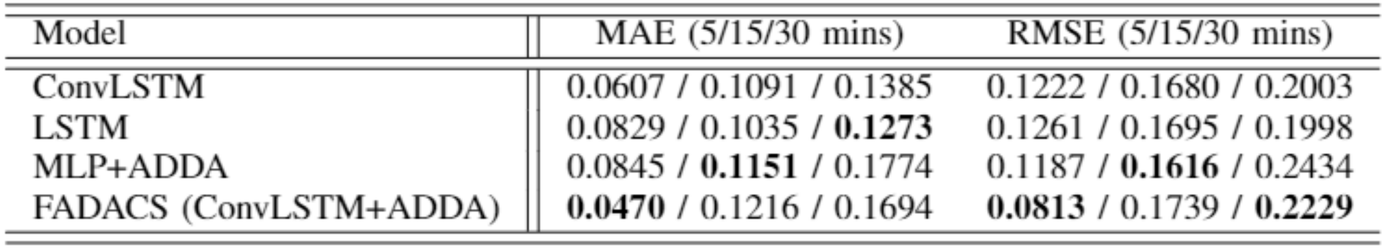
Our results

Full parking data before domain adaptation



Model	MAE (5/15/30 mins)	RMSE (5/15/30 mins)
HA	0.0600	0.1219
MLP	0.0536 / 0.0895 / 0.1188	0.0988 / 0.1456 / 0.1771
LSTM	0.0419 / 0.0767 / 0.1011	0.0942 / 0.1443 / 0.1765
ConvLSTM	**0.0374 / 0.0677 / 0.1005**	**0.0894 / 0.1402 / 0.1714**

6 days parking data with domain adaptation (MelbCity to Rye)



Model	MAE (5/15/30 mins)	RMSE (5/15/30 mins)
ConvLSTM	0.0607 / 0.1091 / 0.1385	0.1222 / 0.1680 / 0.2003
LSTM	0.0829 / 0.1035 / **0.1273**	0.1261 / 0.1695 / 0.1998
MLP+ADDA	0.0845 / **0.1151** / 0.1774	0.1187 / **0.1616** / 0.2434
FADACS (ConvLSTM+ADDA)	**0.0470** / 0.1216 / 0.1694	**0.0813** / 0.1739 / **0.2229**

- Our approach performed the best in general.
- ConvLSTM with parameter transfer perform better than MLP with parameter transfer.
- Adversarial learning is good at learning shared feature spaces.

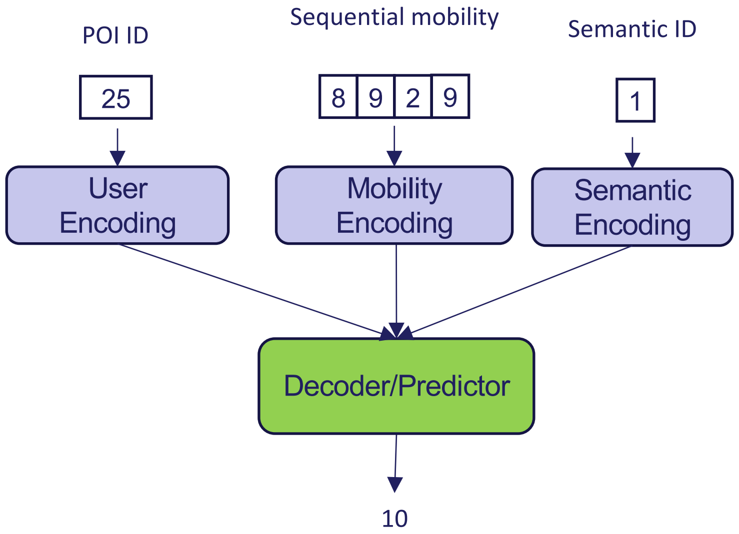
temporal dependency > spatial dependency > domain adaptation

Generative AI and LLMs for mobility and multimodal sensor data

UNSW
AI Institute

UNSW
SYDNEY

Forecasting with Natural Language Prompts



POI ID

25

Sequential mobility

8 9 2 9

Semantic ID

1

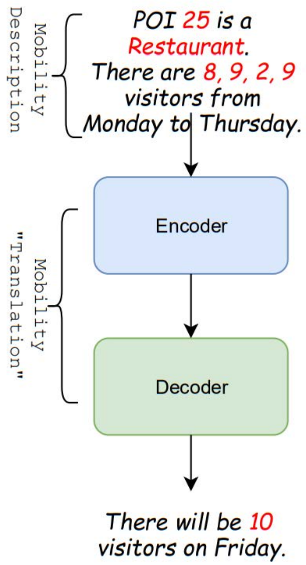
User Encoding

Mobility Encoding

Semantic Encoding

Decoder/Predictor

10



Mobility Description

POI 25 is a Restaurant. There are 8, 9, 2, 9 visitors from Monday to Thursday.

Mobility "Translation"

Encoder

Decoder

There will be 10 visitors on Friday.

Typical Time-series Forecasting Framework:

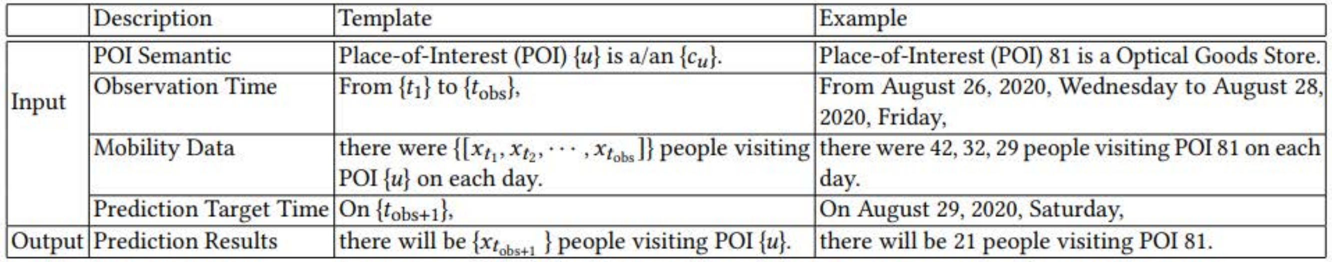
- Numbers in, number(s) out
- Difficult to design different encoders for encoding different information
- How to merge these contextual encoded features? Is concatenation optimal?

Proposed Forecasting via Language Generation:

- ✓ Sentences in, sentence(s) out
- ✓ Easy to leverage existing "Translation" model architectures from NLP
- ✓ No need to worry merging various contextual information – just pass them as sentences!

Hao Xue, Flora D. Salim, Yongli Ren, Charles Clarke. Translating Human Mobility Forecasting through Natural Language Generation, WSDM 2022

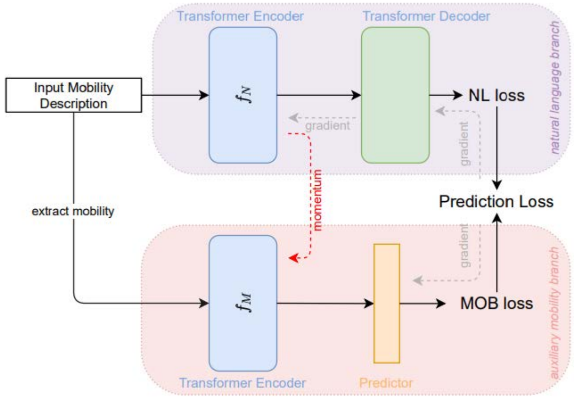
Our Solution: SHIFT (mobility to text)



	Description	Template	Example
Input	POI Semantic	Place-of-Interest (POI) $\{u\}$ is a/an $\{c_u\}$.	Place-of-Interest (POI) 81 is a Optical Goods Store.
	Observation Time	From $\{t_1\}$ to $\{t_{\text{obs}}\}$,	From August 26, 2020, Wednesday to August 28, 2020, Friday,
	Mobility Data	there were $\{[x_{t_1}, x_{t_2}, \cdots, x_{t_{\text{obs}}}]\}$ people visiting POI $\{u\}$ on each day.	there were 42, 32, 29 people visiting POI 81 on each day.
	Prediction Target Time	On $\{t_{\text{obs+1}}\}$,	On August 29, 2020, Saturday,
Output	Prediction Results	there will be $\{x_{t_{\text{obs+1}}}\}$ people visiting POI $\{u\}$.	there will be 21 people visiting POI 81.

Template-based mobility description

Mobility-to-text Description: transform numerical mobility data and other information (e.g., semantic) into natural language sentences, to address the gap between mobility forecasting and language generation.

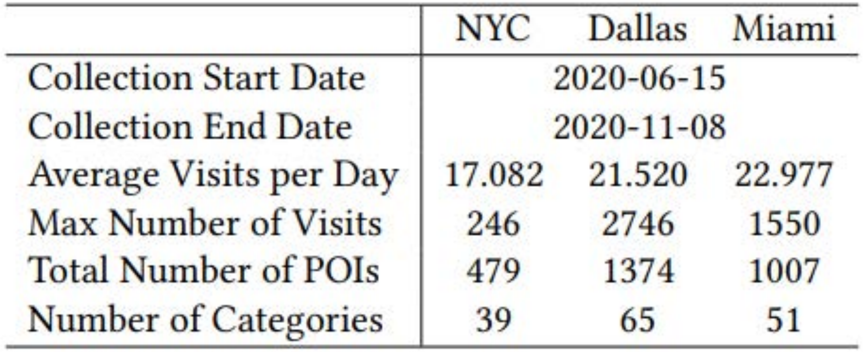


SHIFT Model Design:

- Natural Language Branch (NL): a branch with the sequence-to-sequence structure, which is the main branch of SHIFT to translate the input prompt to generate output sentences
- Auxiliary Mobility Branch (Mob): an auxiliary branch to strengthen the ability of SHIFT in learning mobility patterns for forecasting.

Hao Xue, Flora D. Salim, Yongli Ren, Charles Clarke. Translating Human Mobility Forecasting through Natural Language Generation, WSDM 2022

Dataset and Evaluation



	NYC	Dallas	Miami
Collection Start Date	2020-06-15		
Collection End Date	2020-11-08		
Average Visits per Day	17.082	21.520	22.977
Max Number of Visits	246	2746	1550
Total Number of POIs	479	1374	1007
Number of Categories	39	65	51

Real-world Human Mobility Data

- *Collected from SafeGraph Weekly Patterns*
- *Visitor and demographic aggregations for POIs in the US*
- *Three major cities selected*

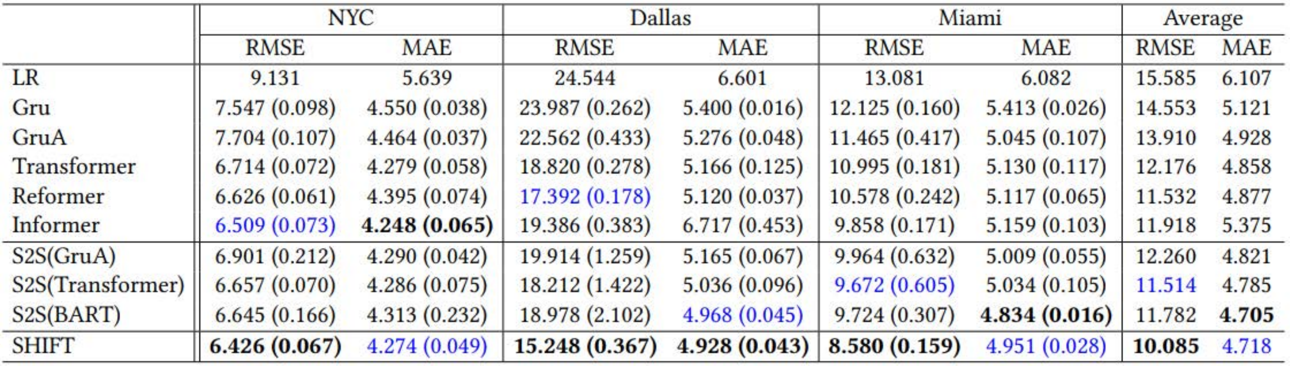
Hao Xue, Flora D. Salim, Yongli Ren, Charles Clarke. Translating Human Mobility Forecasting through Natural Language Generation, WSDM 2022

Dataset and Evaluation

Numerical Time-series Forecasting

Language Seq2Seq

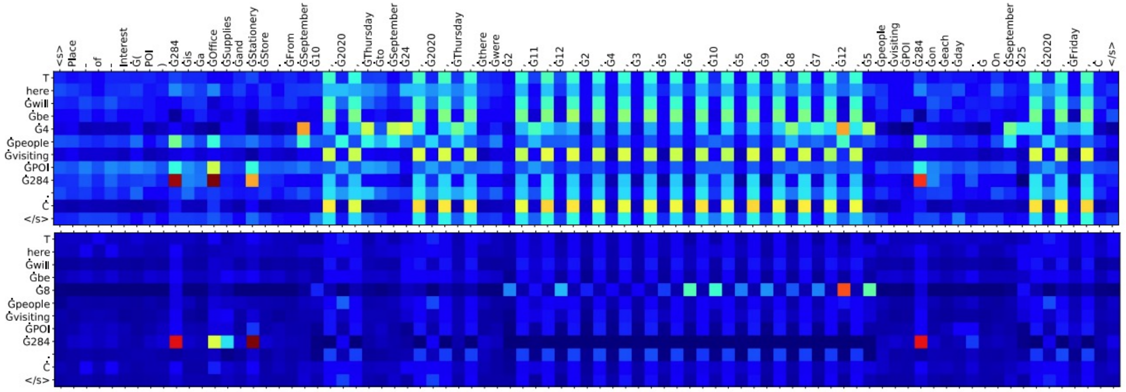
Ours



	NYC		Dallas		Miami		Average	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
LR	9.131	5.639	24.544	6.601	13.081	6.082	15.585	6.107
Gru	7.547 (0.098)	4.550 (0.038)	23.987 (0.262)	5.400 (0.016)	12.125 (0.160)	5.413 (0.026)	14.553	5.121
GruA	7.704 (0.107)	4.464 (0.037)	22.562 (0.433)	5.276 (0.048)	11.465 (0.417)	5.045 (0.107)	13.910	4.928
Transformer	6.714 (0.072)	4.279 (0.058)	18.820 (0.278)	5.166 (0.125)	10.995 (0.181)	5.130 (0.117)	12.176	4.858
Reformer	6.626 (0.061)	4.395 (0.074)	17.392 (0.178)	5.120 (0.037)	10.578 (0.242)	5.117 (0.065)	11.532	4.877
Informer	6.509 (0.073)	**4.248 (0.065)**	19.386 (0.383)	6.717 (0.453)	9.858 (0.171)	5.159 (0.103)	11.918	5.375
S2S(GruA)	6.901 (0.212)	4.290 (0.042)	19.914 (1.259)	5.165 (0.067)	9.964 (0.632)	5.009 (0.055)	12.260	4.821
S2S(Transformer)	6.657 (0.070)	4.286 (0.075)	18.212 (1.422)	5.036 (0.096)	9.672 (0.605)	5.034 (0.105)	11.514	4.785
S2S(BART)	6.645 (0.166)	4.313 (0.232)	18.978 (2.102)	4.968 (0.045)	9.724 (0.307)	**4.834 (0.016)**	11.782	**4.705**
SHIFT	**6.426 (0.067)**	4.274 (0.049)	**15.248 (0.367)**	**4.928 (0.043)**	**8.580 (0.159)**	4.951 (0.028)	**10.085**	4.718

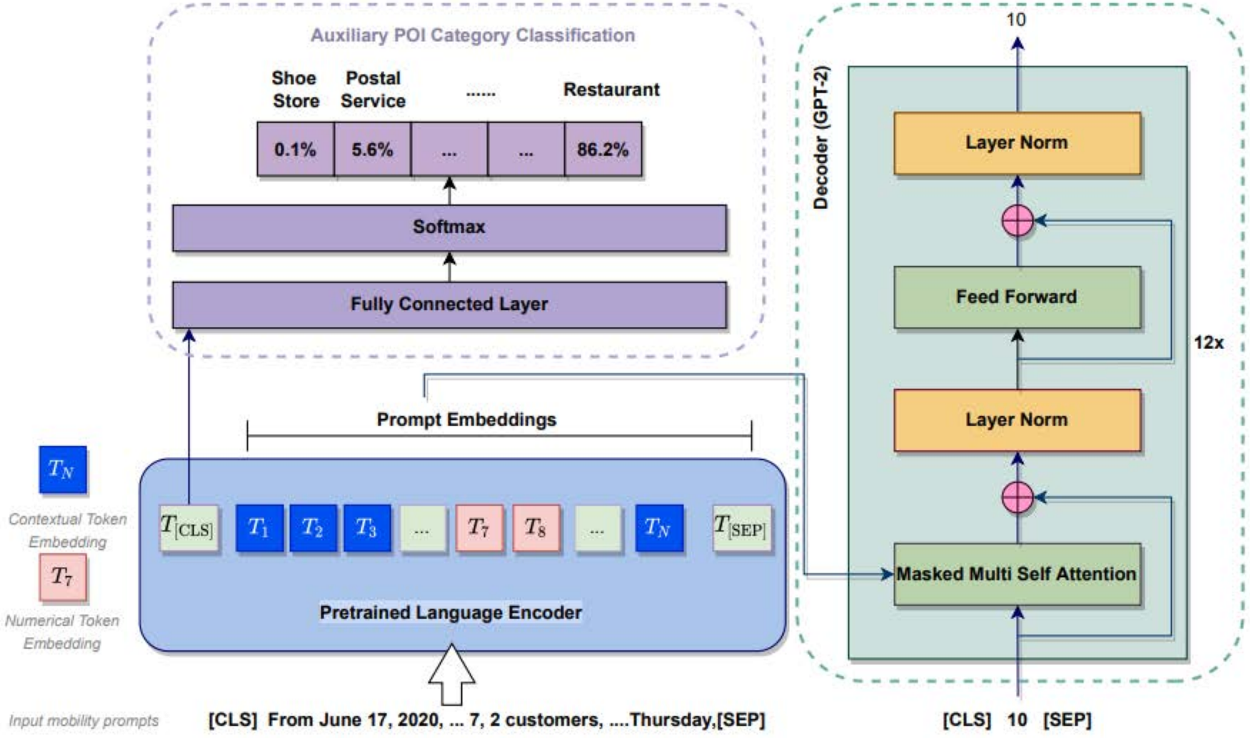
On average

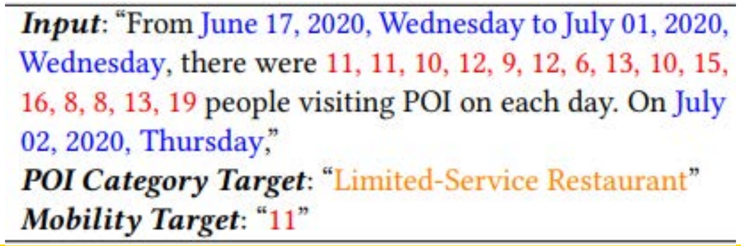
- RMSE: improve 12.4% compared to the send best
- MAE: only about 0.2% worse than the best performer BART



Hao Xue, Flora D. Salim, Yongli Ren, Charles Clarke. Translating Human Mobility Forecasting through Natural Language Generation, WSDM 2022

Leveraging Large Language Models (LLM) with Mobility Prompting for Forecasting





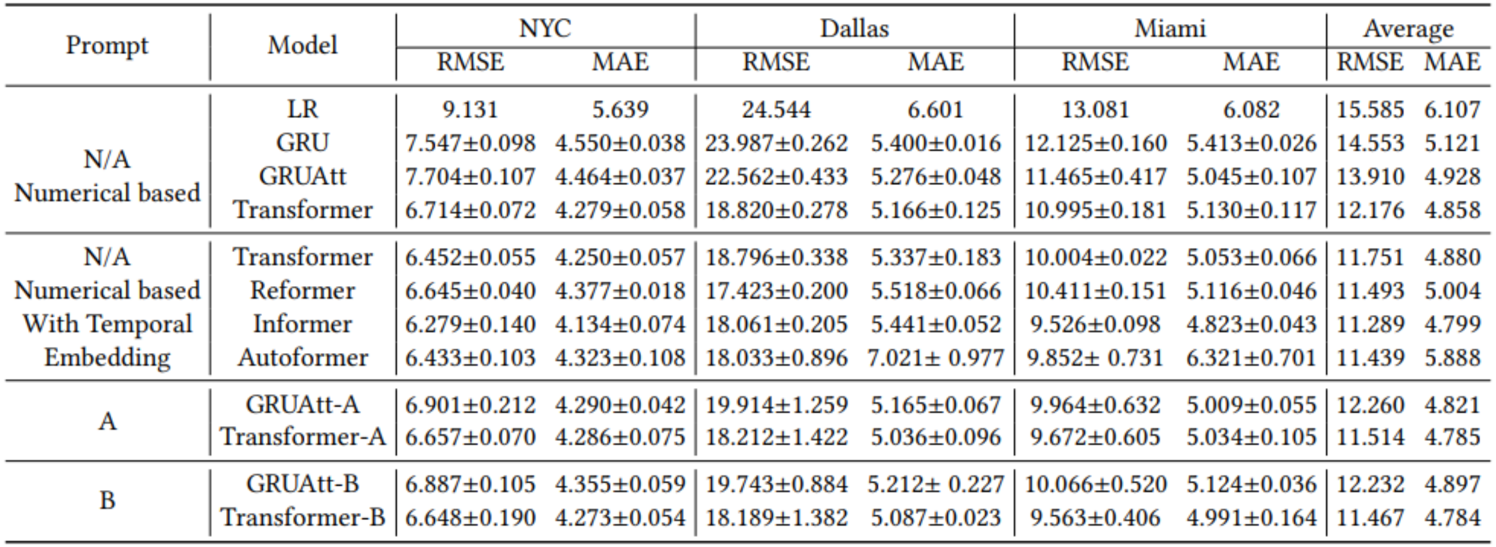
Input: "From June 17, 2020, Wednesday to July 01, 2020, Wednesday, there were 11, 11, 10, 12, 9, 12, 6, 13, 10, 15, 16, 8, 8, 13, 19 people visiting POI on each day. On July 02, 2020, Thursday,"

POI Category Target: "Limited-Service Restaurant"

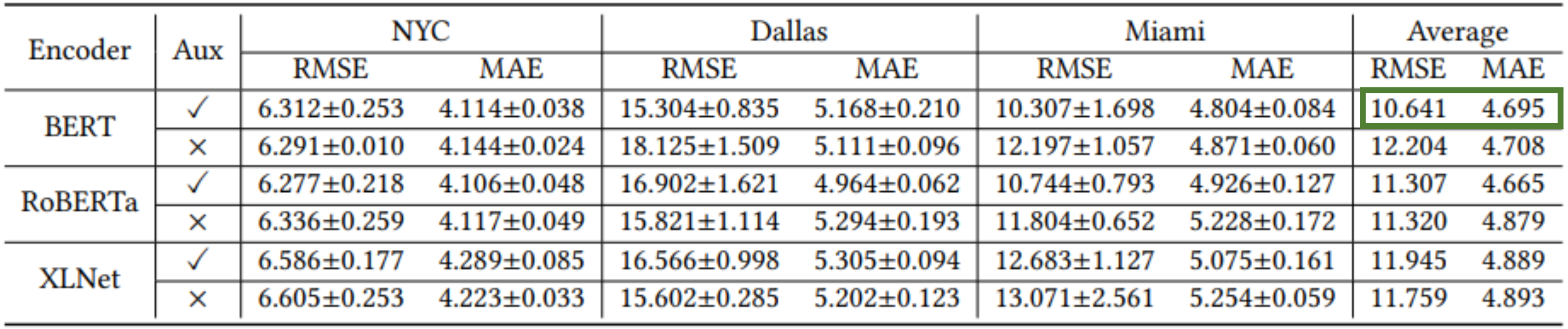
Mobility Target: "11"

Hao Xue, Bhanu Prakash Voutharoja, Flora D. Salim. Leveraging Language Foundation Models for Human Mobility Forecasting, SIGSPATIAL2022

Evaluation



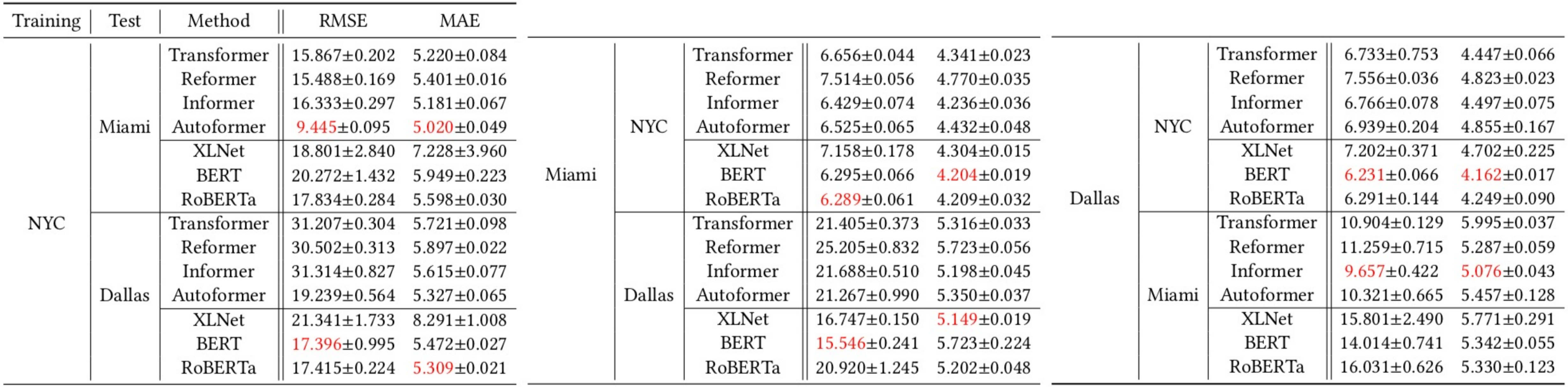
Prompt	Model	NYC		Dallas		Miami		Average	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
N/A Numerical based	LR	9.131	5.639	24.544	6.601	13.081	6.082	15.585	6.107
	GRU	7.547±0.098	4.550±0.038	23.987±0.262	5.400±0.016	12.125±0.160	5.413±0.026	14.553	5.121
	GRUAtt	7.704±0.107	4.464±0.037	22.562±0.433	5.276±0.048	11.465±0.417	5.045±0.107	13.910	4.928
	Transformer	6.714±0.072	4.279±0.058	18.820±0.278	5.166±0.125	10.995±0.181	5.130±0.117	12.176	4.858
N/A Numerical based With Temporal Embedding	Transformer	6.452±0.055	4.250±0.057	18.796±0.338	5.337±0.183	10.004±0.022	5.053±0.066	11.751	4.880
	Reformer	6.645±0.040	4.377±0.018	17.423±0.200	5.518±0.066	10.411±0.151	5.116±0.046	11.493	5.004
	Informer	6.279±0.140	4.134±0.074	18.061±0.205	5.441±0.052	9.526±0.098	4.823±0.043	11.289	4.799
	Autoformer	6.433±0.103	4.323±0.108	18.033±0.896	7.021± 0.977	9.852± 0.731	6.321±0.701	11.439	5.888
A	GRUAtt-A	6.901±0.212	4.290±0.042	19.914±1.259	5.165±0.067	9.964±0.632	5.009±0.055	12.260	4.821
	Transformer-A	6.657±0.070	4.286±0.075	18.212±1.422	5.036±0.096	9.672±0.605	5.034±0.105	11.514	4.785
B	GRUAtt-B	6.887±0.105	4.355±0.059	19.743±0.884	5.212± 0.227	10.066±0.520	5.124±0.036	12.232	4.897
	Transformer-B	6.648±0.190	4.273±0.054	18.189±1.382	5.087±0.023	9.563±0.406	4.991±0.164	11.467	4.784



Encoder	Aux	NYC		Dallas		Miami		Average	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
BERT	✓	6.312±0.253	4.114±0.038	15.304±0.835	5.168±0.210	10.307±1.698	4.804±0.084	10.641	4.695
	×	6.291±0.010	4.144±0.024	18.125±1.509	5.111±0.096	12.197±1.057	4.871±0.060	12.204	4.708
RoBERTa	✓	6.277±0.218	4.106±0.048	16.902±1.621	4.964±0.062	10.744±0.793	4.926±0.127	11.307	4.665
	×	6.336±0.259	4.117±0.049	15.821±1.114	5.294±0.193	11.804±0.652	5.228±0.172	11.320	4.879
XLNet	✓	6.586±0.177	4.289±0.085	16.566±0.998	5.305±0.094	12.683±1.127	5.075±0.161	11.945	4.889
	×	6.605±0.253	4.223±0.033	15.602±0.285	5.202±0.123	13.071±2.561	5.254±0.059	11.759	4.893

Hao Xue, Bhanu Prakash Voutharoja, Flora D. Salim. Leveraging Language Foundation Models for Human Mobility Forecasting, SIGSPATIAL2022

Results: Zero-shot Setting



Training	Test	Method	RMSE	MAE
NYC	Miami	Transformer	15.867±0.202	5.220±0.084
		Reformer	15.488±0.169	5.401±0.016
		Informer	16.333±0.297	5.181±0.067
		Autoformer	9.445±0.095	5.020±0.049
		XLNet	18.801±2.840	7.228±3.960
		BERT	20.272±1.432	5.949±0.223
		RoBERTa	17.834±0.284	5.598±0.030
	Dallas	Transformer	31.207±0.304	5.721±0.098
		Reformer	30.502±0.313	5.897±0.022
		Informer	31.314±0.827	5.615±0.077
		Autoformer	19.239±0.564	5.327±0.065
		XLNet	21.341±1.733	8.291±1.008
		BERT	17.396±0.995	5.472±0.027
		RoBERTa	17.415±0.224	5.309±0.021
Miami	NYC	Transformer	6.656±0.044	4.341±0.023
		Reformer	7.514±0.056	4.770±0.035
		Informer	6.429±0.074	4.236±0.036
		Autoformer	6.525±0.065	4.432±0.048
		XLNet	7.158±0.178	4.304±0.015
		BERT	6.295±0.066	4.204±0.019
		RoBERTa	6.289±0.061	4.209±0.032
	Dallas	Transformer	21.405±0.373	5.316±0.033
		Reformer	25.205±0.832	5.723±0.056
		Informer	21.688±0.510	5.198±0.045
		Autoformer	21.267±0.990	5.350±0.037
		XLNet	16.747±0.150	5.149±0.019
		BERT	15.546±0.241	5.723±0.224
		RoBERTa	20.920±1.245	5.202±0.048
Dallas	NYC	Transformer	6.733±0.753	4.447±0.066
		Reformer	7.556±0.036	4.823±0.023
		Informer	6.766±0.078	4.497±0.075
		Autoformer	6.939±0.204	4.855±0.167
		XLNet	7.202±0.371	4.702±0.225
		BERT	6.231±0.066	4.162±0.017
		RoBERTa	6.291±0.144	4.249±0.090
	Miami	Transformer	10.904±0.129	5.995±0.037
		Reformer	11.259±0.715	5.287±0.059
		Informer	9.657±0.422	5.076±0.043
		Autoformer	10.321±0.665	5.457±0.128
		XLNet	15.801±2.490	5.771±0.291
		BERT	14.014±0.741	5.342±0.055
		RoBERTa	16.031±0.626	5.330±0.123

- train each method on one dataset and test the trained model on the test set of the rest two datasets
- Cold-start scenario
- A good way to evaluate the generalization performance
- A promising direction for future work

Hao Xue, Bhanu Prakash Voutharoja, Flora D. Salim. Leveraging Language Foundation Models for Human Mobility Forecasting, SIGSPATIAL2022

PromptCast: Overview

- **Dataset**
  - 3 forecasting scenarios
  - 311,932 data instances in total

- **Benchmark**
  - numerical-based forecasting methods: 10
  - popular language generation models: 10

https://arxiv.org/abs/2210.08964

Dataset: PISA

*PISA: **P**rompt-based t**I**me **S**eries forec**A**sting*

- City Temperature (CT): provides the daily average temperature (in Fahrenheit degrees) of multiple cities globally. 110 international cities are randomly selected to form the dataset.
- Electricity Consumption Load (ECL): includes the electricity consumption values (in Kwh) of 321 users. Filtered users with missing values and randomly selected 50 users with full records of the entire data collection period.
- SafeGraph Human Mobility Data (SG): contains the daily raw counts of visitors to POIs. Expand the data collection from 5 months in previous papers to almost 15 months and then randomly selected 324 POIs with full records.

https://arxiv.org/abs/2210.08964

https://github.com/HaoUNSW/PISA

Dataset: PISA

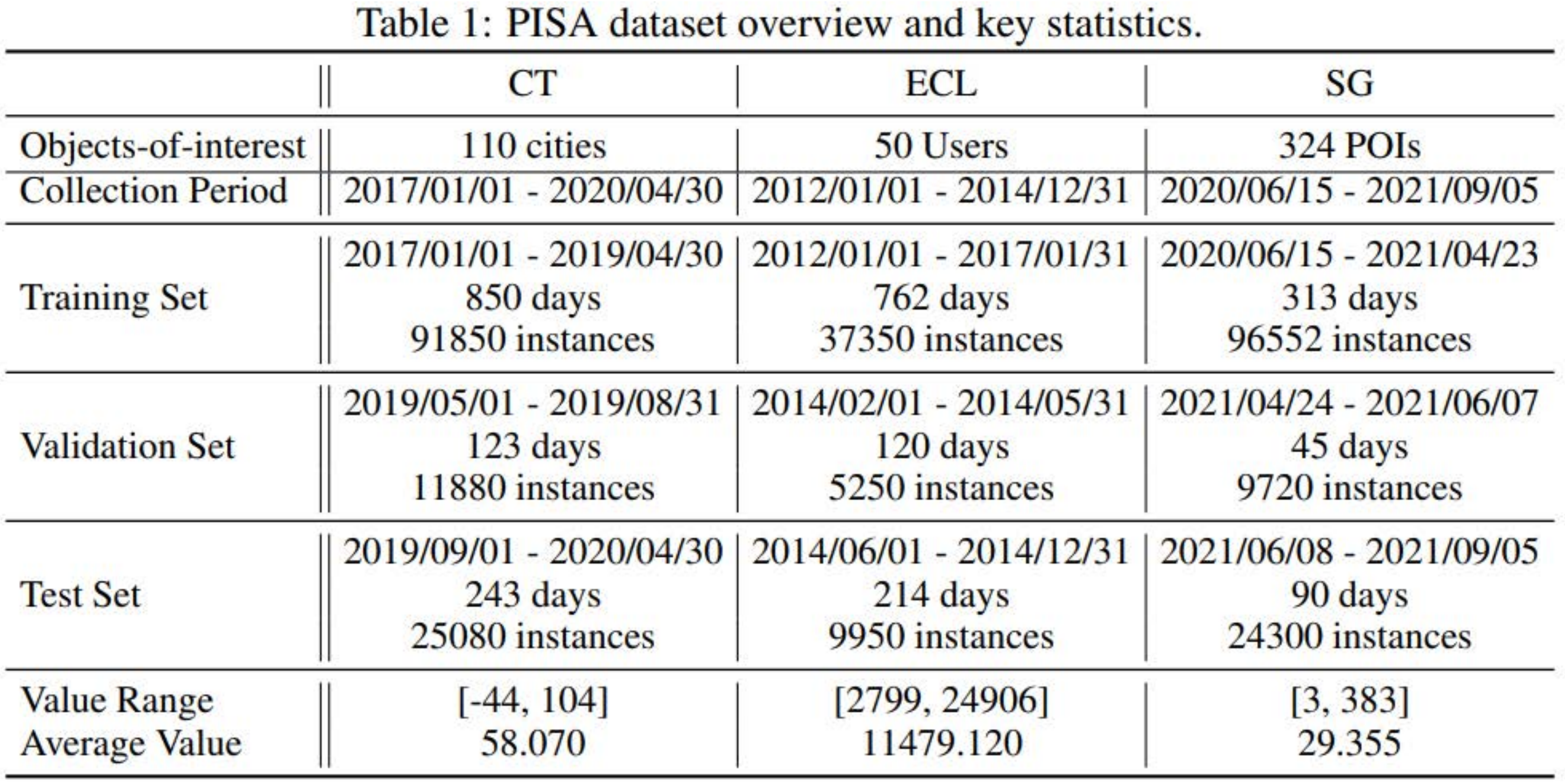


Table 1: PISA dataset overview and key statistics.

	CT	ECL	SG
Objects-of-interest	110 cities	50 Users	324 POIs
Collection Period	2017/01/01 - 2020/04/30	2012/01/01 - 2014/12/31	2020/06/15 - 2021/09/05
Training Set	2017/01/01 - 2019/04/30 850 days 91850 instances	2012/01/01 - 2017/01/31 762 days 37350 instances	2020/06/15 - 2021/04/23 313 days 96552 instances
Validation Set	2019/05/01 - 2019/08/31 123 days 11880 instances	2014/02/01 - 2014/05/31 120 days 5250 instances	2021/04/24 - 2021/06/07 45 days 9720 instances
Test Set	2019/09/01 - 2020/04/30 243 days 25080 instances	2014/06/01 - 2014/12/31 214 days 9950 instances	2021/06/08 - 2021/09/05 90 days 24300 instances
Value Range Average Value	[-44, 104] 58.070	[2799, 24906] 11479.120	[3, 383] 29.355

https://arxiv.org/abs/2210.08964

https://github.com/HaoUNSW/PISA

PISA Template

- Describe Data

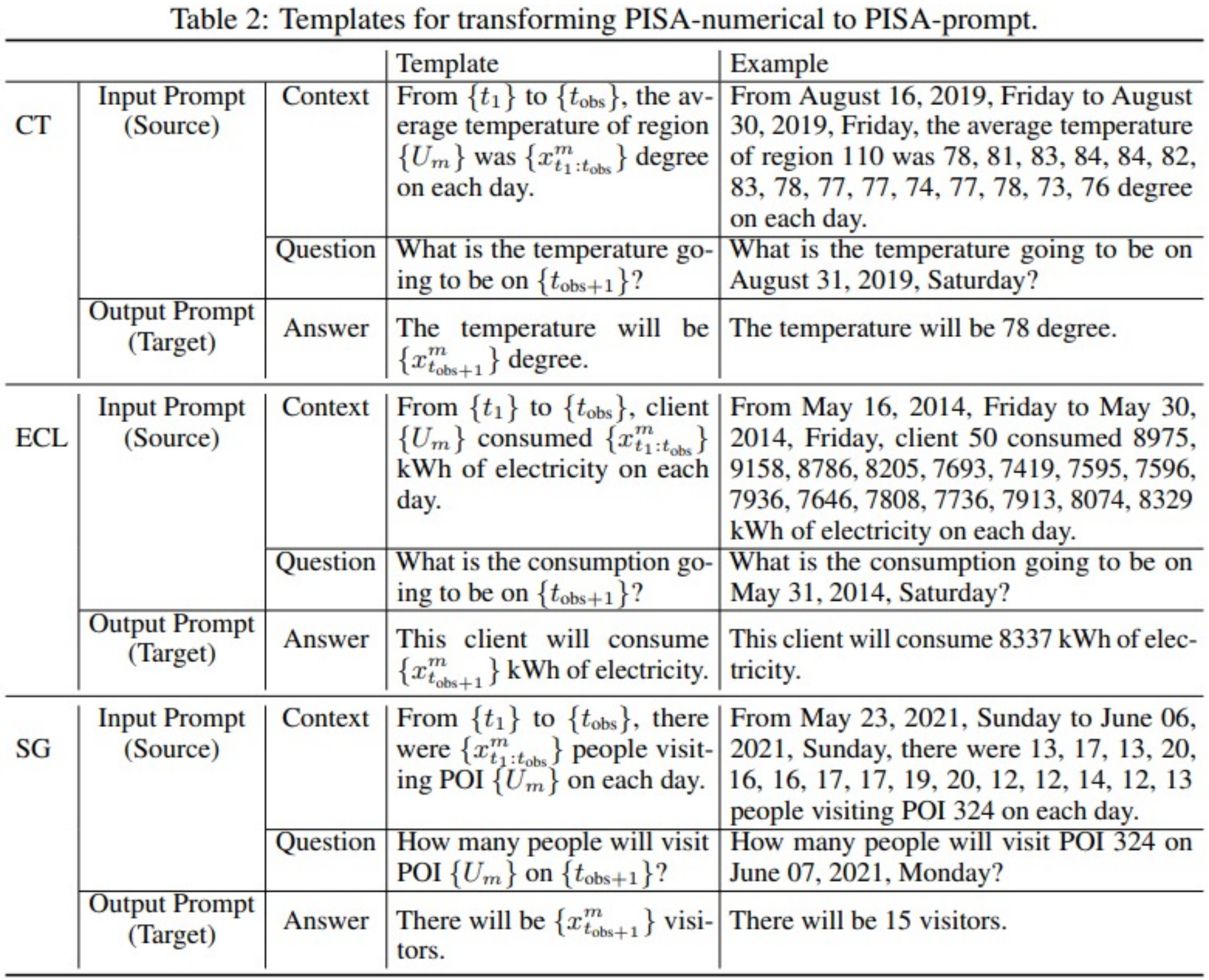


Table 2: Templates for transforming PISA-numerical to PISA-prompt.

			Template	Example
CT	Input Prompt (Source)	Context	From $\{t_1\}$ to $\{t_{\text{obs}}\}$, the average temperature of region $\{U_m\}$ was $\{x^m_{t_1:t_{\text{obs}}}\}$ degree on each day.	From August 16, 2019, Friday to August 30, 2019, Friday, the average temperature of region 110 was 78, 81, 83, 84, 84, 82, 83, 78, 77, 77, 74, 77, 78, 73, 76 degree on each day.
		Question	What is the temperature going to be on $\{t_{\text{obs}+1}\}$?	What is the temperature going to be on August 31, 2019, Saturday?
	Output Prompt (Target)	Answer	The temperature will be $\{x^m_{t_{\text{obs}+1}}\}$ degree.	The temperature will be 78 degree.
ECL	Input Prompt (Source)	Context	From $\{t_1\}$ to $\{t_{\text{obs}}\}$, client $\{U_m\}$ consumed $\{x^m_{t_1:t_{\text{obs}}}\}$ kWh of electricity on each day.	From May 16, 2014, Friday to May 30, 2014, Friday, client 50 consumed 8975, 9158, 8786, 8205, 7693, 7419, 7595, 7596, 7936, 7646, 7808, 7736, 7913, 8074, 8329 kWh of electricity on each day.
		Question	What is the consumption going to be on $\{t_{\text{obs}+1}\}$?	What is the consumption going to be on May 31, 2014, Saturday?
	Output Prompt (Target)	Answer	This client will consume $\{x^m_{t_{\text{obs}+1}}\}$ kWh of electricity.	This client will consume 8337 kWh of electricity.
SG	Input Prompt (Source)	Context	From $\{t_1\}$ to $\{t_{\text{obs}}\}$, there were $\{x^m_{t_1:t_{\text{obs}}}\}$ people visiting POI $\{U_m\}$ on each day.	From May 23, 2021, Sunday to June 06, 2021, Sunday, there were 13, 17, 13, 20, 16, 16, 17, 17, 19, 20, 12, 12, 14, 12, 13 people visiting POI 324 on each day.
		Question	How many people will visit POI $\{U_m\}$ on $\{t_{\text{obs}+1}\}$?	How many people will visit POI 324 on June 07, 2021, Monday?
	Output Prompt (Target)	Answer	There will be $\{x^m_{t_{\text{obs}+1}}\}$ visitors.	There will be 15 visitors.

https://arxiv.org/abs/2210.08964

https://github.com/HaoUNSW/PISA

Results

Naïve Forecasting

Basic Numerical

Basic Numerical

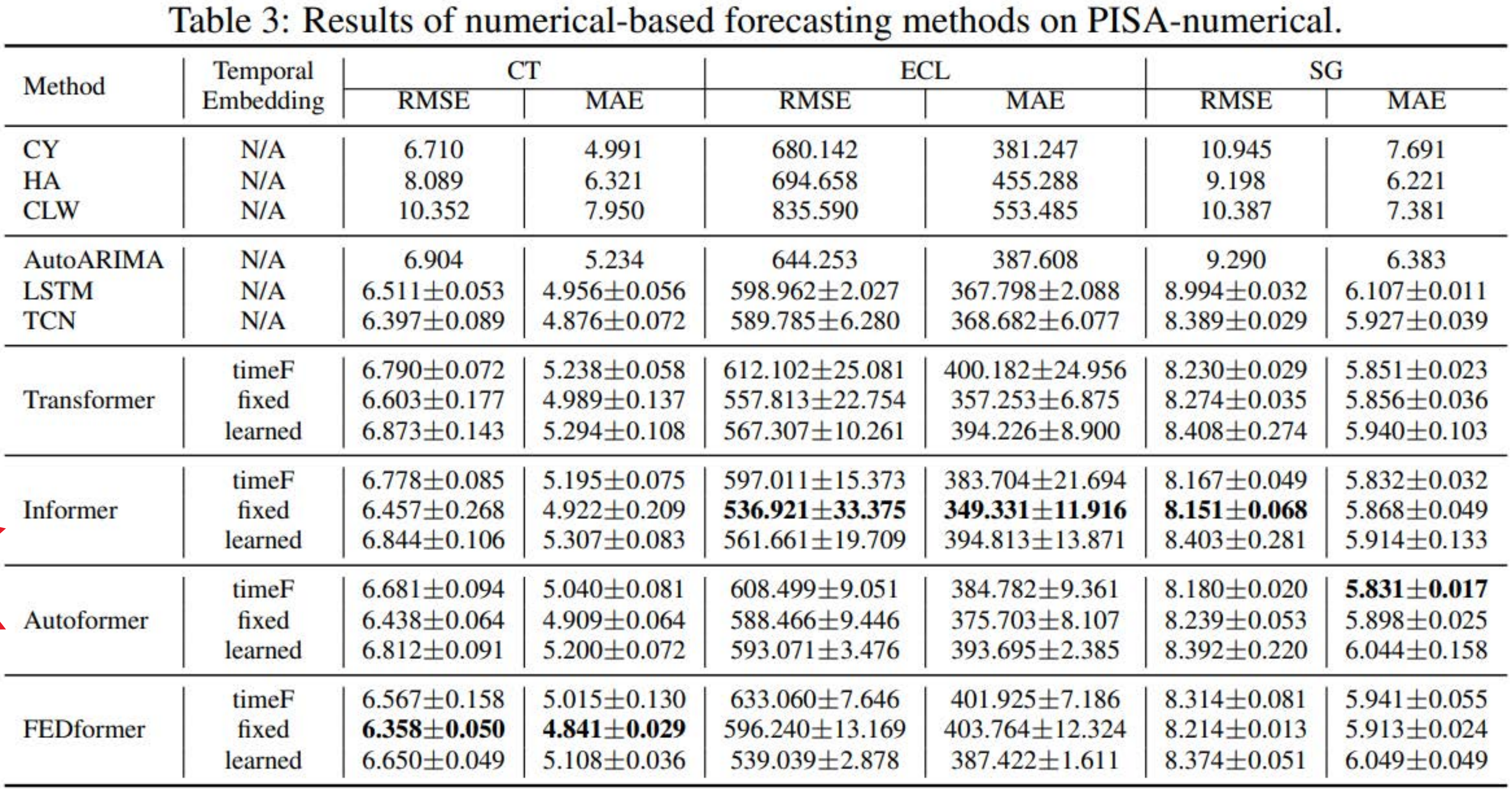


Table 3: Results of numerical-based forecasting methods on PISA-numerical.

Method	Temporal Embedding	CT		ECL		SG	
		RMSE	MAE	RMSE	MAE	RMSE	MAE
CY	N/A	6.710	4.991	680.142	381.247	10.945	7.691
HA	N/A	8.089	6.321	694.658	455.288	9.198	6.221
CLW	N/A	10.352	7.950	835.590	553.485	10.387	7.381
AutoARIMA	N/A	6.904	5.234	644.253	387.608	9.290	6.383
LSTM	N/A	6.511±0.053	4.956±0.056	598.962±2.027	367.798±2.088	8.994±0.032	6.107±0.011
TCN	N/A	6.397±0.089	4.876±0.072	589.785±6.280	368.682±6.077	8.389±0.029	5.927±0.039
Transformer	timeF	6.790±0.072	5.238±0.058	612.102±25.081	400.182±24.956	8.230±0.029	5.851±0.023
	fixed	6.603±0.177	4.989±0.137	557.813±22.754	357.253±6.875	8.274±0.035	5.856±0.036
	learned	6.873±0.143	5.294±0.108	567.307±10.261	394.226±8.900	8.408±0.274	5.940±0.103
Informer	timeF	6.778±0.085	5.195±0.075	597.011±15.373	383.704±21.694	8.167±0.049	5.832±0.032
	fixed	6.457±0.268	4.922±0.209	**536.921±33.375**	**349.331±11.916**	**8.151±0.068**	5.868±0.049
	learned	6.844±0.106	5.307±0.083	561.661±19.709	394.813±13.871	8.403±0.281	5.914±0.133
Autoformer	timeF	6.681±0.094	5.040±0.081	608.499±9.051	384.782±9.361	8.180±0.020	**5.831±0.017**
	fixed	6.438±0.064	4.909±0.064	588.466±9.446	375.703±8.107	8.239±0.053	5.898±0.025
	learned	6.812±0.091	5.200±0.072	593.071±3.476	393.695±2.385	8.392±0.220	6.044±0.158
FEDformer	timeF	6.567±0.158	5.015±0.130	633.060±7.646	401.925±7.186	8.314±0.081	5.941±0.055
	fixed	**6.358±0.050**	**4.841±0.029**	596.240±13.169	403.764±12.324	8.214±0.013	5.913±0.024
	learned	6.650±0.049	5.108±0.036	539.039±2.878	387.422±1.611	8.374±0.051	6.049±0.049

Results

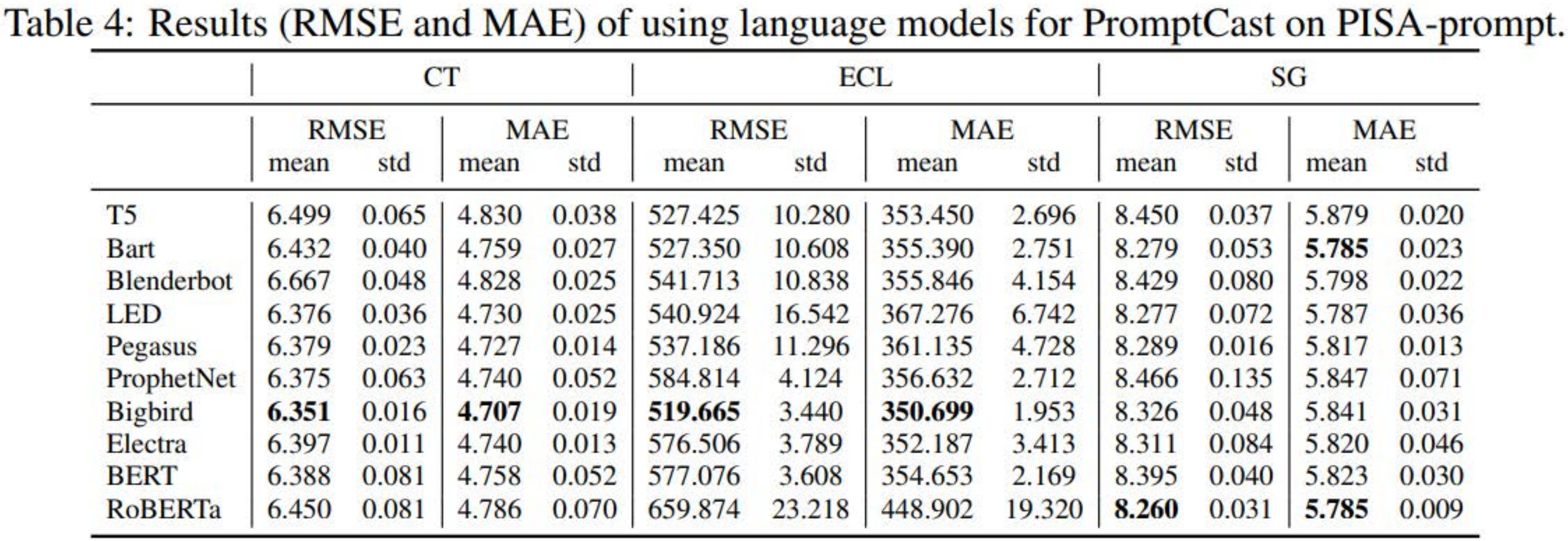


Table 4: Results (RMSE and MAE) of using language models for PromptCast on PISA-prompt.

	CT				ECL				SG			
	RMSE		MAE		RMSE		MAE		RMSE		MAE	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
T5	6.499	0.065	4.830	0.038	527.425	10.280	353.450	2.696	8.450	0.037	5.879	0.020
Bart	6.432	0.040	4.759	0.027	527.350	10.608	355.390	2.751	8.279	0.053	**5.785**	0.023
Blenderbot	6.667	0.048	4.828	0.025	541.713	10.838	355.846	4.154	8.429	0.080	5.798	0.022
LED	6.376	0.036	4.730	0.025	540.924	16.542	367.276	6.742	8.277	0.072	5.787	0.036
Pegasus	6.379	0.023	4.727	0.014	537.186	11.296	361.135	4.728	8.289	0.016	5.817	0.013
ProphetNet	6.375	0.063	4.740	0.052	584.814	4.124	356.632	2.712	8.466	0.135	5.847	0.071
Bigbird	**6.351**	0.016	**4.707**	0.019	**519.665**	3.440	**350.699**	1.953	8.326	0.048	5.841	0.031
Electra	6.397	0.011	4.740	0.013	576.506	3.789	352.187	3.413	8.311	0.084	5.820	0.046
BERT	6.388	0.081	4.758	0.052	577.076	3.608	354.653	2.169	8.395	0.040	5.823	0.030
RoBERTa	6.450	0.081	4.786	0.070	659.874	23.218	448.902	19.320	**8.260**	0.031	**5.785**	0.009

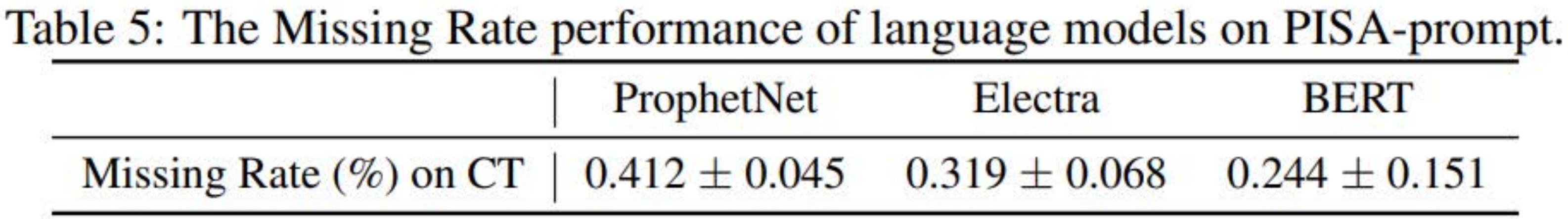


Table 5: The Missing Rate performance of language models on PISA-prompt.

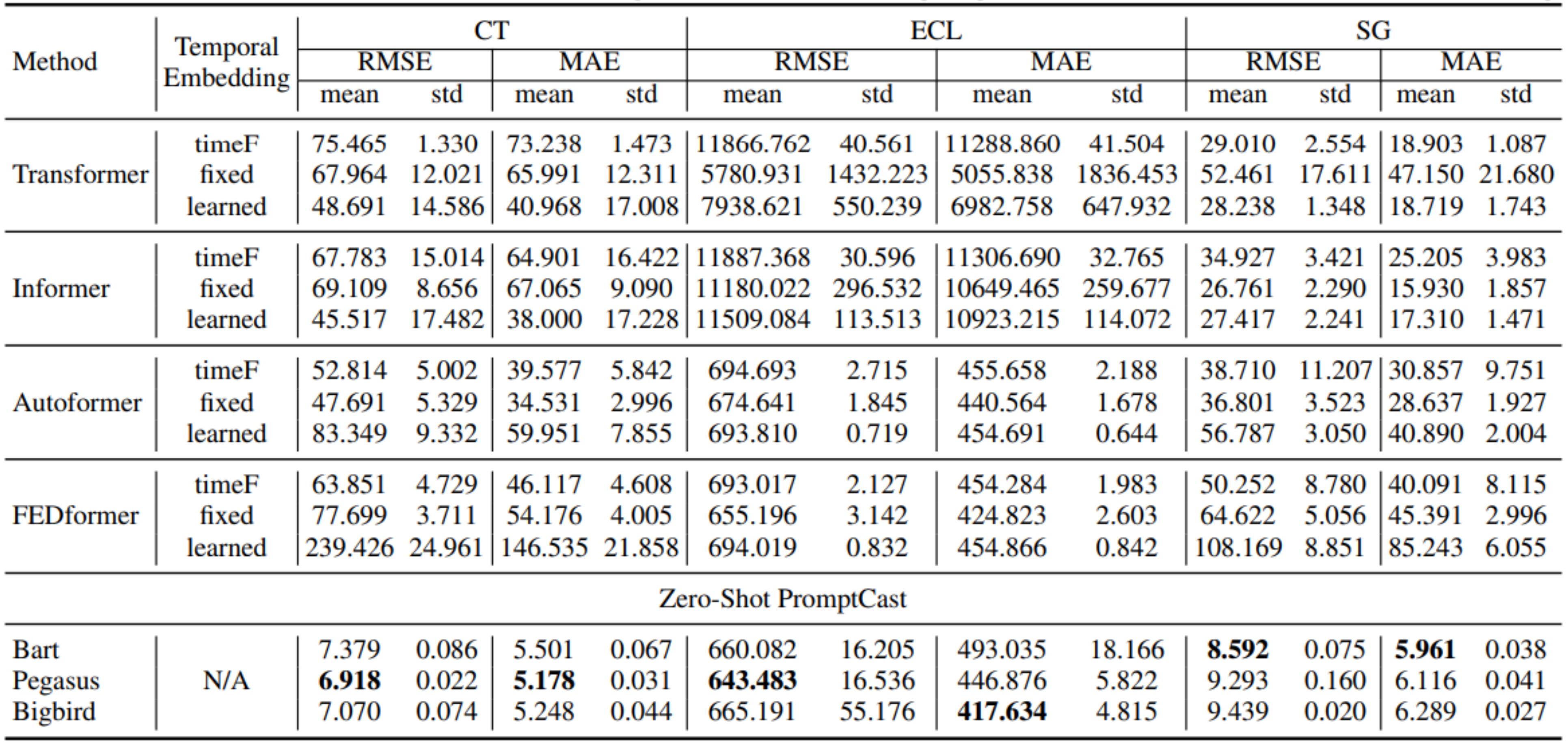
	ProphetNet	Electra	BERT
Missing Rate (%) on CT	0.412 ± 0.045	0.319 ± 0.068	0.244 ± 0.151

https://arxiv.org/abs/2210.08964

https://github.com/HaoUNSW/PISA

Results: Zero-shot Setting

Table 6: Results of numerical forecasting methods and language models under the zero-shot setting.



Method	Temporal Embedding	CT				ECL				SG			
		RMSE		MAE		RMSE		MAE		RMSE		MAE	
		mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
Transformer	timeF	75.465	1.330	73.238	1.473	11866.762	40.561	11288.860	41.504	29.010	2.554	18.903	1.087
	fixed	67.964	12.021	65.991	12.311	5780.931	1432.223	5055.838	1836.453	52.461	17.611	47.150	21.680
	learned	48.691	14.586	40.968	17.008	7938.621	550.239	6982.758	647.932	28.238	1.348	18.719	1.743
Informer	timeF	67.783	15.014	64.901	16.422	11887.368	30.596	11306.690	32.765	34.927	3.421	25.205	3.983
	fixed	69.109	8.656	67.065	9.090	11180.022	296.532	10649.465	259.677	26.761	2.290	15.930	1.857
	learned	45.517	17.482	38.000	17.228	11509.084	113.513	10923.215	114.072	27.417	2.241	17.310	1.471
Autoformer	timeF	52.814	5.002	39.577	5.842	694.693	2.715	455.658	2.188	38.710	11.207	30.857	9.751
	fixed	47.691	5.329	34.531	2.996	674.641	1.845	440.564	1.678	36.801	3.523	28.637	1.927
	learned	83.349	9.332	59.951	7.855	693.810	0.719	454.691	0.644	56.787	3.050	40.890	2.004
FEDformer	timeF	63.851	4.729	46.117	4.608	693.017	2.127	454.284	1.983	50.252	8.780	40.091	8.115
	fixed	77.699	3.711	54.176	4.005	655.196	3.142	424.823	2.603	64.622	5.056	45.391	2.996
	learned	239.426	24.961	146.535	21.858	694.019	0.832	454.866	0.842	108.169	8.851	85.243	6.055
Zero-Shot PromptCast													
Bart	N/A	7.379	0.086	5.501	0.067	660.082	16.205	493.035	18.166	**8.592**	0.075	**5.961**	0.038
Pegasus		**6.918**	0.022	**5.178**	0.031	**643.483**	16.536	446.876	5.822	9.293	0.160	6.116	0.041
Bigbird		7.070	0.074	5.248	0.044	665.191	55.176	**417.634**	4.815	9.439	0.020	6.289	0.027

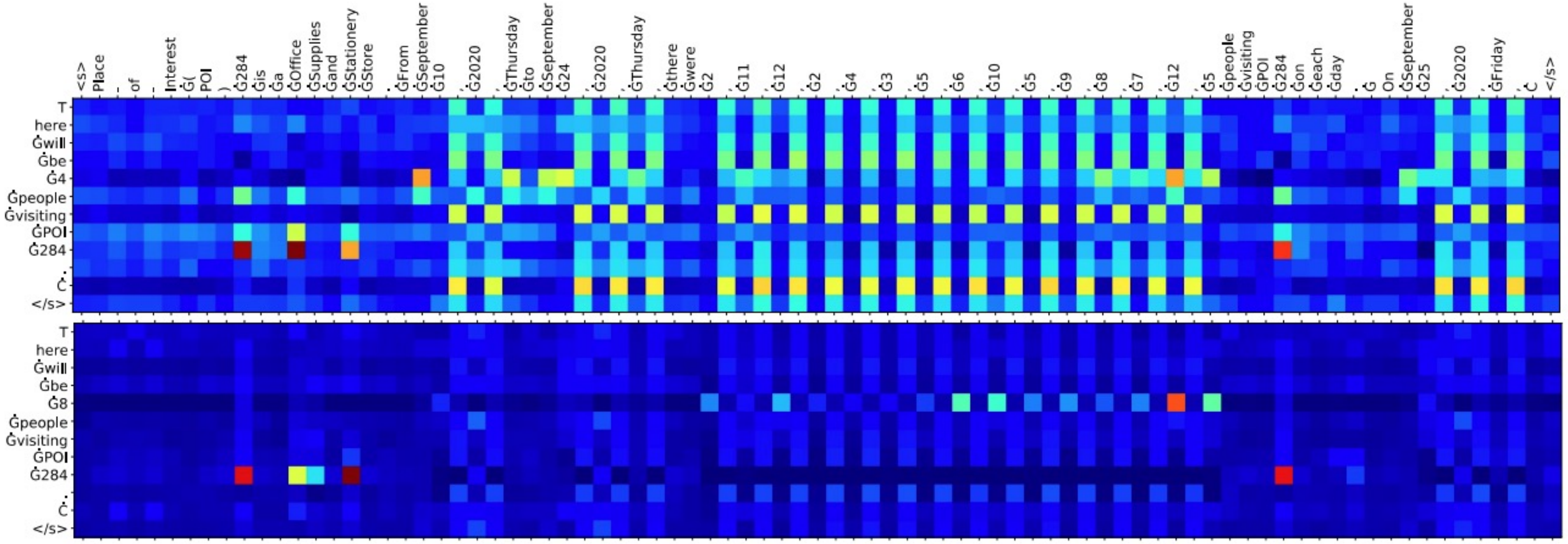
https://arxiv.org/abs/2210.08964

https://github.com/HaoUNSW/PISA

Why it works

- Intra-relation (between numerical values) and inter-relation (numerical values and auxiliary information) modeled simultaneously

- Limitation: Still need further investigations



Example: attentions in SHIFT model

https://arxiv.org/abs/2210.08964

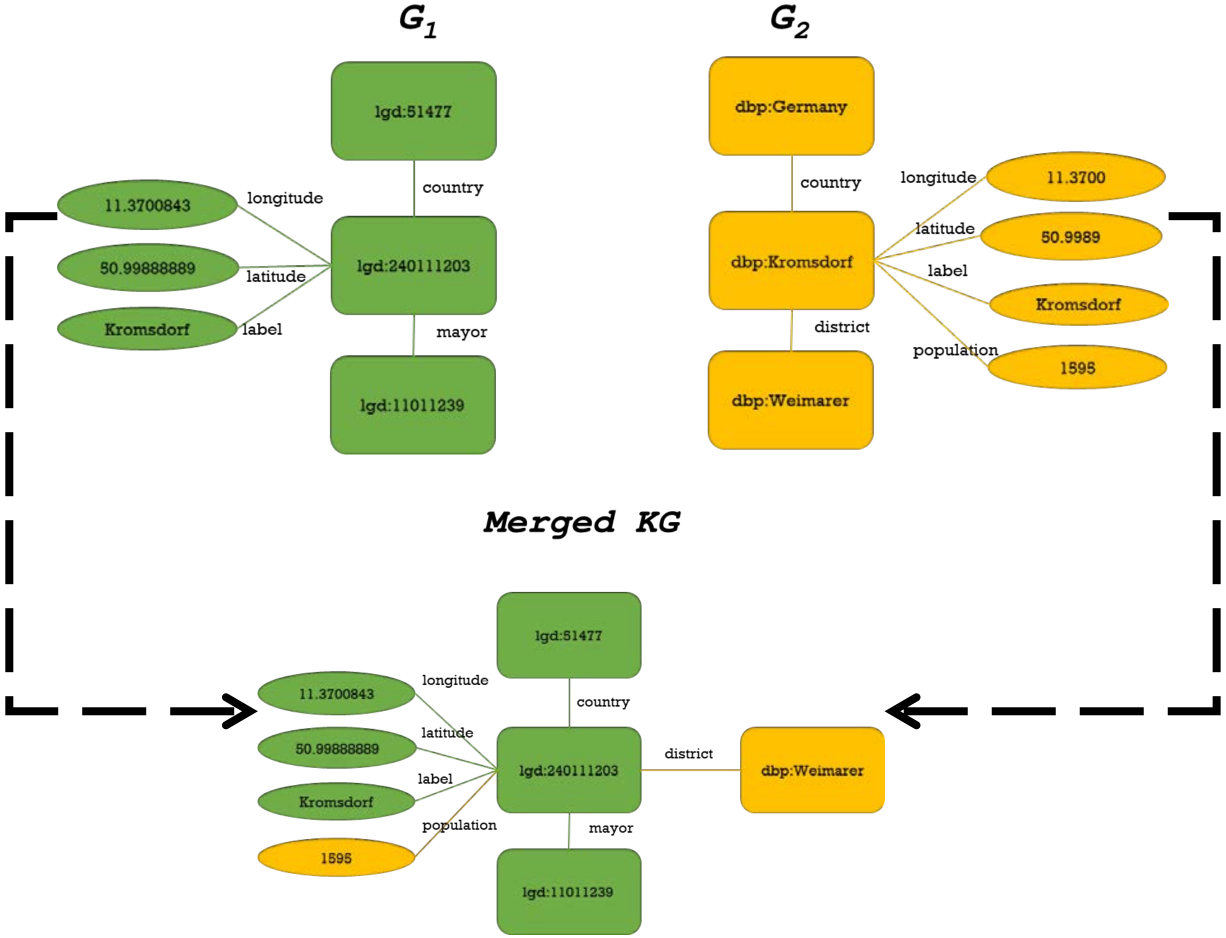
i-Align: An Interpretable Knowledge Graph Alignment Model

Bayu Distiawan Trisedya[1], **Flora D. Salim**[2], Jeffrey Chan[1], Damiano Spina[1], Falk Scholer[1], Mark Sanderson[1]

[1]RMIT University, Melbourne, Australia
[2]University of New South Wales (**UNSW**), Sydney, Australia

ARC Centre of Excellence for Automated Decision Making and Society (**ADM+S**)

KG Alignment: Example



Motivations

KGs are incomplete → the need for enrichment methods

- Manual curation
- Extract new information from text

→ **KG alignment:** merge two or more KGs to have a more comprehensive KG

Many open KGs are constructed for different purposes

- They have different level of details per entity
- Combines them to have a more complete KG

Goals and Challenges

Goals:

- Maintain high alignment performance
- Enable interpretability of alignment results

Challenge-1: The interpretability of the embedding-based KG alignment models is non-trivial

Goals and Challenges

Challenge-2: applying a post-hoc (model-agnostic) explainer is sub-optimal

- Existing GNN post-hoc explainers, such as GNNExplainer and PGExplainer, can only extract the most influential neighbors but not the most influential attributes

Challenge-3: Scalability

- The best-performing embedding-based model is built on top GNN → struggling on processing large KG

Goals and Challenges

Goal:

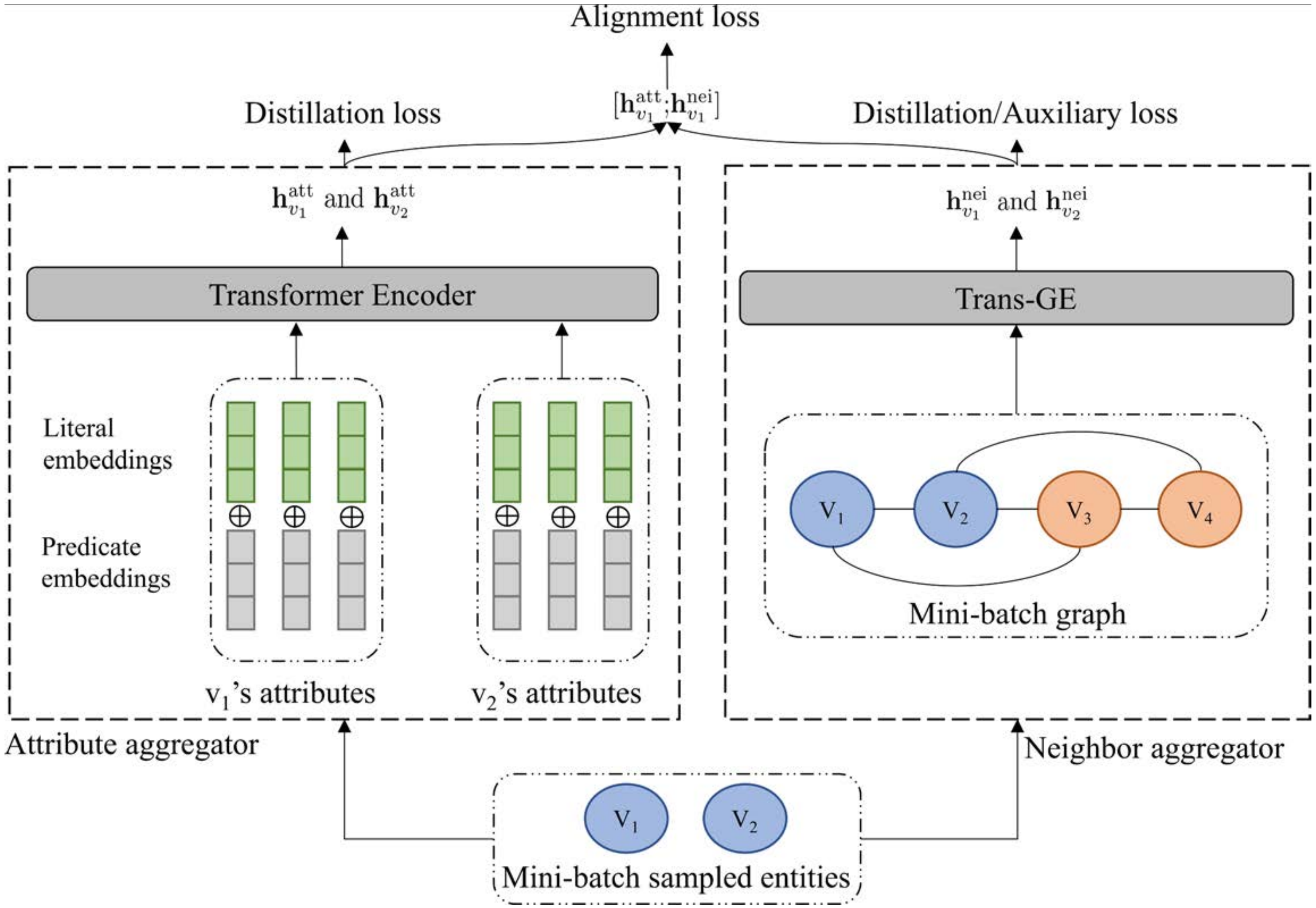
propose an accurate KG alignment model for aligning large KGs that can highlight top-n influential neighbours and attributes

→ ***i-Align***

Contributions

1. An interpretable KG alignment model is proposed, where an explanation of the alignment prediction can be automatically derived..
2. Along with the proposed model, a novel Transformer-based graph encoder is proposed for controllable information aggregation.
3. Extensive experiments and analyses are conducted to show the model’s effectiveness in predicting the alignments and providing explanations.

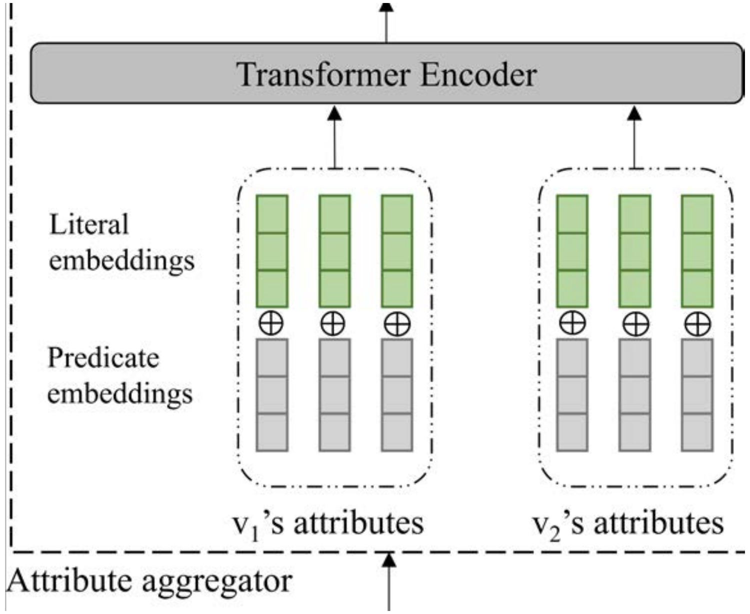
Proposed *i-Align* Model



Proposed Model

Attribute Aggregator

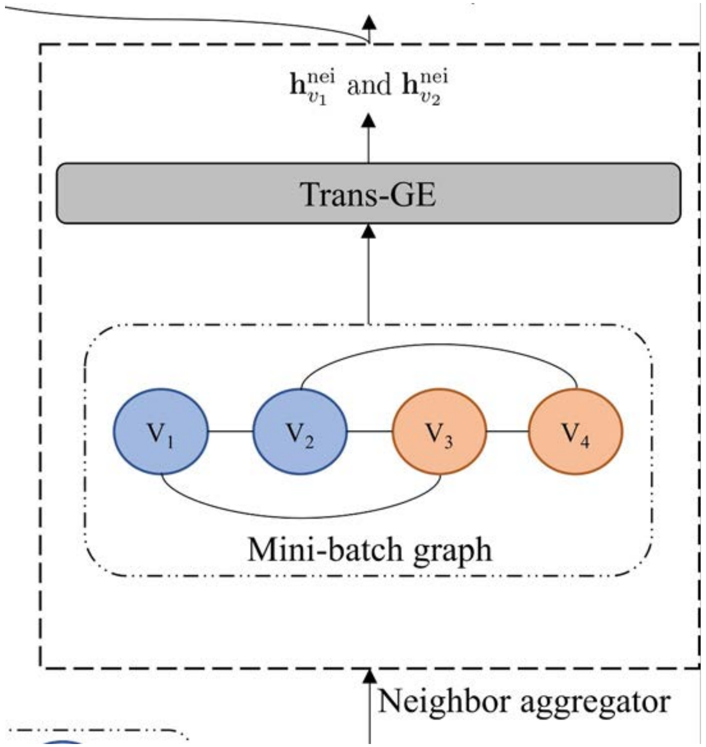
- Compute the attribute embeddings given the attribute triples in a mini-batch.
- Combines attribute key & value using GRU
- Compute attribute importance using Transformer Encoder → used for interpretability



Proposed Model

Neighbor Aggregator

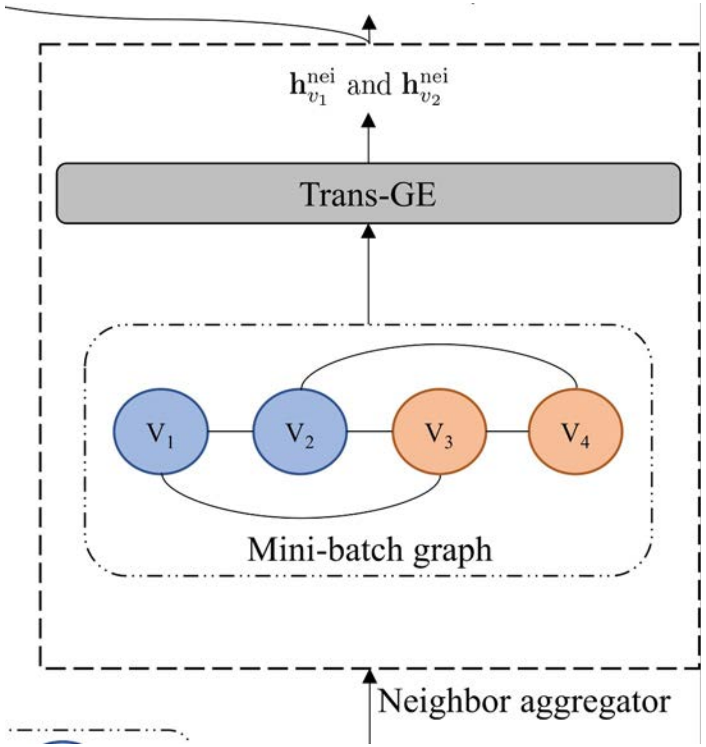
- Compute the neighbourhood embeddings given the attribute triples in a mini-batch.
- Components:
  - **Trans-GE →** return neighbour importance for interpretability
  - **Historical Embeddings**



Proposed Model

Trans-GE

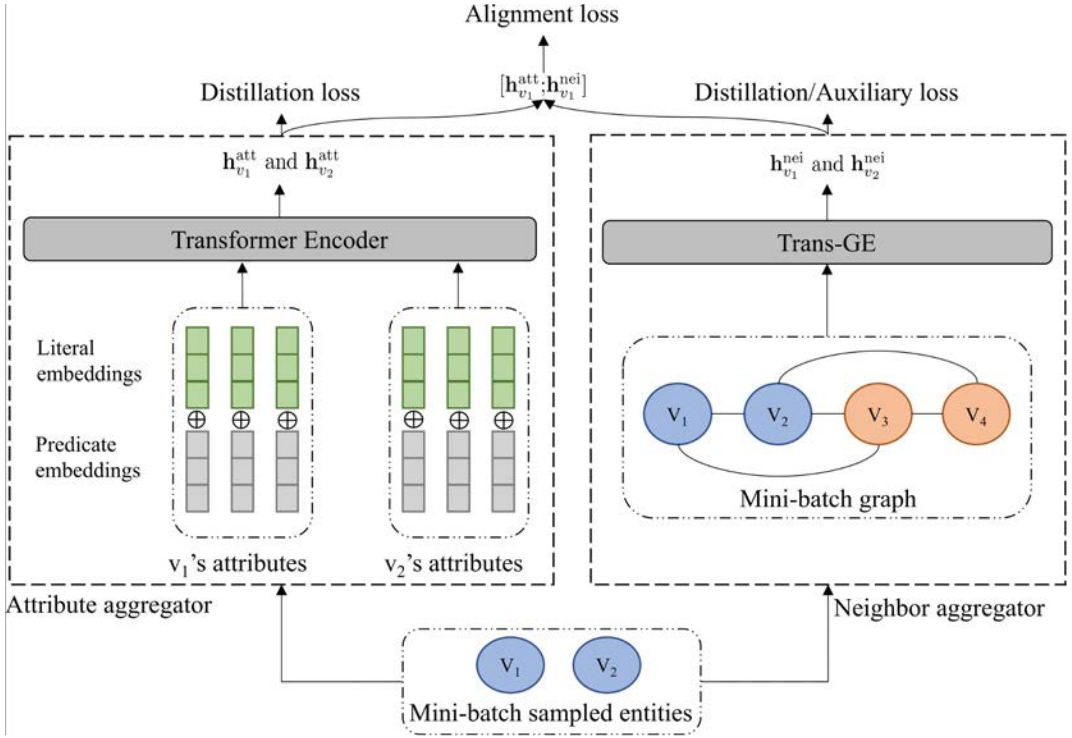
- Edge-gated attention for capturing the structural information of a sub-graph.
- A Transformer encoder applied on entities' neighbours.



Proposed Model

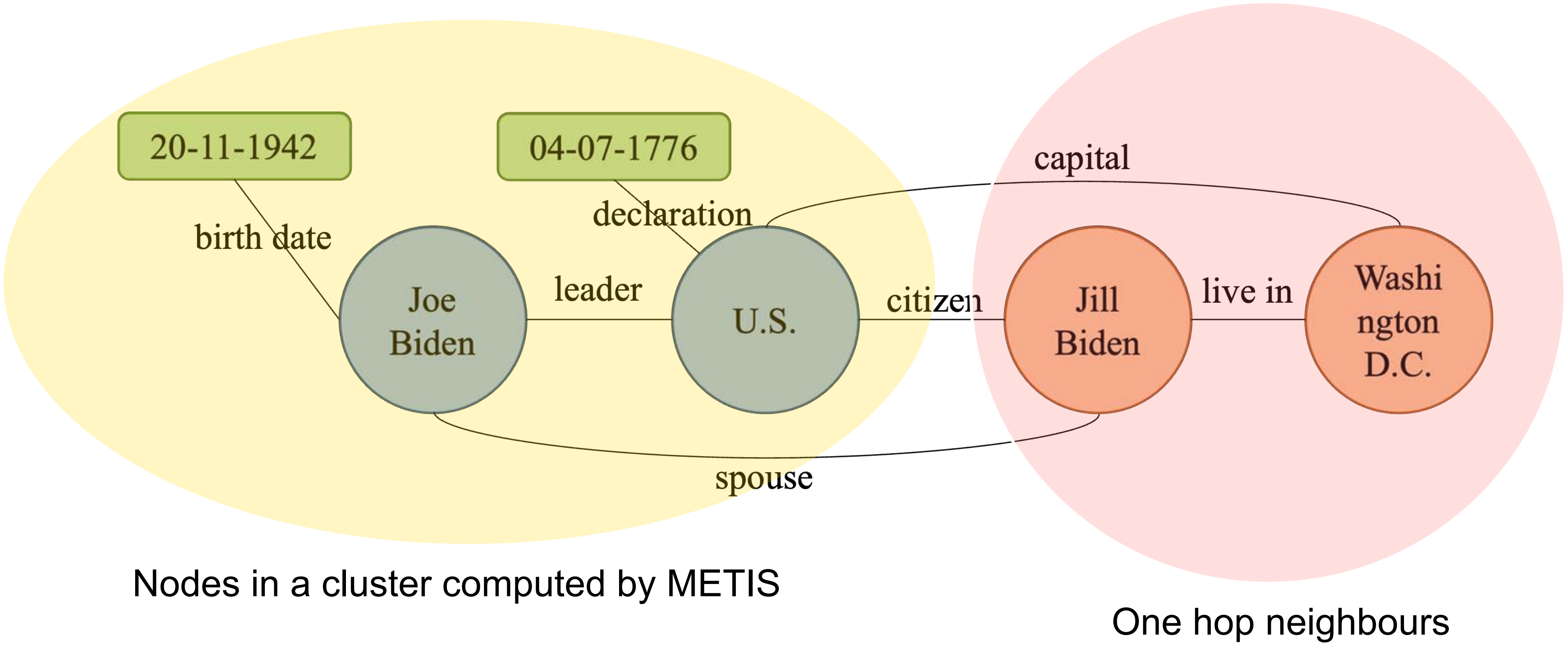
Historical Embeddings

- To approximate the full computational graph of a KG in a mini-batch
- Borrows the idea from GNNAutoScale
  - Split graph using a graph clustering algorithm, e.g., METIS.
  - The mini-batch takes the nodes in the cluster and their first hop neighbours.
  - Uses a linear transformation learned to approximate the current node embeddings based on the embeddings in the previous state → Learned by distillation loss



Proposed Method

Mini-batch illustration:



Nodes in a cluster computed by METIS

One hop neighbours

Experiments

Dataset:

- DWY-NB → a KG alignment benchmark dataset containing two pairs of KGs
  - DBpedia - Wikidata (DW-NB) → 50, 000 aligned entities
  - DBpedia - YAGO (DY-NB) → 15, 000 aligned entities
  - 30% of the aligned entities are used as seed alignment
- DBP-LGD, contain 10,000 aligned entities between DBpedia and LinkGeoData.
- DBP-GEO, contain 10,000 aligned entities between DBpedia and Geonames.

Experiments

Comparisons of KG alignment models performance

Model	DW-NB		DY-NB		LGD-DBP		GEO-DBP	
	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10	Hits@1	Hits@10
MTransE	7.88	25.75	0.08	0.68	33.59	35.76	33.14	34.75
JAPE	12.57	19.96	1.4	3.27	33.47	34.42	33.35	34.27
GCN-Align	24.76	48.52	24.36	53.43	48.57	52.74	46.12	51.32
MRAEA	81.54	85.97	73.71	78.52	78.98	83.13	72.11	75.32
NMN	84.03	88.21	75.87	80.54	78.88	82.35	75.87	80.18
MultiKE	84.06	90.05	84.97	90.84	83.12	90.55	79.33	85.22
AtrrE	87.98	**95.8**	90.44	**94.23**	84.17	92.05	86.91	92.32
Proposed	**88.35**	94.22	**91.21**	93.44	**87.21**	**94.22**	**88.87**	**93.87**

Experiments

Manual Evaluation of Alignment Explanation

- Randomly take 50 correct alignments and 50 incorrect alignments
- Given an alignment result, annotators are asked to predict whether it is correct and their confidence level.

Model	Correct Prediction		Incorrect Prediction	
	Prec	Conf	Prec	Conf
GCN-Align + GNNExplainer	0.39	0.47	0.80	0.75
i-Align (Neighbors only)	0.68	0.71	0.87	0.81
i-Align	**0.95**	**0.90**	**0.93**	**0.93**

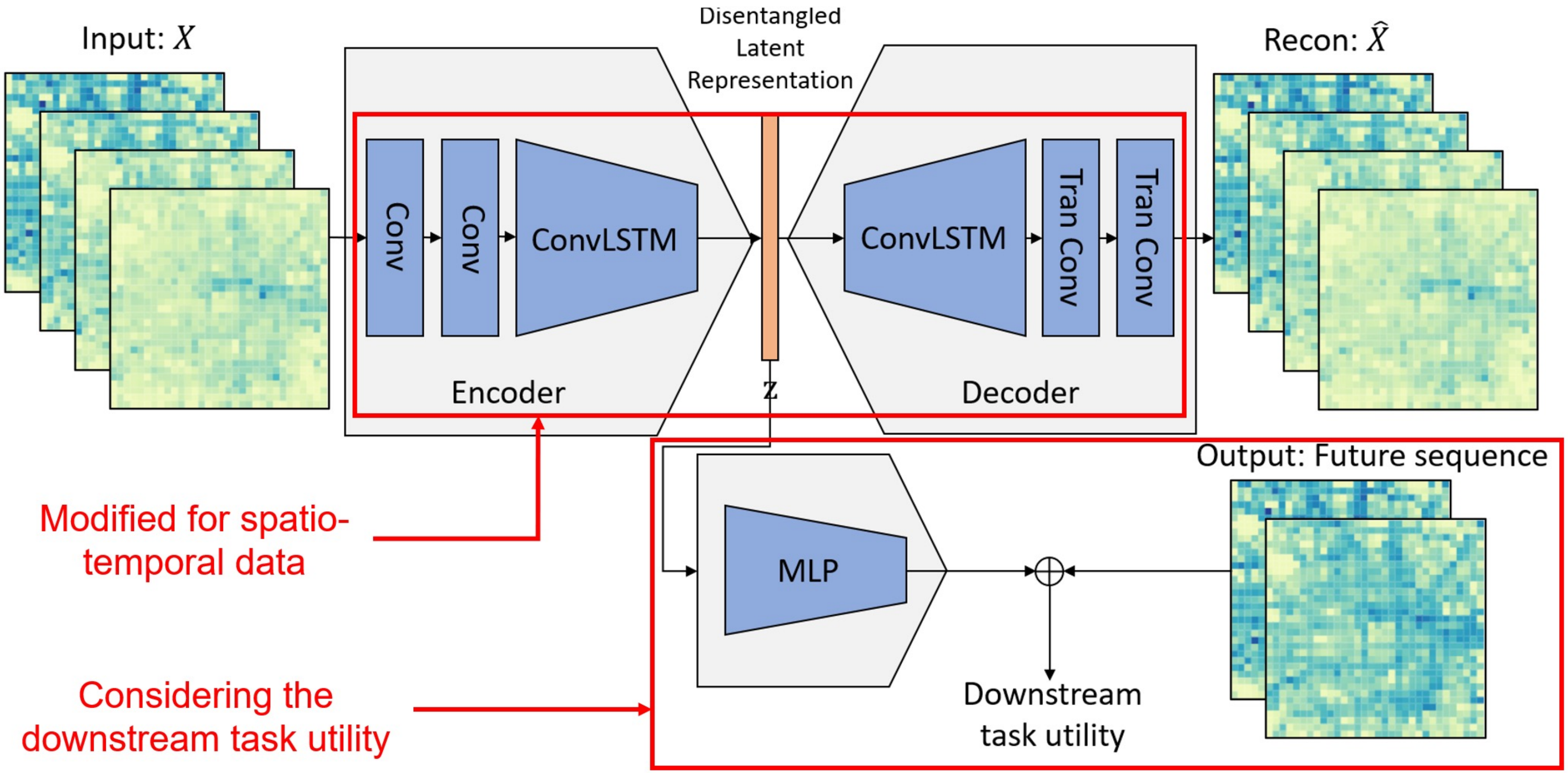
Experiments

Example of Attributes/Neighbors Explanation of A **Wrong** Alignment Prediction

- i-Align provides better explanations in terms of the top 5 aligned attributes and neighbours
- Expert can easily detect the wrong alignments

	i-Align prediction		GCN-Align + GNNExplainer prediction	
Entities	Carl Ferdinand Cori (Q78501)	Ferdinand I of Bulgaria (Q151667)	Carl Ferdinand Cori (Q78501)	Tomáš Cihlář (Q88483170)
Top-5 Attributes	(given_name, Ferdinand) (date_of_death, 1984-10-20) (married, 1920) (last_name, Cori) (date_of_birth, 1896-12-05)	(given_name, Ferdinand) (date_of_death, 1948-09-10) (position_end_time, 1918) (last_name, Bulgaria) (date_of_birth, 1861-02-26)	N/A	N/A
Top-5 Neighbors	(occupation, biochemist) (citizenship, Czechoslovakia) (place_of_death, Cambridge) (member_of, Royal Society) (place_of_birth, Prague)	(occupation, entomologist) (native_language, German) (place_of_birth, Vienna) (place_of_burial, St. Augustine's Church) (member_of, Academy of Sciences)	(occupation, biochemist) (language, English) (gender, male) (citizenship, Czechoslovakia) (employer, Harvard University)	(occupation, biochemist) (employer, Gilead Sciences) (country_of_citizenship, Czech Republic) (language, English) (gender, male)

Explainability of Generative AI



Modified for spatio-temporal data

Considering the downstream task utility

Zhao, S., Shao, W., Chan, J., & Salim, F. D. (2022). Measuring disentangled generative spatio-temporal representation. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)* (pp. 522-530).

UNSW
AI Institute

Thank you

Professor Flora Salim
UNSW Computer Science and Engineering
Cisco Chair of Digital Transport
Deputy Director of Engagement, UNSW AI Institute
Chief Investigator, ARC Centre of Excellence of Automated Decision Making and Society

UNSW Sydney

Email: flora.salim@unsw.edu.au
Twitter: @flosalim