

ICONIP 2024



31st International Conference on Neural Information Processing

December 2–6, 2024 • Auckland, New Zealand iconip2024.org

DynaTNet: Enhanced Transformer-Based Spatiotemporal Dynamic Graph Neural Network for Multivariate Time Series Classification

Zijun Dou¹, Xiaosong Han¹, Zhelun Peng¹, Heng Li¹, Bingyi Xiang¹, Yanchun Liang^{1,2}

¹ Jilin University ² Zhuhai College of Science and Technology

Presenter Email: douzijunabc@gmail.com



Sponsors:



100% PURE
NEW ZEALAND



AUT KNOWLEDGE ENGINEERING &
DISCOVERY RESEARCH INNOVATION



Supported by:



Nottingham Trent
University





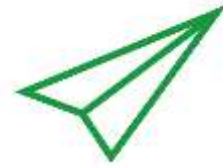
Background



Design



Experiment



Conclusion



2



Sponsors:



AUT KNOWLEDGE ENGINEERING & DISCOVERY RESEARCH INNOVATION



Springer

Supported by:



Nottingham Trent University



Time Series Classification

Wide Applications



Stock trading



Disease diagnosis

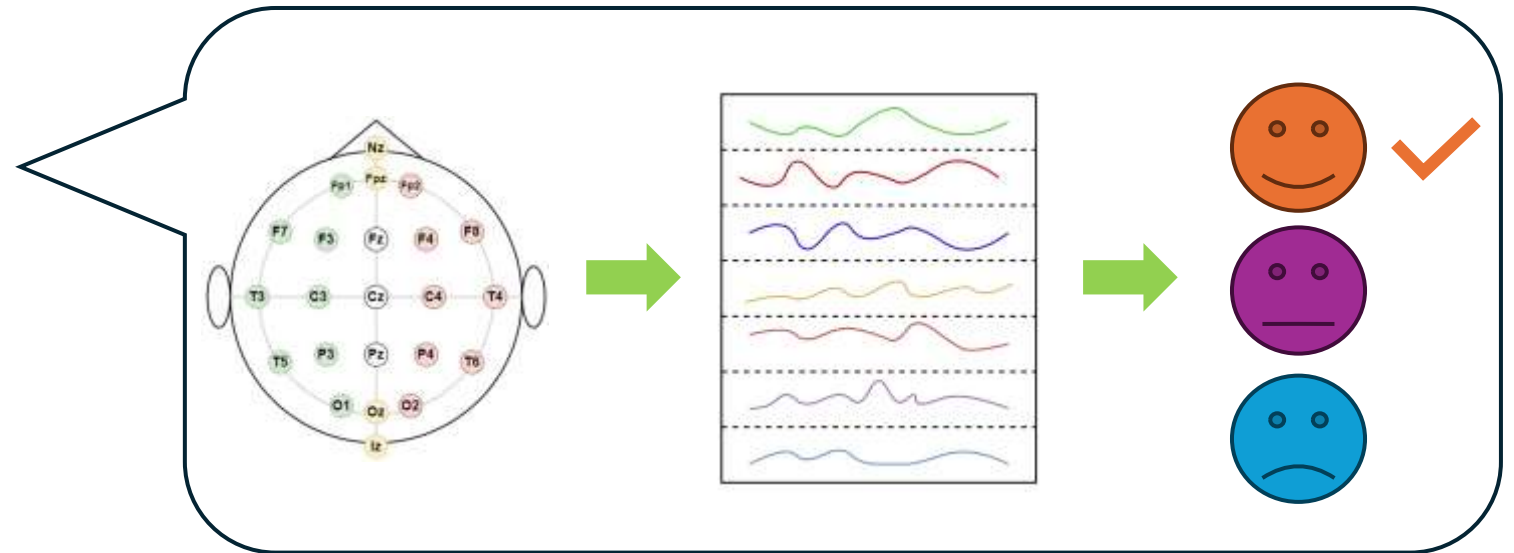


Industry manufacture



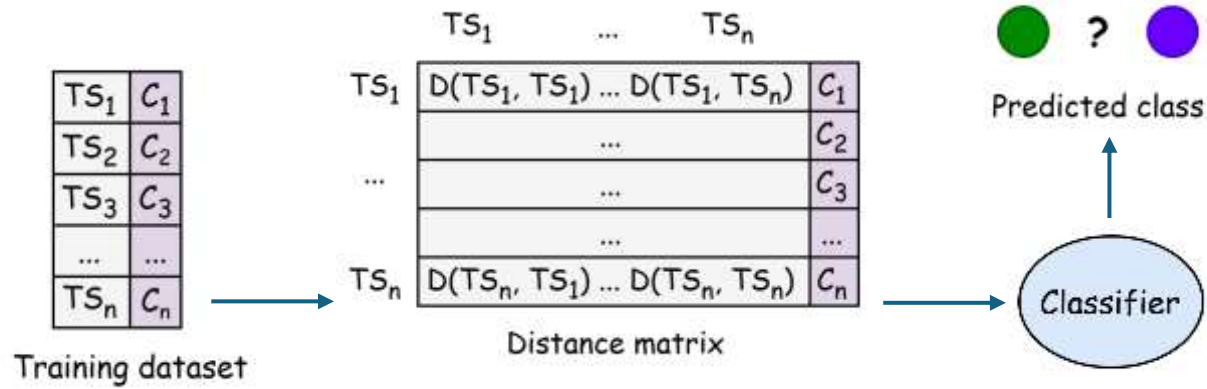
Weather estimation

Electroencephalogram Analysis



Time Series classification tasks exist in various domains and affect human life.

Approaches of TSC

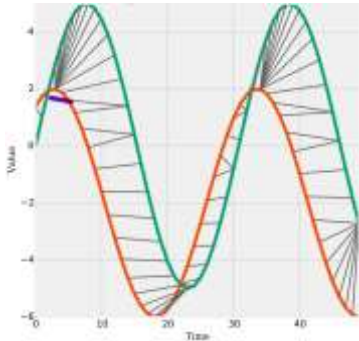


Distance based classifiers

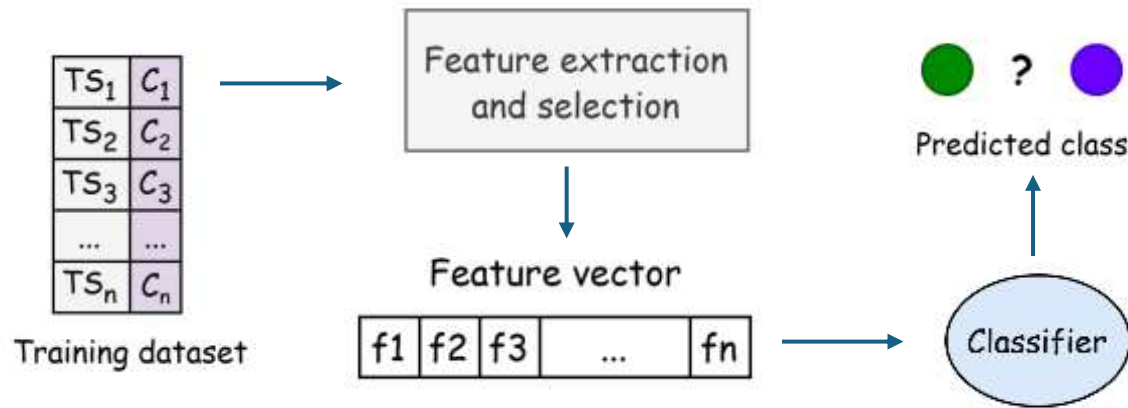


Feature based classifiers

High Dimension Data



Approaches of TSC



catch22: **CA**nonical
Time-series
CHaracteristics



Distance based classifiers



Feature based classifiers

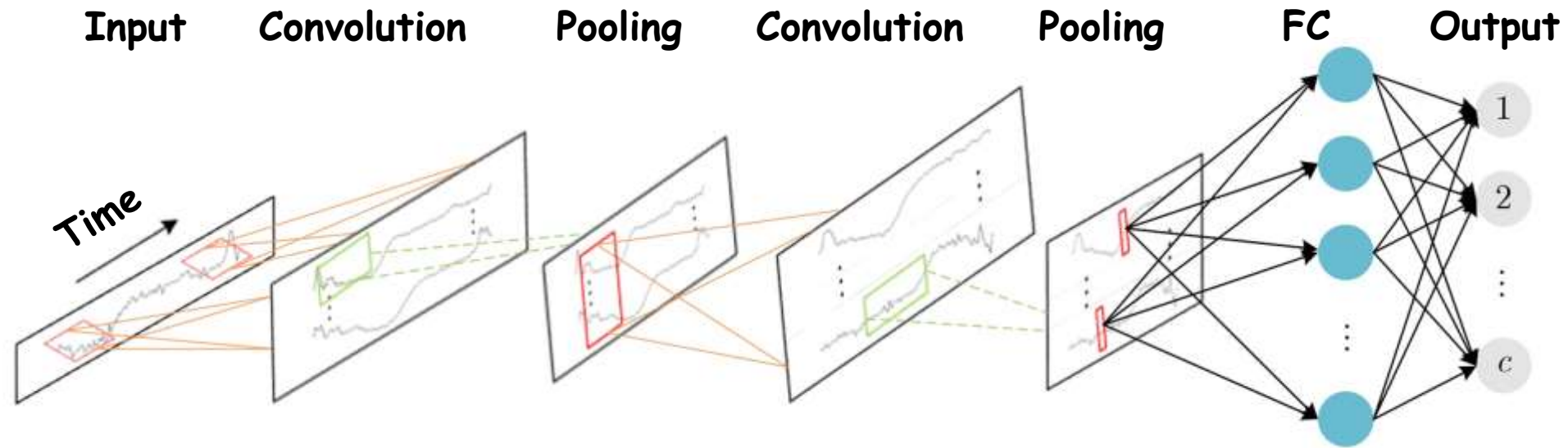


Deep learning based classifiers

High Dimension
Data

Handcrafted
Feature Design

Improving Neural Network Structure



Challenge 1: Address the limitations in modeling long-range dependencies in time series.



6

Sponsors:



100% PURE NEW ZEALAND



AUT KNOWLEDGE ENGINEERING & DISCOVERY RESEARCH INNOVATION



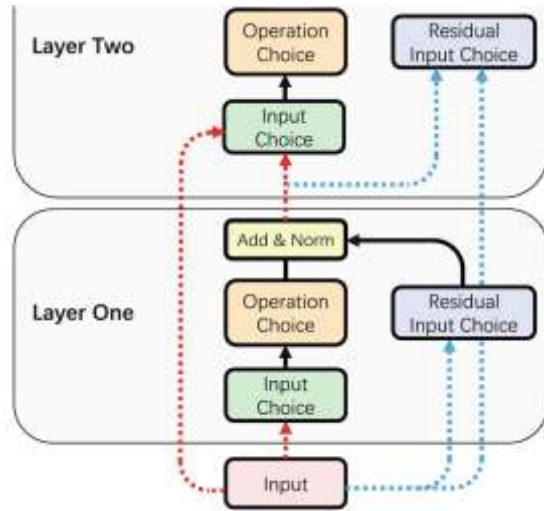
Supported by:



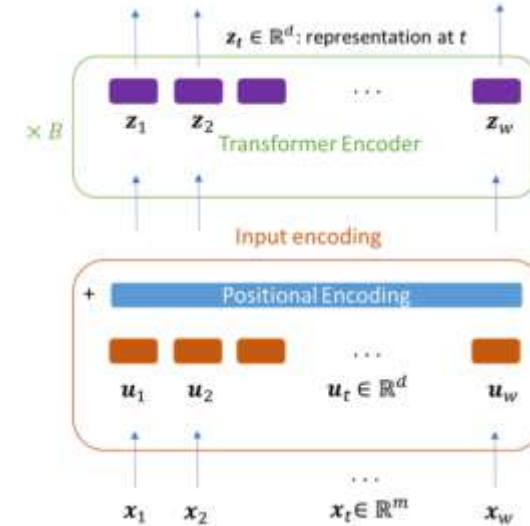
Nottingham Trent University



Transformer-Based Methods



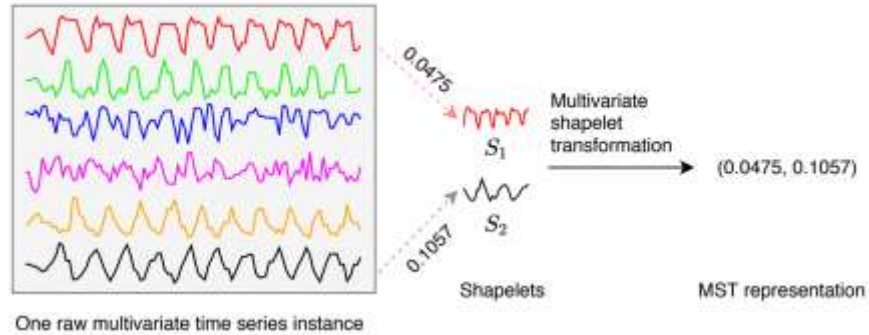
- ⚠ Data dependency
- ⚠ Substantial training time
- 😞 High computational cost



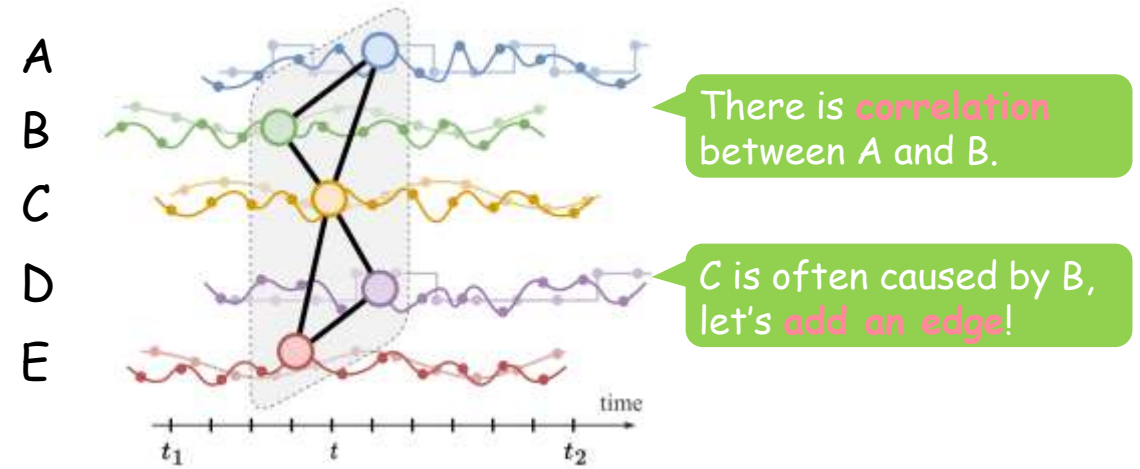
- ⚠ Lack in short-range features
- 😞 Limited local feature modeling

Challenge 2: Ensure the capture of local features with reduced training time.

Shape-Based Methods



Fail to adequately capture the relationships between multiple variables



Challenge 3: Multivariate time series modeling.



8

Sponsors:



100% PURE
NEW ZEALAND



AUT KNOWLEDGE ENGINEERING &
DISCOVERY RESEARCH INNOVATION



Supported by:



Nottingham Trent
University



Challenges

Address the limitations in modeling long-range dependencies in time series.

- ✓ Long-range dependencies often contain patterns that are stable over time.

Ensure the capture of local features with reduced training time.

- ✓ Local features, such as short-term patterns, often carry critical information for accurate classifications.
- ✓ Faster model training makes the real-time applications more efficient.

Multivariate time series modeling.

- ✓ The model can lead better performance and insights by modeling multiple variables together.



Background



Design



Experiment



Conclusion

Sponsors:



AUT KNOWLEDGE ENGINEERING &
DISCOVERY RESEARCH INNOVATION

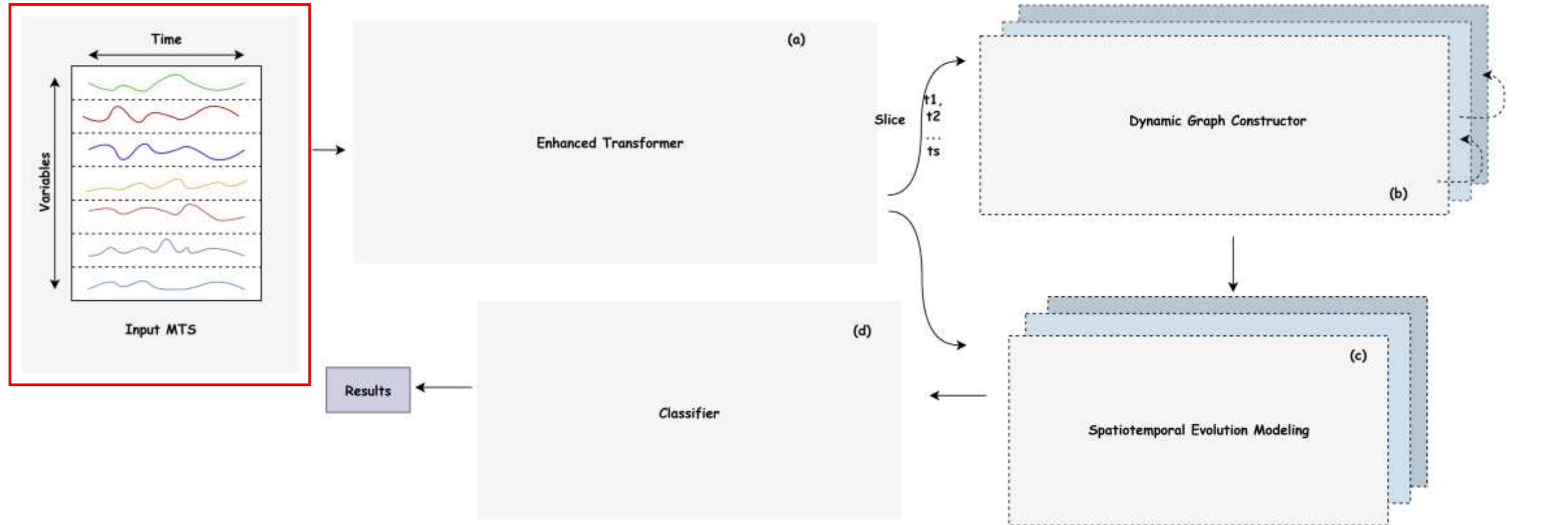


Supported by:

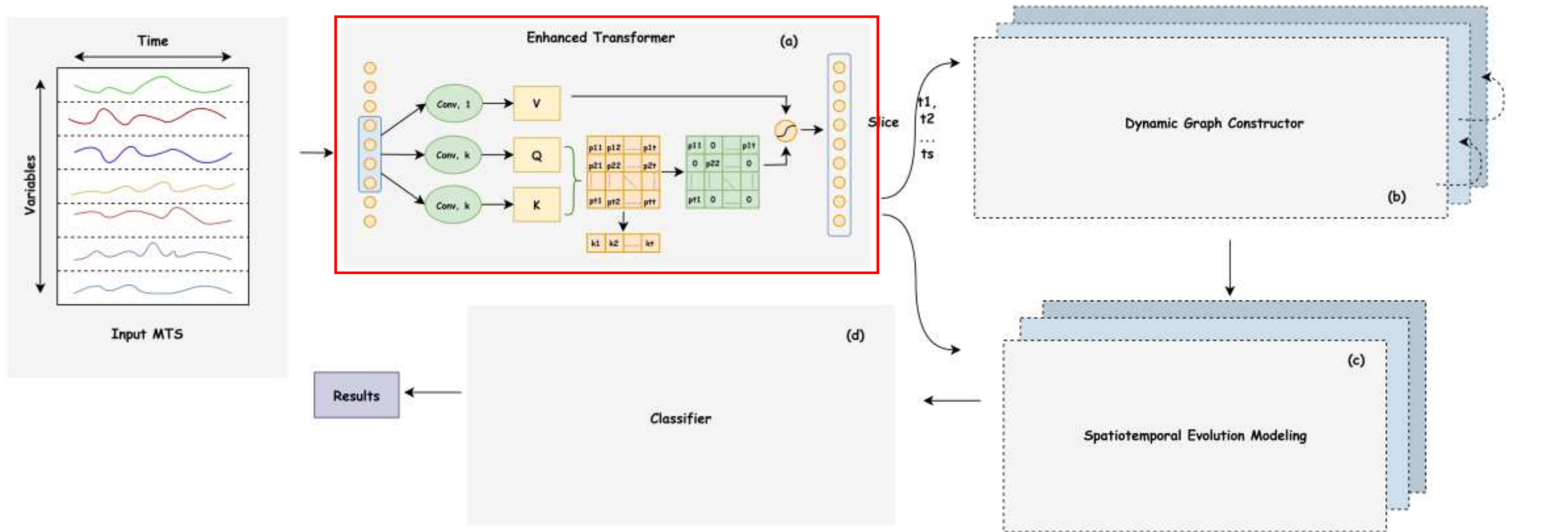


10

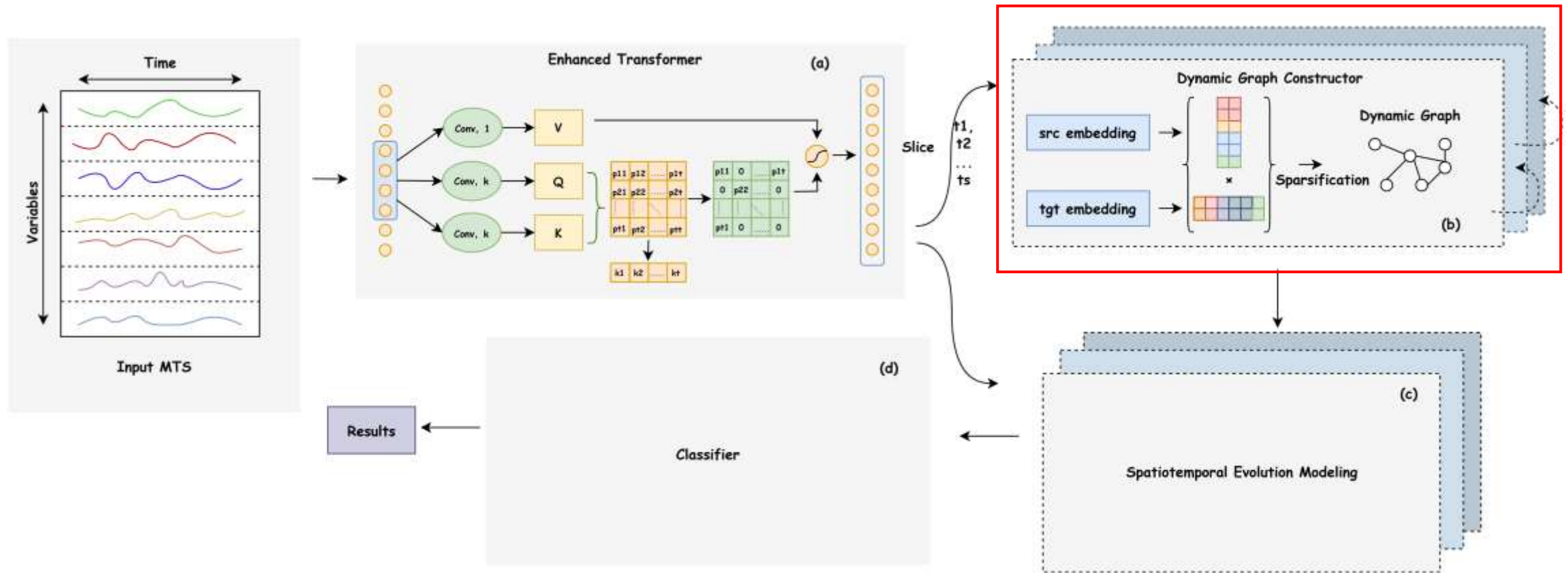
DynaTNet Design Overview



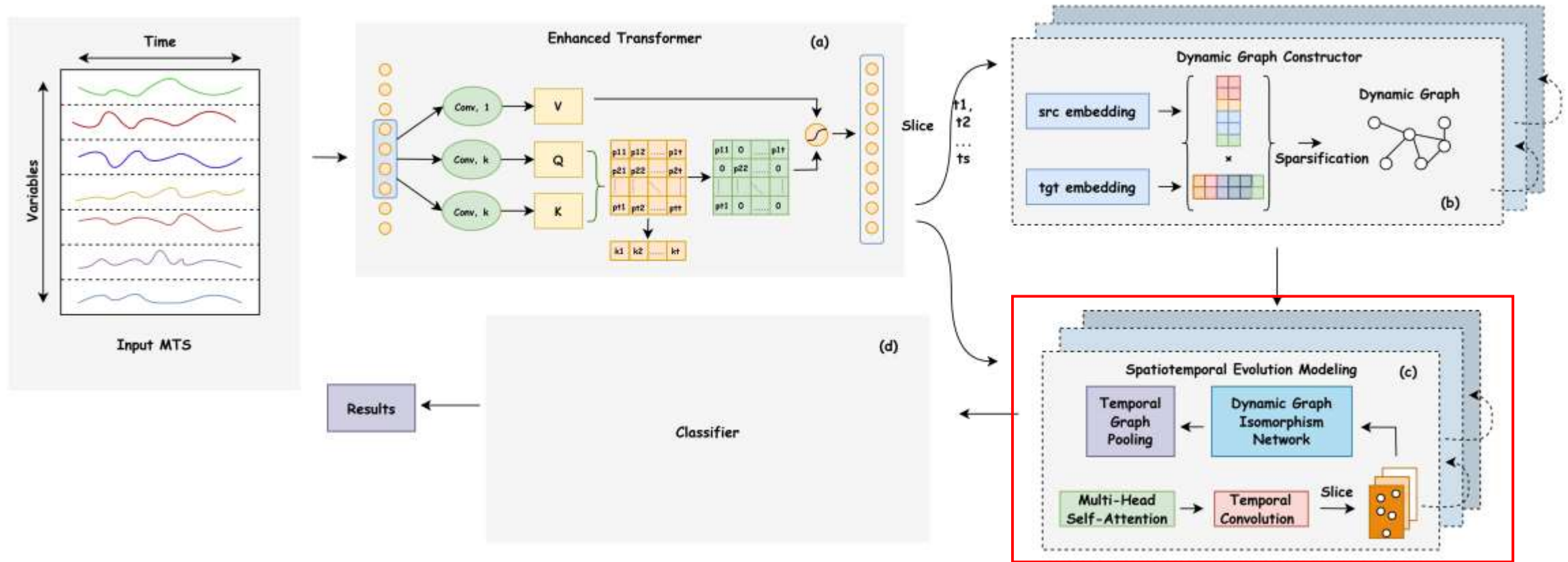
DynaTNet Design Overview



DynaTNet Design Overview



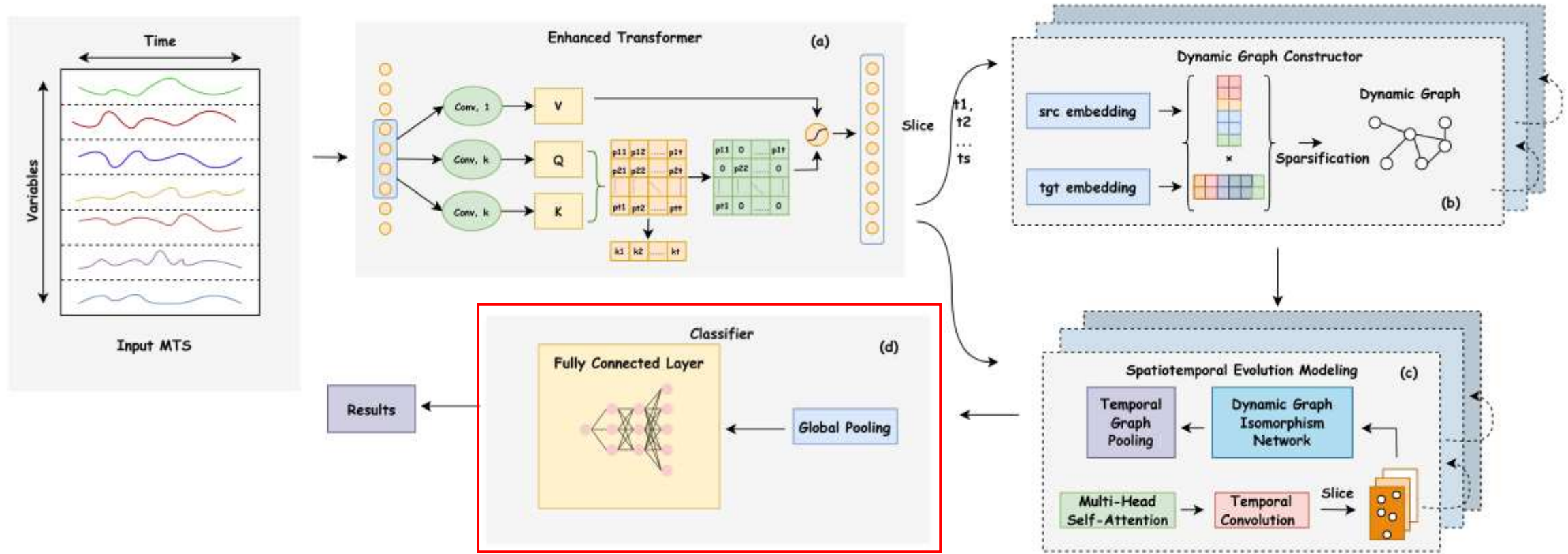
DynaTNet Design Overview



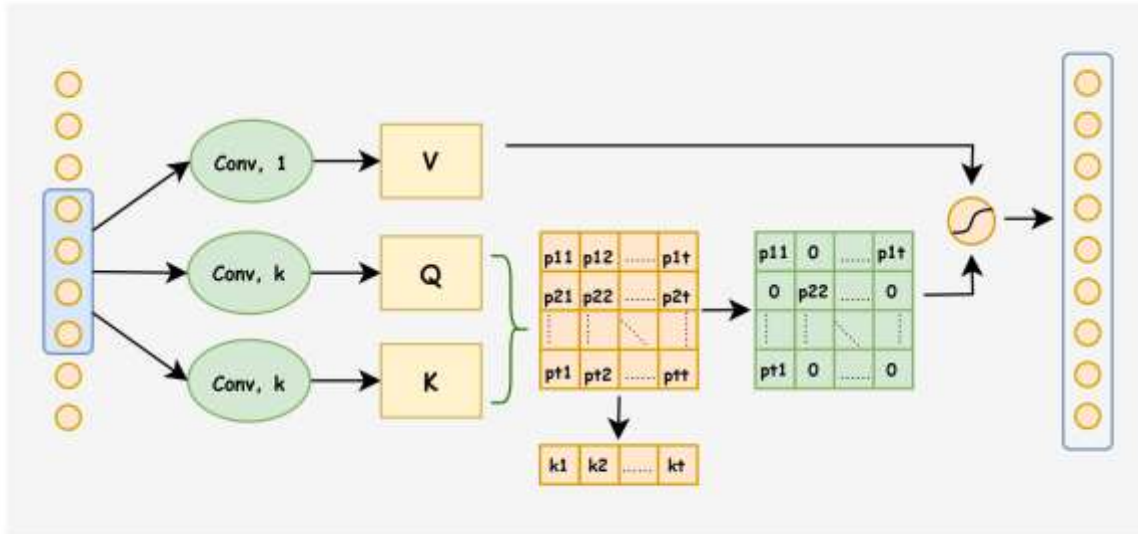
14

ICONIP 2024
31st International Conference on Neural Information Processing
December 2-6, 2024 - Auckland, New Zealand - iconip2024.org

DynaTGNet Design Overview



Capturing Global Dependencies in TS



Enhanced Transformer



Transformer

- ✓ Capture long-range dependencies in time series.

Convolution attention

- ✓ $Q_h = \text{Conv}_h^Q(X), K_h = \text{Conv}_h^K(X), V_h = \text{Conv}_h^V(X).$
- ✓ $P = Q_h K_h^T / \sqrt{d_k}.$
- ✓ Balance the use of convolution to extract local features.

Top- k strategy

- ✓ Retain the top k largest weights of attention matrix.
- ✓ Reduce computational cost.

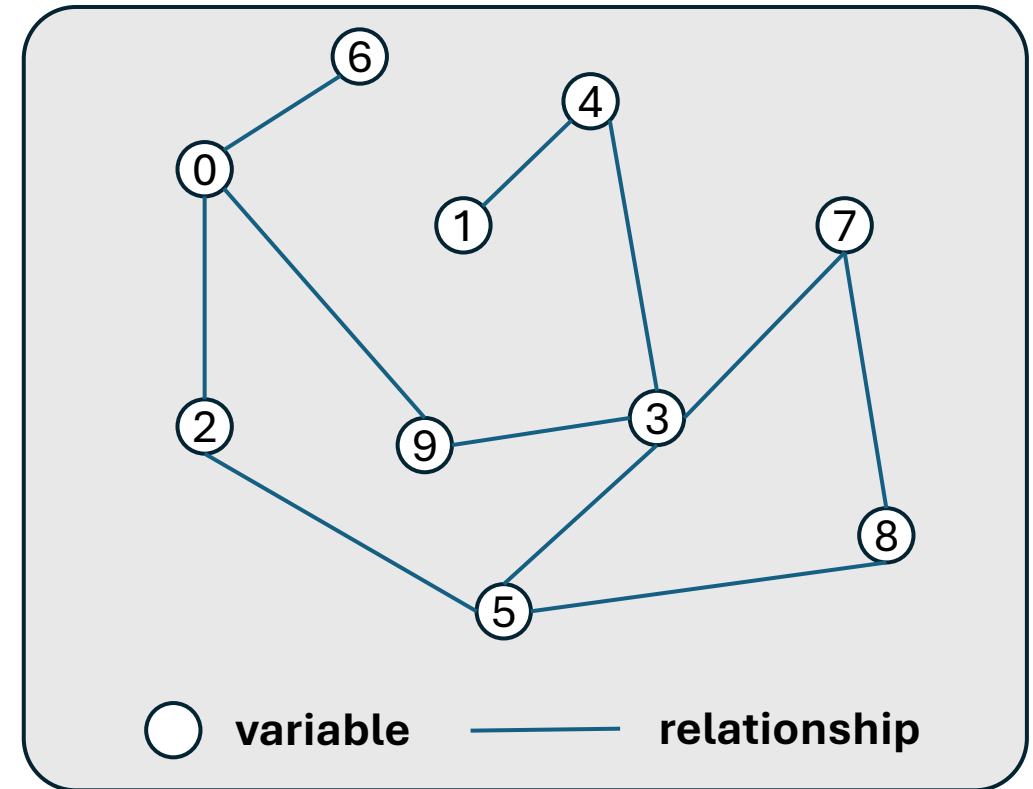


Multiple Variables Modeling

Multiple variables

- ✓ The variable represent the entity in real world.
- ✓ Multivariate time series record the changes of different entities over time.
- ✓ Complex relationships exist between different variables.

Step $T - 1$



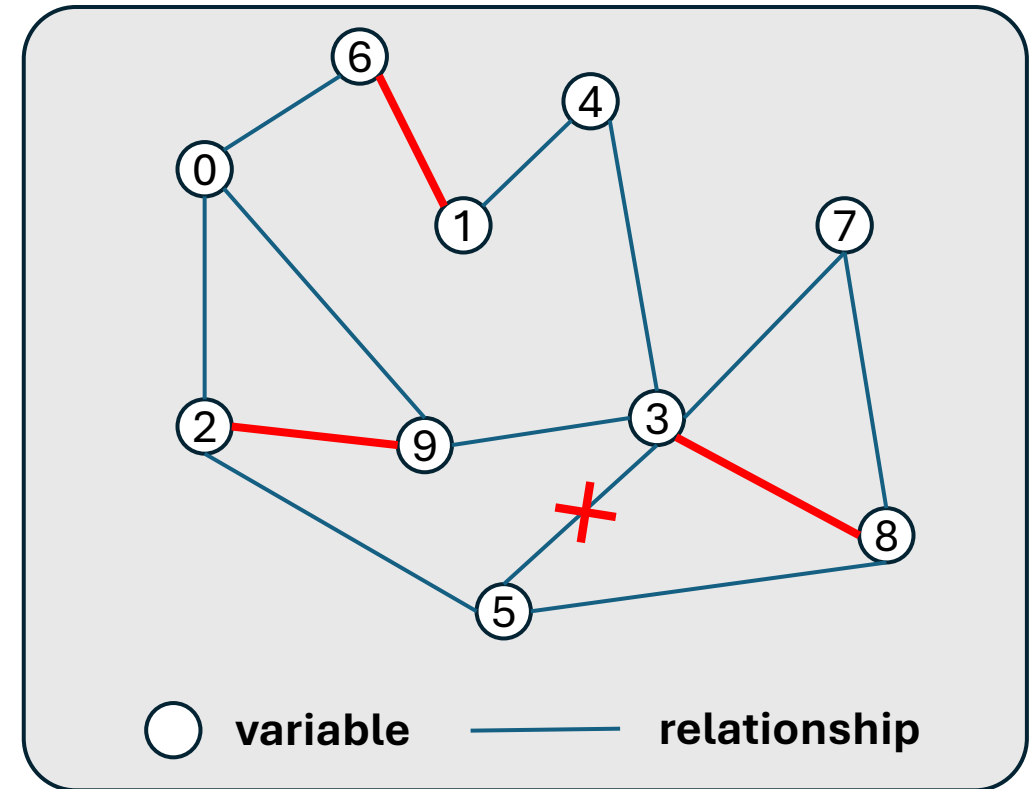
Multiple Variables Modeling

Multiple variables

- ✓ The variable represent the entity in real world.
- ✓ Multivariate time series record the changes of different entities over time.
- ✓ Complex relationships exist between different variables.
- ✓ As time moves forward, the relationships between variables change.

Capture dynamic relationships

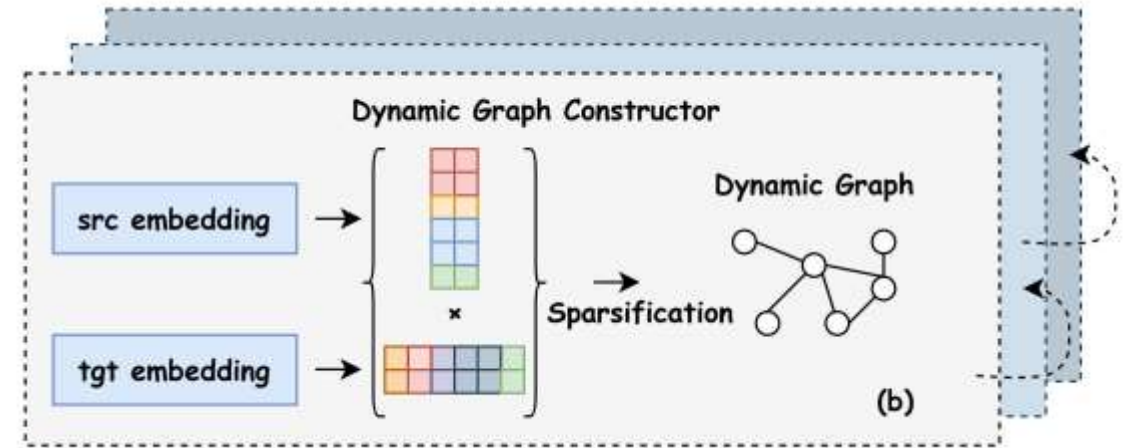
Step T



Multiple Variables Modeling

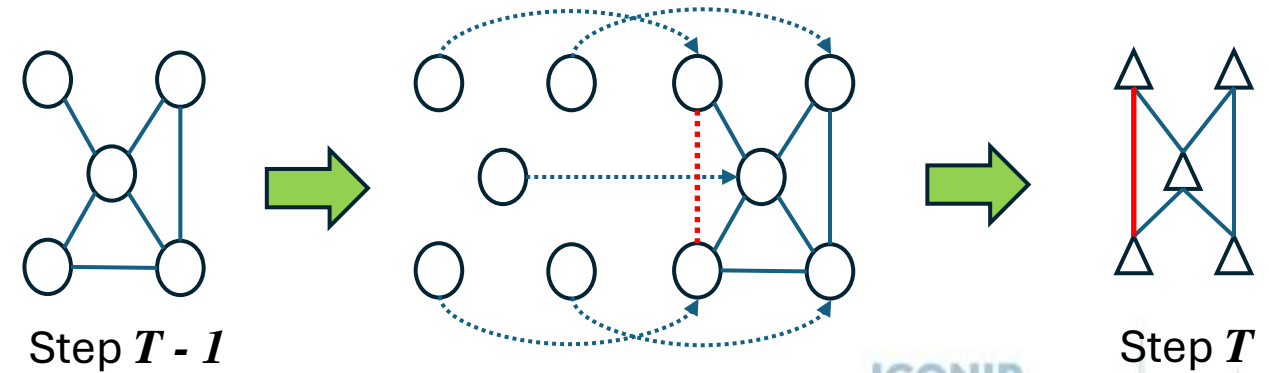
Dynamic Graph Constructor

- ✓ **Automatically** learn the relationship.
- ✓ **Sparsification**: reduce computational cost.



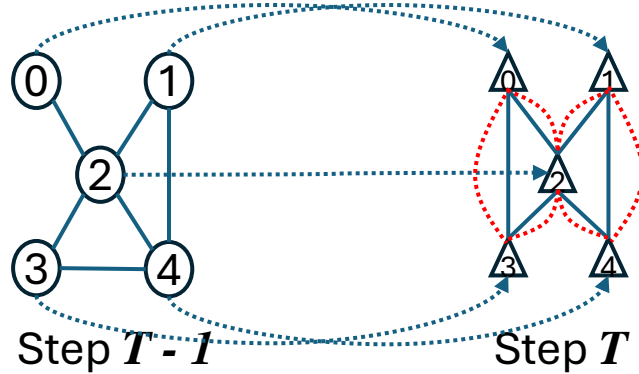
Connect two time slot's graph

- ✓ The latter graph **aggregates information** from previous one.
- ✓ The **foundation** of Graph Isomorphism Network.



Graph Isomorphism Network

Message Passing

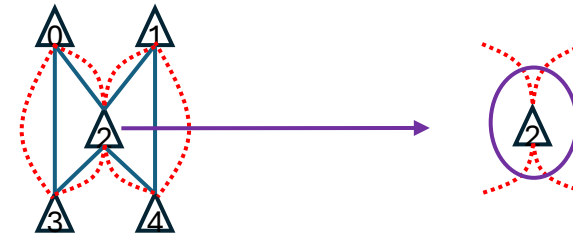


.....> Unidirectional passing

..... Bidirectional passing

Passing a node's **historical state** and **current neighbor information**.

Feature Aggregation



Step T

Step T

Node 2 **aggregates the feature** from nodes 0, 1, 3, 4, and itself.

Ensure the capture of local features.



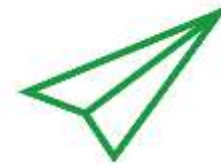
Background



Design



Experiment



Conclusion

Sponsors:



AUT KNOWLEDGE ENGINEERING &
DISCOVERY RESEARCH INNOVATION



Supported by:



21

Experiment Setup

Dataset

- ✓ **26** multivariate datasets, each with distinct characteristics and challenges.
- ✓ Each dataset contains between **2** and **1,345** variables.
- ✓ Time series lengths vary from **8** up to over **17,984** time points.
- ✓ Each dataset typically has between **2** and **39** classes for classification purposes.

#Datasets	#Variables	#Time Points	#Classes
26	2 - 1345	8 - 17984	2 - 39

Datasets	Code	Classes	Dims	Length	Test Size	Train Size
ArticulatoryWordRecognition	AWR	25	9	144	300	275
AtrialFibrillation	AF	3	2	640	15	15
BasicMotions	BM	4	6	100	40	40
Cricket	CR	12	6	1197	72	108
DuckDuckGeese	DDG	5	1345	270	50	50
EigenWorms	EW	5	6	17984	131	128
Epilepsy	EP	4	3	206	138	137
EthanolConcentration	EC	4	3	1751	263	261
ERing	ER	6	4	65	270	30
FaceDetection	FD	2	144	62	3524	5890
FingerMovements	FM	2	28	50	100	316
HandMovementDirection	HMD	4	10	400	74	160
Handwriting	HW	26	3	152	850	150
Heartbeat	HB	2	61	405	205	204
Libras	LIB	15	2	45	180	180
LSST	LSST	14	6	36	2466	2459
MotorImagery	MI	2	64	3000	100	278
NATOPS	NATO	6	24	51	180	180
PenDigits	PD	10	2	8	3498	7494
PEMS-SF	PEMS	7	963	144	173	267
PhonemeSpectra	PS	39	11	217	3353	3315
RacketSports	RS	4	6	30	152	151
SelfRegulationSCP1	SRS1	2	6	896	293	268
SelfRegulationSCP2	SRS2	2	7	1152	180	200
StandWalkJump	SWJ	3	4	2500	15	12
UWaveGestureLibrary	UW	8	3	315	320	120



Experiment Setup

Metric

✓
$$\text{Accuracy} = \frac{\text{Number of Correct Classifications}}{\text{Total Number of Samples}}$$

Experiment Environment

- ✓ PyTorch 1.11.0 with Python 3.9.12.
- ✓ Training continue for 2000 epohs.
- ✓ Xeon(R) Platinum 8352V CPU, an RTX 4090 GPU, and 24GB memory.

Accuracy Evaluation

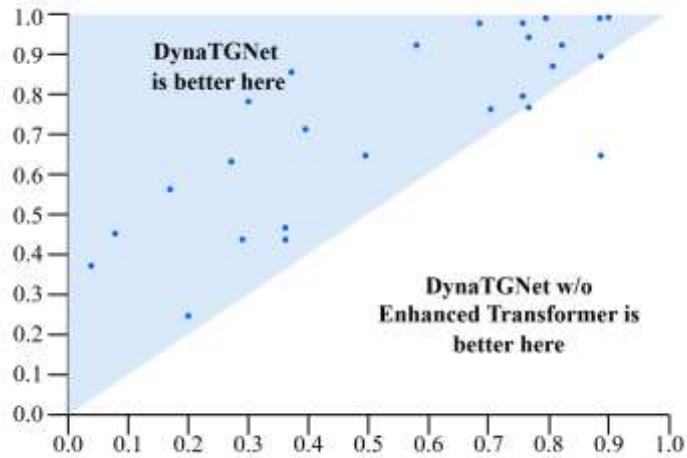
Dataset	CIF	ROCKET	IT	HC	MOS -CNN	ShapeNet	TapNet	WEASEL +MUSE	DynaTGNet
AWR	0.978	0.995	<u>0.991</u>	0.979	<u>0.991</u>	0.987	0.971	0.990	0.983
AF	0.251	0.248	0.220	0.293	0.183	<u>0.400</u>	0.302	0.333	0.487
BM	<u>0.997</u>	0.990	1.000	1.000	1.000	1.000	0.991	1.000	1.000
CR	0.982	1.000	<u>0.994</u>	0.992	0.990	0.986	0.975	1.000	1.000
DDG	0.560	0.461	0.634	0.476	0.615	0.725	0.582	0.575	<u>0.711</u>
EW	0.903	0.862	N/A	0.781	0.508	0.878	N/A	<u>0.890</u>	0.880
EP	0.983	0.990	0.986	1.000	0.996	0.987	0.960	1.000	<u>0.997</u>
EC	<u>0.728</u>	0.446	0.279	0.806	0.382	0.312	0.289	0.133	0.393
ER	<u>0.956</u>	0.980	0.921	0.942	0.915	0.133	0.894	0.430	0.895
FD	0.688	0.694	<u>0.772</u>	0.691	0.568	0.589	0.528	0.490	0.782
FM	0.539	0.552	<u>0.561</u>	0.537	0.568	<u>0.589</u>	0.513	0.490	0.625
HMD	0.522	0.445	0.423	0.377	0.361	0.338	0.323	0.365	<u>0.448</u>
HW	0.351	0.566	<u>0.657</u>	0.504	0.667	0.451	0.329	0.605	0.454
HB	0.765	0.717	0.732	0.721	0.604	<u>0.756</u>	0.739	0.727	<u>0.756</u>
LIB	<u>0.916</u>	0.906	0.887	0.902	0.965	0.856	0.836	0.878	0.859
LSST	0.561	0.631	0.339	0.538	0.521	0.590	0.463	0.590	<u>0.615</u>
MI	0.518	0.531	0.511	0.521	0.515	<u>0.610</u>	N/A	0.500	0.640
NATO	0.844	0.885	<u>0.966</u>	0.828	0.951	0.883	0.903	0.870	0.968
PD	0.989	0.995	<u>0.996</u>	0.971	0.983	0.977	0.936	0.948	0.997
PEMS	0.998	0.856	0.828	<u>0.979</u>	0.764	0.751	0.792	N/A	0.780
PS	0.265	0.283	0.367	<u>0.328</u>	0.295	0.298	N/A	0.190	0.263
RS	0.893	0.927	0.916	0.906	<u>0.929</u>	0.882	0.858	0.934	0.934
SRS1	0.859	0.865	0.846	0.860	0.829	0.782	0.956	0.710	<u>0.925</u>
SRS2	0.488	0.513	0.520	0.516	0.510	<u>0.578</u>	0.534	0.460	0.579
SWJ	0.451	0.455	0.420	0.406	0.383	0.553	0.351	0.333	<u>0.467</u>
UW	0.924	0.944	0.912	0.913	<u>0.926</u>	0.906	0.883	0.916	0.800
Avg.	<u>0.727</u>	0.721	0.707	0.722	0.689	0.685	0.692	0.654	0.740
Total best accuracy	<u>4</u>	<u>4</u>	2	3	3	3	1	<u>4</u>	9

The performance improvement is particularly significant for long sequences or small sample datasets, such as MI, HB and SWJ.

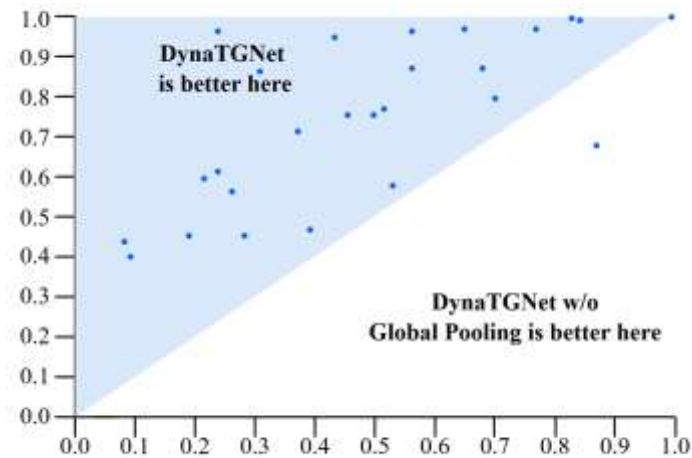
Compared with baseline methods, DynaTGNet has the best performance of 0.740.

It achieves the highest classification accuracy on 9 datasets.

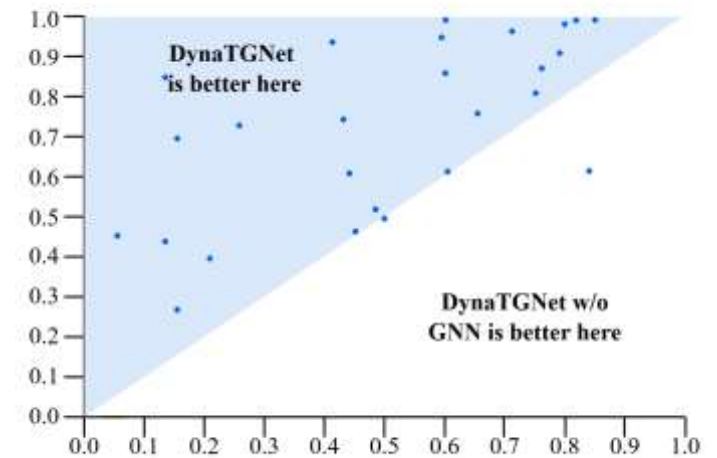
Ablation Study



Without Enhanced Transformer



Without Global Pooling



Without GNN

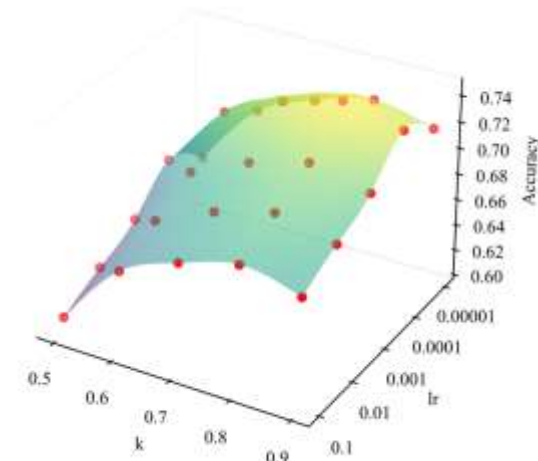
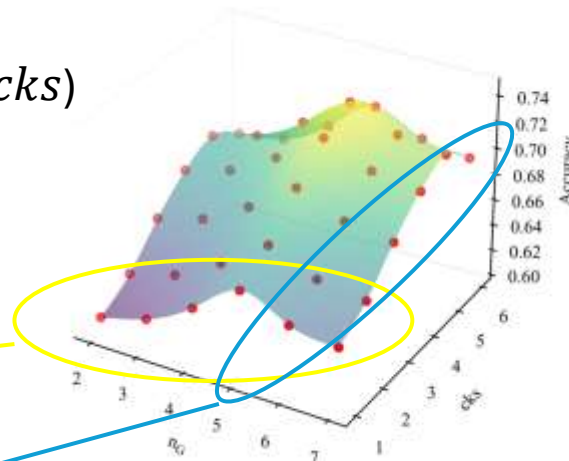
Each component has its impact, while Enhanced Transformer and GNN are the most significant.

Hyperparameter Stability

- ✓ The number of dynamic graphs (n_G)
- ✓ *Top-k* in the Dynamic Graph Constructor (k)
- ✓ The convolution kernel size in the self-attention (cks)
- ✓ Learning rate (lr)
- ✓ 13 randomly selected datasets.

n_G \uparrow \rightarrow Accuracy \uparrow first and then \downarrow

cks \uparrow \rightarrow Accuracy \uparrow first and then \downarrow



The best performance:

#k	#lr	# n_G	#cks
0.8	10^{-4}	5	5



Background



Design



Experiment



Conclusion

Sponsors:



AUT KNOWLEDGE ENGINEERING &
DISCOVERY RESEARCH INNOVATION



Supported by:



Nottingham Trent
University



27

Conclusion

DynaTGNNet

- ✓ A novel end-to-end framework for **multivariate time series classification**.
- ✓ Effectively capture **global correlations**, **local hidden information** and **spatiotemporal dynamics**.

Key Designs of DynaTGNNet

- ✓ Enhanced Transformer: convolutional self-attention mechanism and **top k** selection strategy.
- ✓ Multiple Variables Modeling: Dynamic Graph Constructor and connect two time slot's graph.
- ✓ Graph Isomorphism Network: message passing and feature aggregation.



吉林大学
JILIN UNIVERSITY



珠海科技学院
ZHUHAI COLLEGE OF SCIENCE AND TECHNOLOGY

Thank You!

Q & A

DynaTGNet: Enhanced Transformer-Based Spatiotemporal Dynamic Graph Neural Network for Multivariate Time Series Classification

29



Sponsors:



100% PURE
NEW ZEALAND



AUT KNOWLEDGE ENGINEERING &
DISCOVERY RESEARCH INNOVATION



Supported by:



Nottingham Trent
University

