MPRE: Multi-perspective Patient Representation Extractor for Disease Diagnosis

Ziyue Yu[1], Jiayi Wang[1], Wuman Luo[1], Rita Tse[1] and Giovanni Pau[2,3]

[1]Macao Polytechnic University, Macao, China
[2]University of Bologna, Bologna, Italy
[3]University of California, Los Angeles, CA, USA

Speaker: Ziyue Yu (Julius)

澳門理工大學
Universidade Politécnica de Macau
Macao Polytechnic University

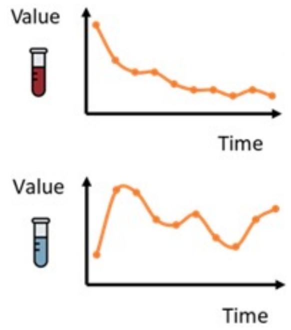
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

UCLA
University of California, Los Angeles

What is Electronic Health Records (EHR) Data?

- **A list of temporally ordered visit data**
  - ❑ Static features (e.g., gender, ethnicity)
  - ❑ Dynamic features (e.g., hemoglobin, creatinine)
  - ❑ Diagnostic result (e.g., chronic heart disease)

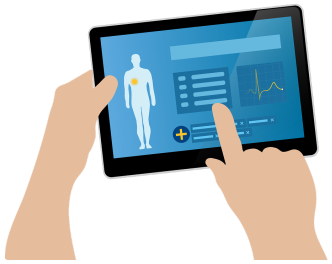
Static Features	Value
Gender	Female
Ethnicity	Black
…	…



…

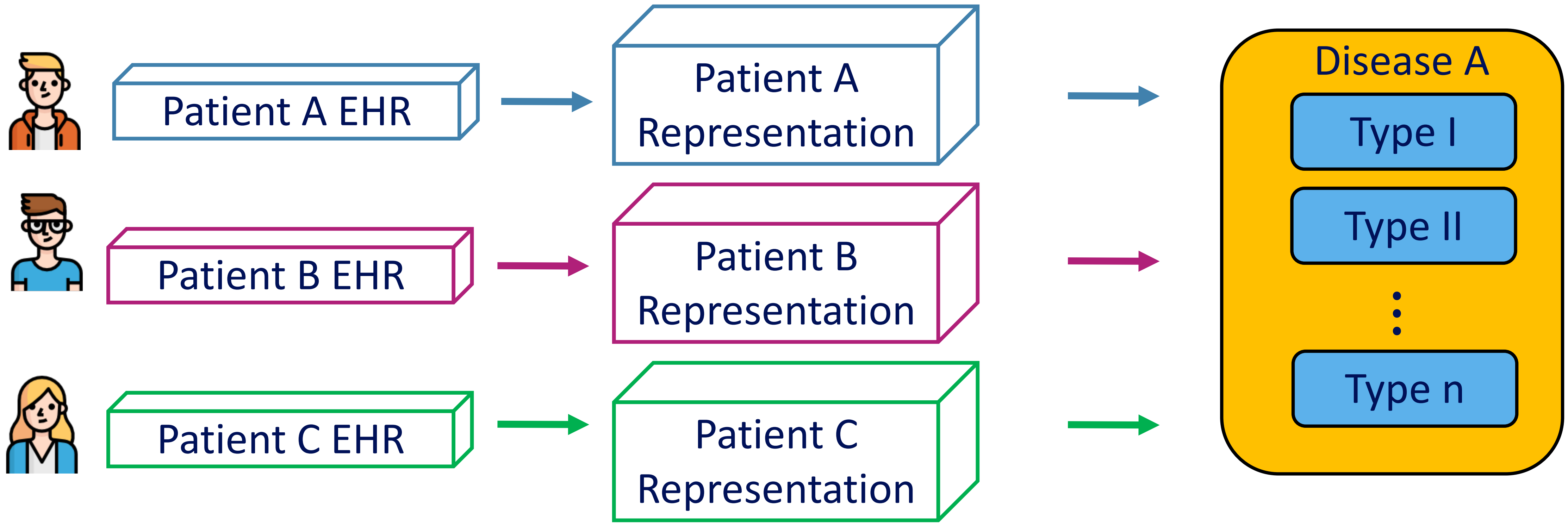
Dynamic Features

Diagnostic Result



What We Do?

- **Given electronic health records data**
  - ❑ Effectively learn the **patient representation** for disease prediction



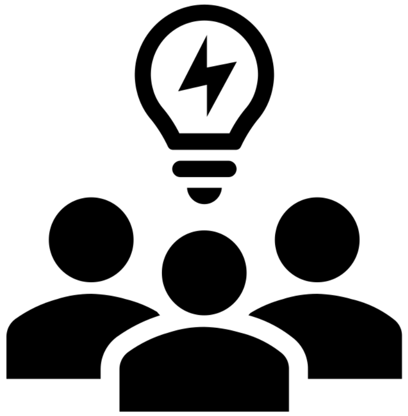
Related Works

Variation Pattern Detection Methods

- ❑ RETAIN (NIPS)
- ❑ Dipole (KDD)
- ❑ SAnD (AAAI)

Time-aware Methods

- ❑ TCN
- ❑ AdaCare (AAAI)
- ❑ T-LSTM (KDD)
- ❑ StageNet (WWW)
- ❑ ConCare (AAAI)



What Drives Us?

- **Existing works still have much room for improvement**

❑ Long-term and short-term trend, variations

† Upward trend in creatinine indicates the risk of kidney disease

† Abnormal increase in bicarbonate indicates the risk of metabolic alkalosis

❑ Correlation between trend and variation

† In blood albumin levels: positive correlation indicates an upward trend with a gradual increasing pattern of variation, which causes acute inflammation

❑ Contributions of differences in adjacent variation to the disease diagnosis

† Alternating positive and negative fluctuations in blood glucose indicate abnormal insulin secretion

Challenge: Data Sparsity (1/2)

- **EHR is a time series data with limited patient visit records**
  - ❑ Average patient visit is only **10 in 2 years**
  - ❑ Intervals between visits are **irregular**
    - † Average interval between two contiguous visits is as large as **2.5** months

- **Traditional time series decomposition methods are usually suitable for periodic time-series data**
  - ❑ However, **poor Cyclicality** due to data sparsity, which makes **traditional time series decomposition** methods **inapplicable**

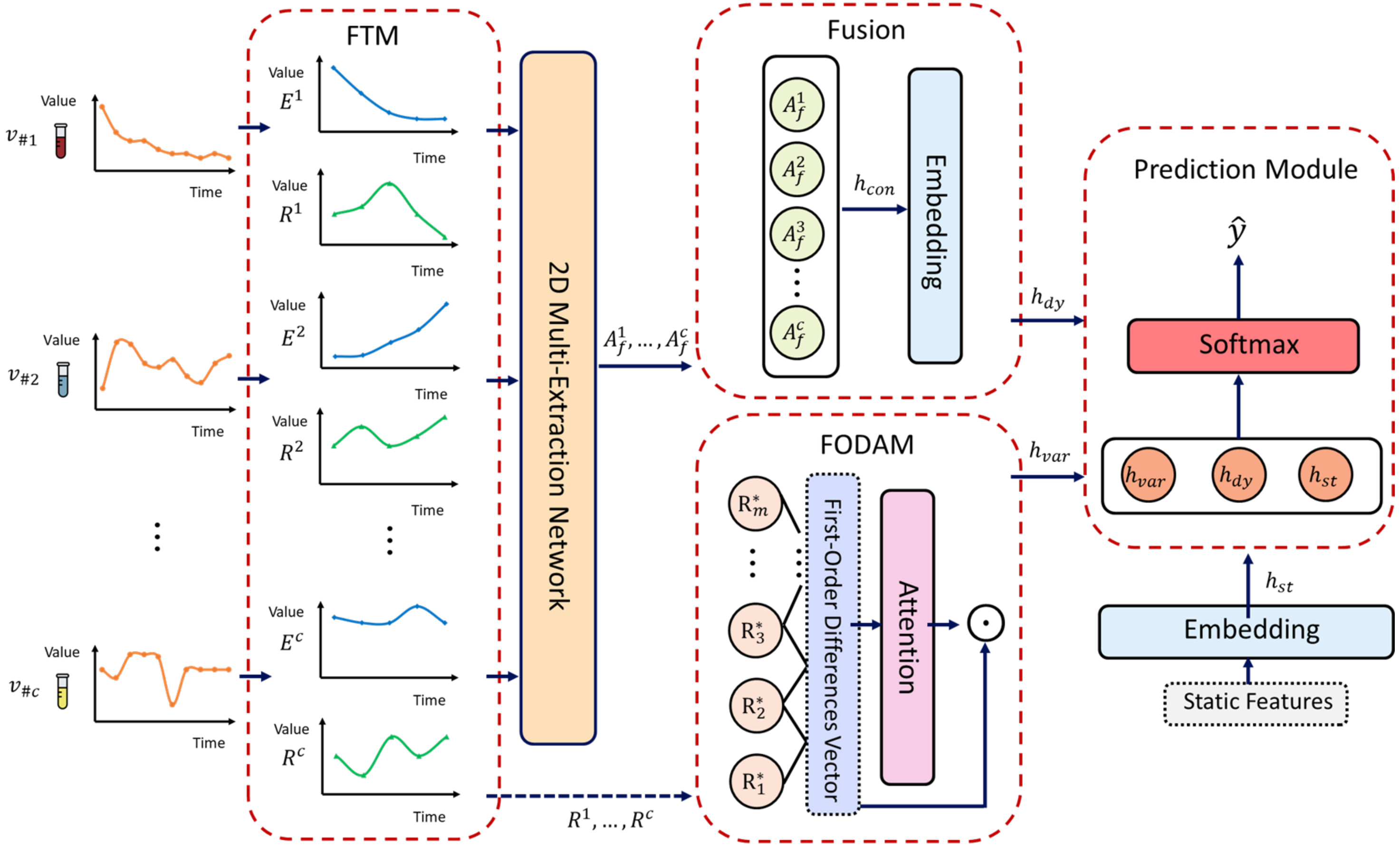
Challenge: Data Sparsity (2/2)

- **EHR is a time series data with limited patient visit records**
  - ❑ Average patient visit is only **10 in 2 years**
  - ❑ Intervals between visits are **irregular**
    - † Average interval between two contiguous visits is as large as **2.5** months

Limit the ability of the **deep learning models** to detect the hidden useful information of medical features

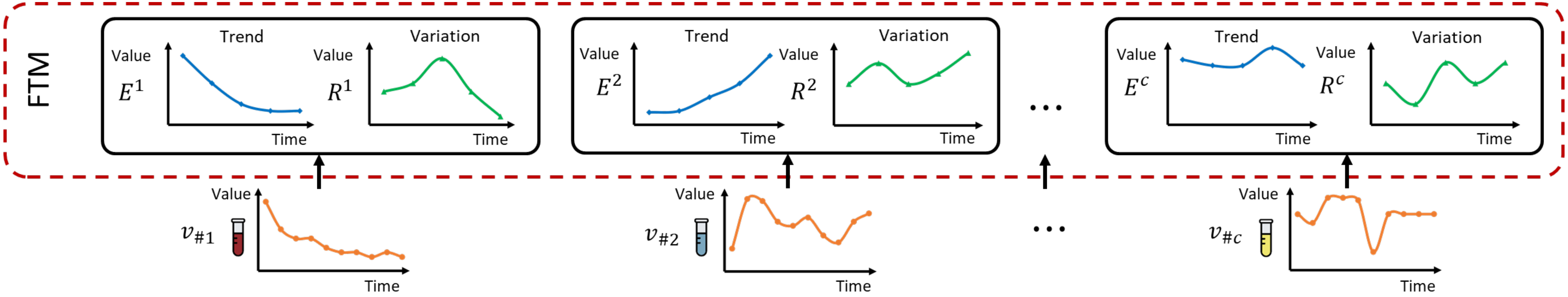
Our Solution

Multi-perspective Patient Representation Extractor (MPRE)



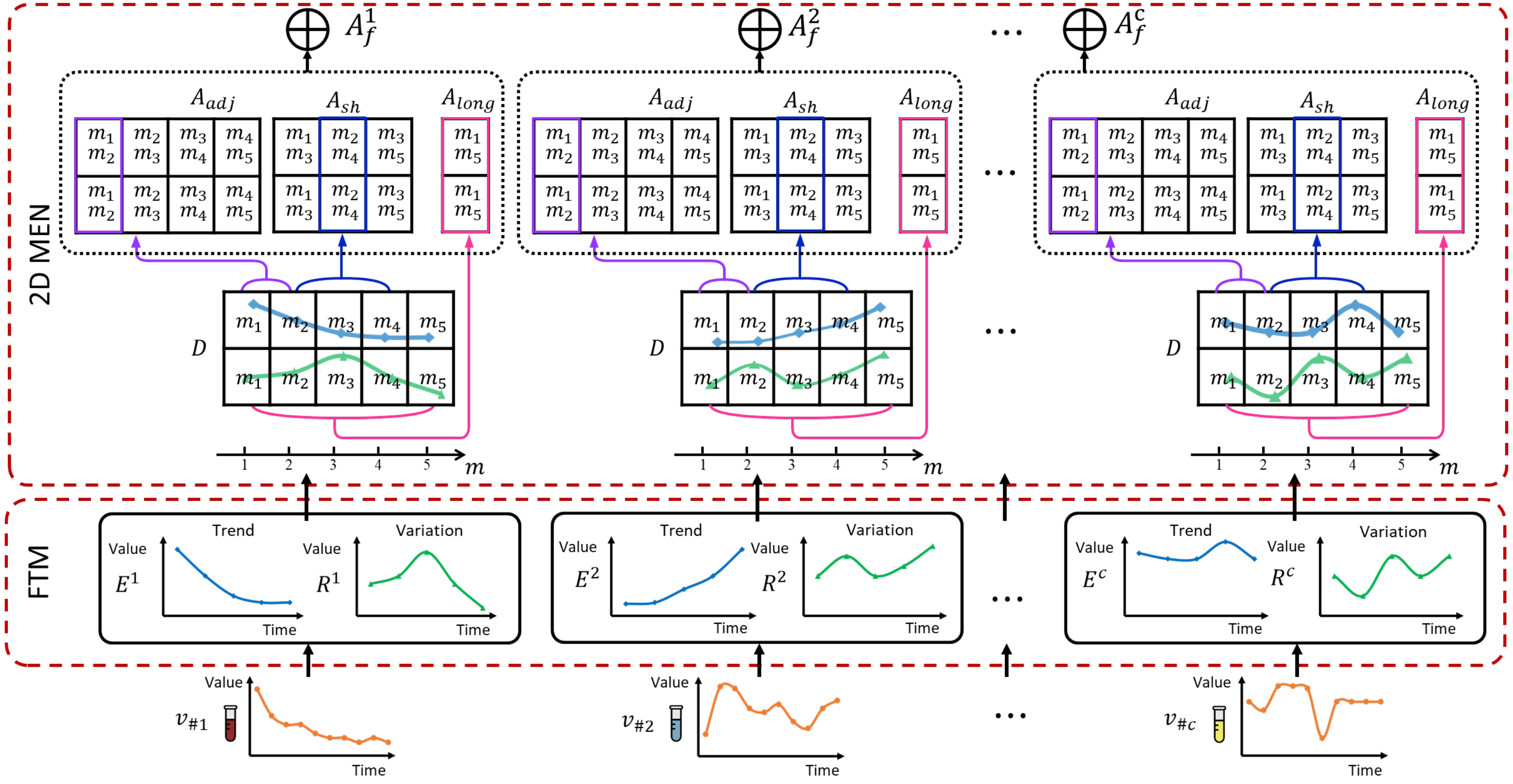
- **Frequency Transformation Module (FTM)**
  - ❑ Extract the trend and variation
- **2D Multi-Extraction Network (2D MEN)**
  - ❑ Capture the correlation between the trend and variation
- **First-Order Difference Attention Mechanism (FODAM)**
  - ❑ Calculate the contribution of differences in adjacent variations

MPRE - FTM



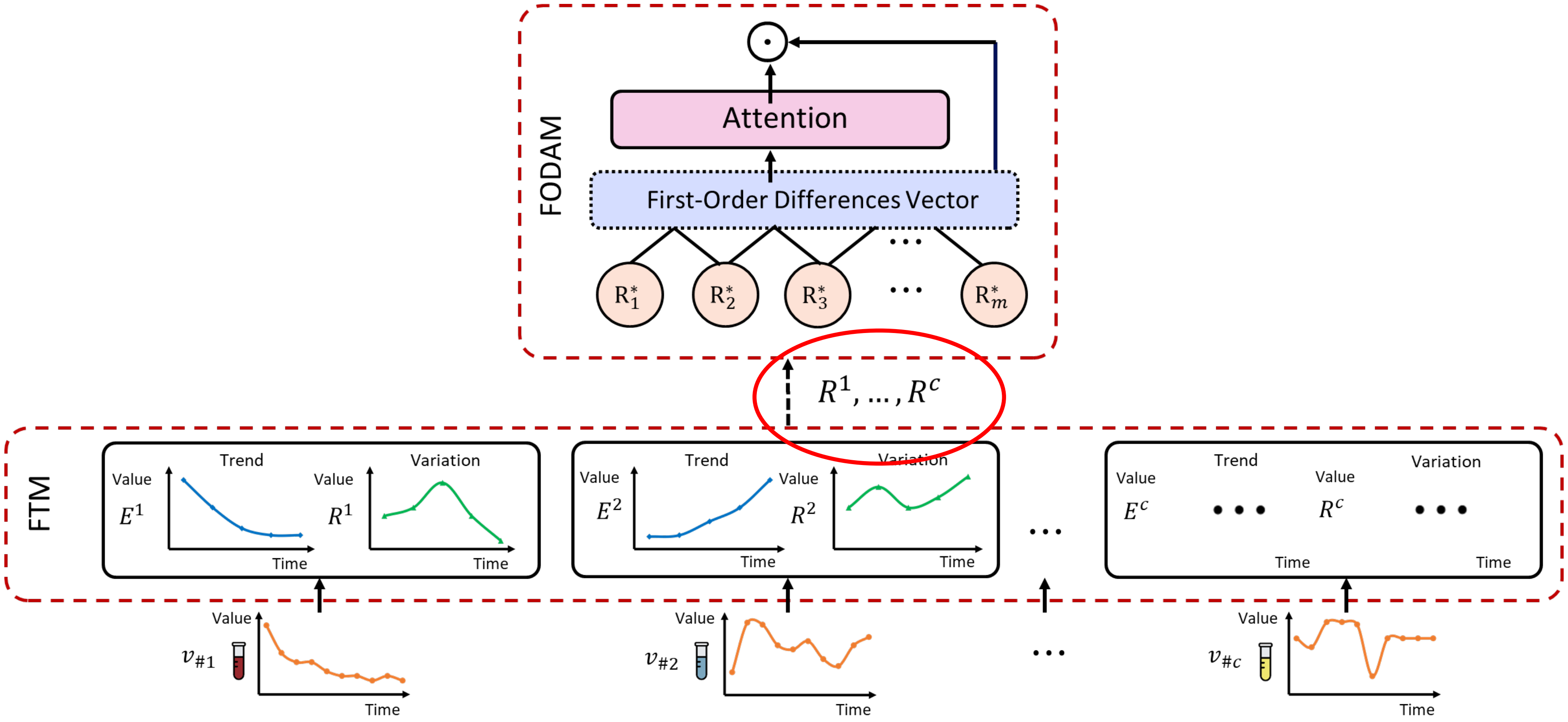
- Symlets wavelet used to decompose each dynamic feature separately
  - ❑ Low-frequency components indicate trend information
  - ❑ High-frequency components express variation information

MPRE - 2D MEN



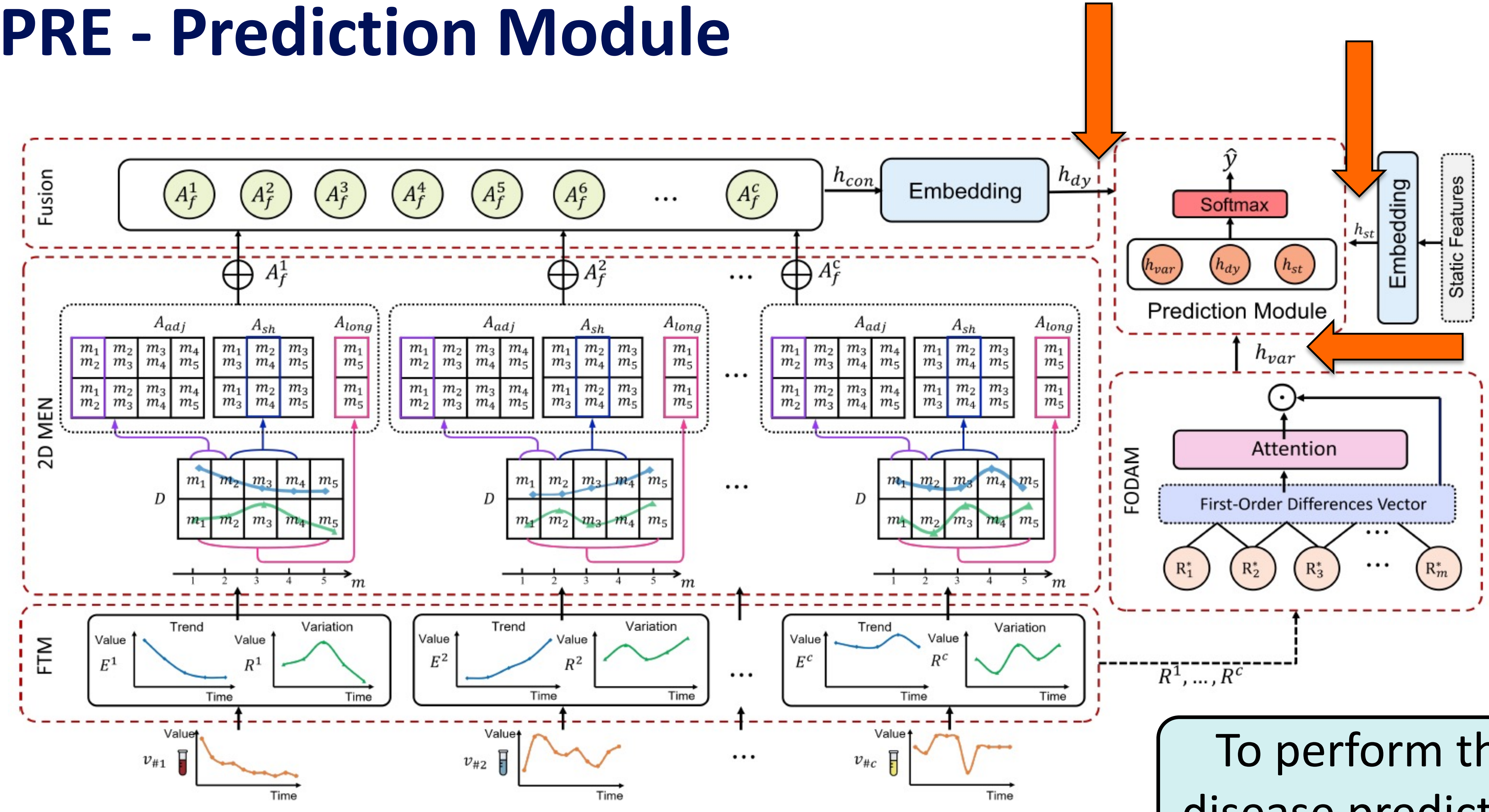
- Reshape trend and variation to form the **2D temporal tensor**
- 2D temporal dilated convolution
  - ❑ Based on different trend and variation spans
- Concatenation operation
  - ❑ To form the representation of dynamic features

MPRE - FODAM



FODAM is used to **adaptively** compute the **contributions of differences** in adjacent variations to the disease progression

MPRE - Prediction Module



To perform the disease prediction

Performance Evaluation

Datasets

- **SCRIPT CarpeDiem Dataset**
  - 12,495 visit records from 585 patients between June 2018 to March 2022.
  - 190 patients had COVID-19, 50 had respiratory viral pneumonia, 252 had bacterial pneumonia, and 93 had respiratory failure.
- **Health Facts Databas**
  - 101,767 visit records f 999 and 2008.
  - Diabetic patients will s in the future, 30,389 visits for 26,744 patient

TABLE I
ICD-9 CODE FOR CIRCULATORY DISEASE

ICD-9 Code	Label
393 - 398	chronic rheumatic heart disease
401 - 405	hypertensive disease
410 - 414	ischemic heart disease
415 - 417	diseases of pulmonary circulation
420 - 429	other forms of heart disease
430 - 438	cerebrovascular disease
440 - 449	diseases of arteries, arterioles, and capillaries
451 - 459	diseases of veins and lymphatics

Performance of MPRE and Baseline Methods

↑ 5.84% ↑ 8.94% ↑ 12.70% ↑ 9.74%

Model	SCRIPT CarpeDiem Dataset		Health Facts Database	
	AUROC	AUPRC	AUROC	AUPRC
GRU [43]	0.7528	0.6405	0.7377	0.6234
TCN [30]	0.8009	0.6751	0.7209	0.6325
RETAIN [26]	0.7612	0.6524	0.7431	0.6190
T-LSTM [34]	0.7338	0.6274	0.7014	0.5978
Dipole [27]	0.8324	0.7428	0.7398	0.6284
SAnD [9]	0.7482	0.6316	0.7263	0.6271
AdaCare [37]	0.7641	0.6449	0.7106	0.6092
StageNet [10]	0.8183	0.7232	0.7326	0.6297
ConCare [35]	0.8425	0.7531	0.7573	0.6507
Ours	**0.8948**	**0.8270**	**0.8675**	**0.7209**

MPRE achieves the **best performance** than baseline methods

Ablation Study

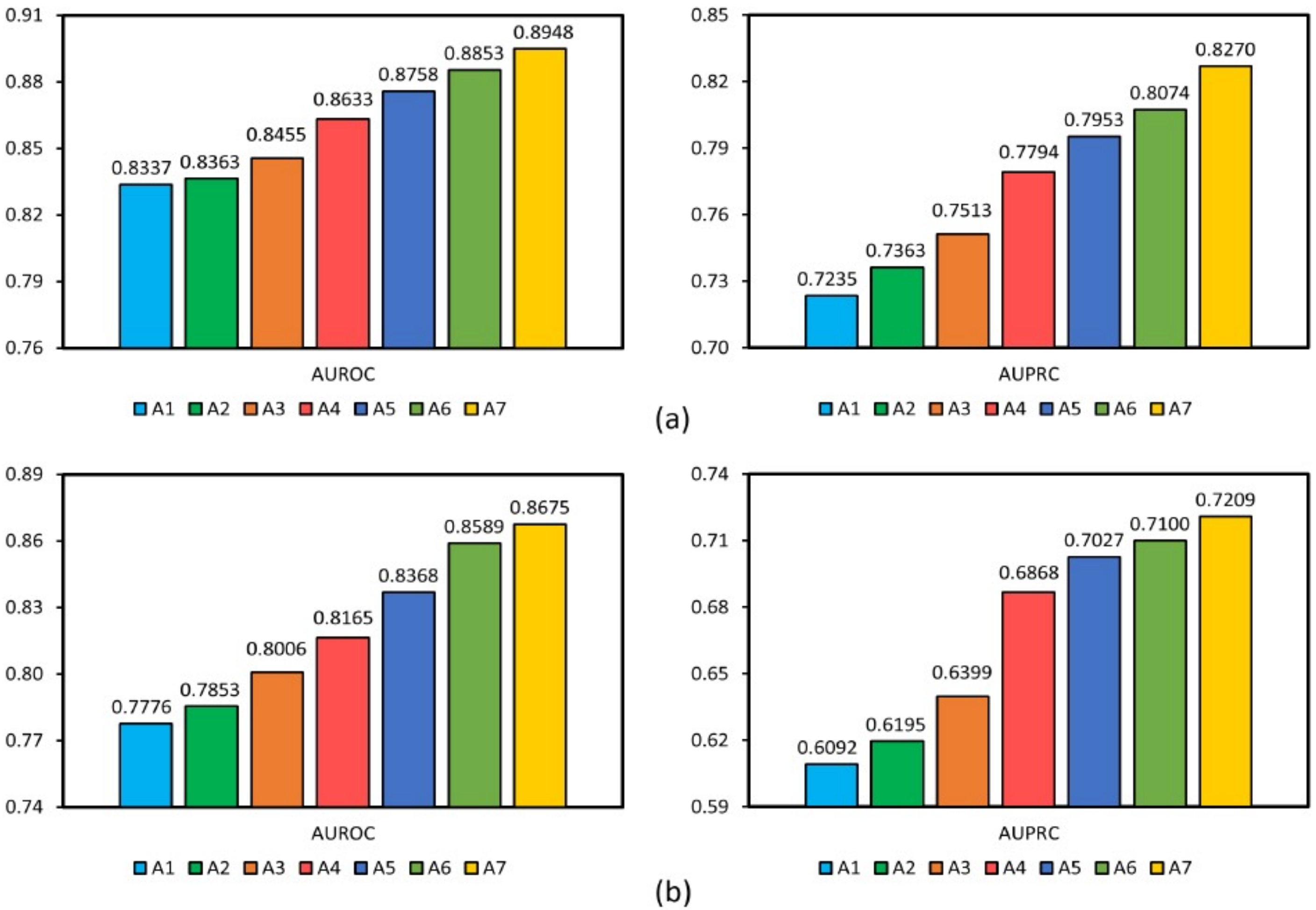
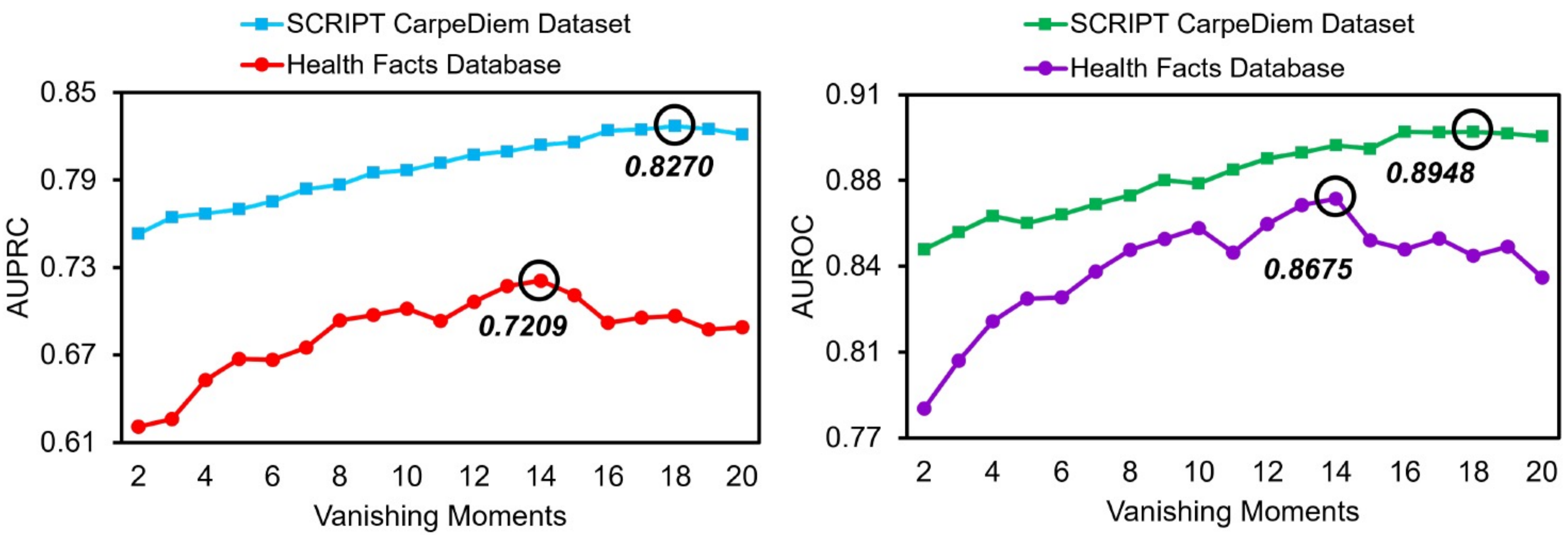


TABLE III
CONFIGURATIONS FOR ABLATION STUDIES

Configurations	FTM	2D MEN	FODAM	Trend	Variation
A1	✓	×	×	✓	×
A2	✓	×	×	×	✓
A3	✓	×	✓	×	✓
A4	✓	×	×	✓	✓
A5	✓	×	✓	✓	✓
A6	✓	✓	×	✓	✓
A7	✓	✓	✓	✓	✓

- (a) shows the average performance on SCRIPT CarpeDiem Dataset
- (b) shows the average performance on Health Facts Database

Analysis of Symlets



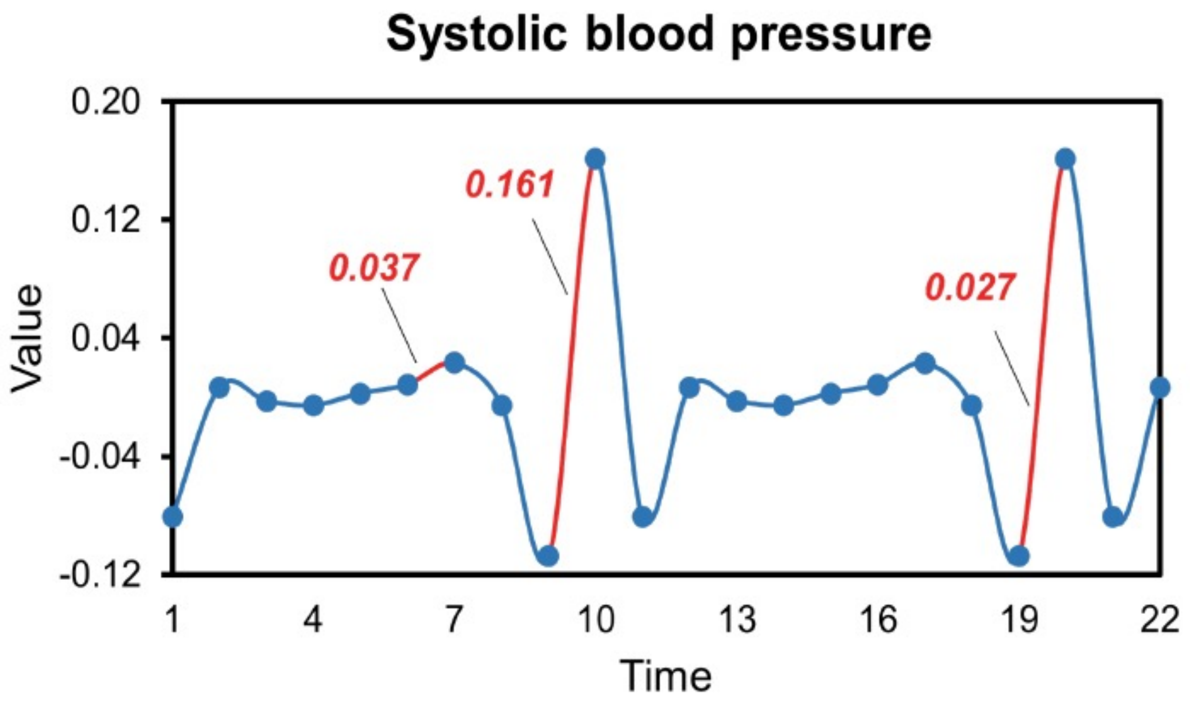
- Symlet-18 for SCRIPT CarpeDiem Dataset
- Symlet-14 for Health Facts Database

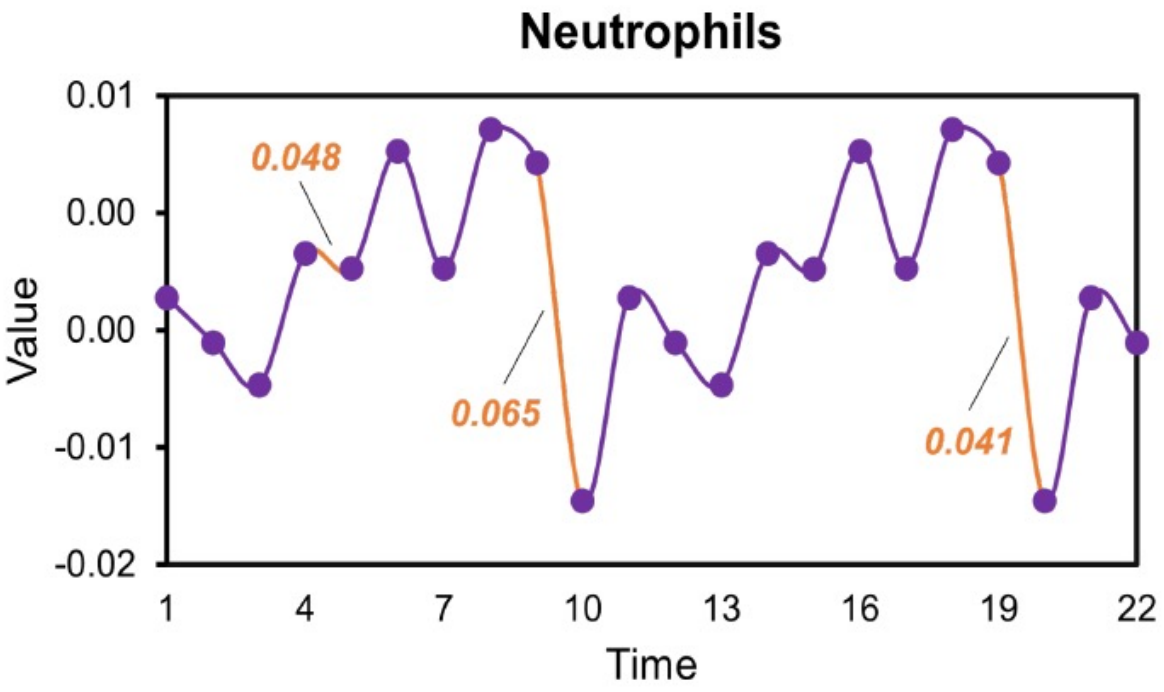
Correlation between Trend and Variation

Types of Respiratory Disease	Top 5 Correlations between the trend and variation
Bacterial Pneumonia	diastolic blood pressure (0.85) hemoglobin (0.83) mean arterial pressure (0.81) systolic blood pressure (0.81) heart rate (0.81)
Respiratory Viral Pneumonia	Platelets (0.80) blood pressure (0.73) diastolic blood pressure (0.73) respiratory rate (0.73) heart rate (0.73)
COVID-19	lymphocytes (0.87) peep changes (0.74) fio2 (0.72) peep (0.70) respiratory rate changes (0.68)
Respiratory Failure	bicarbonate (0.79) heart rate (0.59) urine output (0.58) platelets (0.55) diastolic blood pressure (0.54)

- **SCRIPT CarpeDiem Dataset**
  - ❑ Top 5 correlations in dynamic features among four respiratory diseases

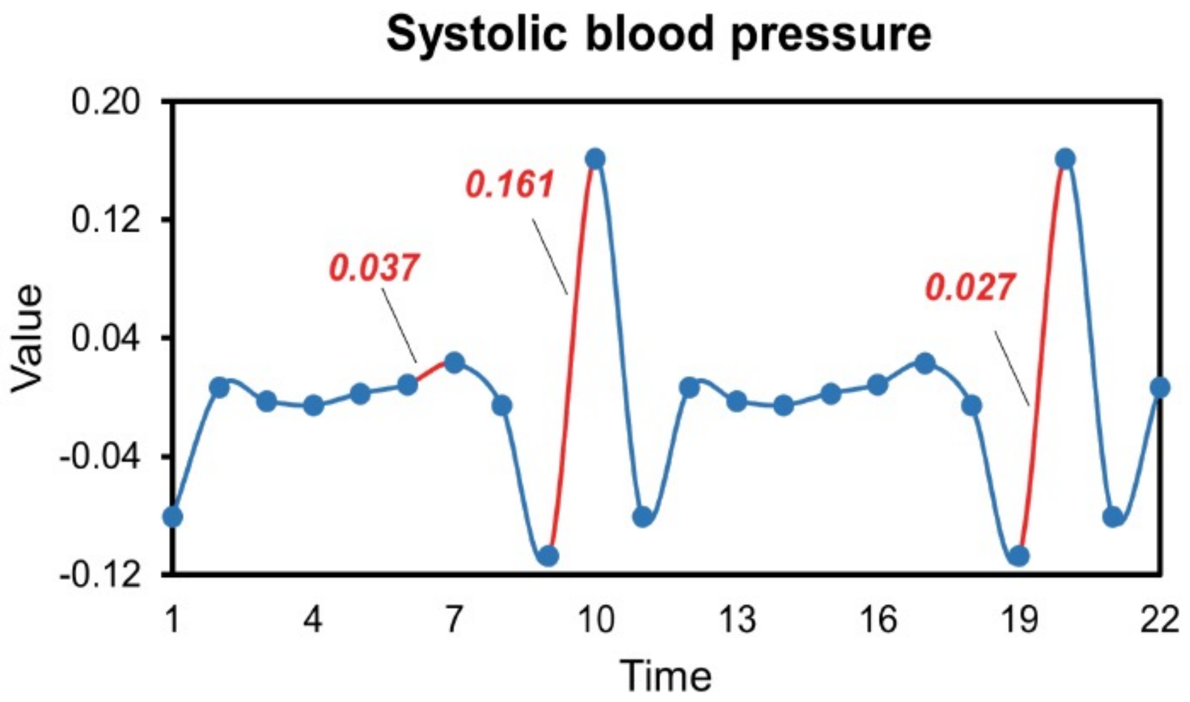
Attention Scores for Differences (1/2)

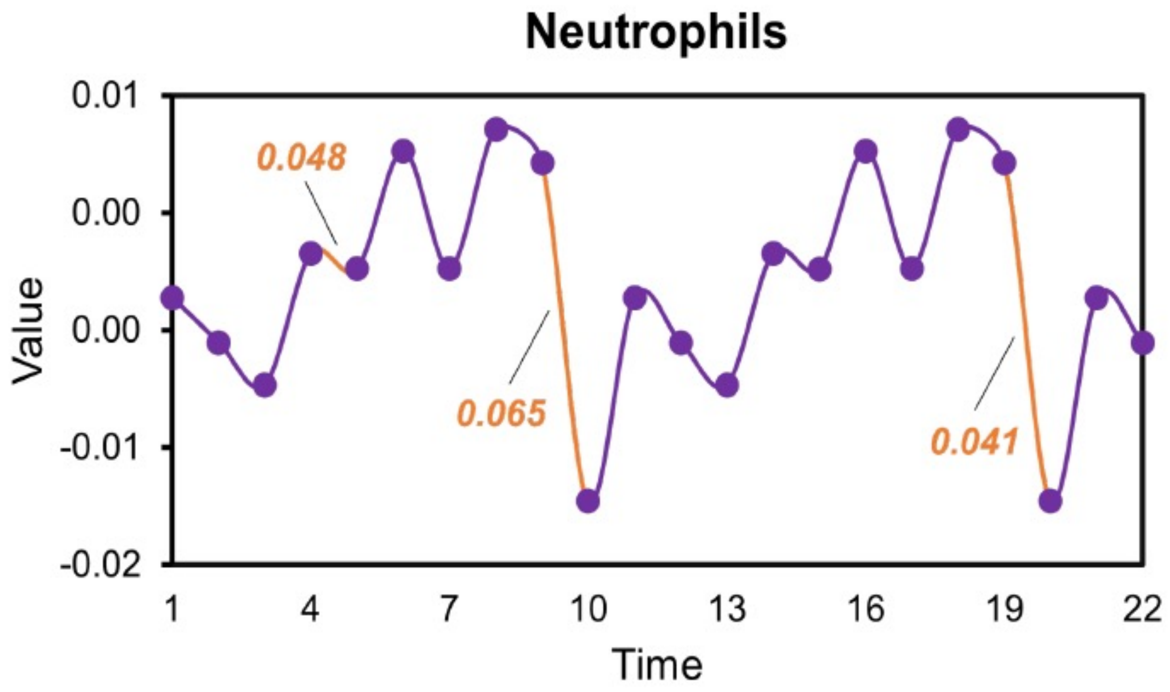




Attention scores reach their **pinnacle** when the patient undergoes the **first** substantial rise or fall

Attention Scores for Differences (2/2)





Physicians should pay attention to **early changes** in the patient’s health status

Conclusion

- We propose MPRE for disease prediction
  - ❑ Extract the trend and variation information
  - ❑ Capture the correlation between the trend and variation
  - ❑ Detect the contributions of differences in dynamic features' variations
- We compare the performance of MPRE and state-of-the-art baseline methods on the two real-world public datasets
  - ❑ The experiment results show that MPRE outperforms all baseline methods in terms of AUROC and AUPRC

Thank you for your listening!

Questions?