StageNet: Stage-Aware Neural Networks for Health Risk Prediction

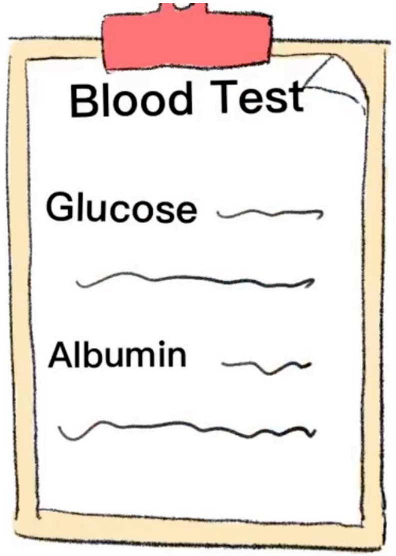
Junyi Gao, Cao Xiao, Yasha Wang, Wen Tang, Lucas M. Glass, Jimeng Sun

Content

- Clinical Background
- Challenges
- Method
- Experiment Results

Clinical Background 1: What is Electronic Health Records (EHR) data?

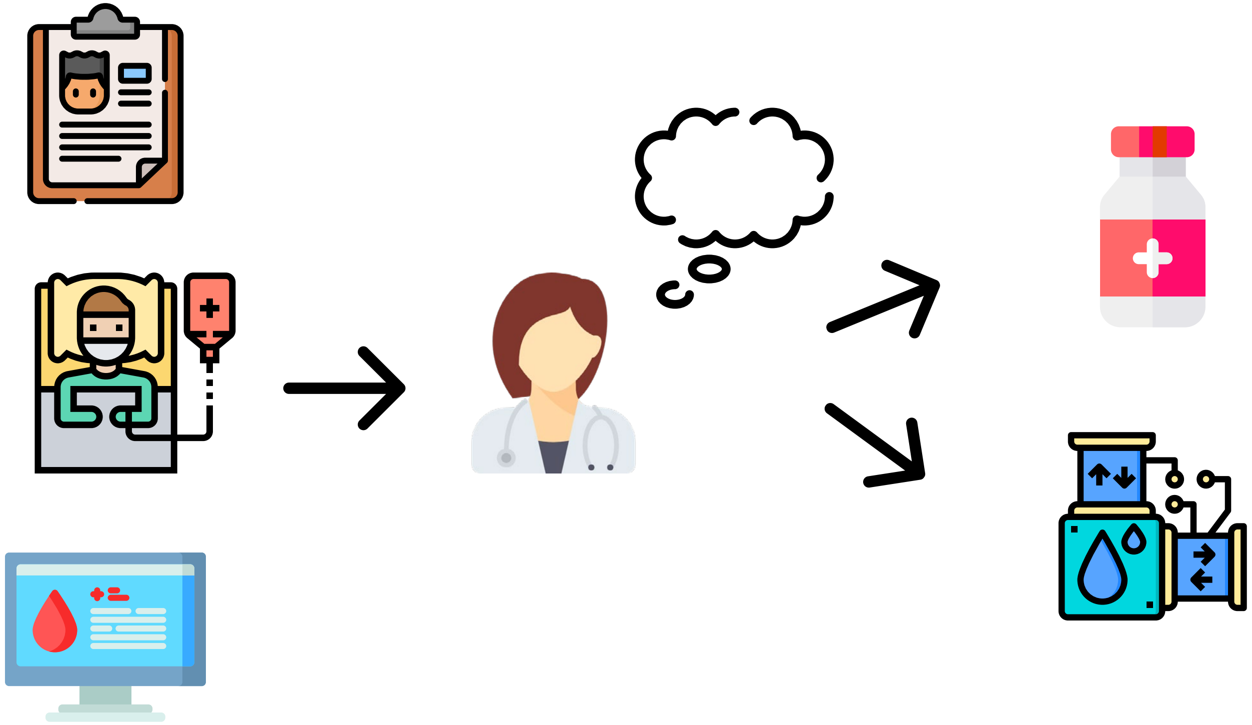
- A type of high-dimensional sequence data
  - Lab test
  - Diagnosis
  - Drugs



		Blood lab test		Physical lab test		New-onset complications			...
Patient ID	Date	Albumin	...	Weight	...	URTI	Cancer	...	...
98	19/12/2015	35.0		66.3		0	0		
98	22/01/2016	38.2		65.2		1	0		
118	09/07/2015	32.8		49.8		0	1		

Clinical Background 2: Why do we need to evaluate the health risk of patients?

- Make personalized treatment plan
- Prevent adverse outcomes



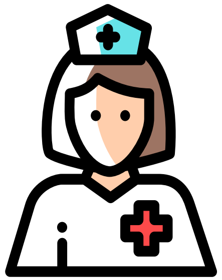
Icons are downloaded from *flaticon.com*.

Clinical Background 2: Why do we need to evaluate the health risk of patients?

- Statistic facts for COVID-19 (according to WHO):

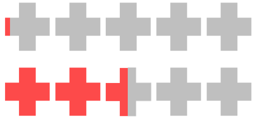
Severe condition patients in China

20%



Overall Mortality Rate

0.7%~4%



50%

Severe Condition Mortality Rate

Normal Bed Shortage

17%

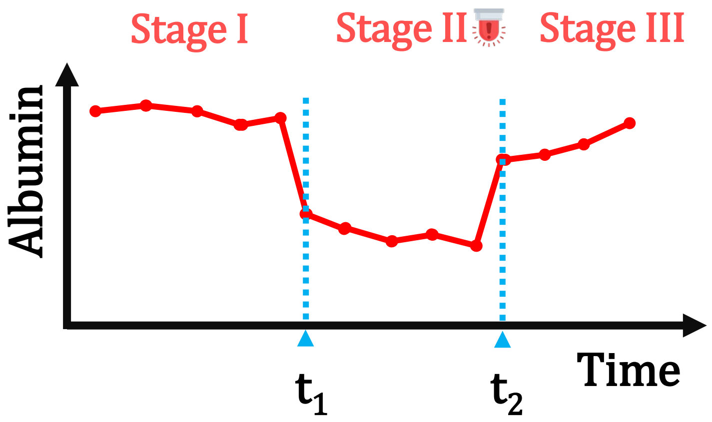
47%

ICU Bed Shortage

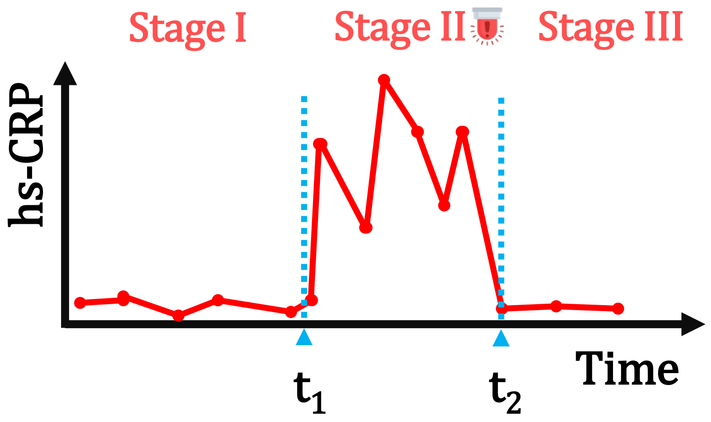
- Allocate limited medical resources
- Reduce mortality rate

Clinical Background 3: What is disease stage and why it matters?

- A time period with consistent status progression
  - Not specific to a single disease



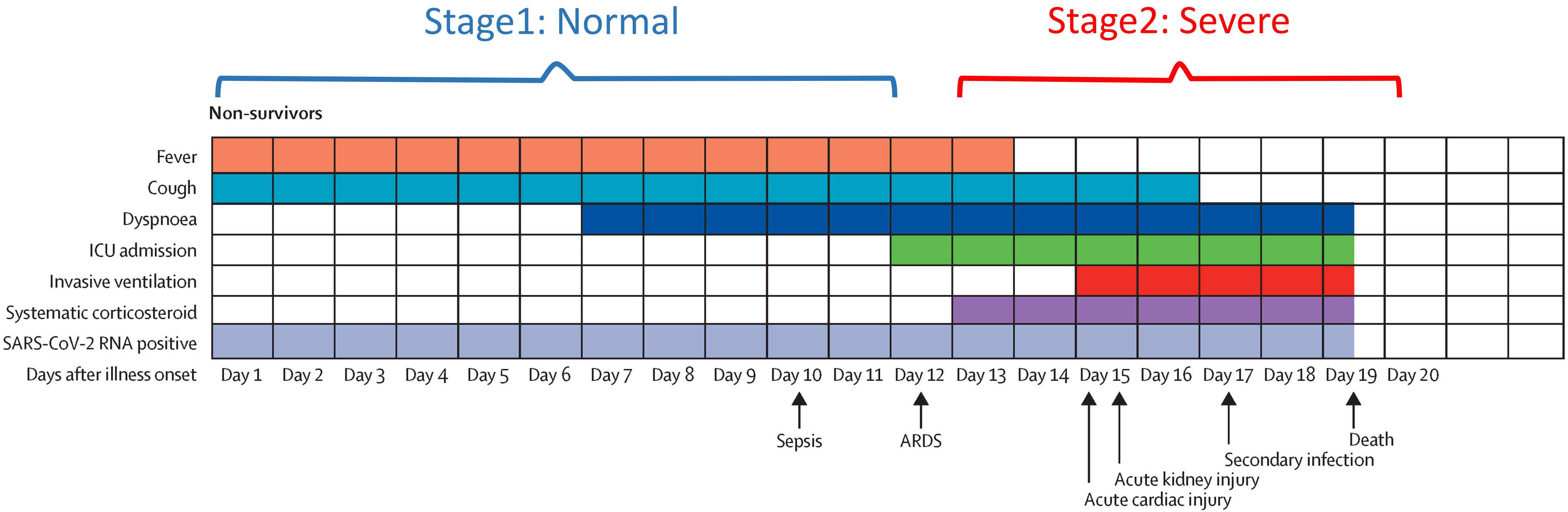
Albumin



hs-CRP

Clinical Background 3: What is disease stage and why it matters?

- Severe symptoms & Rapid deteriorate status for COVID patients



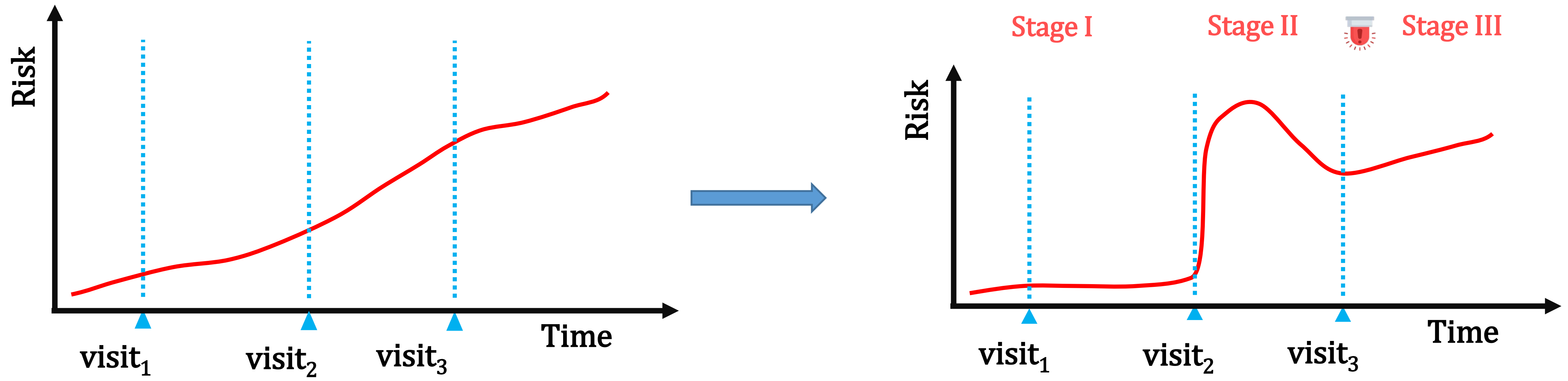
Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study, The Lancet

Content

- **Clinical Background**
- **Challenges**
  - Extract disease progression stages
  - Leverage disease progression stage information
- **Method**
- **Experiment Results**

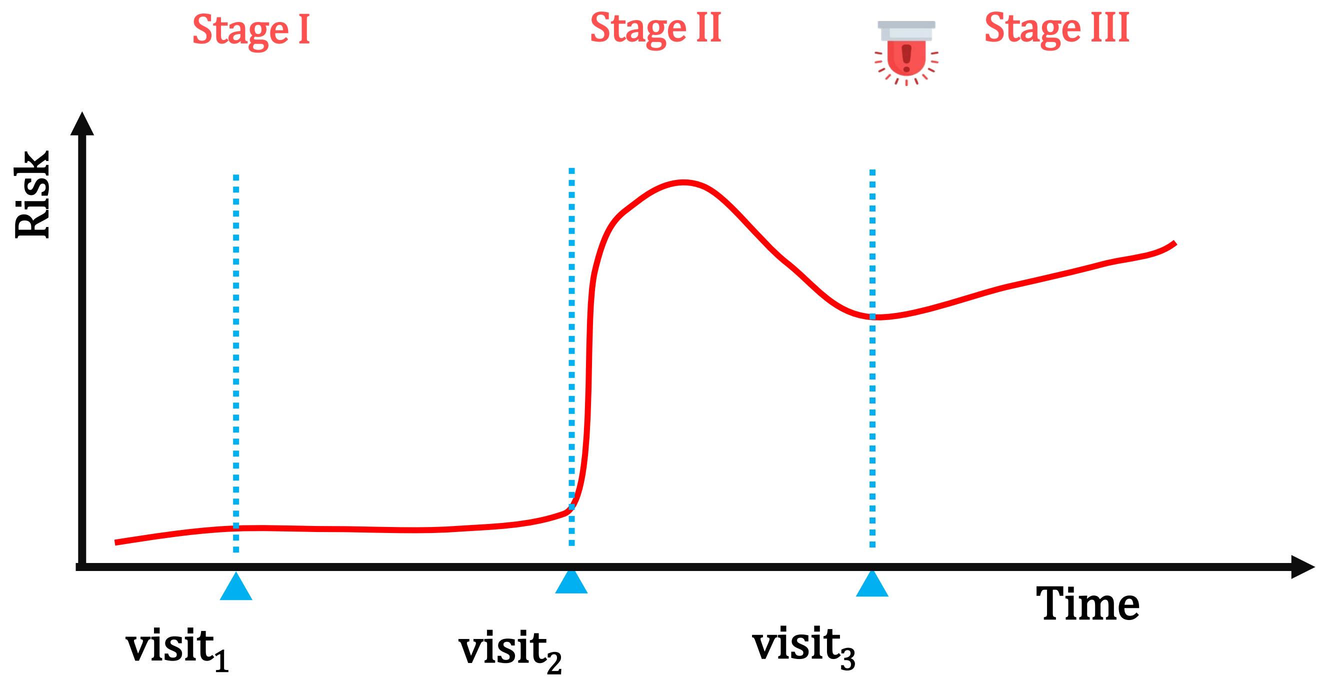
Challenge 1: Extract disease progression stages

- Disease progression speed vary significantly
    - depending on the underlying *disease stage*.



Challenge 2: Leverage disease progression stage information

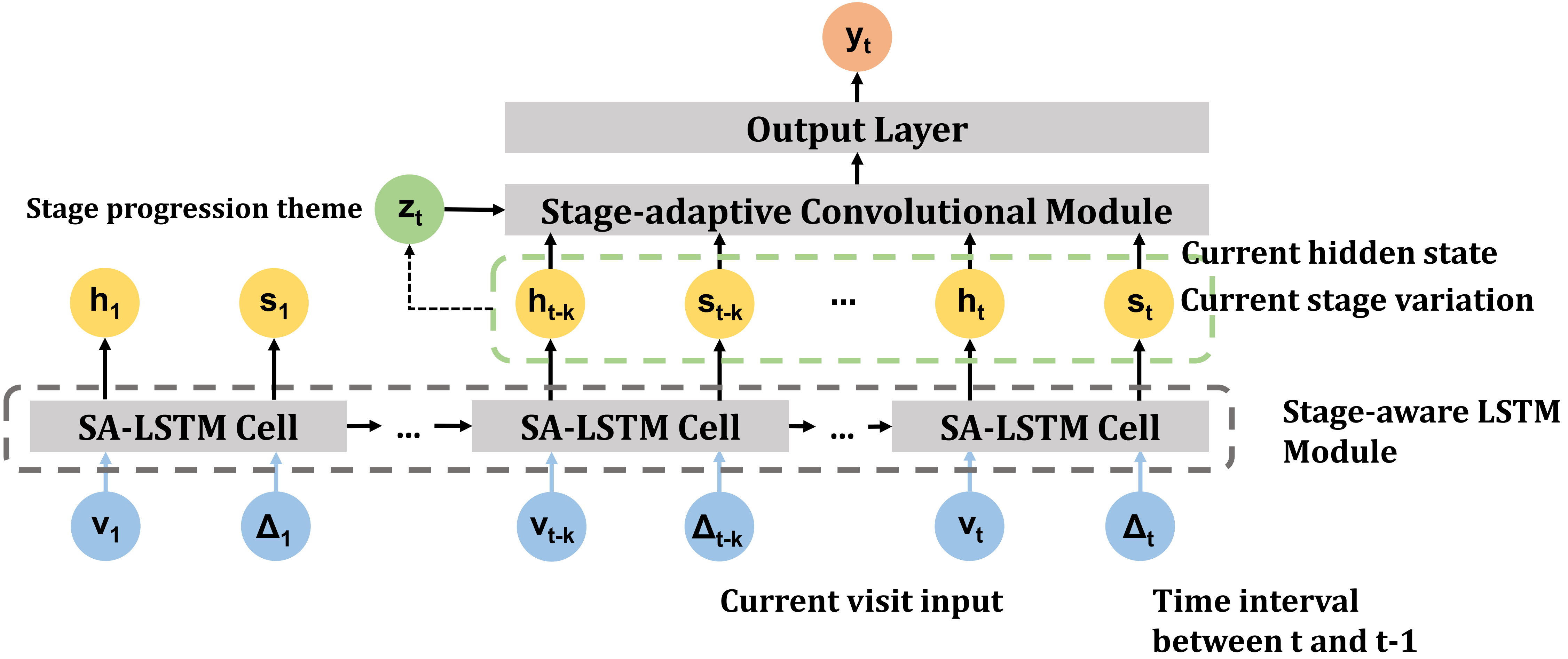
- Disease progression has different patterns
  - Highly related to patients' health risk.
  - Similar within one stage but different across stages.



Content

- **Clinical Background**
- **Challenges**
- **Method**
  - Stage-aware LSTM module
  - Stage-adaptive convolutional module
- **Experiment Results**

Method Overview: StageNet



Method: Stage-aware LSTM module

- Ordered Neuron mechanism (Shen et al.)
- Two new gates to store historical and latest information in c_t
  - $\tilde{f}_t$ (master forget gate) for old history;
    e.g. [0.2, 0.4, 0.8, 1]
  - $\tilde{i}_t$ (master input gate) for recent history;
    [1, 0.8, 0.4, 0.2]

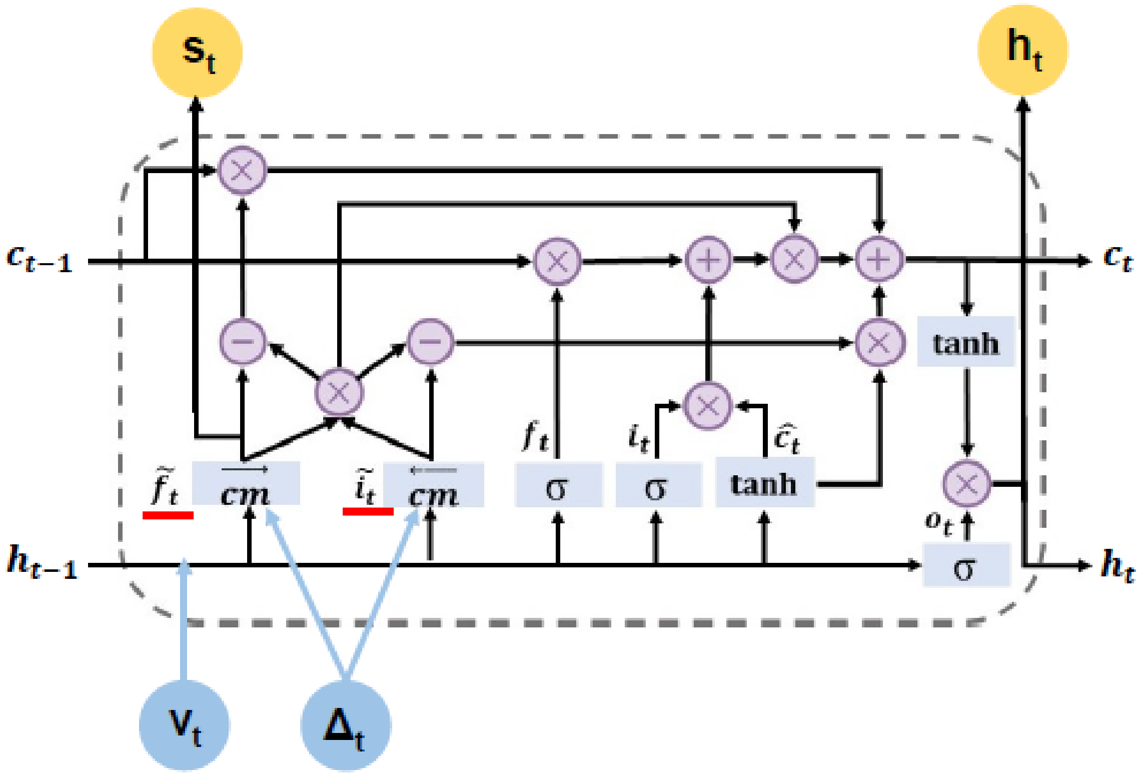
$$p_{\tilde{f}} = softmax(W_{\tilde{f}}(\boldsymbol{v}_t \oplus \Delta_t) + U_{\tilde{f}}(\boldsymbol{h}_{t-1} \oplus \Delta_t) + b_{\tilde{f}})$$

$$p_{\tilde{i}} = softmax(W_{\tilde{i}}(\boldsymbol{v}_t \oplus \Delta_t) + U_{\tilde{i}}(\boldsymbol{h}_{t-1} \oplus \Delta_t) + b_{\tilde{i}})$$

$$\tilde{f}_t = \overrightarrow{\mathrm{cm}}(\boldsymbol{p}_{\tilde{f}})$$

$$\tilde{i}_t = \overleftarrow{\mathrm{cm}}(\boldsymbol{p}_{\tilde{i}})$$

e.g. $\overrightarrow{\mathrm{cm}}([0,1,2,3]) = [0,1,3,6]$
$\overleftarrow{\mathrm{cm}}([0,1,2,3]) = [6,6,5,3]$



Method: Stage-aware LSTM module

- Latest information ($\hat{c}_t$) -> High ranking neurons
- Historical information (c_{t-1}) -> Low ranking neurons
- Update cell state $\boldsymbol{c_t}$ as:

$$\hat{c}_t = tanh(W_c v_t + U_c h_{t-1} + b_c)$$

$$\underline{w_t = \widetilde{f}_t \odot \widetilde{i}_t}$$

$$c_t = w_t \odot (f_t \odot \underline{c_{t-1}} + \underline{i_t} \odot \hat{c}_t)$$

$$+ (\underline{\widetilde{f}_t - w_t}) \odot \underline{c_{t-1}} + (\underline{\widetilde{i}_t - w_t}) \odot \underline{\hat{c}_t}$$

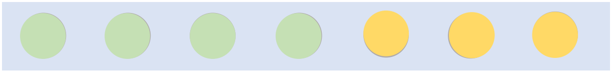
$$h_t = o_t \odot tanh(c_t)$$

- The stage variation:

$$s_t = argmax(p_{\widetilde{f}}) \approx \sum_{i=1}^{N_h} i \times p_{\widetilde{f}}(i) = N_h\left(1 - \frac{1}{N_h} \sum_{i=1}^{N_h} \widetilde{f}_t(i)\right) + 1$$

$\hat{c}_t$ c_{t-1}

c_t



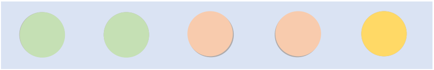
$p_{\tilde{i}} = [0, 0, 0, 1, 0]$ $p_{\tilde{f}} = [0, 0, 1, 0, 0]$

$\widetilde{i}_t = [1, 1, 1, 1, 0]$ $\widetilde{f}_t = [0, 0, 1, 1, 1]$

$w_t = [0, 0, 1, 1, 0]$

$\widetilde{i}_t - w_t = [1, 1, 0, 0, 0]$ $\widetilde{f}_t - w_t = [0, 0, 0, 0, 1]$

c_t



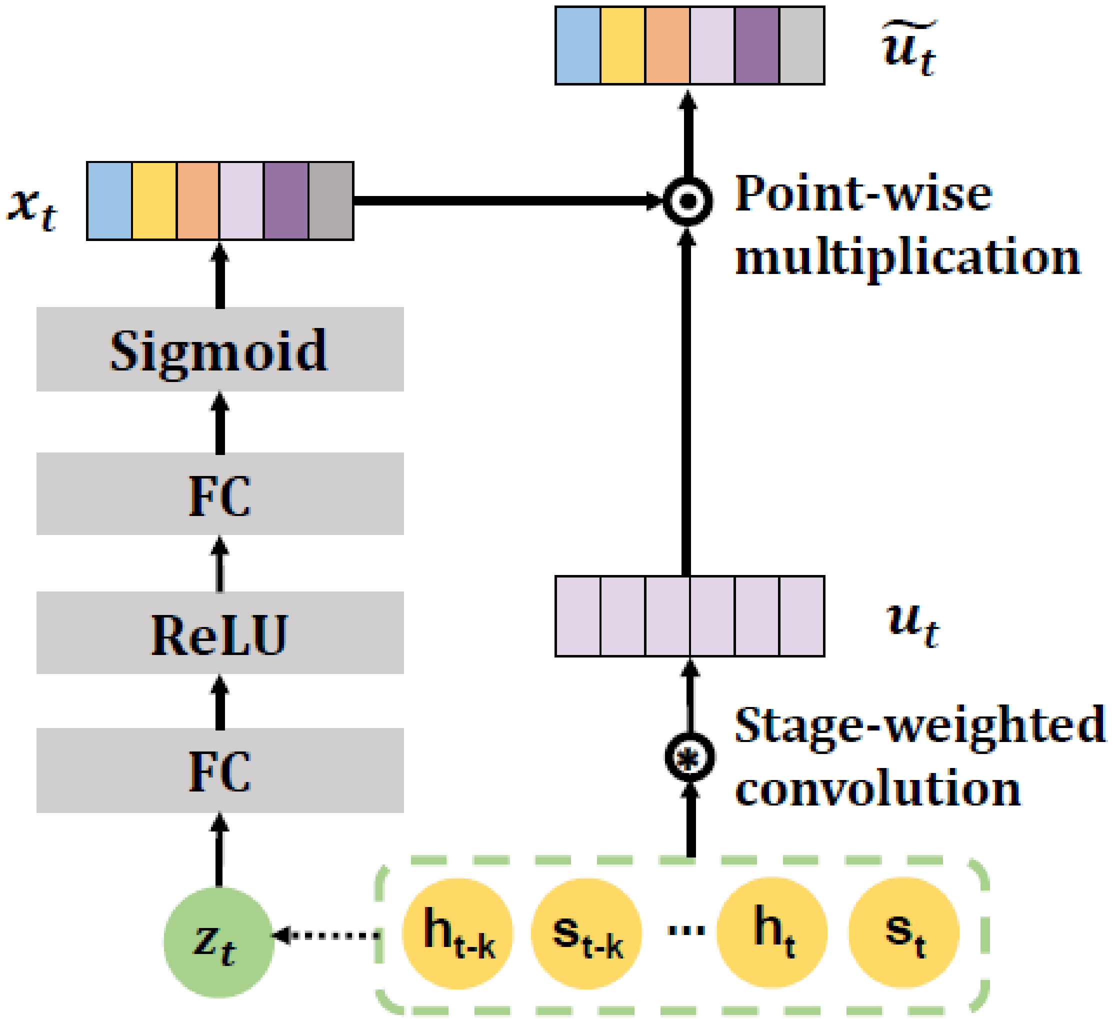
$\hat{c}_t$ c_{t-1}

$f_t \odot c_{t-1} + i_t \odot \hat{c}_t$

$s_t = 2$

Method: Stage-adaptive convolutional module

- Extract patterns that are closely related to current stage
- Adaptively select the **most informative** patterns

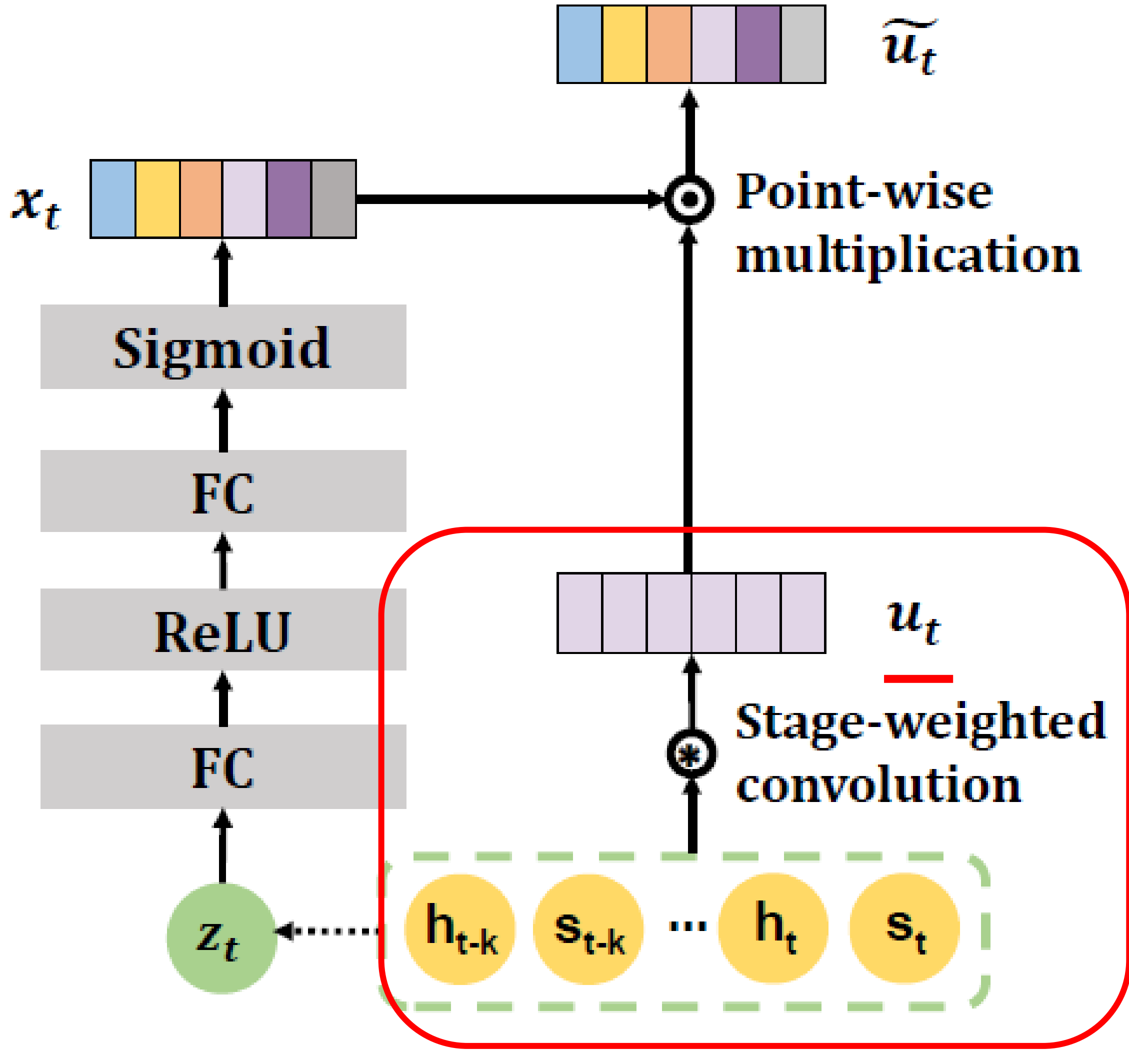


Method: Stage-adaptive convolutional module

- 1. Learning stage progression patterns
  - The distance between stages of historical visits and the stage of current visit:

$$\Delta s^t = \mathrm{softmax}(\overrightarrow{\mathrm{cm}}\ (s_{t-K}, \ldots, s_t))$$

  - Re-weight the hidden state and obtain raw progression patterns u_t

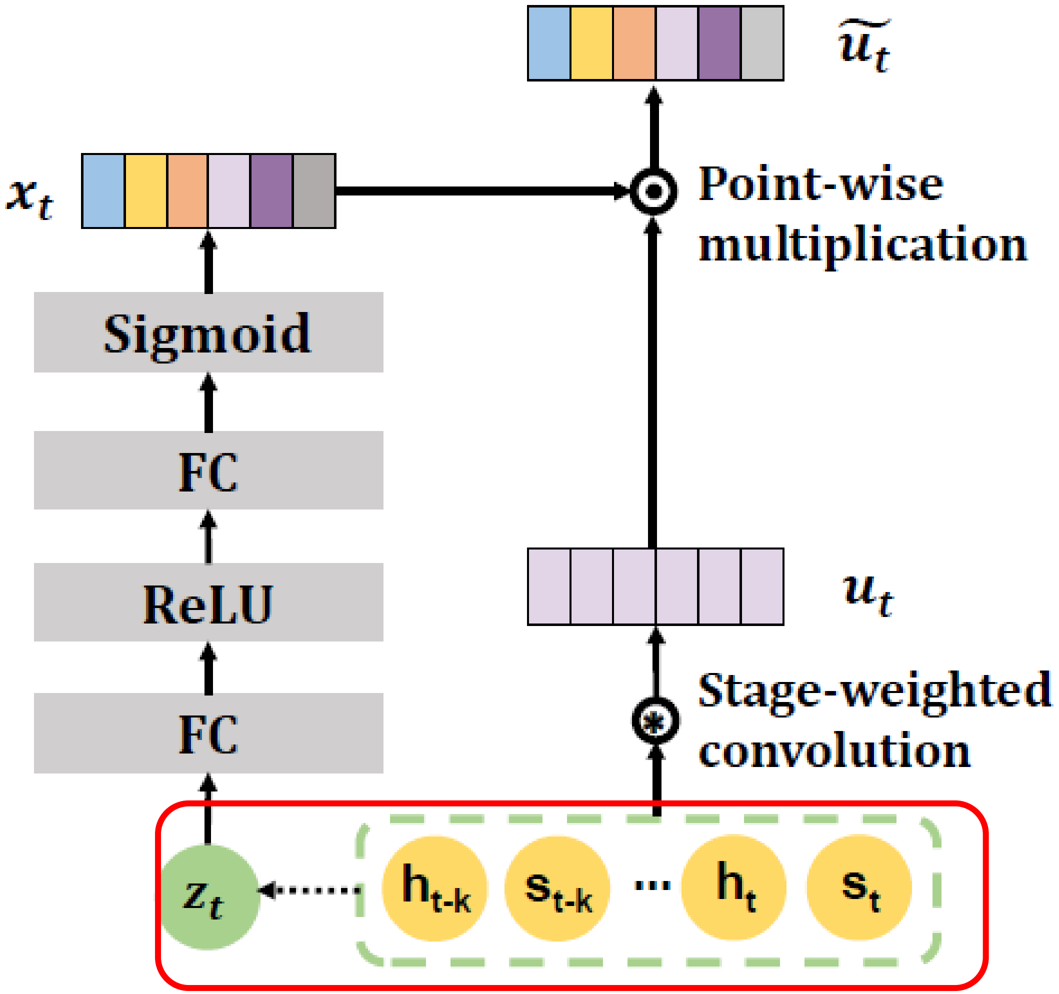


Method: Stage-adaptive convolutional module

- 2. Extracting progression theme at the current stage
  - Global view of patients’ status at the current stage:

$$z_t = \frac{1}{K}\sum_{i=0}^{K} \Delta s_i^t h_{t-K+i}$$

  - z_t - progression theme of the current stage



Method: Stage-adaptive convolutional module

- 3. Re-calibrating progression patterns
  - Calculate pattern importance:

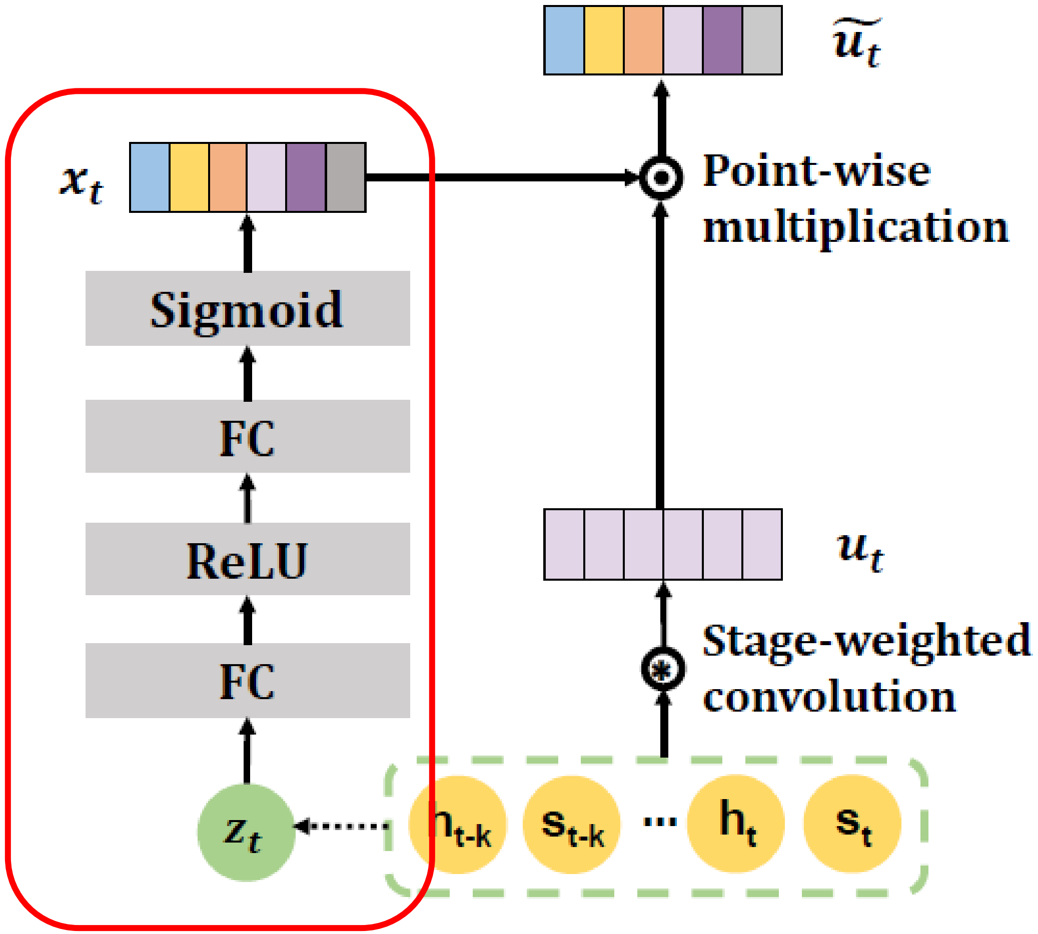
    $$\boldsymbol{x}_t = \sigma(\mathrm{W}_{x1}\delta(\mathrm{W}_{x2}z_t))$$

  - Re-calibrate patterns:

    $$\widetilde{\boldsymbol{u}}_t = \boldsymbol{u}_t \odot \boldsymbol{x}_t$$

- Predict current health risk:

$$\hat{y}_t = \sigma(\mathrm{W}_y(\widetilde{\boldsymbol{u}}_t + \boldsymbol{h}_t) + \mathrm{b}_y)$$



Content

- **Background & Motivation**
- **Problem Formulation**
- **Insights**
- **Solution**
- **Experiment**
    - Health risk prediction
    - Patient subtyping
    - Visualization

Experiment

- **Dataset**
  - MIMIC-III Dataset: Decompensation Prediction
    - ICU data from the Medical Information Mart for Intensive Care (MIMIC-III) database
  - ESRD Dataset: Mortality Prediction
    - End-stage renal disease dataset from Peking University 3rd Hospital
- **Metric**
    - AUPRC, AUROC, Min(Se, p+)

Experiment: Health Risk Prediction

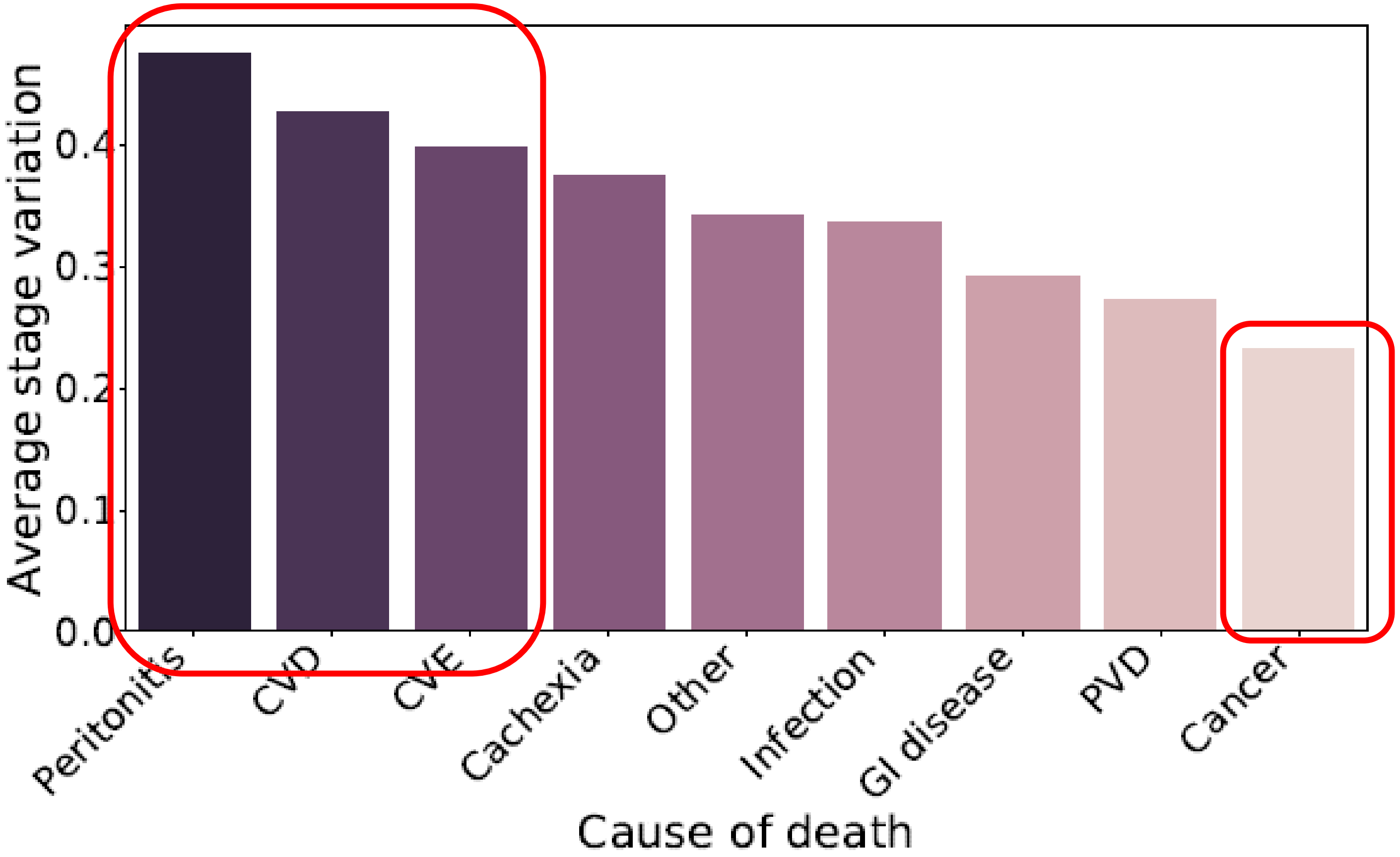
- Outperforms all baseline models across both datasets in all evaluation metrics.

- 10% higher AUPRC on MIMIC-III dataset
- 12% higher AUPRC on ESRD dataset

		MIMIC-III			ESRD		
	Model	AUPRC	AUROC	min(Re, P+)	AUPRC	AUROC	min(Re, P+)
Baseline	LSTM	0.280 (0.003)	0.897 (0.002)	0.324 (0.003)	0.270 (0.029)	0.805 (0.026)	0.318 (0.015)
	ON-LSTM	0.304 (0.002)	0.895 (0.003)	0.343 (0.004)	0.291 (0.021)	0.810 (0.021)	0.333 (0.034)
	T-LSTM	0.282 (0.004)	0.895 (0.002)	0.322 (0.005)	0.276 (0.027)	0.812 (0.026)	0.331 (0.031)
	Decay-LSTM	0.294 (0.002)	0.893 (0.003)	0.330 (0.004)	0.289 (0.020)	0.808 (0.022)	0.328 (0.021)
	Health-ATM^{-}	0.291 (0.002)	0.897 (0.003)	0.325 (0.003)	0.287 (0.021)	0.810 (0.039)	0.331 (0.025)
Reduced Model	StageNet-I	0.313 (0.003)	0.899 (0.003)	0.360 (0.002)	0.296 (0.014)	0.814 (0.031)	0.333 (0.018)
	StageNet-II	0.311 (0.003)	0.897 (0.002)	0.358 (0.003)	0.302 (0.029)	0.812 (0.027)	0.334 (0.017)
Proposed	**StageNet**	**0.323 (0.002)**	**0.903 (0.002)**	**0.372 (0.003)**	**0.327 (0.022)**	**0.821 (0.024)**	**0.352 (0.019)**

Experiment: Health Risk Prediction

- **Health status stability vs. Cause of death**
  - A patient with stable health status will have a low s_t



Experiment: Health Risk Prediction

- **Health status stability vs. Health risk**
  - Divide patients' visits into three groups:
    - Low risk (risk score <= 0.4)
    - Medium risk (0.4 < risk score <= 0.7)
    - High risk (risk score >= 0.7)

Risk level	Low risk	Medium risk	High risk
Avg. stage var.	0.354 (0.003)	0.393 (0.003)	0.437 (0.005)

Experiment: Patient Subtyping

- **Subtyping on ESRD dataset**
  - Calinski-Harabasz score and average groud truth mortality risk

  - 58% higher Calinski-Harabasz score
  - Two high-risk groups (Cluster III and IV) & Two low-risk groups (Cluster I and II).

	C-H score	Cluster I	Cluster II	Cluster III	Cluster IV
LSTM	74	0.09	0.19	0.33	0.59
ON-LSTM	104	0.08	0.20	0.25	0.48
T-LSTM	31	0.03	0.08	0.09	0.85
Decay-LSTM	43	0.08	0.23	0.24	0.54
Health-ATM^{-}	81	0.07	0.23	0.28	0.47
StageNet	**165**	0.05	0.06	0.60	0.69

Experiment: Patient Subtyping

- **Subtyping on ESRD dataset**
  - Using T-test to identify discriminative features (p-value < 0.05)

  - StageNet can identify more discriminative features

StageNet			
Cluster III	Albumin Appetite	Serum creatinine C-rp	Blood urea
Cluster IV	C-rp Albumin	DBP Hemoglobin	Blood potassium

LSTM			
Cluster II	Albumin DBP	C-rp Glucose	Blood chlorine
Cluster III	Albumin Glucose	C-rp DBP	Blood chlorine
Cluster IV	Albumin Serum creatinine	C-rp Glucose	DBP

Thank you!

StageNet: Stage-Aware Neural Networks for Health Risk Prediction

Junyi Gao, Cao Xiao, Yasha Wang, Wen Tang, Lucas M. Glass, Jimeng Sun

- Scan the QR code to try the visualization prototype system!

The visualization system is developed by *Zhongliang Yu*.