Addressing Asynchronicity in Clinical Multimodal Fusion via Individualized Chest X-ray Generation

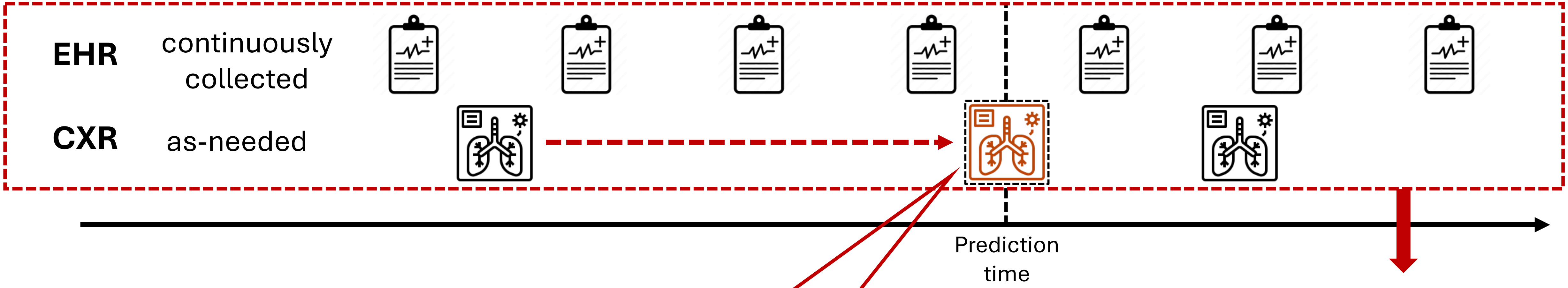
Diffusion-based **D**ynamic **L**atent **C**hest **X**-ray Image Generation **(DDL-CXR)**

Wenfang Yao*[1], Chen Liu *[1,3], Kejing Yin[2]✉, William K. Cheung[2], Jing Qin[1]

[1]School of Nursing, The Hong Kong Polytechnic University
[2] Department of Computer Science, Hong Kong Baptist University
[3] School of Software Engineering, South China University of Technology

https://github.com/Chenliu-svg/DDL-CXR

Challenge 1 - Clinical data are inherently highly asynchronous



ICU settings: rapid changes

"Carry-forward" : outdated CXR

Sub-optimal prediction performance

(a) Initial Chest X-ray (b) CXR taken after 34 hours (c) Generated by DDL-CXR

Challenge 2 - Patient-specific CXR generation

Text-to-audio / text-to-image generation

Explicit controllable attributes*:

VS

Individual clinical image generation

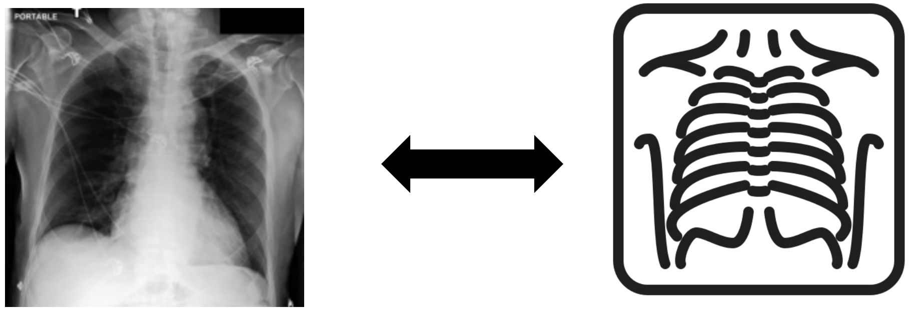
Explicit description of:

A cat in Monet style

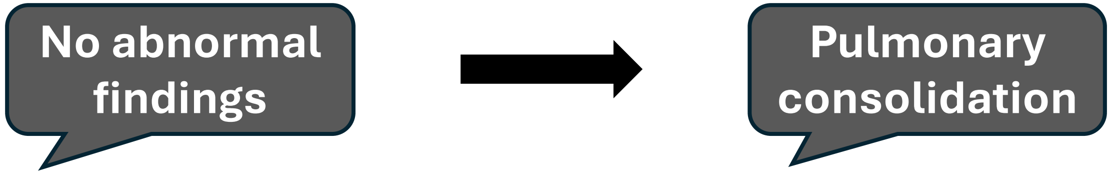
A cat in Van Gogh style

A happy blue cat

A sad orange cat



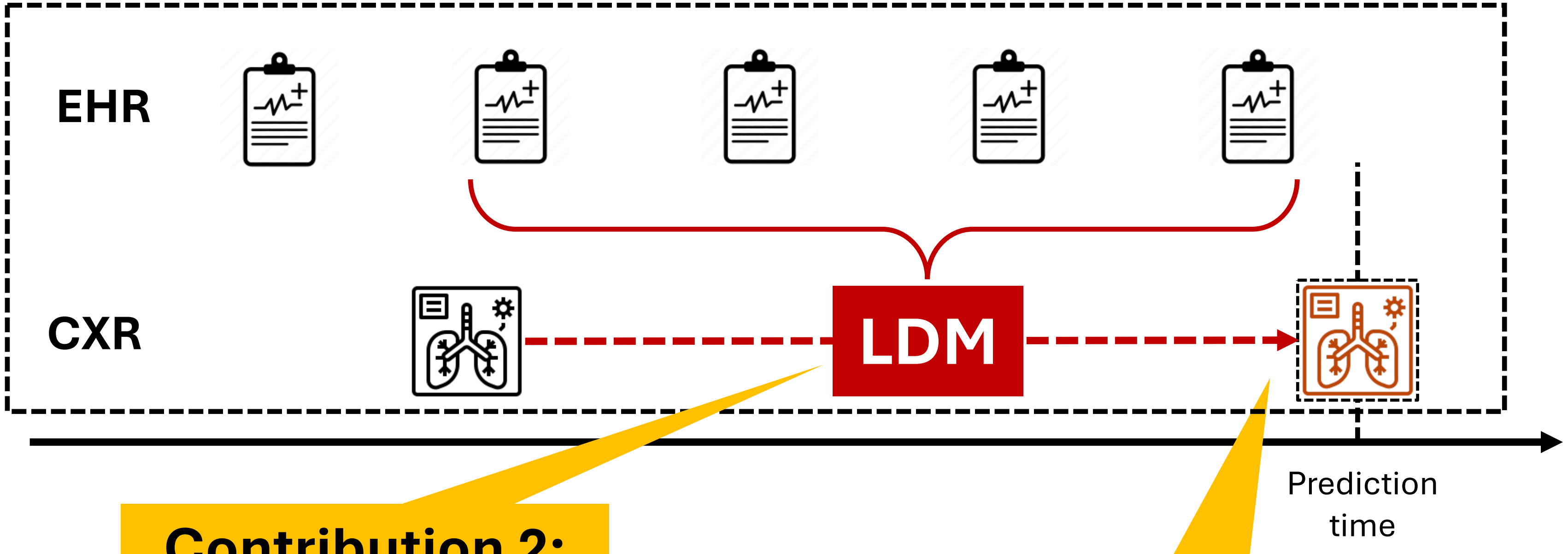
anatomical structures ✗



disease progression ✗

*Generated by Stable Diffusion.

Contributions



Contribution 3 Improved prediction performance

- Outperform SOTA on: mortality prediction, phenotype classification
- Excel in individual CXR generation

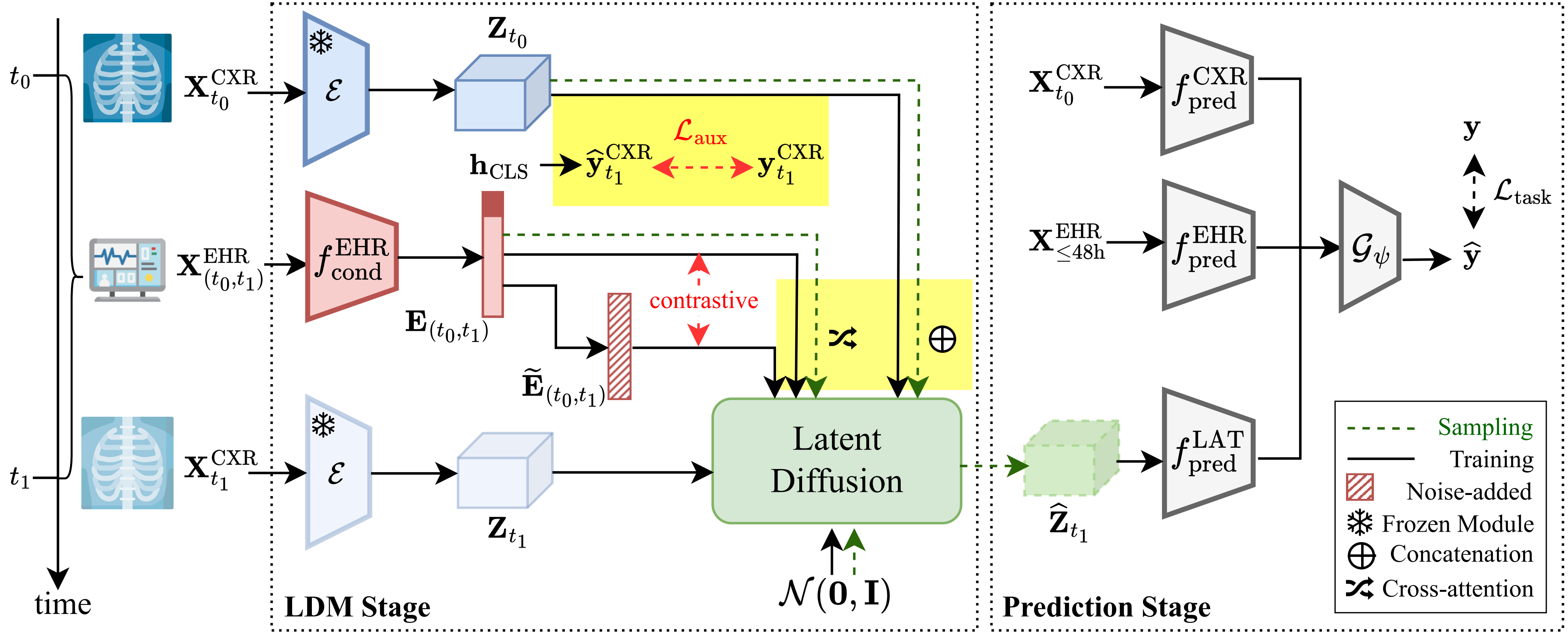
Contribution 2: Contrastive training of LDM

- Capture the disease course in EHR modality
- Enhance cross-modal interaction

Contribution 1: Individualized CXR generation

- Tackle the asynchronicity between EHR and CXR
- Capture interaction in a highly heterogeneous setting

The Proposed Method: DDL-CXR



LDM stage: dynamic latent CXR generation

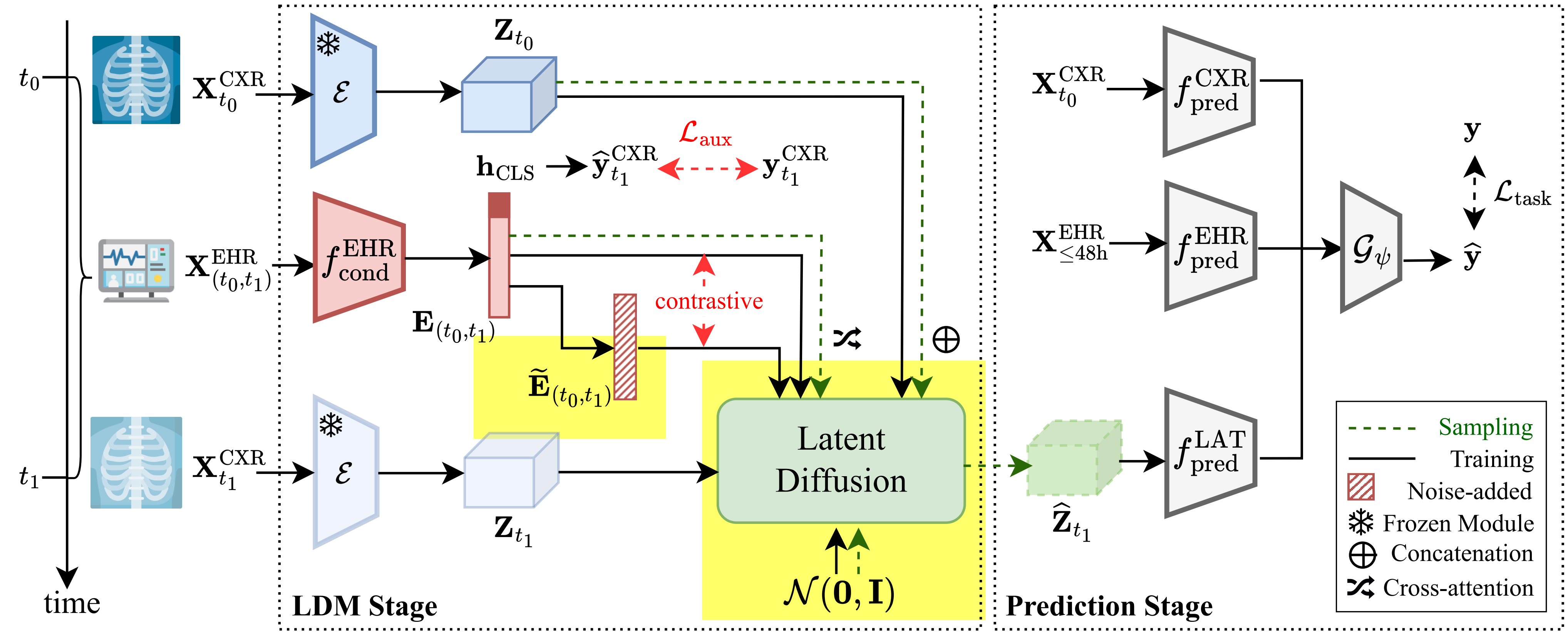
Conditioning mechanisms

$$\text{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\right)\cdot\mathbf{V},$$

$$\text{with } \mathbf{Q} = \mathbf{W}_Q \cdot \varphi\left(\mathbf{Z}_{t_1}^{(n)} || \mathbf{Z}_{t_0}\right), \mathbf{K} = \mathbf{W}_K \cdot f_{\text{cond}}^{\text{EHR}}(\mathbf{X}_{(t_0,t_1)}^{\text{EHR}}), \mathbf{V} = \mathbf{W}_V \cdot f_{\text{cond}}^{\text{EHR}}(\mathbf{X}_{(t_0,t_1)}^{\text{EHR}})$$

Capturing disease course via EHR time series: $\mathcal{L}_{\text{aux}} := \frac{1}{M}\frac{1}{L}\sum_{m=1}^{M}\sum_{l=1}^{L} y_{ml}^{\text{CXR}}\log(\widehat{y}_{ml}^{\text{CXR}}) + (1-y_{ml}^{\text{CXR}})\log(1-\widehat{y}_{ml}^{\text{CXR}})$

The Proposed Method: DDL-CXR



LDM stage: dynamic latent CXR generation

Enhancing semantic multimodal fusion via contrastive LDM learning: $\widetilde{\mathbf{E}}_{(t_0,t_1)} = (1-\beta)\mathbf{E}_{(t_0,t_1)} + \beta\boldsymbol{\delta}$, where $\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$

LDM training loss:

$$\mathcal{L}_{\mathrm{LDM}} := \mathbb{E}_{\mathbf{Z}_{t_1},\mathbf{Z}_{t_0},\mathbf{X}_{(t_0,t_1)}^{\mathrm{EHR}},\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I}),n}\Bigg[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\mathbf{Z}_{t_1}^{(n)}, \mathbf{Z}_{t_0}, f_{\mathrm{cond}}^{\mathrm{EHR}}(\mathbf{X}_{(t_0,t_1)}^{\mathrm{EHR}}), n\right)\right\|_2^2$$

$$+ \lambda_1 \max\left(\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\mathbf{Z}_{t_1}^{(n)}, \mathbf{Z}_{t_0}, \mathbf{E}_{(t_0,t_1)}, n\right)\right\|_2^2 - \left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta\left(\mathbf{Z}_{t_1}^{(n)}, \mathbf{Z}_{t_0}, \widetilde{\mathbf{E}}_{(t_0,t_1)}, n\right)\right\|_2^2 + \alpha, 0\right)\Bigg]$$

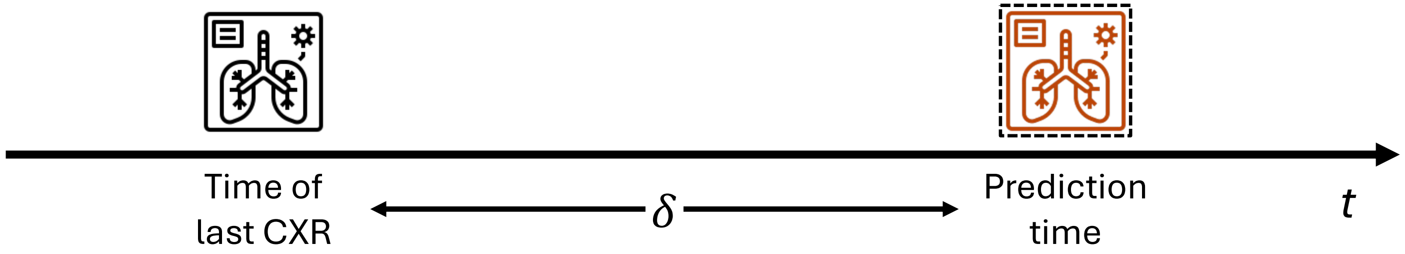
Results – Clinical Prediction (overall performance)

	Phenotyping		Mortality	
	AUPRC	AUROC	AUPRC	AUROC
Uni-EHR [23]	0.434 ±0.009	0.720 ±0.006	0.498 ±0.007	0.815 ±0.007
MMTM [52]	0.430 ±0.005	0.715 ±0.003	0.422 ±0.014	0.785 ±0.004
DAFT [9]	0.435 ±0.002	0.720 ±0.003	0.448 ±0.004	0.800 ±0.003
MedFuse [10]	0.437 ±0.001	0.718 ±0.002	0.443 ±0.009	0.793 ±0.003
DrFuse [13]	0.459 ±0.003	0.729 ±0.004	0.460 ±0.004	0.773 ±0.008
GAN-based [53]	0.453 ±0.010	0.728 ±0.008	0.505 ±0.018	0.816 ±0.010
DDL-CXR (ours)	**0.470** ±0.003	**0.740** ±0.002	**0.523** ±0.011	**0.822** ±0.009

DDL-CXR obtains the best overall performance

- Generating an updated CXR is beneficial for prediction.
- Performance gain in terms of AUPRC: identifying the positive class in imbalanced medical datasets.
- Relative improvements: 2.4% (phenotype classification); 3.56% (mortality prediction)

Results – Mortality prediction with varying time interval



δ increases
Last-CXR: more "outdated"

- **Dynamic generation - different ranges of δ: time interval (hour) between the prediction time and the time of last CXR.**

prevalence	Overall 14.7%	$\delta < 12$ 16.6%	$12 \leq \delta < 24$ 19%	$24 \leq \delta < 36$ 15.9%	$\delta \geq 36$ 9.26%
Uni-EHR [23]	0.815 ±0.007	0.854 ±0.010	0.799 ±0.013	0.756 ±0.019	0.796 ±0.008
MMTM [52]	0.785 ±0.004	0.798 ±0.008	0.763 ±0.004	0.760 ±0.012	0.772 ±0.014
DAFT [9]	0.800 ±0.003	0.803 ±0.010	0.782 ±0.009	**0.776** ±0.006	0.796 ±0.008
MedFuse [10]	0.793 ±0.003	0.812 ±0.004	0.762 ±0.007	0.760 ±0.009	0.800 ±0.010
DrFuse [13]	0.773 ±0.008	0.802 ±0.012	0.717 ±0.023	0.757 ±0.041	0.723 ±0.013
GAN-based [53]	0.816 ±0.010	0.846 ±0.010	**0.800** ±0.011	0.760 ±0.026	0.806 ±0.016
DDL-CXR (ours)	**0.822** ±0.009	**0.867** ±0.015	**0.800** ±0.008	0.753 ±0.015	**0.830** ±0.011

- **DDL-CXR receives a noticeable performance increase (in AUROC) when $\delta \geq 36$h.**

More details can be found at

Project Page

ArXiv

Poster session: Dec 12, 4:30pm – 7:30pm

Thank you!