Towards Boosting Performance of Healthcare Analytics: Resolving Challenges in Electronic Medical Records

Presenter: Kaiping Zheng

Oct 3rd, 2018

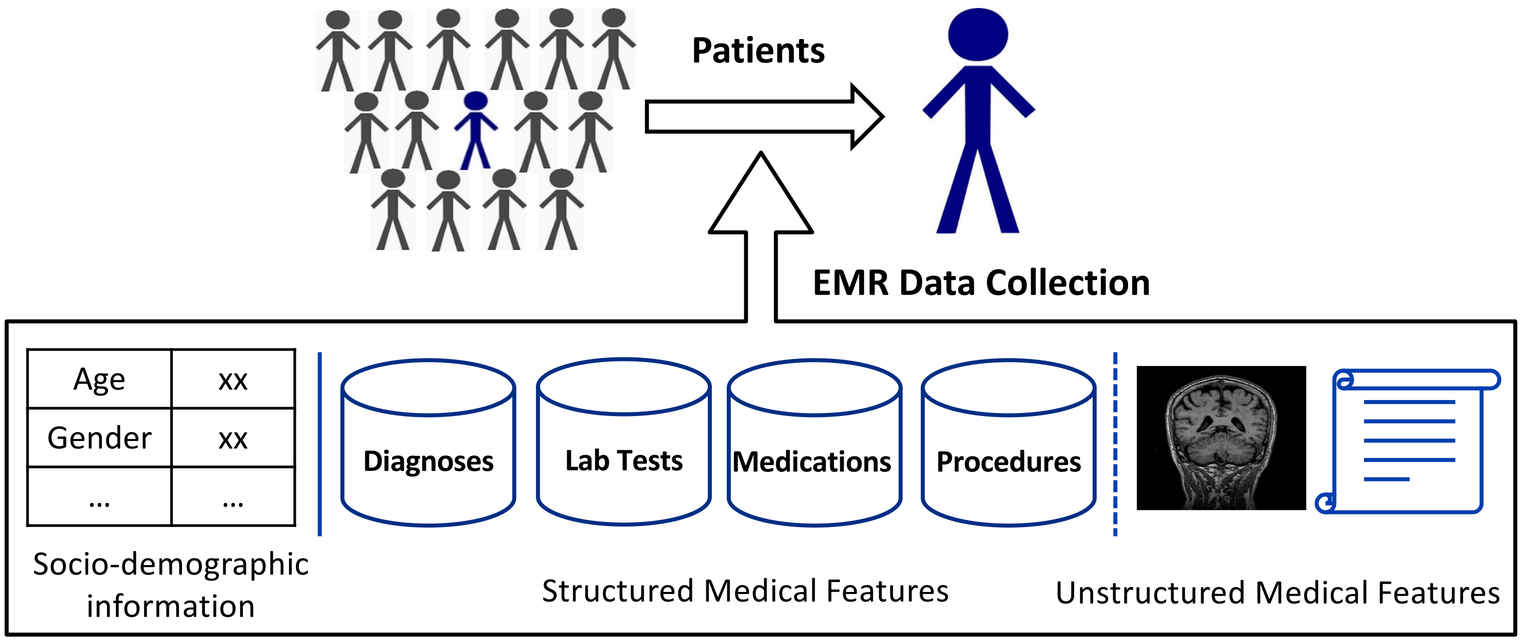
NUS National University of Singapore

Outline

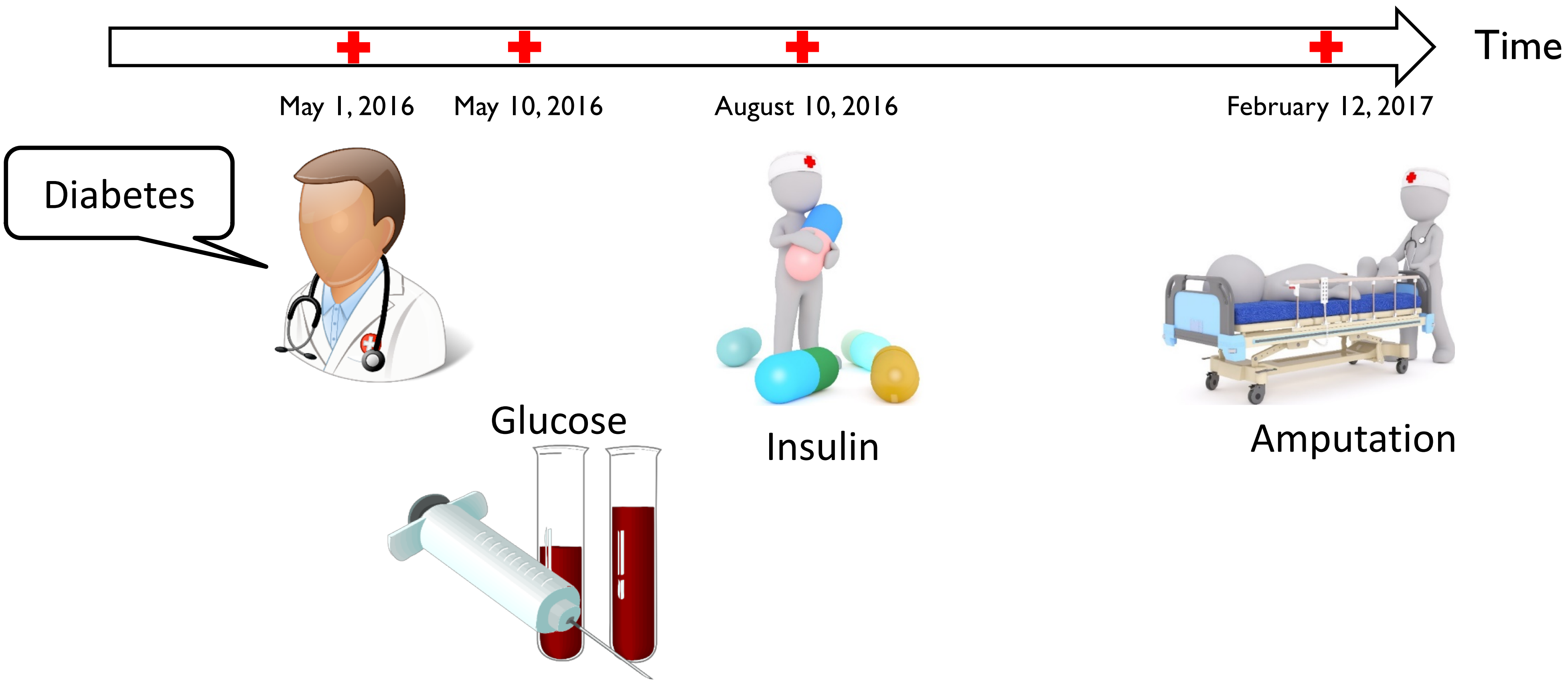
- Electronic Medical Records
- Disease Progression Modelling
- Resolving the Irregularity Challenge
- Resolving the Bias Challenge
- GEMINI Platform

Electronic Medical Records (EMR)

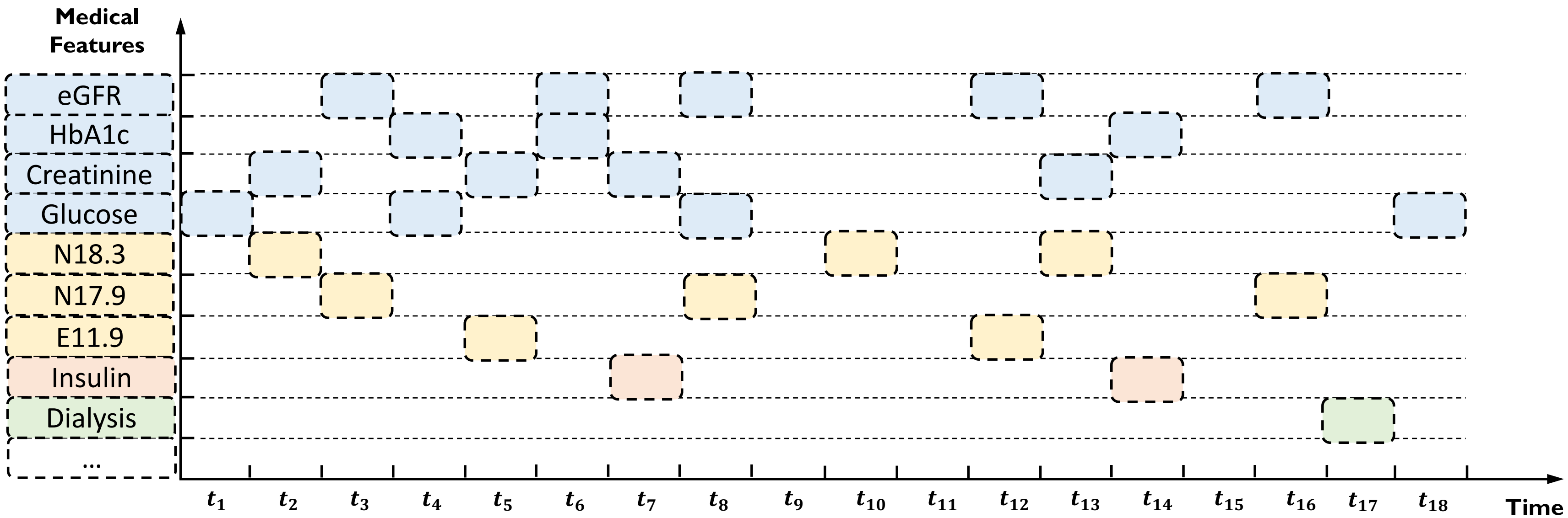
- Time series data that records patients' visits to hospitals
- Including a wide range of medical data



Electronic Medical Records (EMR)



Challenges in EMR



An example patient's time series EMR data with lab tests (eGFR, HbA1c, Creatinine, Glucose), diagnoses (N18.3, N17.9, E11.9), medications (Insulin) and procedures (Dialysis). This longitudinal patient matrix denotes different challenges in EMR data.

Irregularity | **Bias** | **High Dimensionality** | **Missing Data** | . . .

Outline

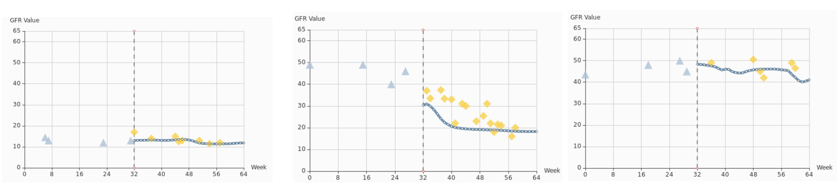
- Electronic Medical Records
- Disease Progression Modelling
- Resolving the Irregularity Challenge
- Resolving the Bias Challenge
- GEMINI Platform

EMR Data Analytics

NUS National University of Singapore

Predictive Application

Stable & Severe | Deteriorating | Stable & Mild

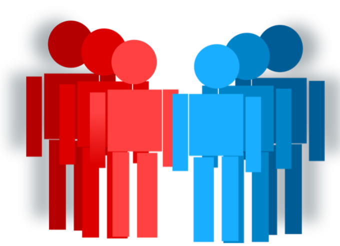


Disease Progression Modeling

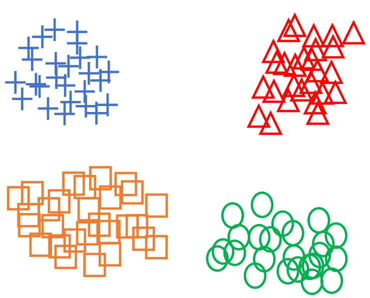
ICU In-hospital Mortality Prediction
ICU Diagnosis by Category Prediction
Readmission Prediction
…

Others

Basic Application



Cohort Analysis



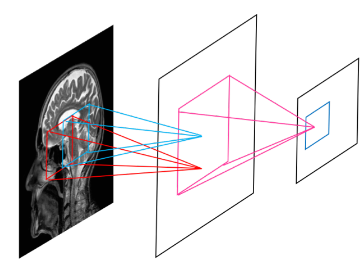
Medical Feature Embedding

Diabetes Phenotype

Diabetes with ketoacidosis
Diabetes with renal manifestations
…

Abnormal HbA1C
Abnormal Blood Pressure
Abnormal Cholesterol
…

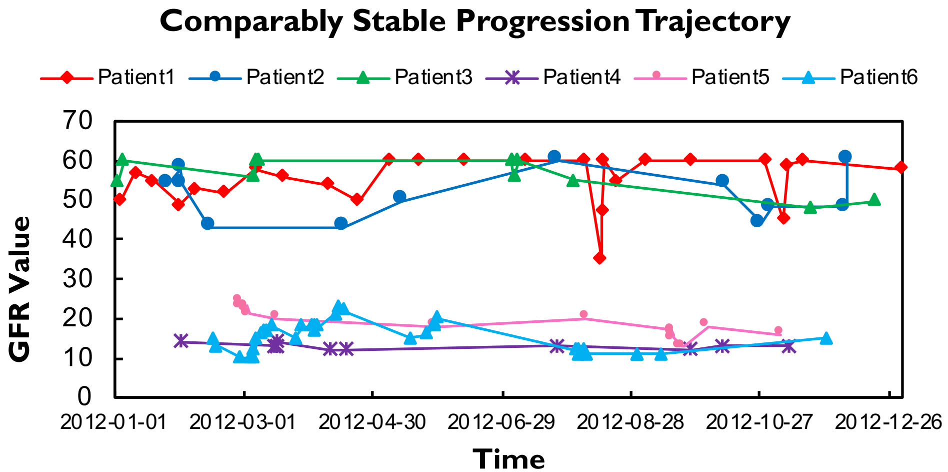
Phenotyping

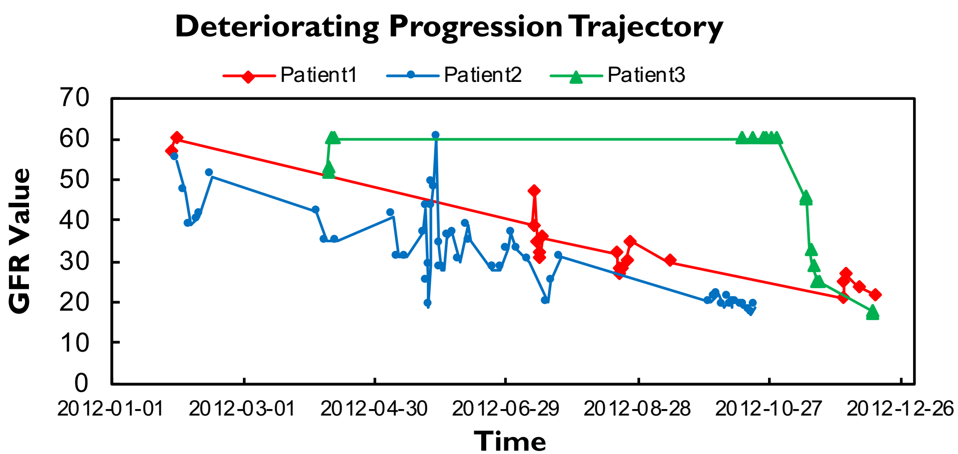


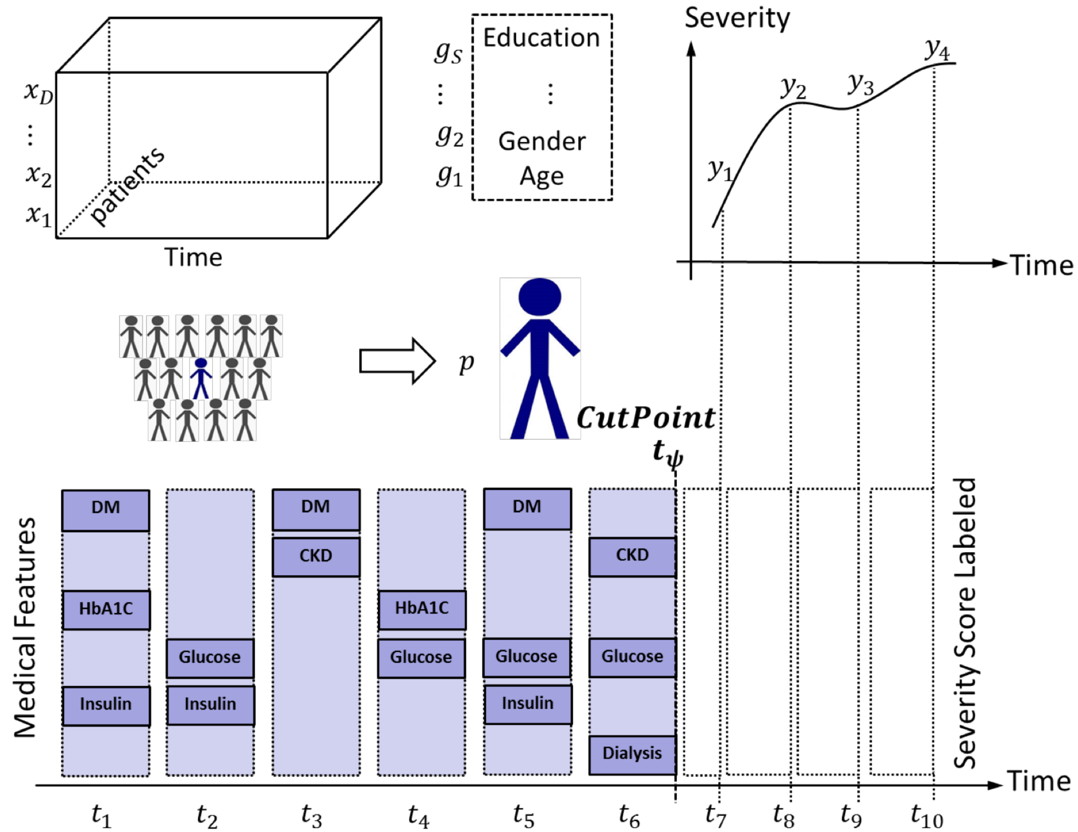
Input images **Deep learning model**

Image Analysis

Disease Progression Modelling







Outline

- Electronic Medical Records
- Disease Progression Modelling
- Resolving the Irregularity Challenge
- Resolving the Bias Challenge
- GEMINI Platform

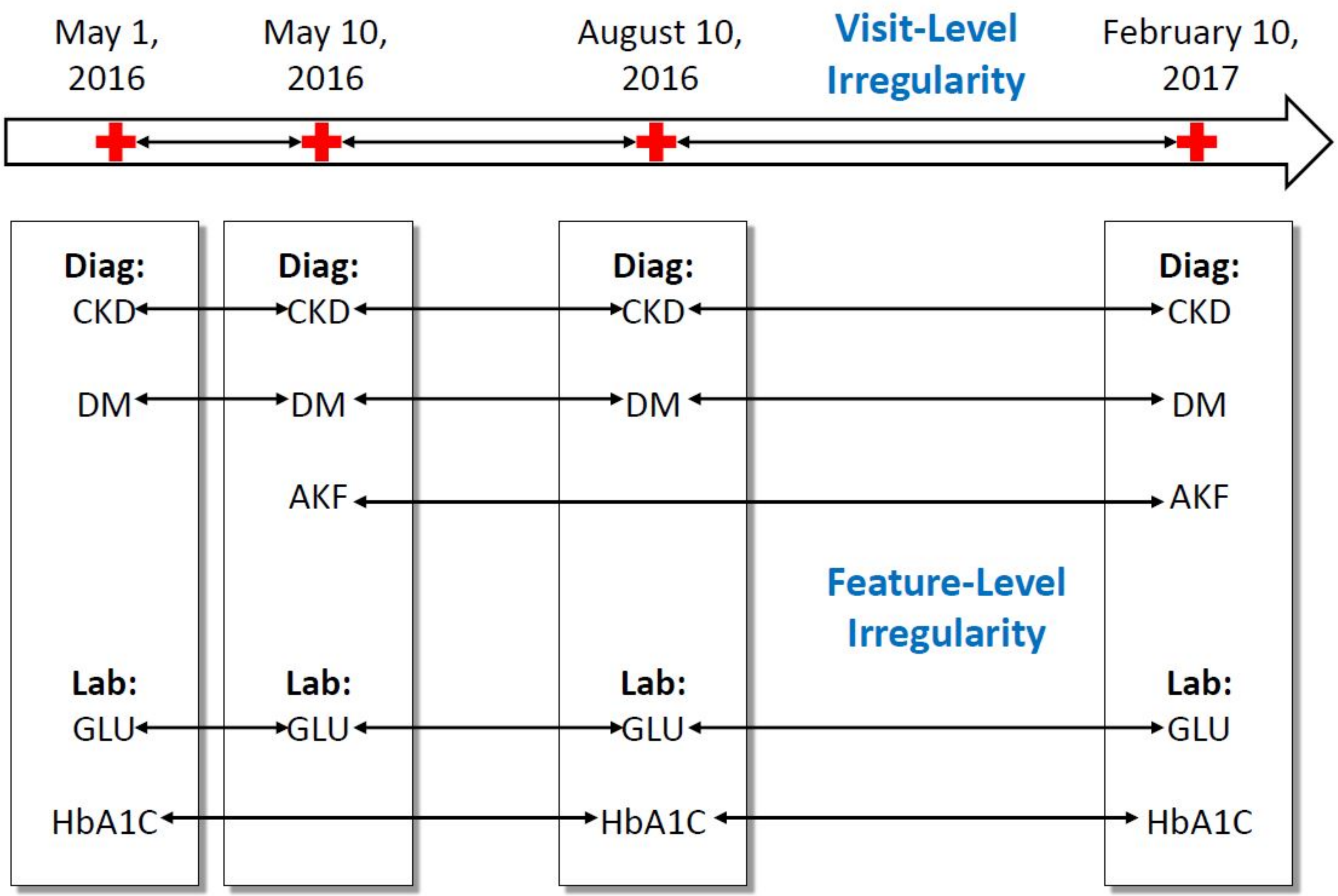
Irregularity Challenge

Two-Level Irregularity

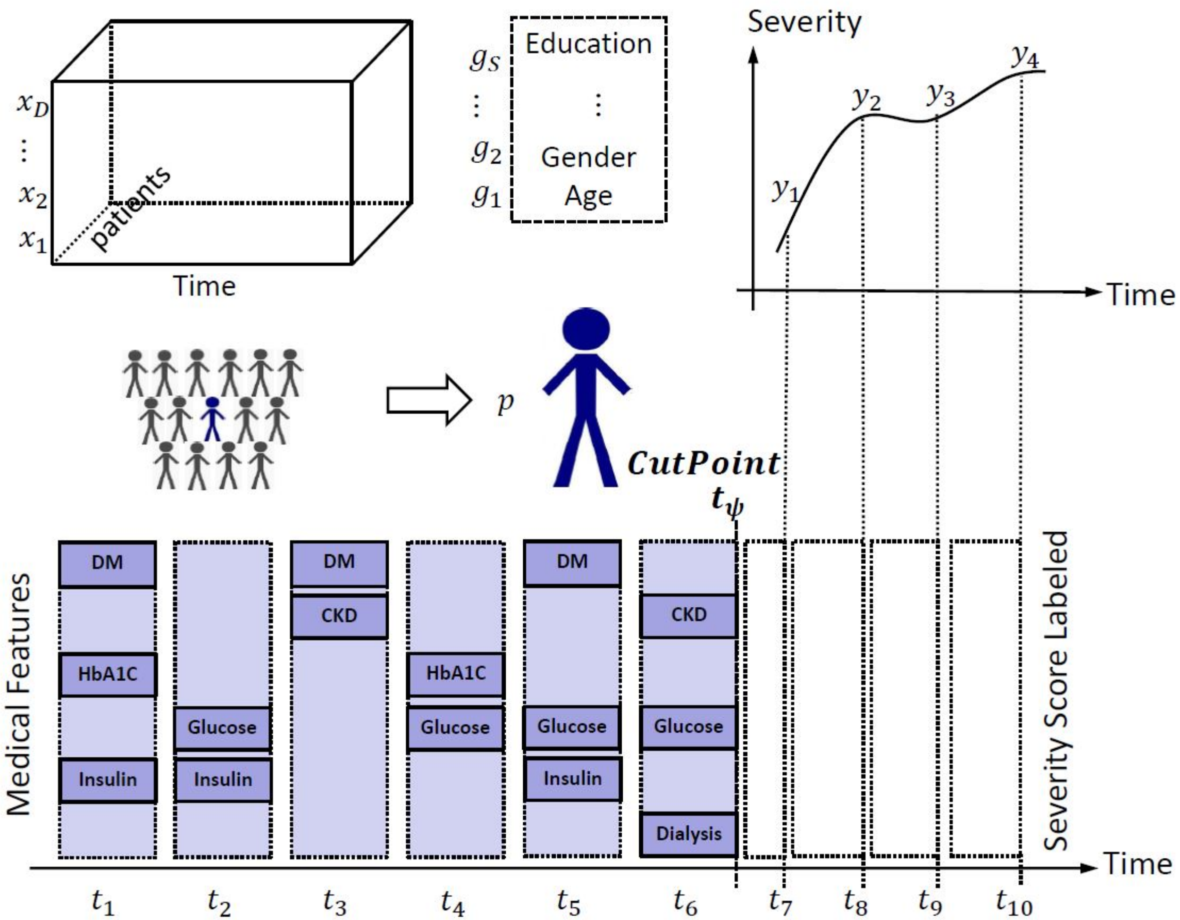
- Visit-level irregularity, Feature-level irregularity
  - Visit-Level Irregularity
    - EMR data appears irregularly with time
    - Time span between consecutive visits is irregular
  - Feature-Level Irregularity
    - Same feature appears irregularly in EMR data with time
    - Time span between a feature's consecutive occurrences is irregular

Irregularity Challenge

Electronic Medical Records (EMR)



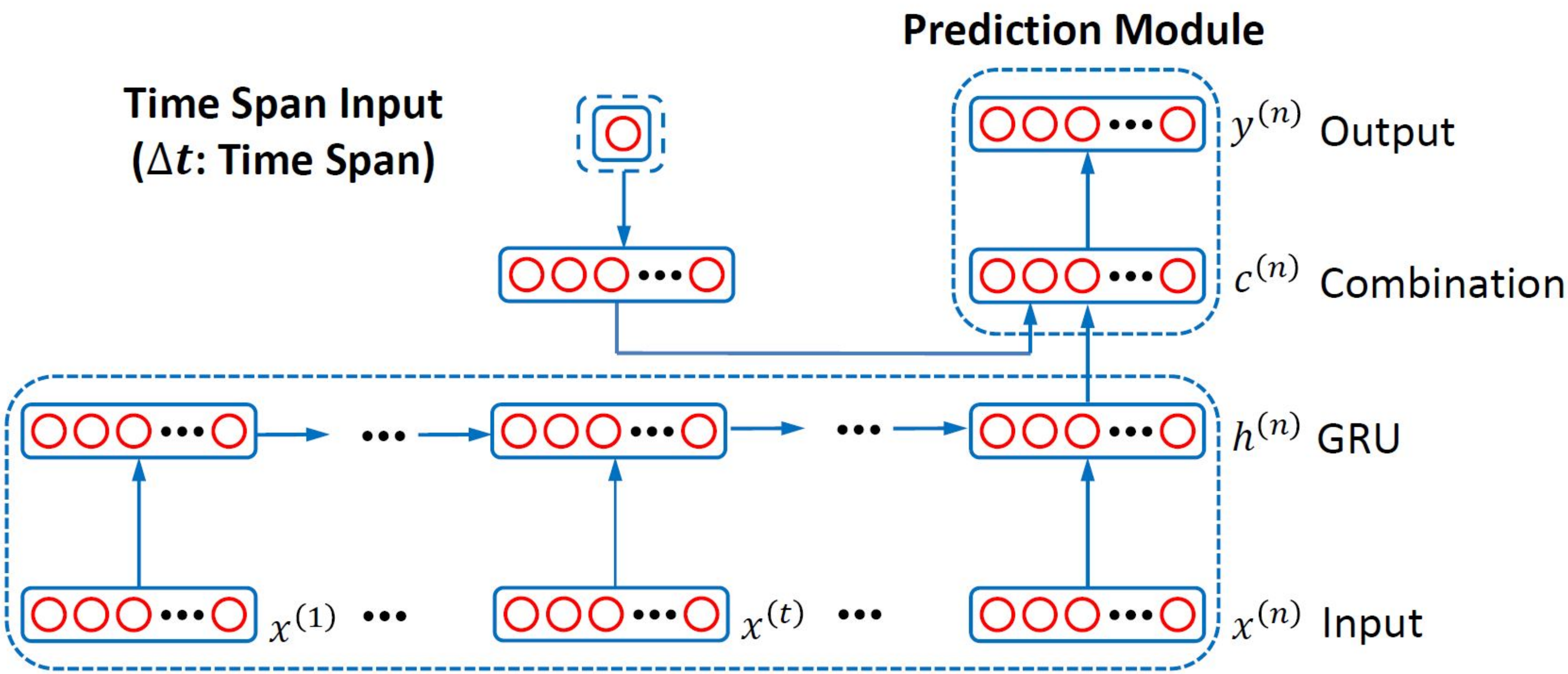
Methodology



Disease Progression Modeling (DPM)

Given a set of training samples $\{< x, y, \Delta t >\}$, the objective of DPM is to obtain a mapping function Φ that minimizes the following loss function over all samples: $L(\Phi(x, \Delta t), y)$

Methodology

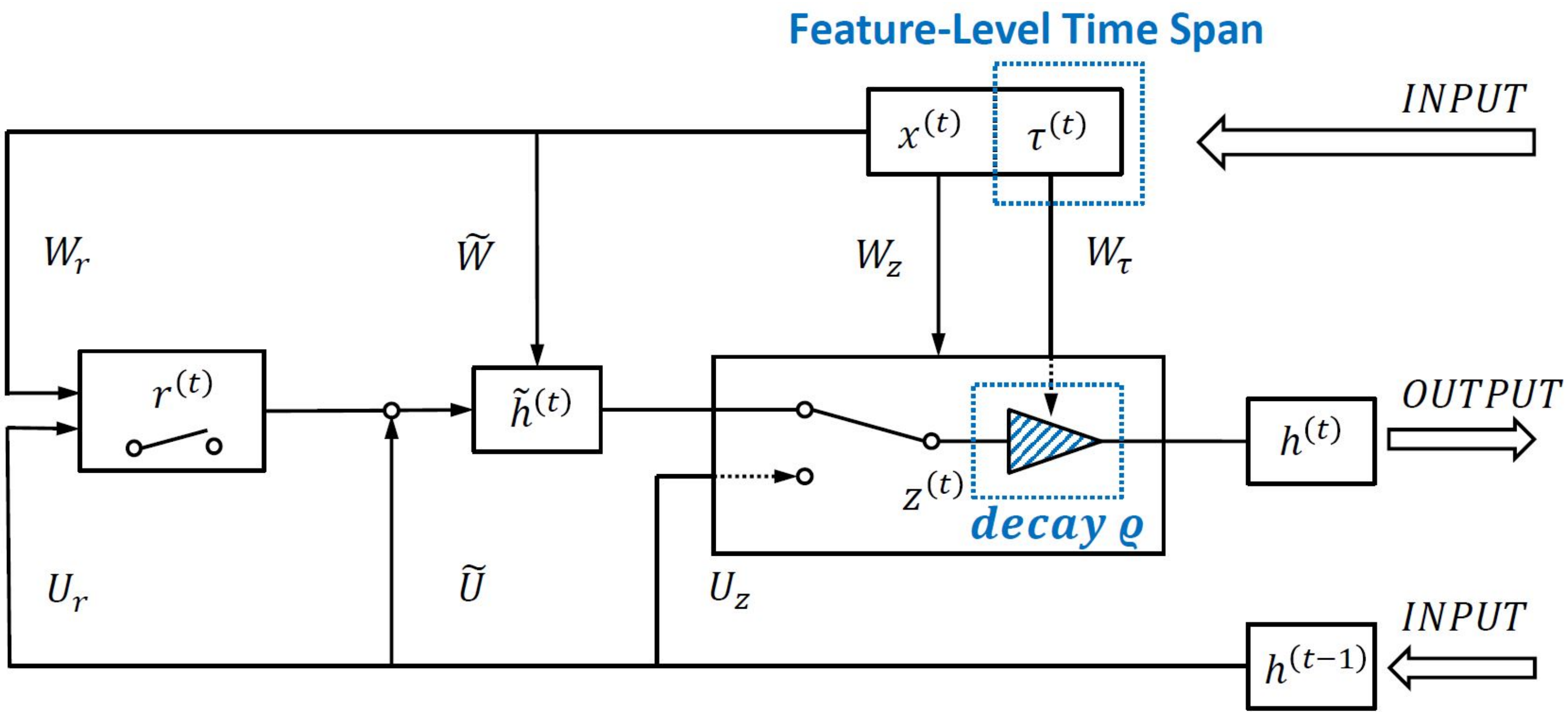


Medical Feature Input ($x^{(t)}$: EMR Data)

Loss function: $L = \frac{1}{|\{<x,y,\Delta t>\}|}\sum\left(y^{(n)} - y\right)^2$

Back-propagation algorithm for updating the model parameters

Methodology



Compute a decay term ϱ using $\tau(t)$ and multiply ϱ to $z(t)$

- $\varrho = 1 - tanh(W_\tau \tau^{(t)} + b_\tau)$
- $z(t) = sigmoid\left(\left(W_z x^{(t)} + U_z h^{(t-1)}\right) \odot \varrho\right)$

Evaluation

ADNI dataset

- Public Alzheimer's disease dataset from Alzheimer's Disease Neuroimaging Initiative
- Severity is measured by Mini-Mental State Examination (MMSE) test ($\in [0,30]$)

NUH-CKD dataset

- Extract from a chronic kidney disease (CKD) dataset from National University Hospital in Singapore
- Choose patients with Stage 3 CKD or higher as cohort, "NUH-CKD" dataset
- Severity is measured by Glomerular Filtration Rate (GFR) test ($\in [0,60]$)

Evaluation metrics

- Mean squared error (MSE)
- Pearson product-moment correlation coefficient (R) value

Evaluation

Dataset	ADNI1 Dataset	NUH-CKD Dataset
# of medical features	591	603
# of demo. features	3 — age, gender, education time	2 — age, gender
# of patients	819	2740
Time span	4 years, M00 to M48, ("M" — "month")	1 year, W00 to W52, ("W" — "week")
# of time steps	7 (aggregated by every 6 months)	52 (aggregated by every week)
CutPoint (t_ψ) setting	M12, M18, M24	W16, W24, W32
# of samples	t_ψ =M12: 1529 t_ψ =M18: 1200 t_ψ =M24: 558	t_ψ =W16: 3601 t_ψ =W24: 2793 t_ψ =W32: 1585

Evaluation

GRU-based baselines

- Window-Based Model
- Visit-Level Model
- Visit-Level Time Decay Model

Multi-task learning (MTL) methods (Zhou et al., 2012)

- Least Convex Fused Group Lasso (cFSGL)
- Least Non-Convex Fused Group Lasso (nFSGL), denote two formulations as nFSGL-1 and nFSGL-2 in experiments

Our proposed method

- Feature-Level Time Decay Model

Evaluation

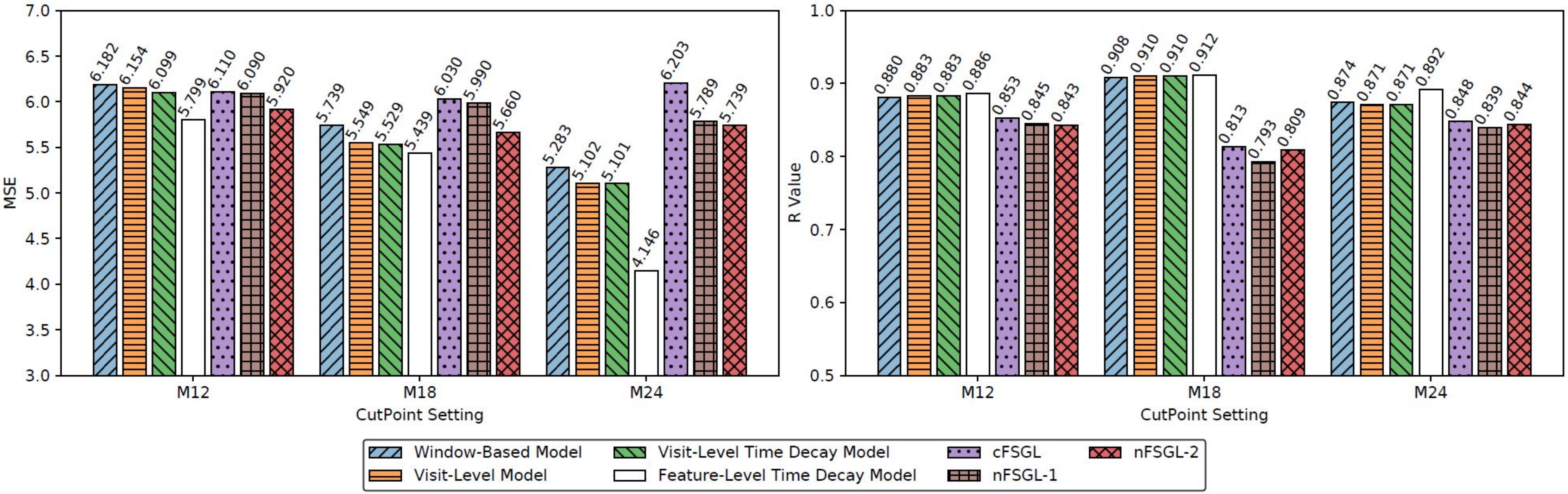


Figure: Experimental results in the ADNI dataset

- For the same CutPoint setting, from Window-Based Model to Feature-Level Time Decay Model, performance is mainly on the ascending trend; Feature-Level Time Decay Model more accurate than MTL-based methods;
- When CutPoint becomes larger, MSE values of GRU-based models decrease

K. Zheng, W. Wang, J. Gao, K.Y. Ngiam, B.C. Ooi and W.L.J. Yip: Capturing Feature-Level Irregularity in Disease Progression Modeling. CIKM 2017.

Evaluation

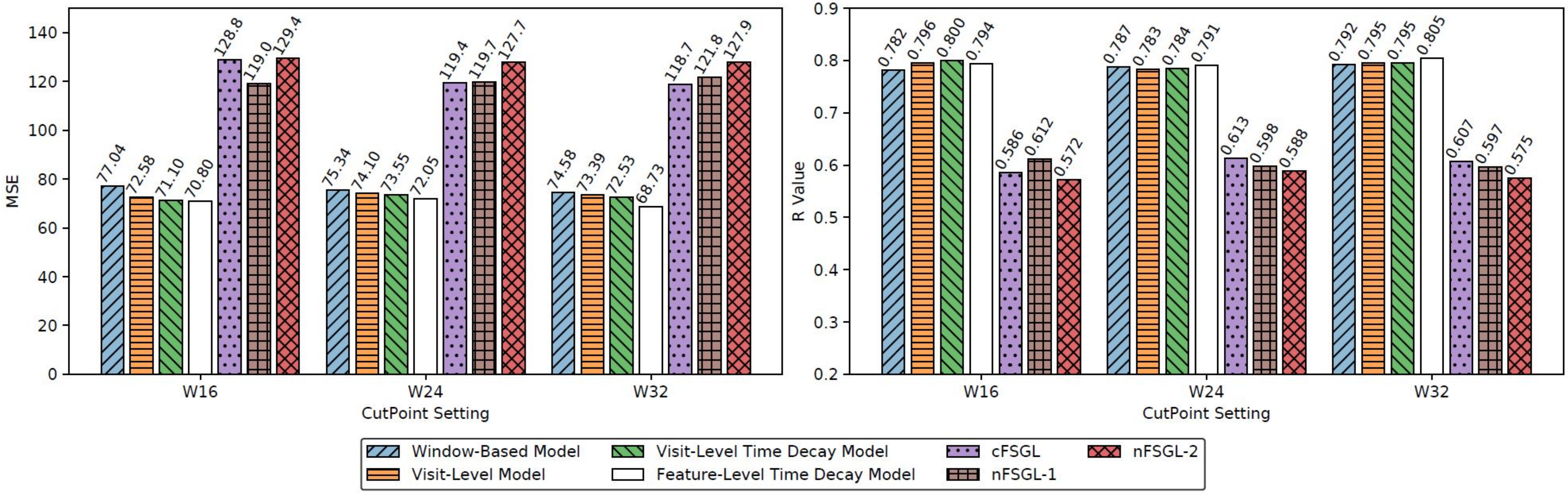


Figure: Experimental results in the NUH-CKD dataset

- From W16 to W24, GRU-based models achieve larger MSE values **- decreasing number of samples**
- From W24 to W32, GRU-based models achieve smaller MSE values **- more time series features**
- Both the sample length and sample number affect the model performance

K. Zheng, W. Wang, J. Gao, K.Y. Ngiam, B.C. Ooi and W.L.J. Yip: Capturing Feature-Level Irregularity in Disease Progression Modeling. CIKM 2017.

Summary

I. Identify the irregularity characteristic residing in EMR data both at the visit level and at the feature level

II. Capturing feature-level irregularity can benefit EMR data analytics through Feature-Level Time Decay Model

- Handle feature-level irregularity
- Decay the influence of previous information on patients' current state
- Learn decaying parameters for different features

III. Evaluate proposed Feature-Level Time Decay Model in both a public ADNI dataset and a private NUH-CKD dataset for two chronic disease cohorts

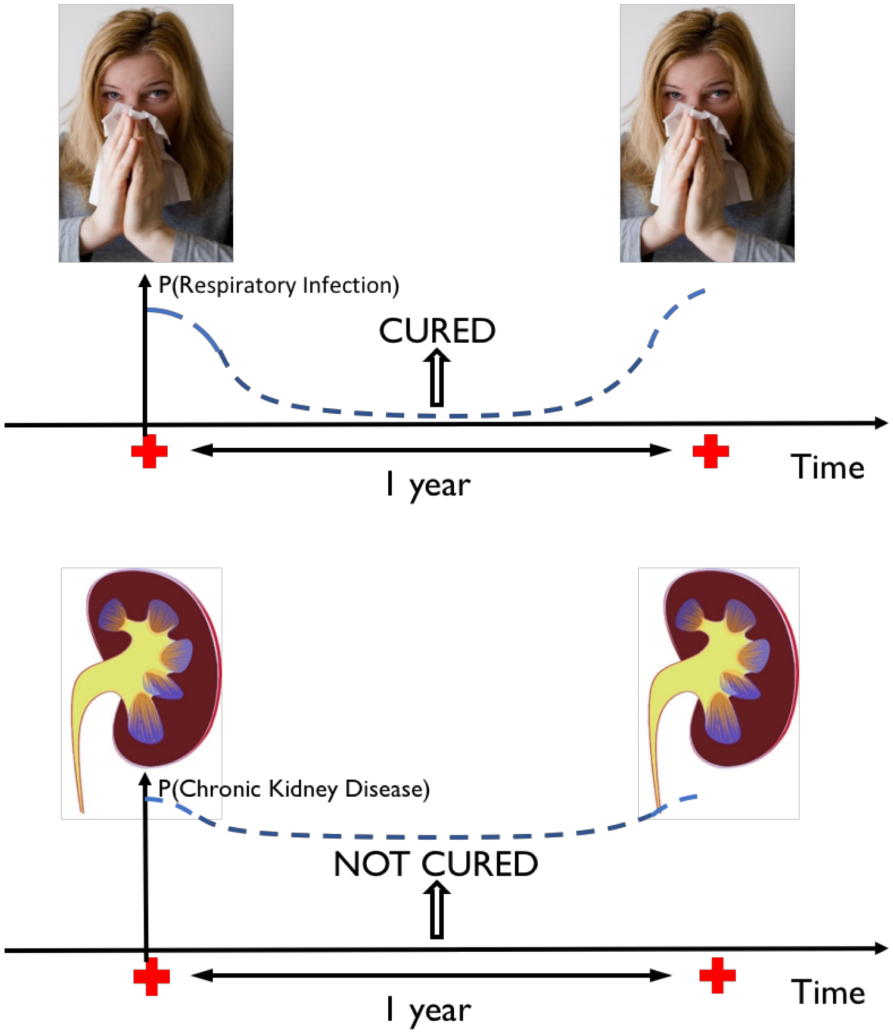
Outline

- Electronic Medical Records
- Disease Progression Modelling
- Resolving the Irregularity Challenge
- Resolving the Bias Challenge
- GEMINI Platform

How Is EMR Data Generated?

- Say Patient1 visits hospital 12 times per year
- Regularly sampled?
  - Patient1 visits hospital on the first day of every month?
- Randomly sampled?
  - Everyday, Patient1 tosses 5 coins, if all heads (1/32 probability), visits hospital?
- No, Patient1 visits hospital only when Patient1 feels sick
- EMR data is not regularly or randomly sampled

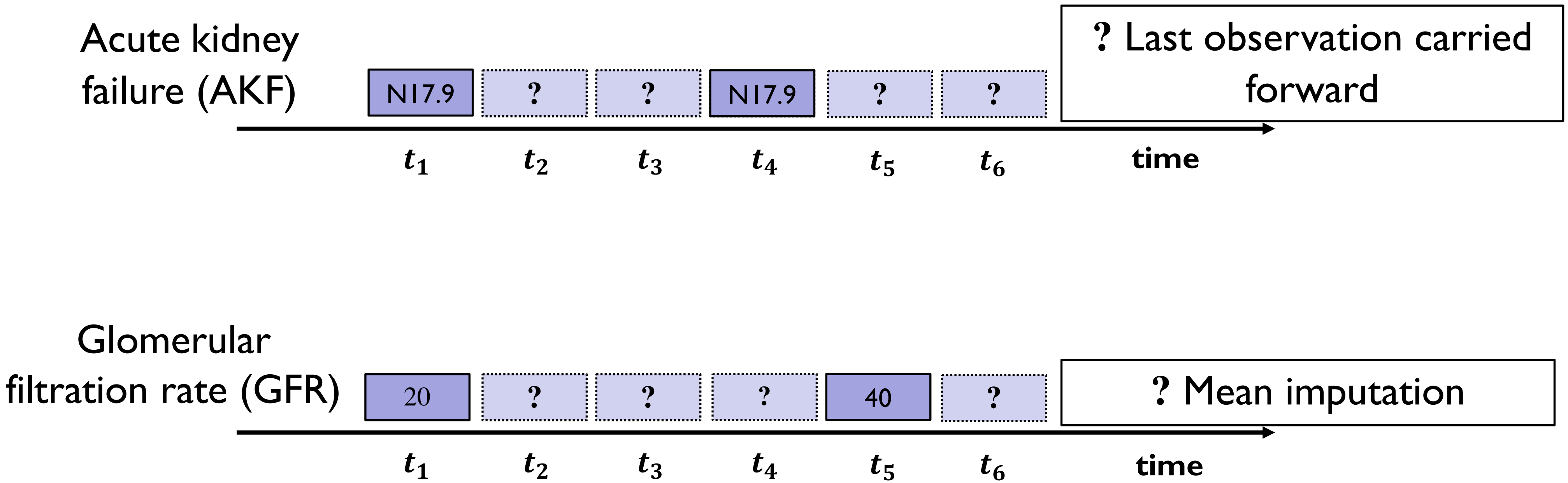
How Is EMR Data Generated?



- Patient1 always visits hospital due to respiratory infection
  - Can we conclude that Patient1 has respiratory infection every day?
- Patient2 always visits hospital due to chronic kidney disease
  - Can we conclude that Patient2 has chronic kidney disease every day?
- What is the difference?

Bias in EMR Data

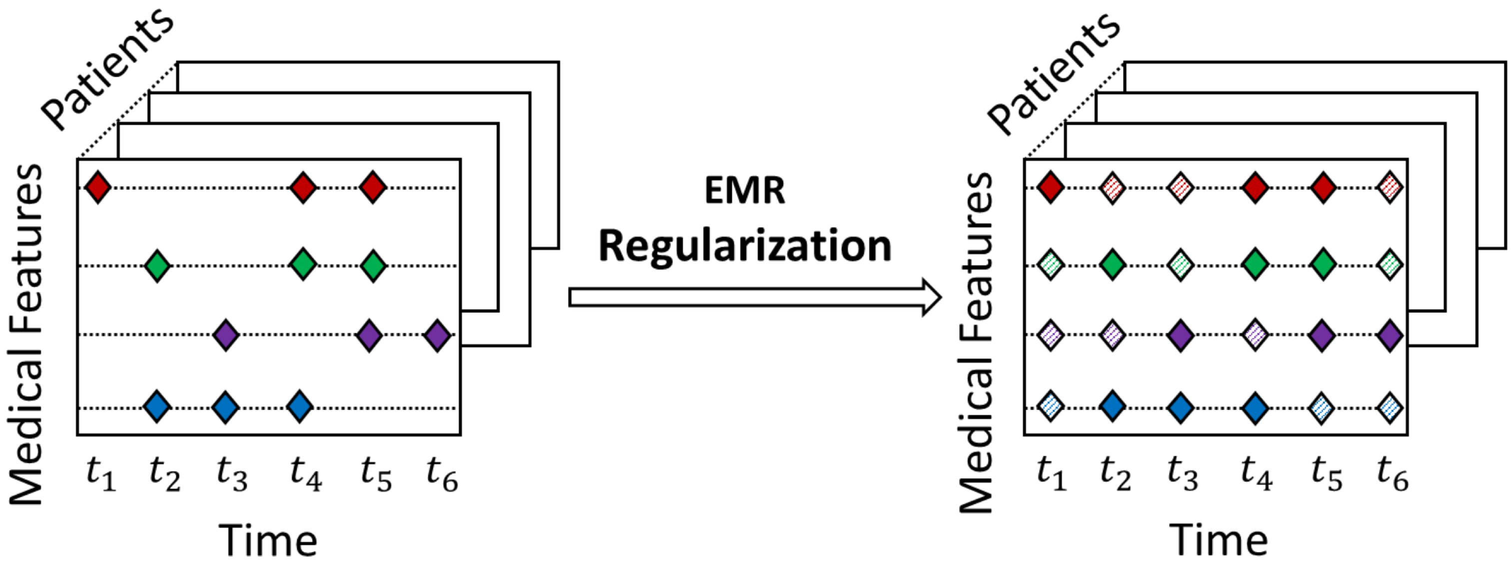
- If a doctor or analyst want to analyze the EMR data with missing values, they may employ traditional imputation methods directly
- → Misinterpretation



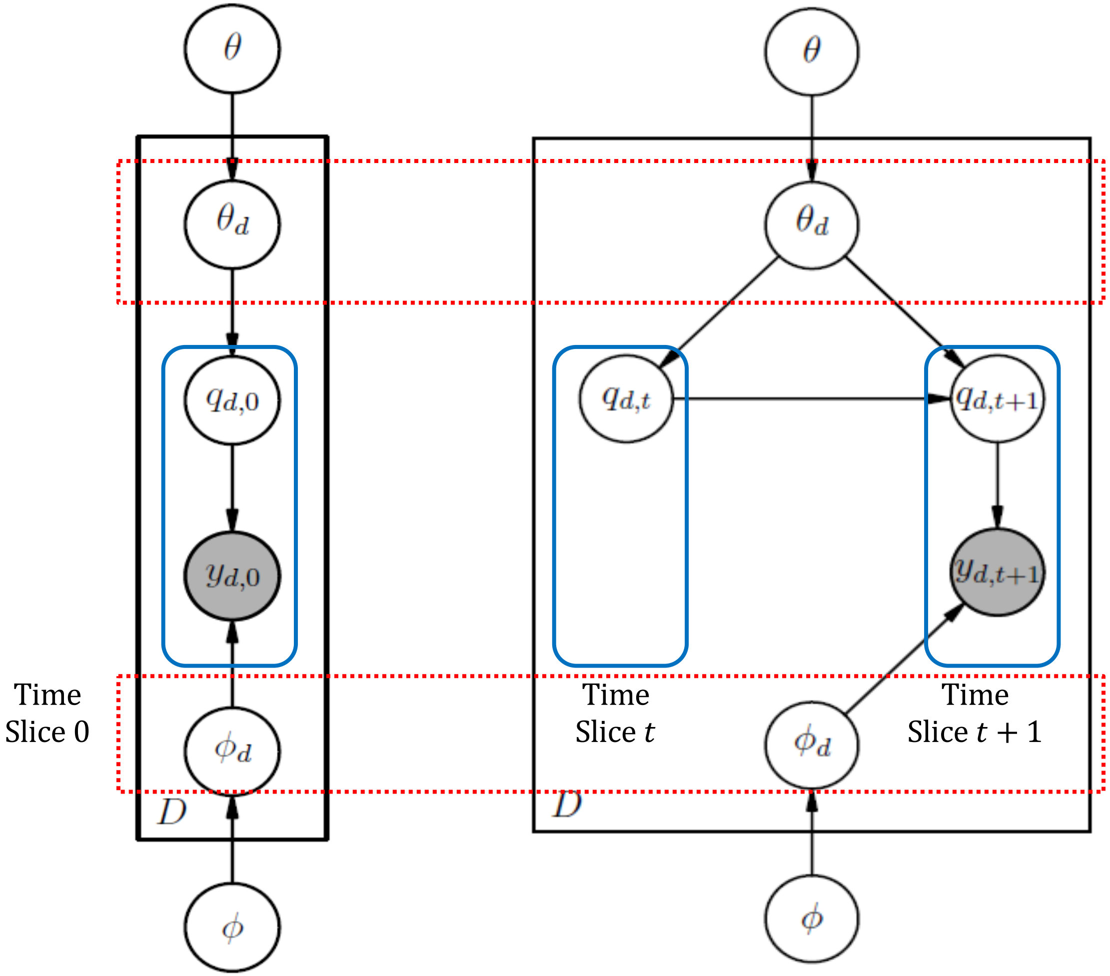
Bias in EMR Data

- Bias – recorded EMR series is different from patients' actual hidden conditions
  - Patients tend to visit hospital more often when they feel sick
  - Doctors tend to prescribe the lab examinations that show abnormality

- ***To Solve Bias Challenge – EMR Regularization***
  - Transform the biased EMR series into unbiased EMR series



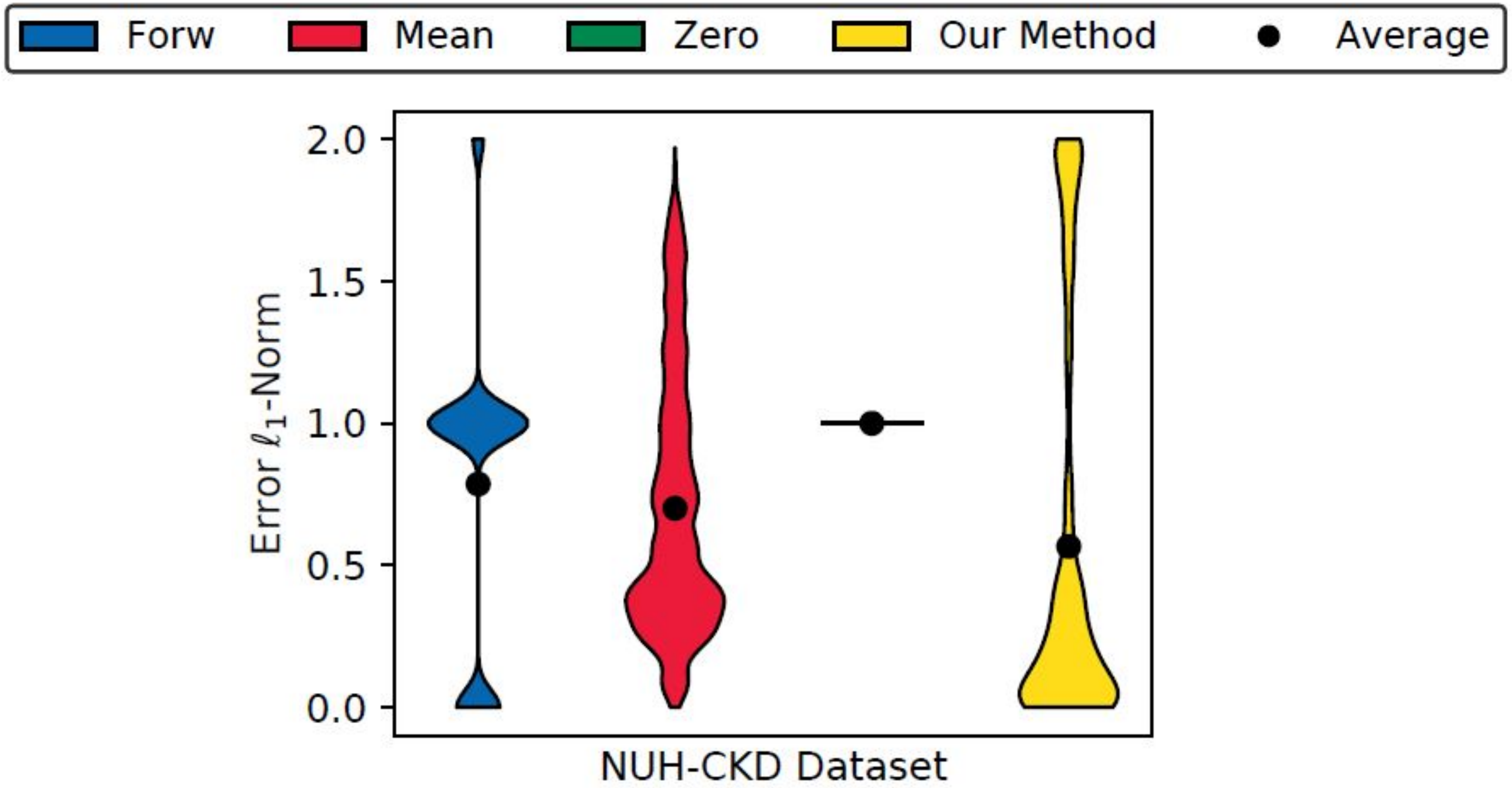
Resolving Bias in EMR Data



- Condition Change Rate (CCR)
  - Measure how a medical feature is likely to change from its condition in the previous observation
- Observation Rate (OR)
  - Measure the probability that a medical feature is exposed at a time point based on its actual condition at that time point

Resolving Bias in EMR Data

- Imputation accuracy evaluation



K. Zheng, J. Gao, K.Y. Ngiam, B. C. Ooi and W.L.J. Yip: Resolving the Bias in Electronic Medical Records. ACM KDD, 2017

Resolving Bias in EMR Data

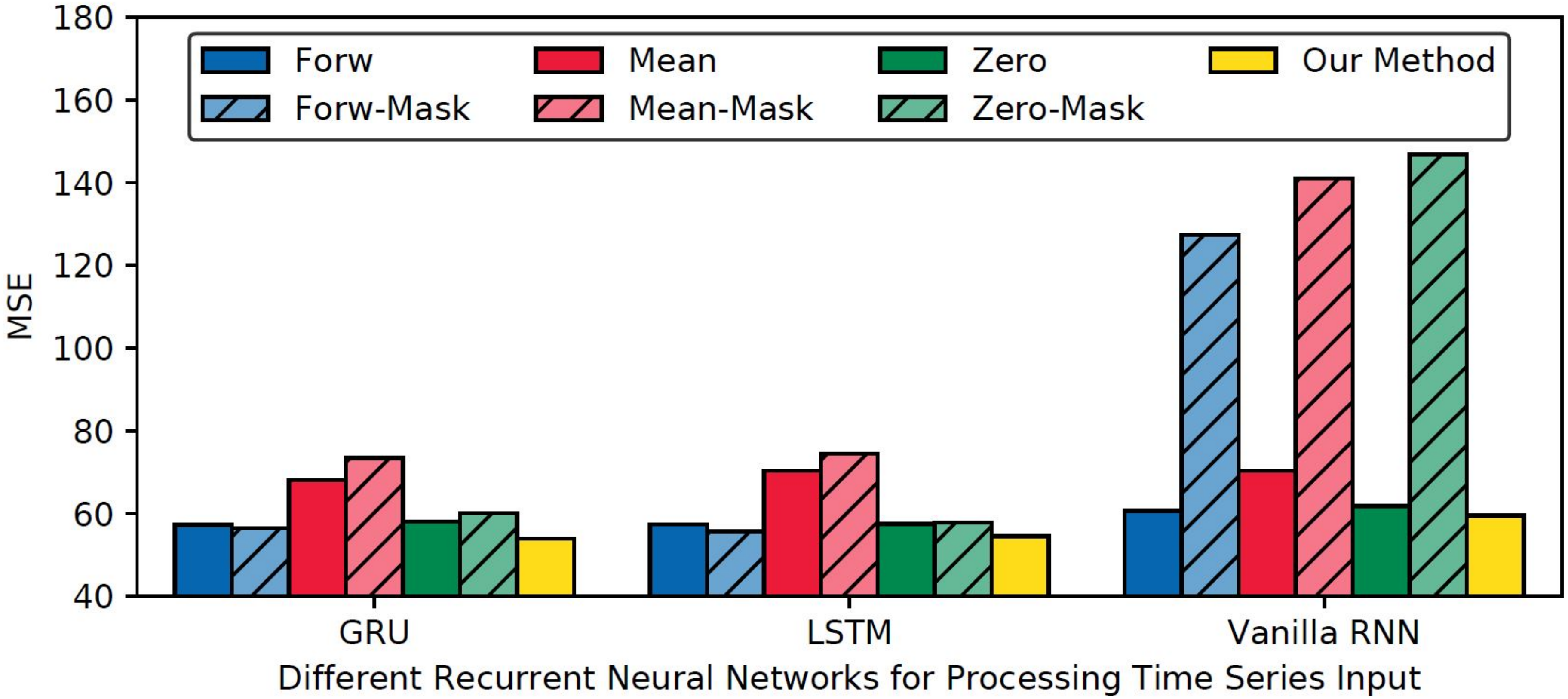


Figure: MSE for NUH-CKD disease progression modelling

K. Zheng, J. Gao, K.Y. Ngiam, B. C. Ooi and W.L.J. Yip: Resolving the Bias in Electronic Medical Records. ACM KDD, 2017

Resolving Bias in EMR Data

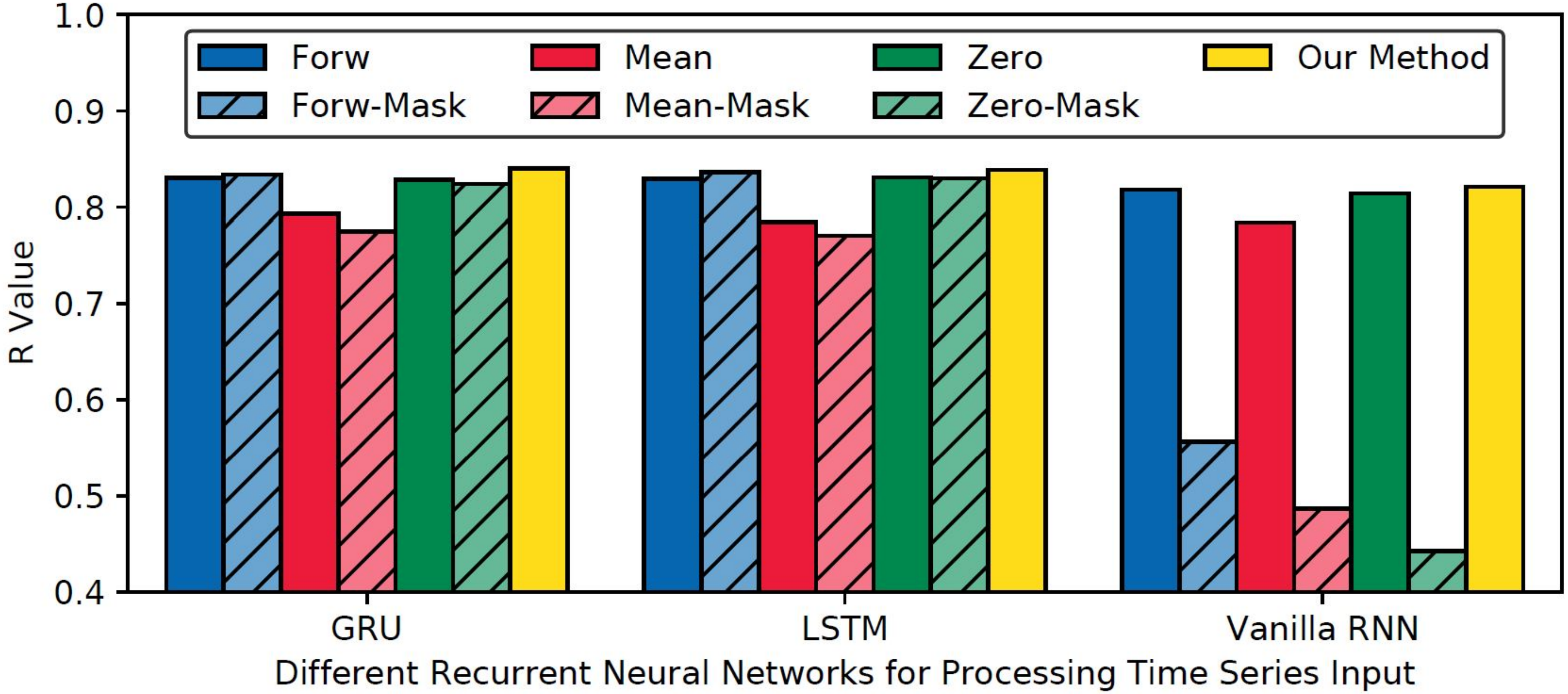


Figure: R value for NUH-CKD disease progression modelling

K. Zheng, J. Gao, K.Y. Ngiam, B. C. Ooi and W.L.J. Yip: Resolving the Bias in Electronic Medical Records. ACM KDD, 2017

Summary

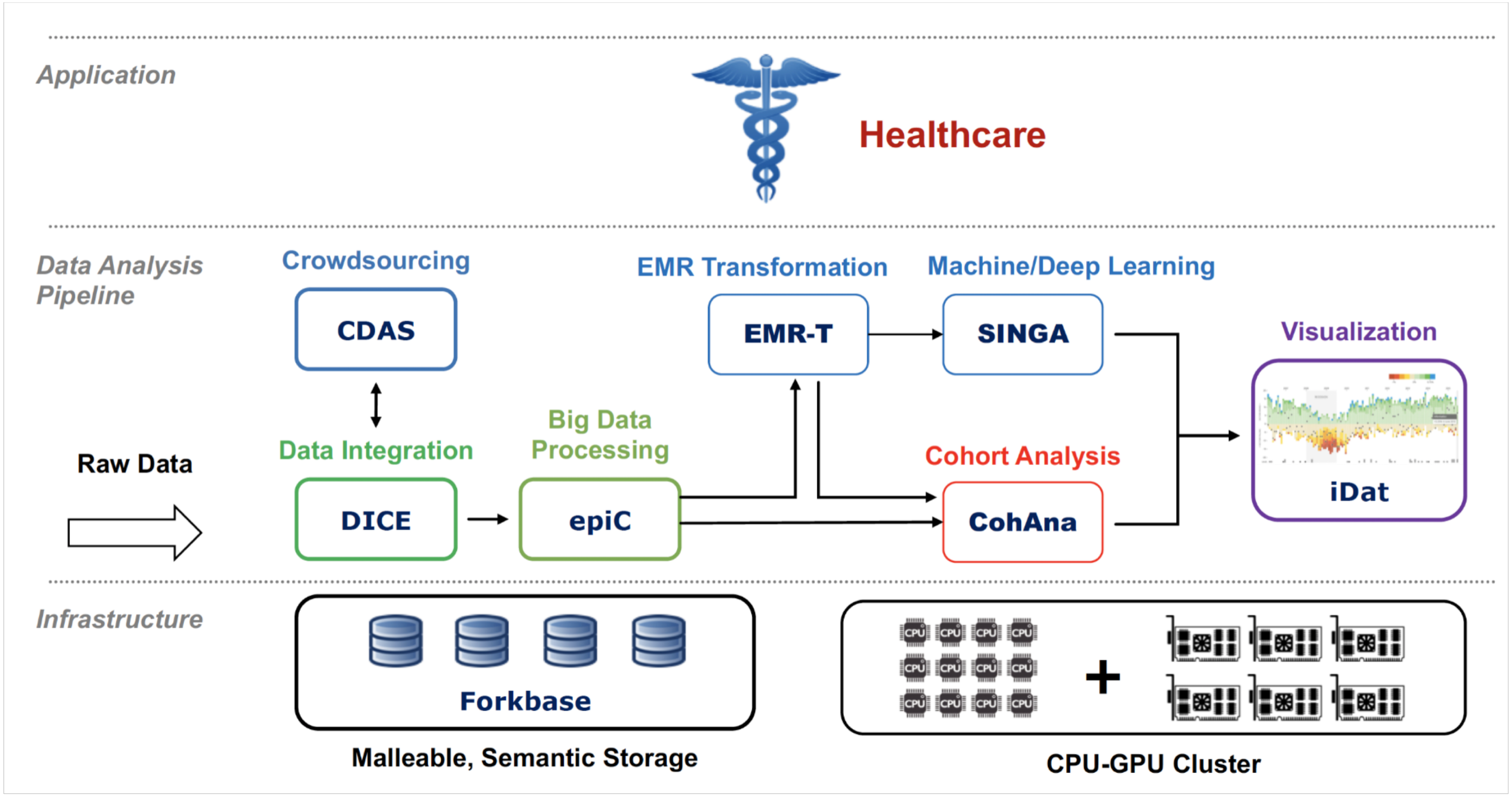
- EMR Regularization to Resolve Bias
  - Consider CCR and OR as characteristics of medical features
  - Employ an HMM variant for learning and inference
  - Impute missing values in EMR data more accurately
  - Improve the analytic performance after resolving the bias

- Possible Extensions:
  - Model different diseases jointly in the probabilistic graphical model for capturing the relationships in between
  - Model the patient personalization as different patients might behave differently in terms of CCR and OR

Outline

- Electronic Medical Records
- Disease Progression Modelling
- Resolving the Irregularity Challenge
- Resolving the Bias Challenge
- **GEMINI Platform**

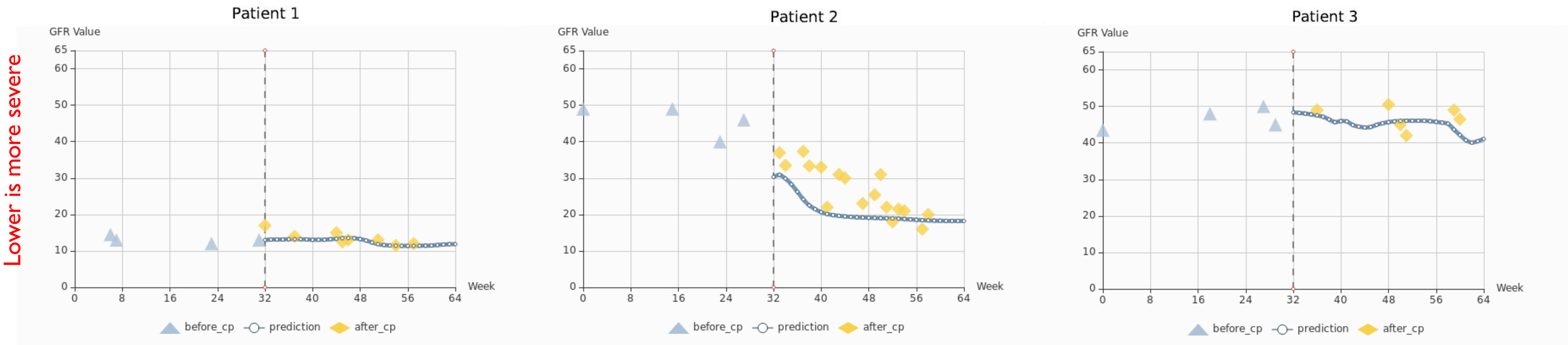
GEMINI Platform



Overview of GEMINI

https://www.comp.nus.edu.sg/~dbsystem/gemini/

Advice to Doctors on Intervention



Powered by GEMINI

- Suggest to guarantee the monitoring for Patient 1 → may need dialysis or kidney transplant
- Suggest healthcare workers to provide more aggressive interventions to Patient 2 in advance
- Suggest to guarantee the monitoring for Patient 3

Thank you!

NUS National University of Singapore