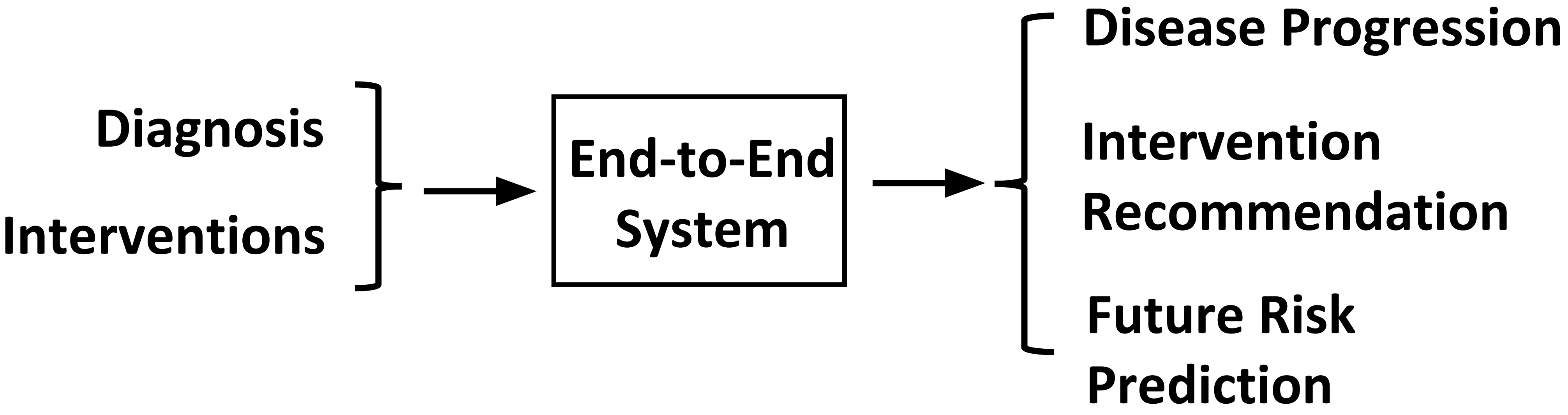
DeepCare: A Deep Dynamic Memory Model for Predictive Medicine

Trang Pham, Truyen Tran, Dinh Phung and Svetha Venkatesh

April 12, 2017

Introduction

- “What is happening?”
- “What happens next?"



Introduction

- This system needs to address four challenges:
  - Long-term dependencies in healthcare

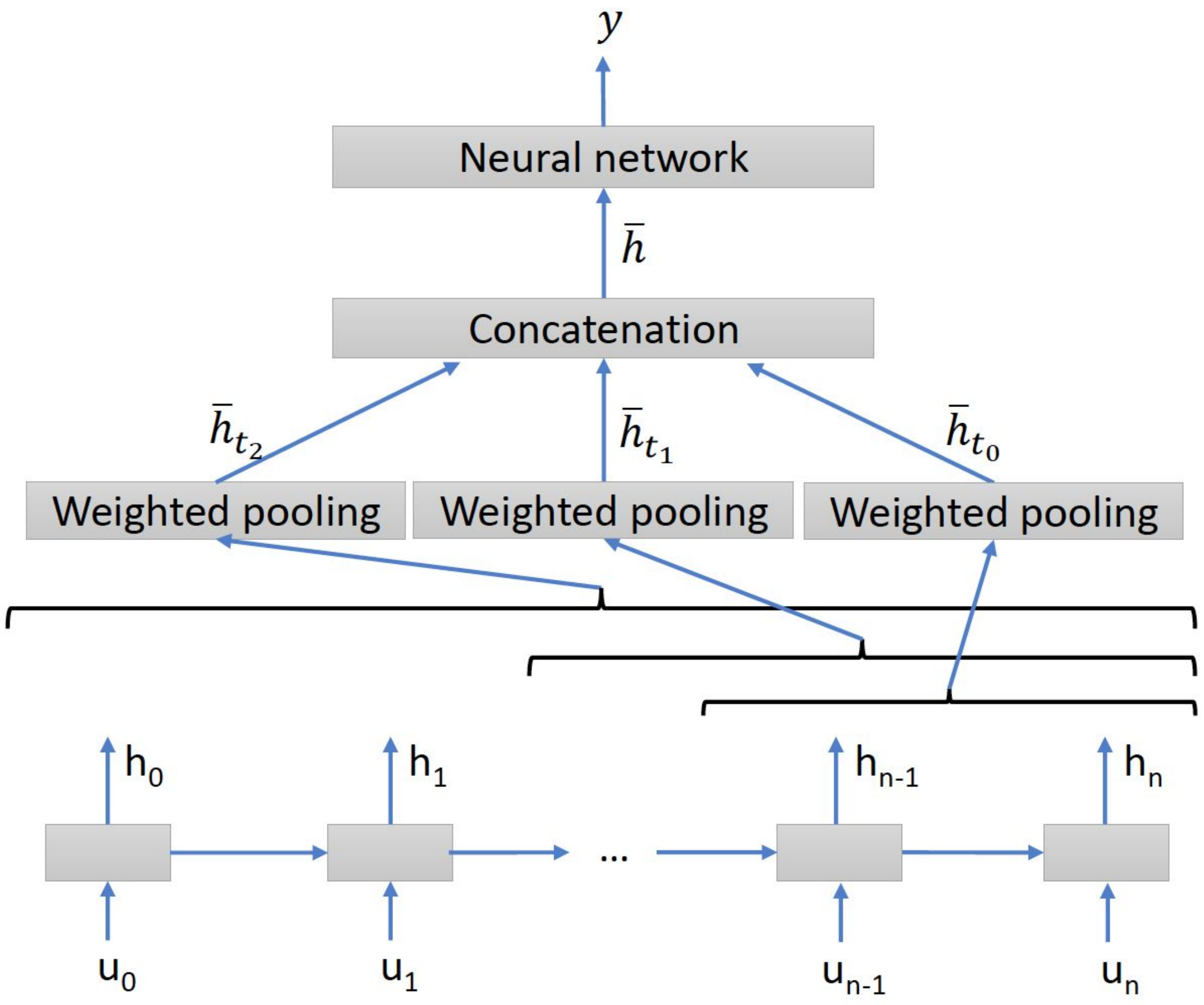
    LSTM
  - Representation of admission

    Vector Embedding
  - Episodic recording and irregular timing

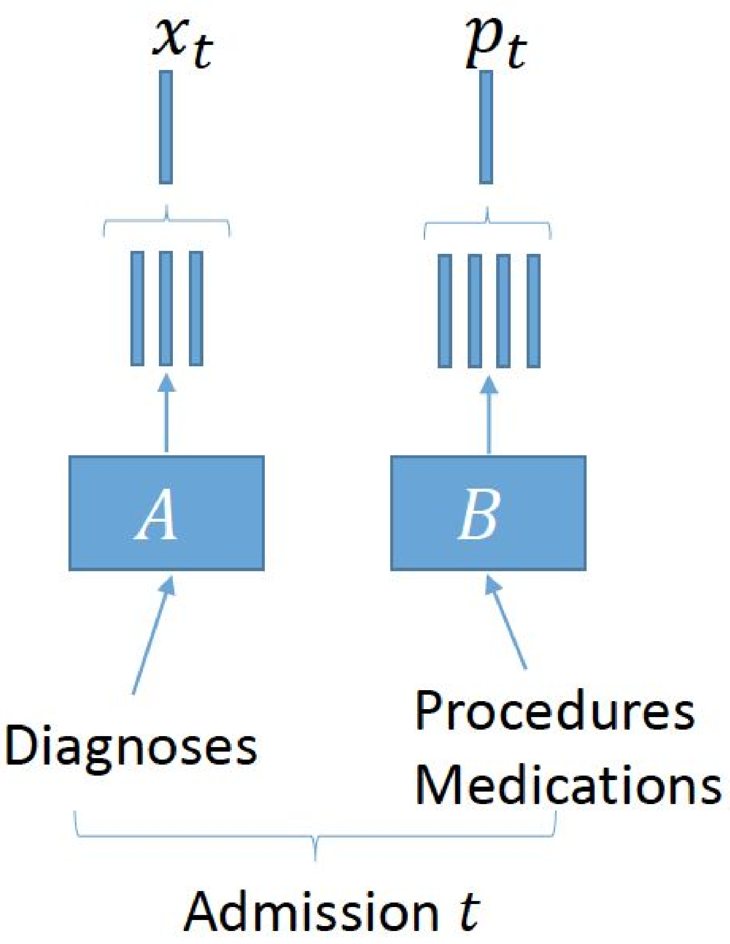
    Modified LSTM to incorporate time intervals
  - Confounding interactions between disease progression and intervention

    Modified LSTM to input diagnosis and intervention

Model Architecture



Admission Representation



Max pooling

$$\boldsymbol{x}_t^i = \max\left(A_i^{d_1}, A_i^{d_2}, ..., A_i^{d_h}\right)$$

$$\boldsymbol{p}_t^i = \max\left(B_i^{s_1}, B_i^{s_2}, ..., B_i^{s_k}\right)$$

Normalized sum pooling

$$\boldsymbol{x}_t^i = \frac{A_i^{d_1} + A_i^{d_2} + ... + A_i^{d_h}}{\sqrt{| A_i^{d_1} + A_i^{d_2} + ... + A_i^{d_h} |}}$$

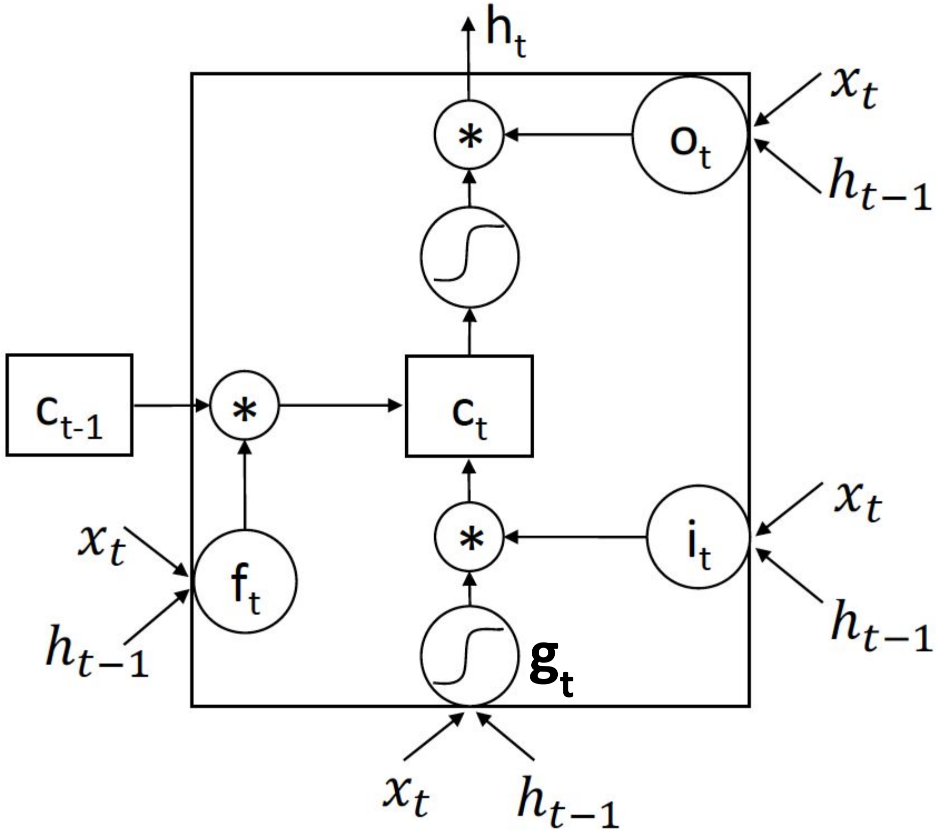
$$\boldsymbol{p}_t^i = \frac{B_i^{s_1} + B_i^{s_2} + ... + B_i^{s_k}}{\sqrt{| B_i^{s_1} + B_i^{s_2} + ... + B_i^{s_k} |}}$$

Mean pooling

$$\boldsymbol{x}_t = \frac{A^{d_1} + A^{d_2} + ... + A^{d_h}}{h}$$

$$\boldsymbol{p}_t = \frac{B^{s_1} + B^{s_2} + ... + B^{s_k}}{k}$$

A Recall of Standard LSTM



$$\boldsymbol{i}_t = \sigma\left(W_i \boldsymbol{x}_t + U_i \boldsymbol{h}_{t-1} + \boldsymbol{b}_i\right)$$

$$\boldsymbol{f}_t = \sigma\left(W_f \boldsymbol{x}_t + U_f \boldsymbol{h}_{t-1} + \boldsymbol{b}_f\right)$$

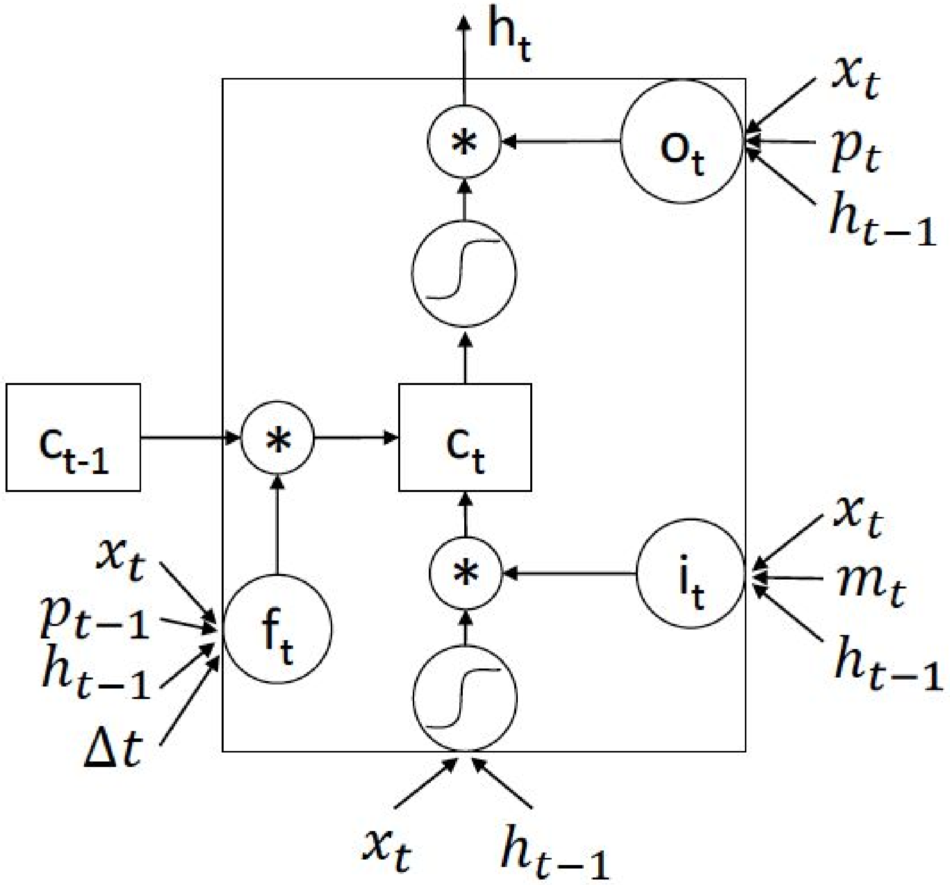
$$\boldsymbol{o}_t = \sigma\left(W_o \boldsymbol{x}_t + U_o \boldsymbol{h}_{t-1} + \boldsymbol{b}_o\right)$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t * \boldsymbol{c}_{t-1} + \boldsymbol{i}_t * \boldsymbol{g}_t$$

$$\boldsymbol{g}_t = \tanh\left(W_c \boldsymbol{x}_t + U_c \boldsymbol{h}_{t-1} + \boldsymbol{b}_c\right)$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t * \tanh\left(\boldsymbol{c}_t\right)$$

Modified LSTM



$$\boldsymbol{i}_t = \frac{1}{m_t}\sigma\left(W_i\boldsymbol{x}_t + U_i\boldsymbol{h}_{t-1} + \boldsymbol{b}_i\right)$$

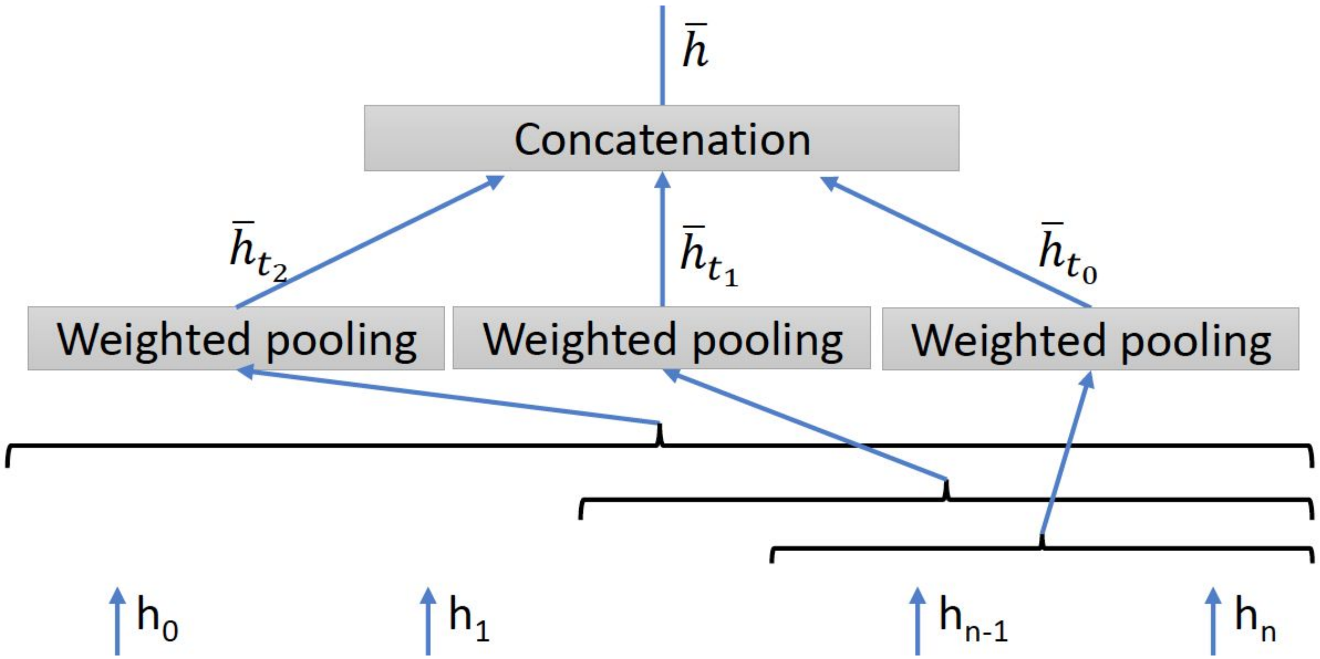
$$\boldsymbol{o}_t = \sigma\left(W_o\boldsymbol{x}_t + U_o\boldsymbol{h}_{t-1} + P_o\boldsymbol{p}_t + \boldsymbol{b}_o\right)$$

$$\boldsymbol{f}_t = \sigma\left(W_f\boldsymbol{x}_t + U_f\boldsymbol{h}_{t-1} + P_f\boldsymbol{p}_{t-1} + \boldsymbol{b}_f\right)$$

$$\boldsymbol{f}_t = \sigma\left(W_f\boldsymbol{x}_t + U_f\boldsymbol{h}_{t-1} + Q_f\boldsymbol{q}_{\Delta_{t-1:t}} + P_f\boldsymbol{p}_{t-1} + \boldsymbol{b}_f\right)$$

$$\boldsymbol{q}_{\Delta_{t-1:t}} = \left(\frac{\Delta_{t-1:t}}{60}, \left(\frac{\Delta_{t-1:t}}{180}\right)^2, \left(\frac{\Delta_{t-1:t}}{365}\right)^3\right)$$

Multiscale Pooling and Recency Attention



$$\bar{\boldsymbol{h}} = \left(\sum_{t=t_0}^{n} r_t \boldsymbol{h}_t\right) / \sum_{t=t_0}^{n} r_t$$

$$r_t = \left[m_t + \log\left(1 + \Delta_{t:n}\right)\right]^{-1}$$

$$\bar{\boldsymbol{h}} = [\bar{\boldsymbol{h}}_{12}, \bar{\boldsymbol{h}}_{24}, \bar{\boldsymbol{h}}_{all}]$$

Data Source and Processing

- 12 years (2002-2013) from a large regional Australian hospital
- Diabetes: 7,191 patients with 53,208 admissions
- Mental health: 6,109 patients with 52,049 admissions
- Preprocessing:
  - Collapse diagnoses that share the first 2 characters into one diagnosis
  - Collapse intervention that share the first digit into one intervention
- 247 diagnosis, 752 procedure and 319 medication codes

Results– Diagnoses Prediction

Table 1: Precision@n_p Diagnoses Prediction.

	Diabetes			Mental		
	$n_p = 1$	$n_p = 2$	$n_p = 3$	$n_p = 1$	$n_p = 2$	$n_p = 3$
Markov	55.1	34.1	24.3	9.5	6.4	4.4
Plain RNN	63.9	58.0	52.0	50.7	45.7	39.5
DeepCare (mean adm.)	**66.2**	**59.6**	**53.7**	**52.7**	**46.9**	**40.2**
DeepCare (sum adm.)	65.5	59.3	53.5	51.7	46.2	39.8
DeepCare (max adm.)	66.1	59.2	53.2	51.5	46.7	**40.2**

Results—Intervention Recommendation

Table 2: Precision@n_p intervention prediction

	Diabetes			Mental		
	$n_p = 1$	$n_p = 2$	$n_p = 3$	$n_p = 1$	$n_p = 2$	$n_p = 3$
Markov	35.0	17.6	11.7	20.7	12.2	8.1
Plain RNN	77.7	54.8	43.1	70.4	55.4	43.7
DeepCare (mean adm.)	77.8	54.9	43.3	70.3	55.7	44.1
DeepCare (sum adm.)	**78.7**	**55.5**	**43.5**	**71.0**	**55.8**	**44.7**
DeepCare (max adm.)	78.4	55.1	43.4	70.0	55.2	43.9

Results—Predicting unplanned readmission

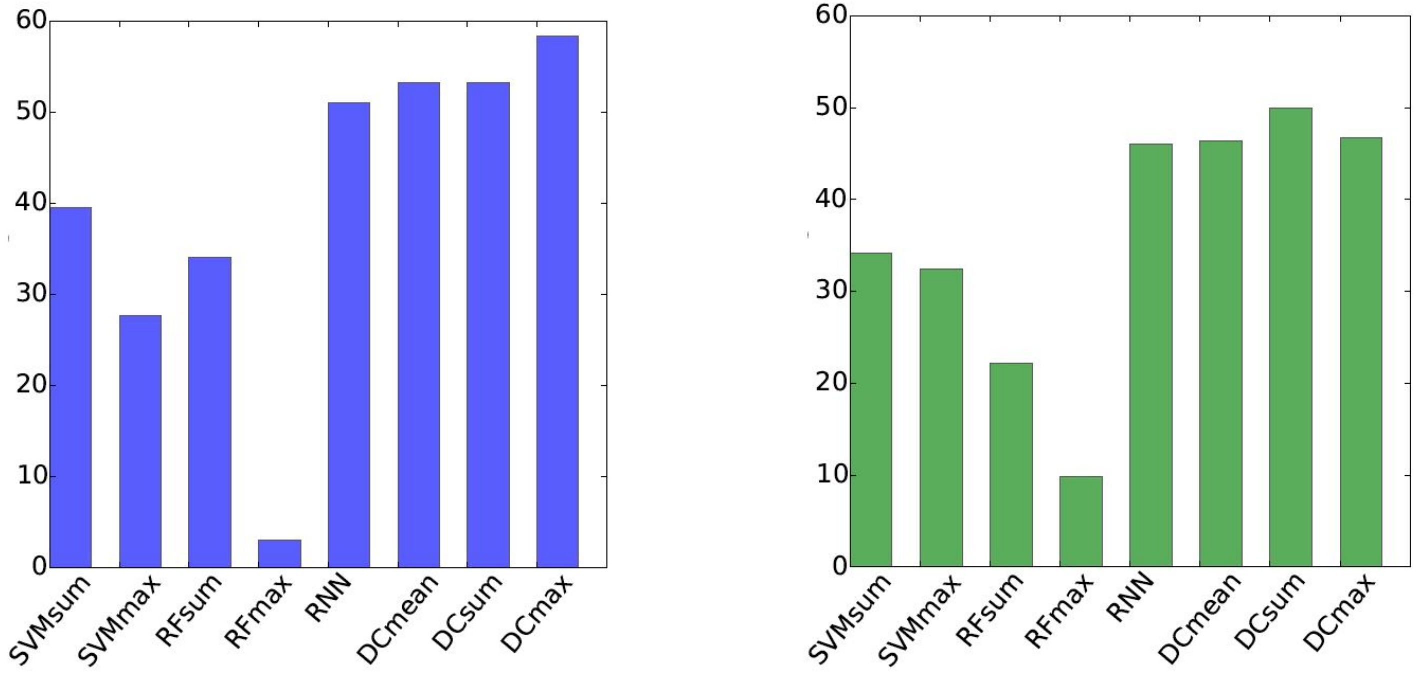
- Risk prediction: For each patient, a discharge is randomly chosen as prediction point, from which unplanned readmission is predicted

Table 4: Results of unplanned readmission prediction in F-score (%) within 12 months for diabetes and 3 months for mental health patients (DC is DeepCare, inv. is intervention).

Model	Diabetes	Mental
1. SVM (*max-pooling*)	64.0	64.7
2. SVM (*sum-pooling*)	66.7	65.9
3. Random Forests (*max-pooling*)	68.3	63.7
4. Random Forests (*sum-pooling*)	71.4	67.9
5. Plain RNN (*logist. regress.*)	75.1	70.5
6. LSTM (*logit. regress.*)	75.9	71.7
7. DC (*nnets + mean adm.*)	76.5	72.8
8. DC ([*inv.+time decay*]+*recent.multi.pool.+nnets+mean adm.*)	77.1	74.5
9. DC ([*inv.+param. time*]+*recent.multi.pool.+nnets+mean adm.*)	**79.0**	**74.7**

Results—Predicting high risk patients

- Risk prediction: For each patient, a discharge is randomly chosen as prediction point, from which high risk patients within X months will be predicted. X=12 for diabetes, X=3 for mental health



(a) Diabetes (b) Mental health

Figure 8: Result of high risk prediction in F-score (%) within 12 months for diabetes (a) and 3 months for mental health (b). DC is DeepCare. Mean, sum, max are 3 admission pooling methods

Summary

- **Innovations:**
- Embedding variable-size discrete admissions into vector space
- Parameterizing time to enable irregular timing
- Incorporating interventions to reflect their targeted influence in the course of illness and disease progression
- Maybe the first one to use LSTM to predict chronic disease