DeepCare: A Deep Dynamic Memory Model for Predictive Medicine

Deep For Haramburger

Overview

A modified LSTM architecture is proposed that can analyze **Electronic Health Records** in order to predict **disease progression** and **unplanned readmissions** of patients.

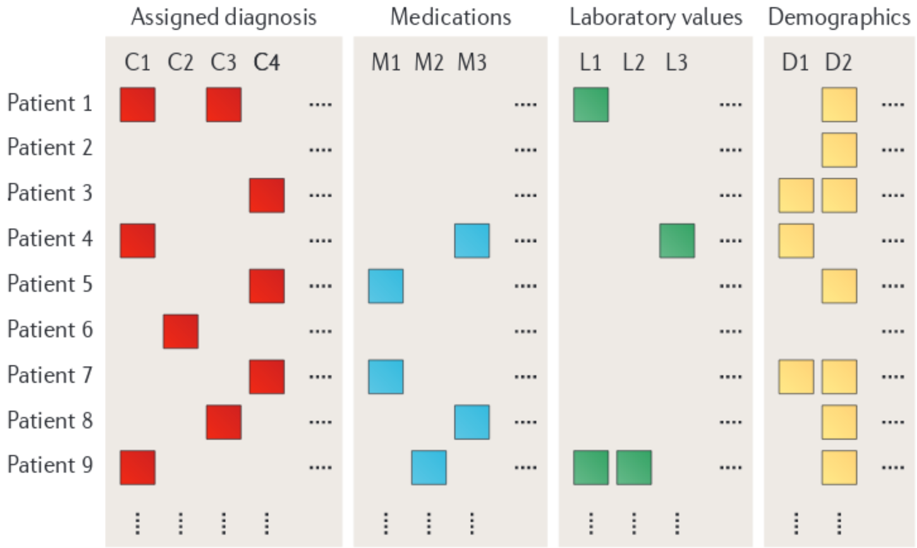
INPUT: N = 12,000 Electronic Health Records from large hospital, 2002-13

OUTPUT: For post-discharge *diabetic* patients: Predicts (1) next stage of disease, (2) intervention recommendation (ARXIV version), and (3) readmission risk

MODEL: Modified LSTM

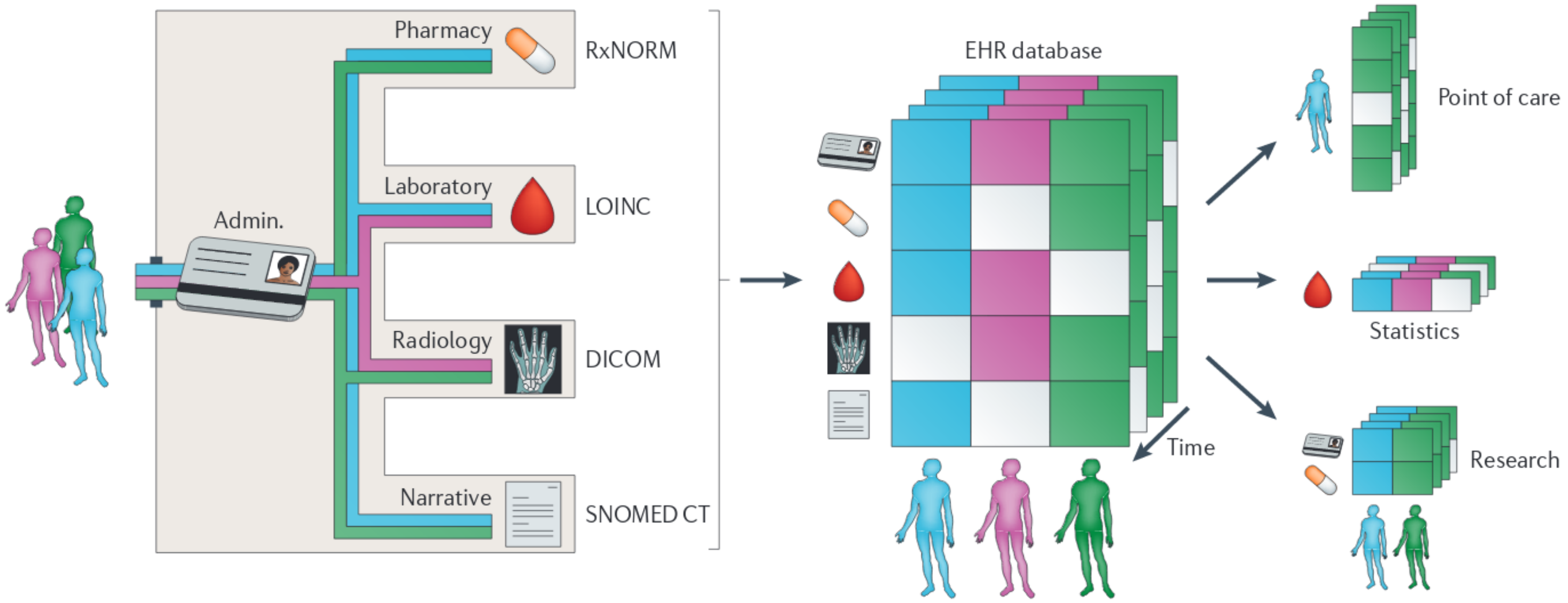
Electronic Health Records (HR)

- Data related to patient admissions, diagnoses, interventions, treatments
    - Disease codes: International Classification of diseases (ICD)
- Created by and stored by institution (e.g. Hospital)
- Contains both structured and unstructured data
    - Structured: drug prescriptions and dosages



Simplified representation of EHR
P. Jensen, L. Jensen, S. Brunak - 2012

Electronic Health Records (EHR)



EHR generation, storage and uses
P. Jensen, L. Jensen, S. Brunak - 2012

Electronic Health Records (EHR)

Characteristics of EHR

- Variable length
- Episodic at Irregular time intervals
- Interactions between disease progression and intervention
- Both structured and unstructured data
  - Structured: disease/procedure codes
  - Unstructured: Doctor’s notes
- Long term dependencies

DeepCare Architecture

End-to-end dynamic memory network for modeling illness trajectories and predicting future outcomes

Extends the basic LSTM to deal with three issues:

1. **Variable-size discrete inputs**
2. **Confounding interactions between disease progression and intervention**
3. **Irregular timing**

DeepCare Model

Input Construction

Input admission contains ***diagnoses codes*** (current condition) and ***interventions*** (procedures and medications)

Codes and interventions are each embedded into embedding matrices*: $A \in \mathbb{R}^{M \times |\mathcal{D}|}$ $B \in \mathbb{R}^{M \times |\mathcal{I}|}$.

Admission *t* has *h* diagnoses and *k* interventions:
$d_1, d_2, ..., d_h \in \{1, 2, .., |\mathcal{D}|\}$ $s_1, s_2, ..., s_k \in \{1, 2, ..., |\mathcal{I}|\}$

$\Rightarrow$ **Admission Pooling**

*unclear what method is actually implemented for learning the embedding.
It could be Continuous BoW or RNN based

DeepCare Model

Input Construction

Max Pooling Admission 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

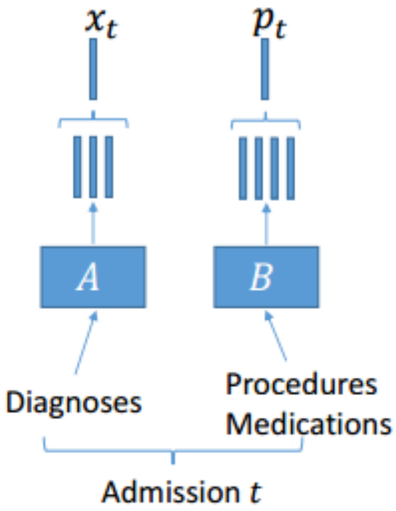
$$\boldsymbol{x}_t = \frac{A^{d_1} + A^{d_2} + ... + A^{d_h}}{h}$$

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

DeepCare Model

Input Construction

Final admission embedding is a 2M dimension vector $[\boldsymbol{x}_t, \boldsymbol{p}_t]$



DeepCare Model

High Level Architecture

Given input admission embedding ($[\boldsymbol{x}_t, \boldsymbol{p}_t]$), admission type (m_t), and elapsed time between current admission and previous one (Δt).

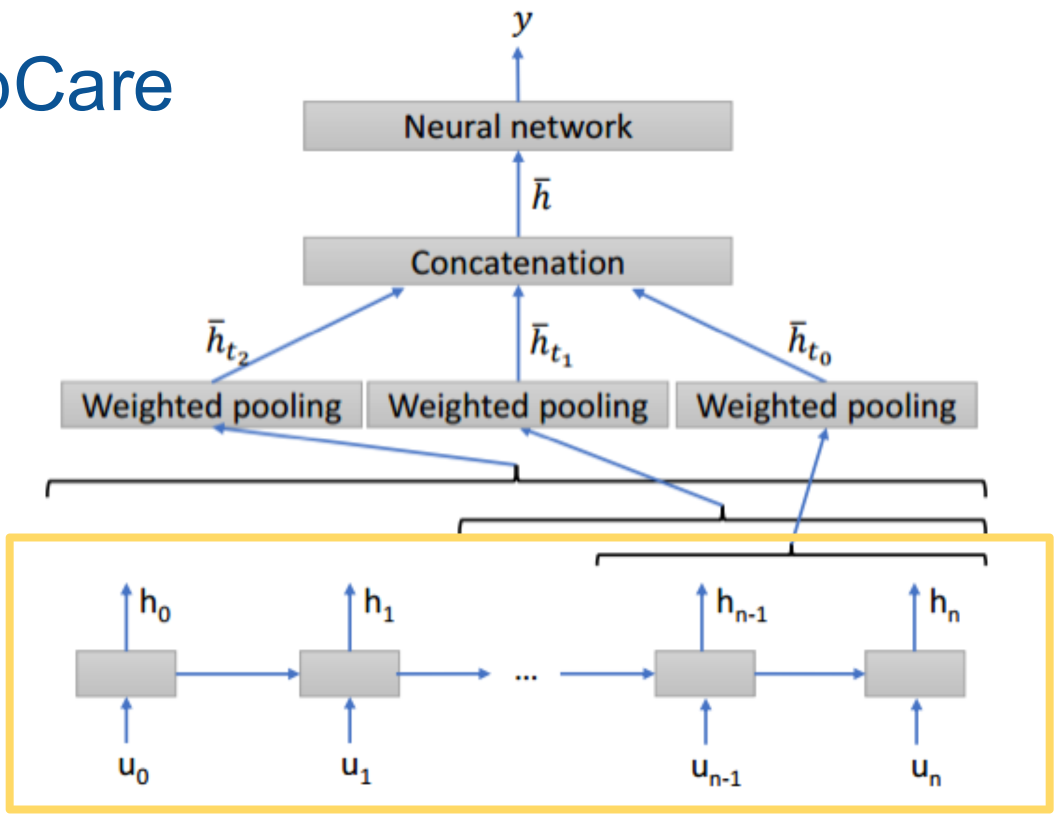
Input sequence: $\boldsymbol{u}_0, \boldsymbol{u}_1, ..., \boldsymbol{u}_n$ $\boldsymbol{u}_t = [\boldsymbol{x}_t, \boldsymbol{p}_t, m_t, \Delta t]$.

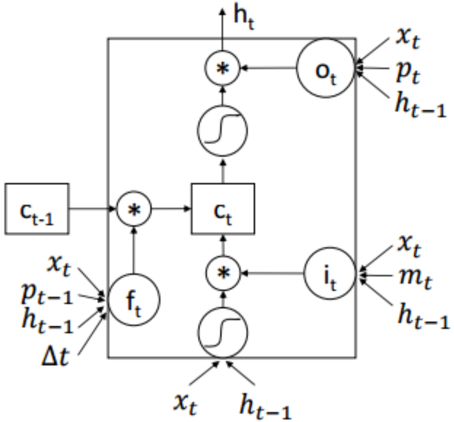
LSTM output sequence: distributed illness states $\boldsymbol{h}_0, \boldsymbol{h}_1, ..., \boldsymbol{h}_n$, where $\boldsymbol{h}_t \in \mathbb{R}^K$

Aggregate illness states: $\bar{\boldsymbol{h}} = \{\boldsymbol{h}_0, \boldsymbol{h}_1, ..., \boldsymbol{h}_n\}$, where $\bar{\boldsymbol{h}} \in \mathbb{R}^{sK}$ for s scales.

Outcome estimate: $P(y \mid \boldsymbol{u}_{0:n}) = P\left(\text{nnet}_y\left(\text{pool}\left\{\text{LSTM}(\boldsymbol{u}_{0:n})\right\}\right)\right)$

DeepCare





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

DeepCare Model

Modified LSTM Structure

Incorporating admission type

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

$m_t = 1$ if the admission method is unplanned, $m_t = 2$ otherwise

Modeling Effect of Interventions

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

P_o is the intervention weight matrix for the output gate and $\boldsymbol{p}_t$ is intervention at time step t.

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

$\boldsymbol{p}_{t-1}$ is intervention embedded vector at time step $t-1$ and P_f is the intervention weight matrix

DeepCare Model

Modified LSTM Structure

Dealing with Time Irregularity

1. Natural Time Decay

$$\boldsymbol{f}_t \leftarrow d\left(\Delta_{t-1:t}\right)\boldsymbol{f}_t$$

$\Delta_{t-1:t}$ is the time passed between step $t-1$ and step t.
$d(\Delta_{t-1:t}) = [\log(e + \Delta_{t-1:t})]^{-1}$, where $\Delta_{t-1:t}$ is measured in days

1. Parametric Time

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

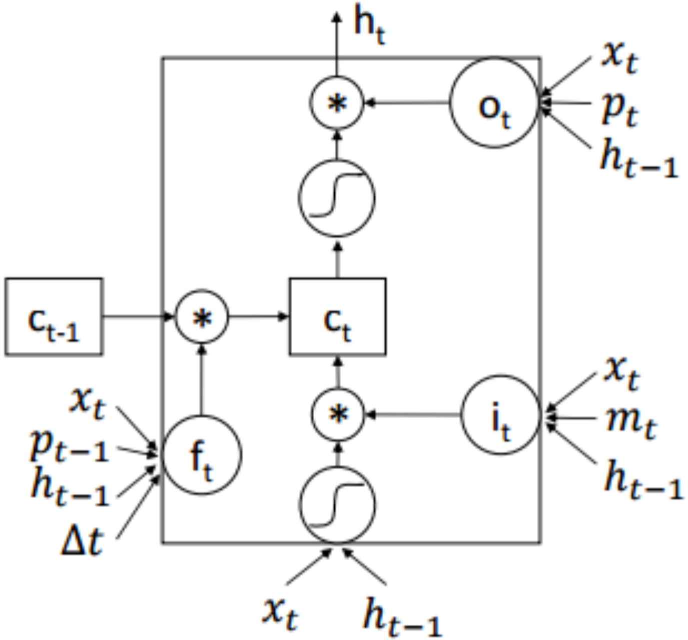
$\boldsymbol{q}_{\Delta_{t-1:t}}$ is a vector derived from the time difference $\Delta_{t-1:t}$:

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

Ex. Third-order effects:

DeepCare Model

Modified LSTM Structure



DeepCare Model

Multiscale Pooling

Problem: max-pooling only reflects a single illness state in patient's history

Solution: DeepCare uses a simple attention mechanism for pooling

Weighted sum over all historical states

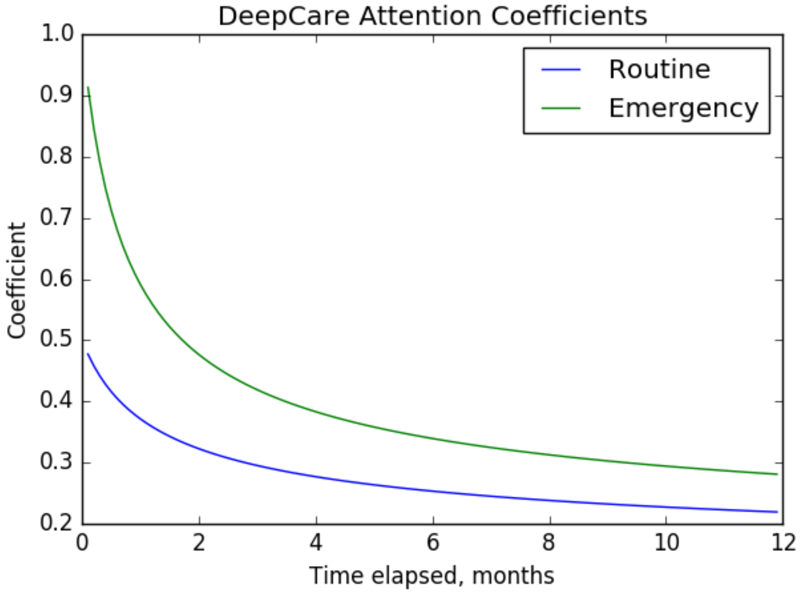
Recent events more heavy

Emergency admissions more heavy

Illness states pooled over multiple

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

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



DeepCare Model

Model FC/Readout layers

Fully-connected hidden layer

$$\boldsymbol{a}_h = \sigma\left(U_h\bar{\boldsymbol{h}} + \boldsymbol{b}_h\right)$$

Softmax readout layer

$$\begin{aligned}\boldsymbol{z}_y &= U_y\boldsymbol{a}_h + \boldsymbol{b}_y\\ P\left(y \mid \boldsymbol{u}_{1:n}\right) &= \mathrm{f}_{prob}\left(\boldsymbol{z}_y\right)\end{aligned}$$

DeepCare Model

Training

- Cross-entropy loss via SGD with minibatch size 16
  - Learning rate adjusts to loss decreases
    - After 5 iterations of no change, cut LR in half
- Regularization
  - L2-regularization on LSTM outputs
  - Dropout on initial embedding layer and hidden FC layer
- Pretraining

Results

Performance of DeepCare assessed on two fronts:

1. Prediction of future disease diagnosis

1. Prediction of readmission within 12 months after discharge

Predicting future disease diagnoses

- Predict the next n_{pred} diagnoses after each discharge
- 243 possible disease diagnoses

Table 1. Precision@n_{pred} diagnoses prediction.

Model	$n_{pred} = 1$	$n_{pred} = 2$	$n_{pred} = 3$
Markov	55.1	34.1	24.3
Plain RNN	63.9	58.0	52.0
DeepCare (interven. + param. time)	**66.0**	**59.7**	**54.1**

Markov model: memoryless disease transition probabilities:

$$P\left(d_t^i \mid d_{t+1}^j\right) \text{ from disease } d^j \text{ to } d^i \text{ at time } t$$

Probability of d_t^i as next disease :

$$\frac{1}{|D_t|} \sum_{j \in D_t} P\left(d_t^i \mid d_{t+1}^j\right)$$

Predicting future disease diagnoses

- Predict the next n_{pred} diagnoses after each discharge
- 243 possible disease diagnoses

Table 1. Precision@n_{pred} diagnoses prediction.

Model	$n_{pred} = 1$	$n_{pred} = 2$	$n_{pred} = 3$
Markov	55.1	34.1	24.3
Plain RNN	63.9	58.0	52.0
DeepCare (interven. + param. time)	**66.0**	**59.7**	**54.1**

Predicting disease diagnoses using LSTM output:

$$\underbrace{\text{output states } \boldsymbol{h}_t}$$

$$P\left(y_t = l \mid \boldsymbol{x}_{1:t}\right) = \text{softmax}\left(\boldsymbol{v}_l^\top \boldsymbol{h}_t\right)$$

$$\text{label specific parameters } \boldsymbol{v}_l$$

Predicting unplanned readmission

- a discharge is randomly chosen as prediction point
- unplanned readmission after 12 months is predicted

	Model	F-score (%)
1	SVM (*max-pooling*)	64.0
2	SVM (*sum-pooling*)	66.7
3	Random Forests (*max-pooling*)	68.3
4	Random Forests (*sum-pooling*)	71.4
5	LSTM (*mean-pooling + logit. regress.*)	75.9
6	DeepCare (*mean-pooling + nnets*)	76.5
7	DeepCare ([*interven. + time decay*]+ *recent.multi.pool. + nnets*)	77.1
8	**DeepCare ([*interven. + param. time*]+ *recent.multi.pool. + nnets*)**	**79.1**

F-score is harmonic mean of the precision and recall: $F_1 = 2\frac{1}{\frac{1}{recall} + \frac{1}{precision}}$

Contribution of time irregularity to predicting unplanned readmission

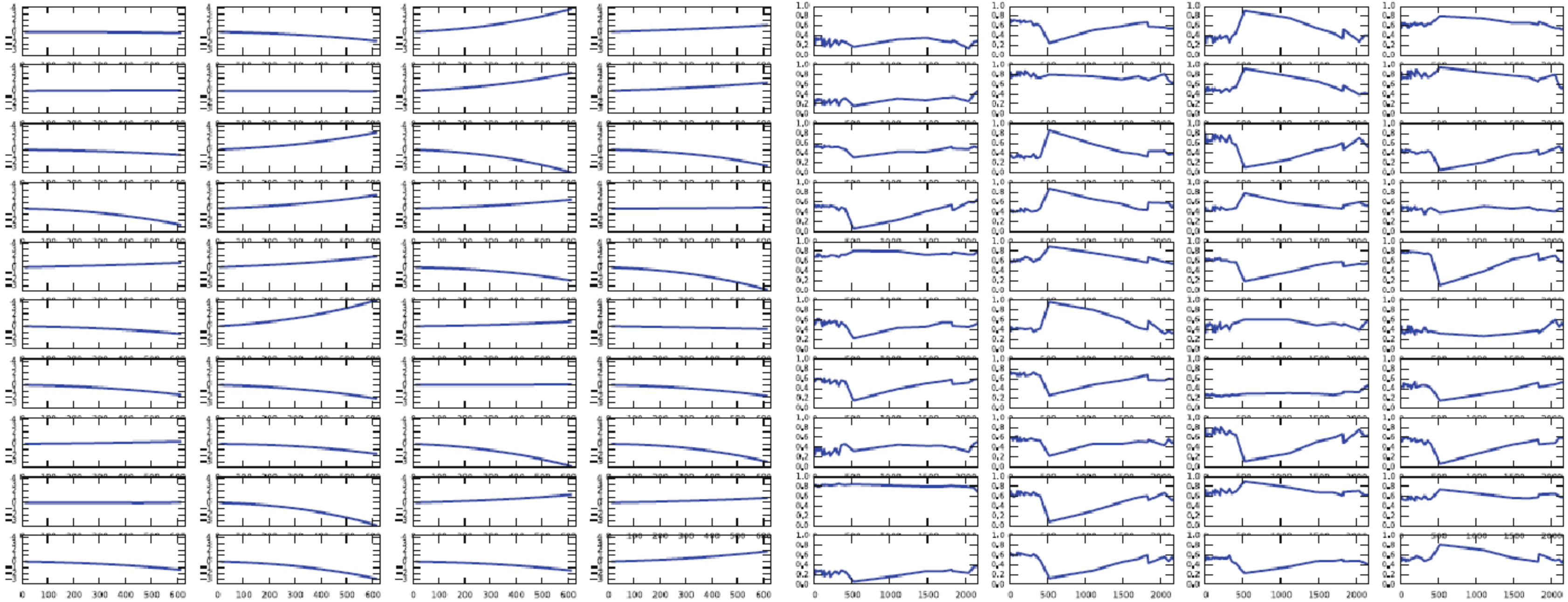
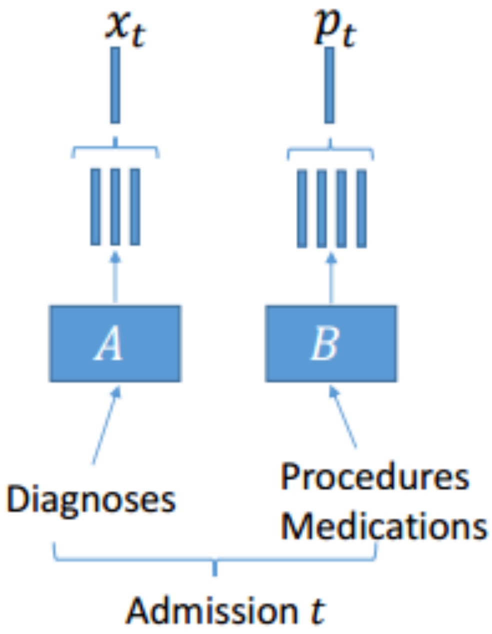


Fig. 4. (Left) 40 channels of forgetting due to time elapsed. (Right) The forget gates of a patient in the course of their illness.

$$\boldsymbol{f}_t = \sigma\left(W_f \boldsymbol{x}_t + U_f \boldsymbol{h}_{t-1} + Q_f \boldsymbol{q}_{\Delta_{t-1:t}} + P_f \boldsymbol{p}_{t-1} + \boldsymbol{b}_f\right) \quad \boldsymbol{q}_{\Delta_{t-1:t}} = \left(\Delta_{t-1:t}, \Delta_{t-1:t}^2, \Delta_{t-1:t}^3\right)$$

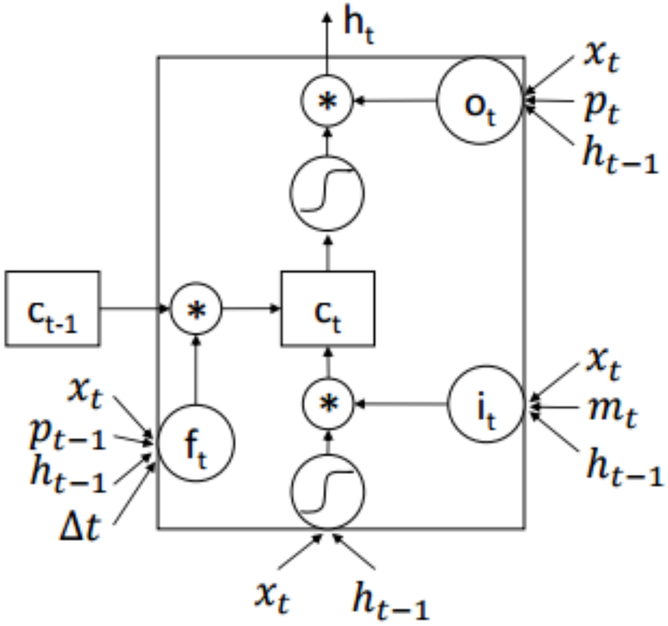
Summary: solve four major problems

- Representation of variable inputs
- Long-term dependencies
- Episodic recording and irregular timing
- Interactions between disease progression and intervention



Summary: solve four major problems

- Representation of variable inputs
- Long-term dependencies
- Episodic recording and irregular timing
- Interactions between disease progression and intervention



Conclusions

- DeepCare's primary value is that it frees model designers from manually designing features from EMRs.
- It uses modified LSTM and pooling units to handle timing irregularities with which existing techniques struggle.
- The full progression of a disease is embedded into a vector which is then used to make risk and readmission predictions.
- Results are competitive against current state-of-the-arts.