



Discovery



Diagnosis



Prognosis



Care

# Deep learning for biomedicine II



**Truyen Tran**  
Deakin University

Seoul, Nov 2017



truyen.tran@deakin.edu.au



truyentran.github.io



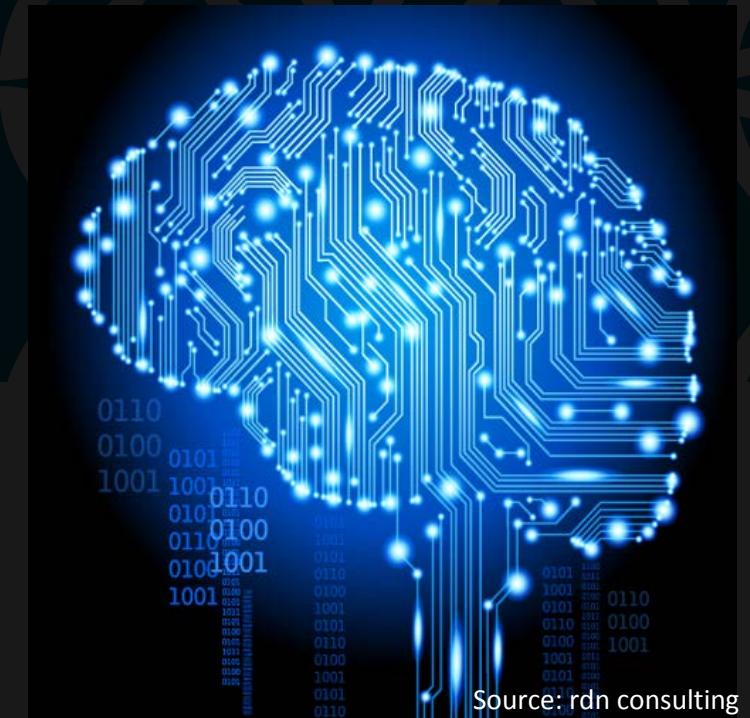
@truyenoz



letdataspeak.blogspot.com



goo.gl/3jJ100



# Resources

Slides and references:

- <https://truyentran.github.io/acml17-tute.html>

Key survey paper (updated frequently):

- **Ching, Travers, et al. "Opportunities And Obstacles For Deep Learning In Biology And Medicine." *bioRxiv* (2017): 142760**

# Agenda

Topic 1: Introduction (20 mins)

Topic 2: Brief review of deep learning (25 mins)

- Classic architectures
- Capsules
- Graphs
- Memory-augmented nets

Topic 3: Genomics (25 mins)

- Nanopore sequencing
- Genomics modelling

QA (10 mins)



**Break (20 mins)**

Topic 4: Biomedical imaging (15 mins)

- Cellular imaging
- Diagnostics imaging
- EEG/ECG

Topic 5: Healthcare (25 mins)

- Time-series of physio measures
- Trajectories prediction

Topic 6: Generative biomed (30 mins)

- Few-shot learning
- Generative models
- Drug design
- Future outlook

QA (10 mins)

# Sensing technologies and data

Raw signals are ideal candidates for deep learning

Speech & vision techniques can be applied with minimal changes

#REF: Ravì, Daniele, et al. "Deep learning for health informatics." *IEEE journal of biomedical and health informatics* 21.1 (2017): 4-21.



# The state of biomedical imaging



As estimated by IBM, 90% data of healthcare is imaging.

Biomedical imaging is perhaps the most ready area for current DL techniques:

- **It typically means CNN!**
- **The game is in the data acquisition and problem definition/transformation**

Examples of application areas:

- Cellular imaging
- Tumor Detection & tracking
- Blood Flow Quantification and Visualization
- Medical Interpretation
- Diabetic Retinopathy

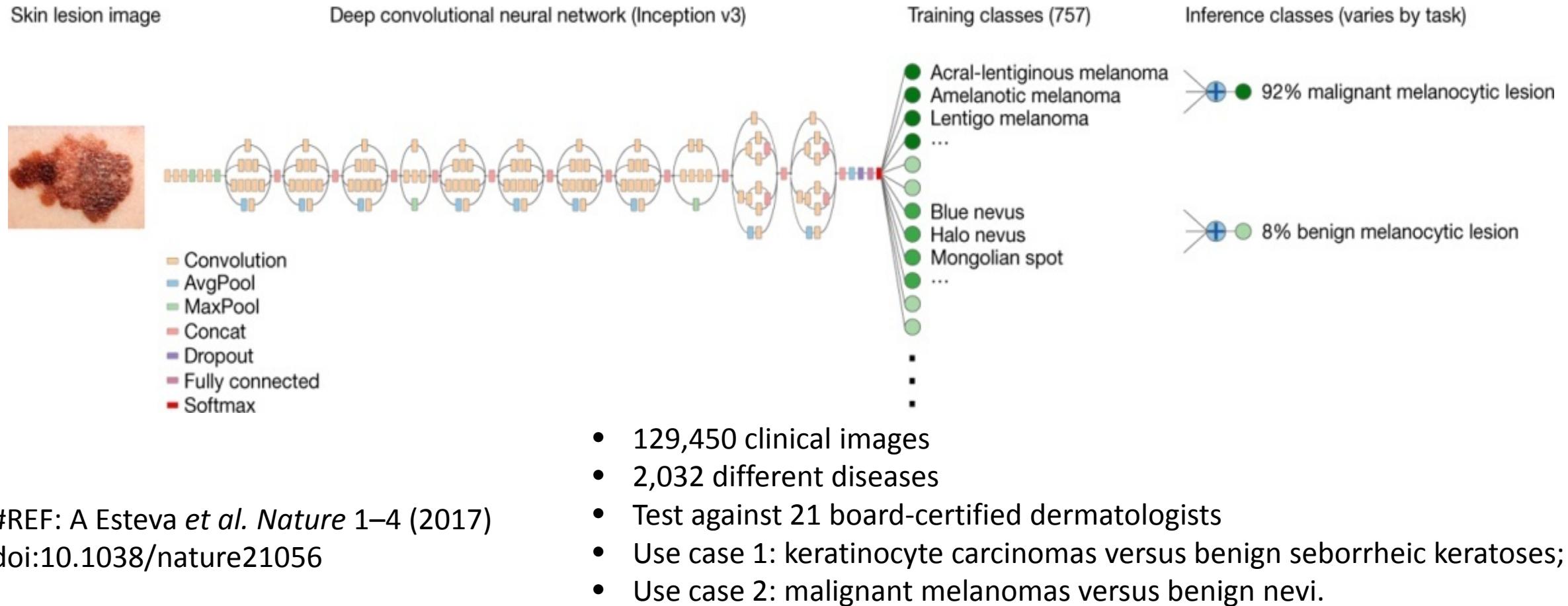
## Tasks

- Classification
- Segmentation
- Localization

## Challenges

- Low quality
- Very high resolution
- Tiny localized areas

# Skin cancer diagnosis using inception-v3



#REF: A Esteva *et al. Nature* 1–4 (2017)  
doi:10.1038/nature21056

# Microscopy + mobile phone + CNN

Highly relevant  
for:

- Developing countries
- Rural areas

#REF: Quinn, John A., et al.  
"Deep convolutional neural  
networks for microscopy-  
based point of care  
diagnostics." *Machine  
Learning for Healthcare  
Conference*. 2016.

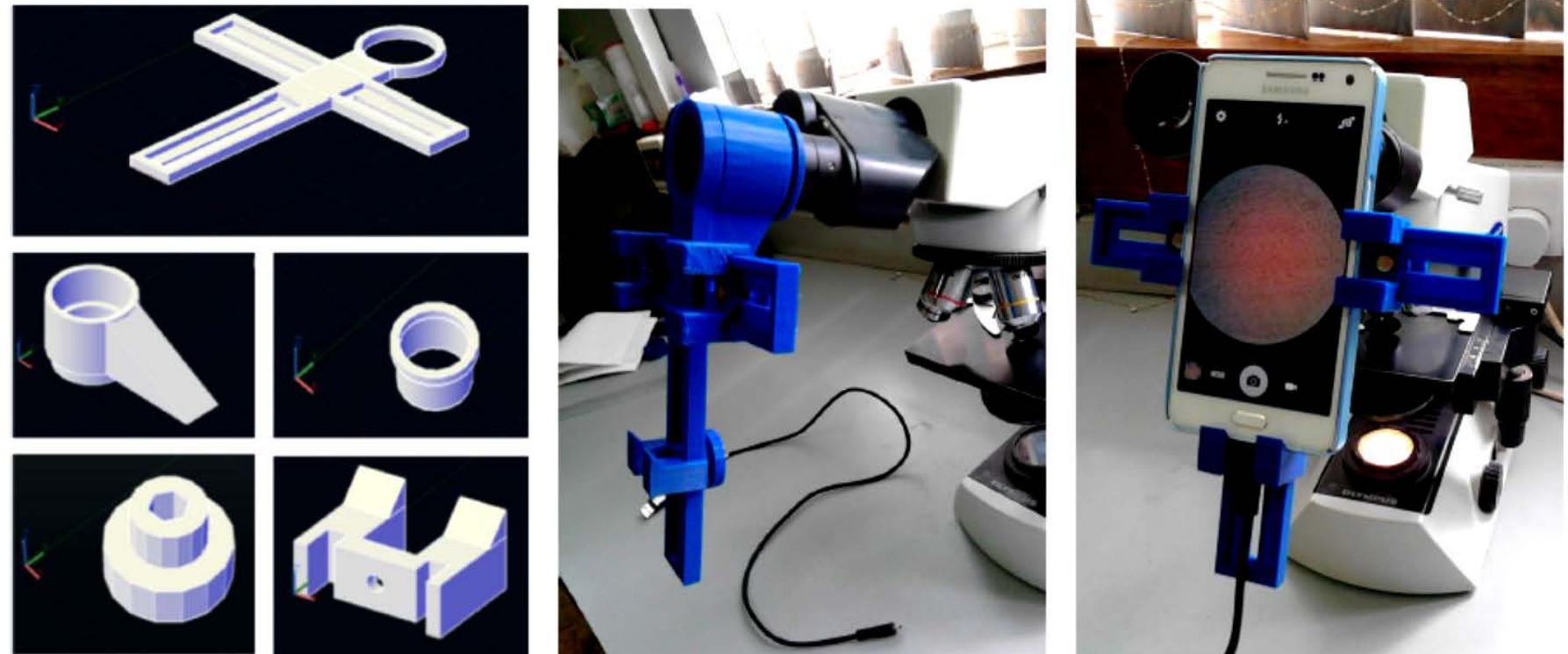
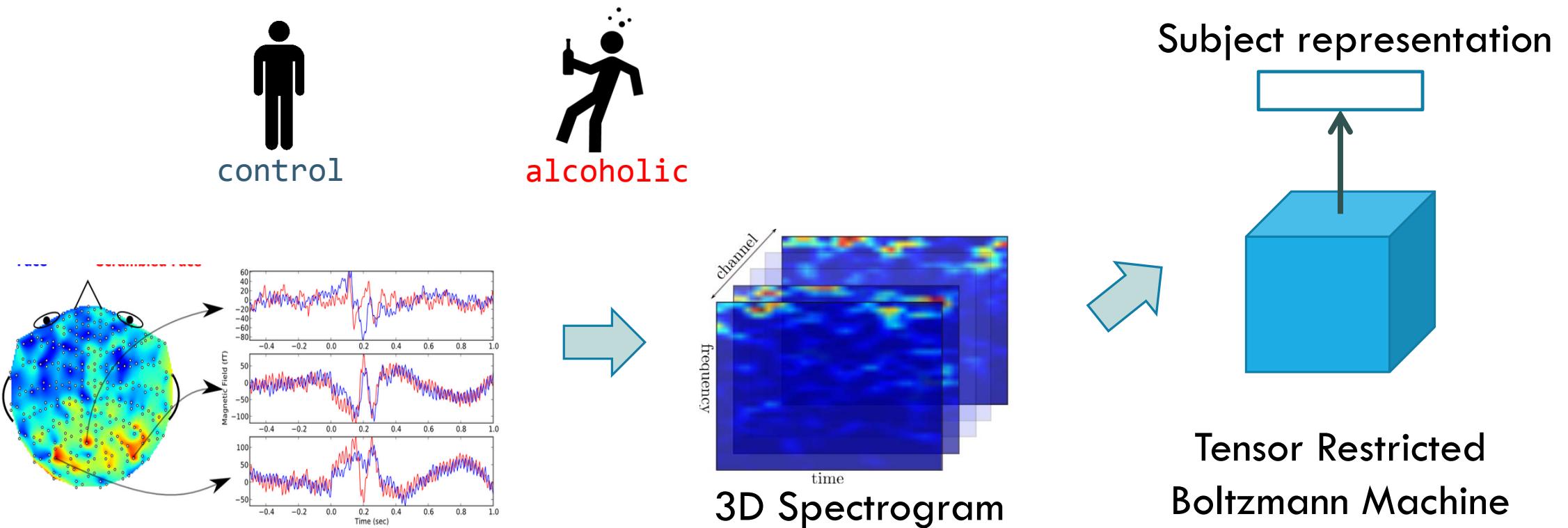


Figure 1: Microscope smartphone adapter: design of components (left), 3D-printed adapter mounted on microscope (center), smartphone inserted into adapter (right).

# EEG → Tensor RBM for alcoholic diagnosis



#Ref: Tu D. Nguyen, Truyen Tran, D. Phung, and S. Venkatesh,  
Tensor-variate Restricted Boltzmann Machines, AAAI 2015.

# EEG → Matrix LSTM → Classification

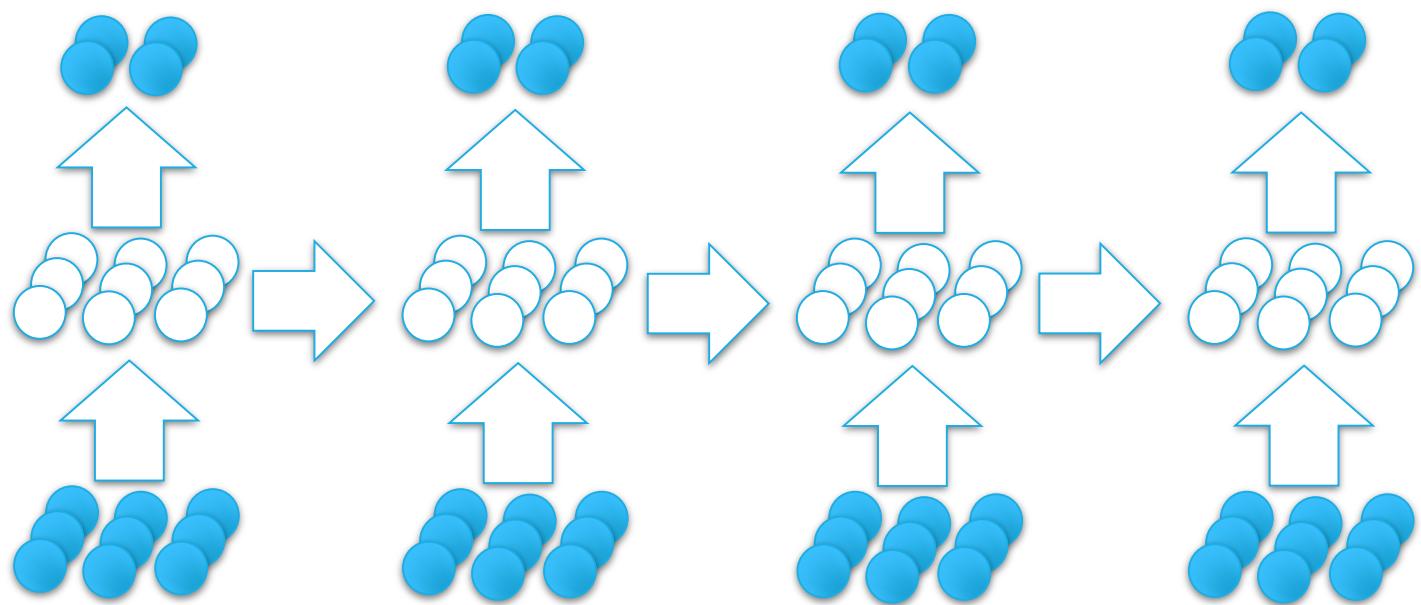
EEG segments as matrices

Temporal dynamics as recurrence

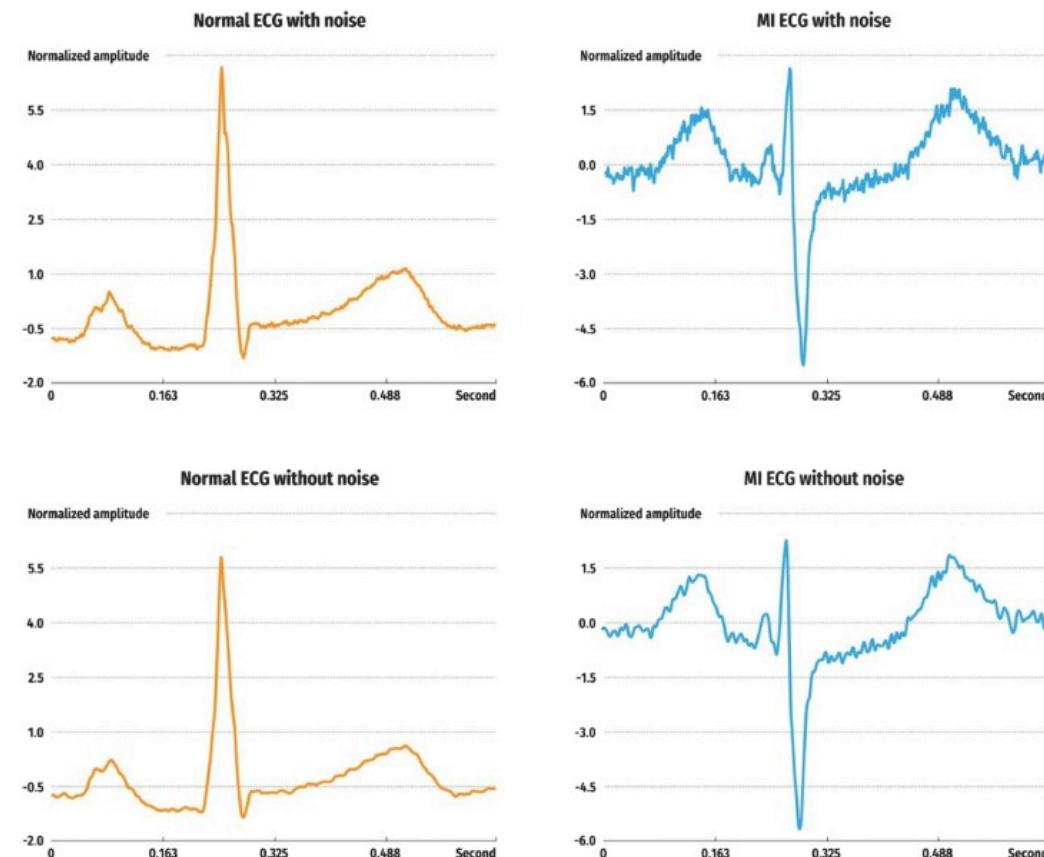
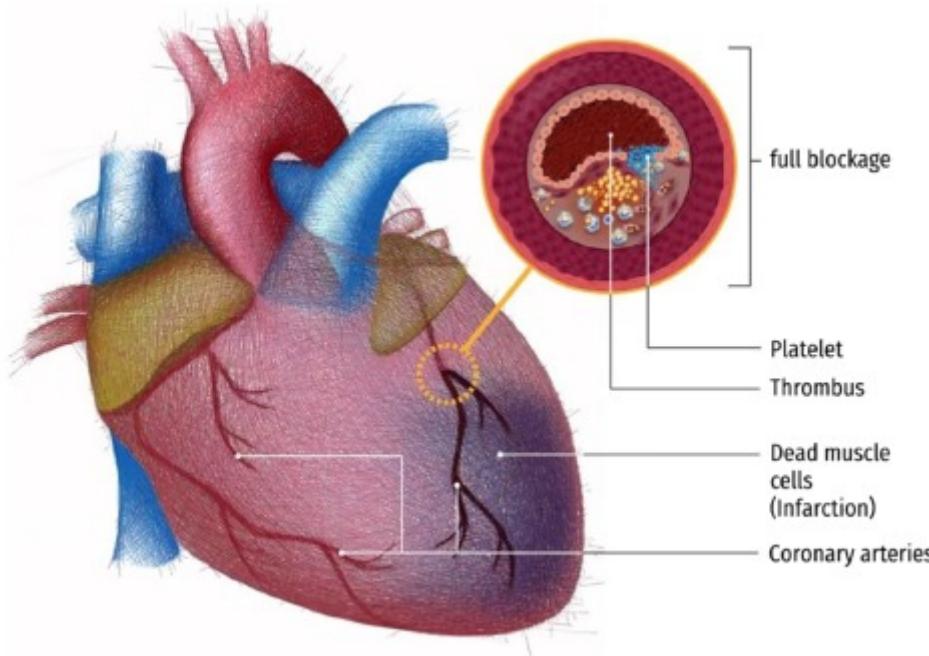
#REF: Kien Do, Truyen Tran, Svetha Venkatesh, "Learning Recurrent Matrix Representation", *Third Representation Learning for Graphs Workshop (ReLiG 2017)*

Recurrent dynamics

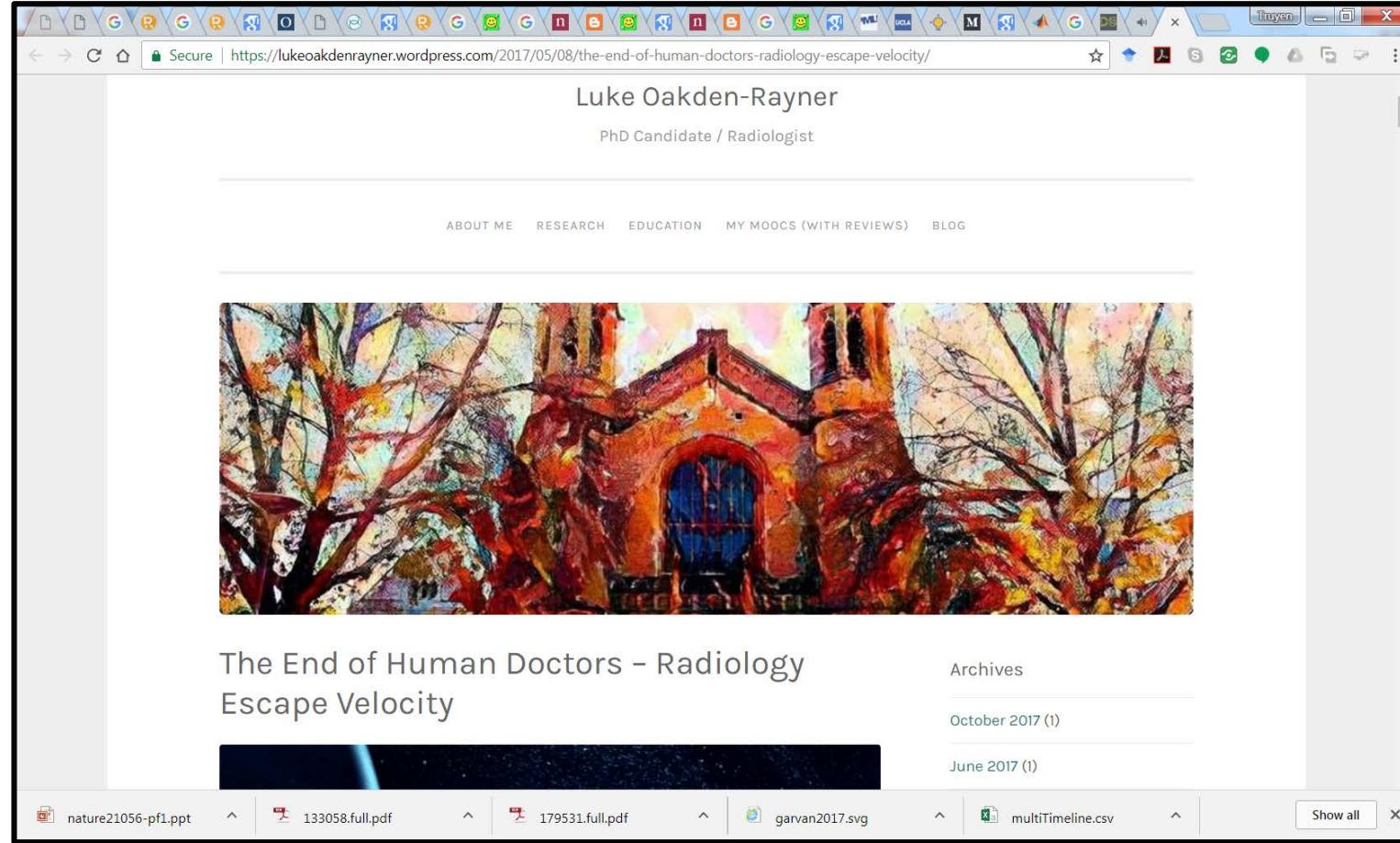
$$H_t = \sigma(U_x^\top X_t V_x + U_h^\top H_{t-1} V_h + B)$$



# ECG → CNN for heart attack detection



#REF: Acharya, U. Rajendra, et al. "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals." *Information Sciences* 415 (2017): 190-198.



“They should stop training radiologists now.”

Geoff Hinton (as of April 2017)

# Agenda

Topic 1: Introduction (20 mins)

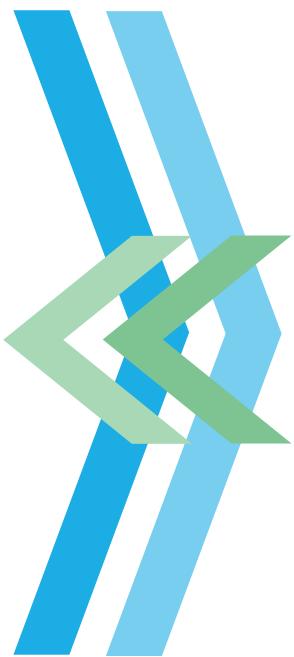
Topic 2: Brief review of deep learning (25 mins)

- Classic architectures
- Capsules
- Graphs
- Memory-augmented nets

Topic 3: Genomics (25 mins)

- Nanopore sequencing
- Genomics modelling

QA (10 mins)



**Break (20 mins)**

Topic 4: Biomedical imaging (15 mins)

- Cellular imaging
- Diagnostics imaging
- EEG/ECG

Topic 5: Healthcare (25 mins)

- Time-series of physio measures
- Trajectories prediction

Topic 6: Generative biomed (30 mins)

- Few-shot learning
- Generative models
- Drug design
- Future outlook

QA (10 mins)

# Modeling electronic medical records (EMR)

Need to model the healthcare processes, which are interactions of:

- Disease progression
- Interventions & care processes
- Recording processes (Electronic Medical/Health Records)



Source: [medicalbillingcodings.org](http://medicalbillingcodings.org)

# EMR Connects Services: System of Systems



## Five main functions

- Integrated view of patient data
- Clinical decision support
- Clinician order entry
- Access to knowledge resources
- Integrated communication and reporting support

# Clinical Decision Supports

Support protocol/planning of treatment/discharge.

Suggest course of actions:

- E.g., medication/dose/duration.

Estimate risk & predict outcomes.

Alert/reminder.

Support (semi) automated diagnosis.

heart failure diabetes  
mental health

COPD

heart attack cancers preterm

## risk prediction (prognosis)

suicide attempts

side effects

death

toxicity

readmission stress

quality-of-life

progression to advanced stages

length-of-stay

# Warning: leakage!

Make sure the patients are counted AFTER first diagnosis

- Often, we have future data as well
- Retrospective nature

Never use outcomes to do anything, except for training the model

Our early suicide attempt classification from assessments was a form of leakage:

- Any attempt in history is considered as an outcome. BUT:
- Previous attempts were accounted in current assessment already!

# Preprocessing: Data normalization & dictionary compression

## Drugs & tests

- Drug companies offer different brand names of the essentially the same drug
- DDD/ATC is the central register for the medication classes, maintained by WHO
- Several test names may be the same

It may not be robust to use the original “vocabularies”

- Tens of thousands of ICD-codes, thousands of procedures, hundreds of DRGs, thousands of medication classes
- Codes are usually organized in hierarchy
- Choosing the right hierarchy is statistical issue

# Attend to risks in Intensive Care Unit (ICU)

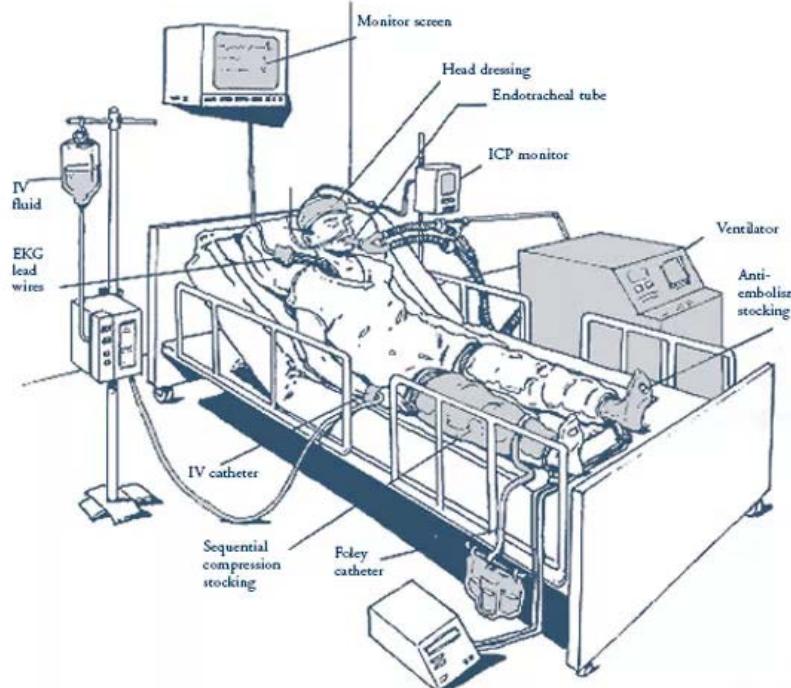
## The needs

- Accuracy
- Interpretability

## The process:

- Irregular time-series → Regular time-steps → Data imputation → Bi-LSTM → **Multiple attentions** → Classification

#REF: Phuoc Nguyen, Truyen Tran, Svetha Venkatesh, "Deep Learning to Attend to Risk in ICU", *IJCAI'17 Workshop on Knowledge Discovery in Healthcare II: Towards Learning Healthcare Systems (KDH 2017)*.



Source: [healthpages.org](http://healthpages.org)

Time	Parameter	Value
00:00	RecordID	132539
00:00	Age	54
00:00	Gender	0
00:00	Height	-1
00:00	ICUType	4
00:00	Weight	-1
00:07	GCS	15
00:07	HR	73
00:07	NIDiasABP	65
00:07	NIMAP	92.33
00:07	NISysABP	147
00:07	RespRate	19
00:07	Temp	35.1
00:07	Urine	900
00:37	HR	77
00:37	NIDiasABP	58
00:37	NIMAP	91
00:37	NISysABP	157
00:37	RespRate	19
00:37	Temp	35.6
00:37	Urine	60

Data: **Physionet 2012**

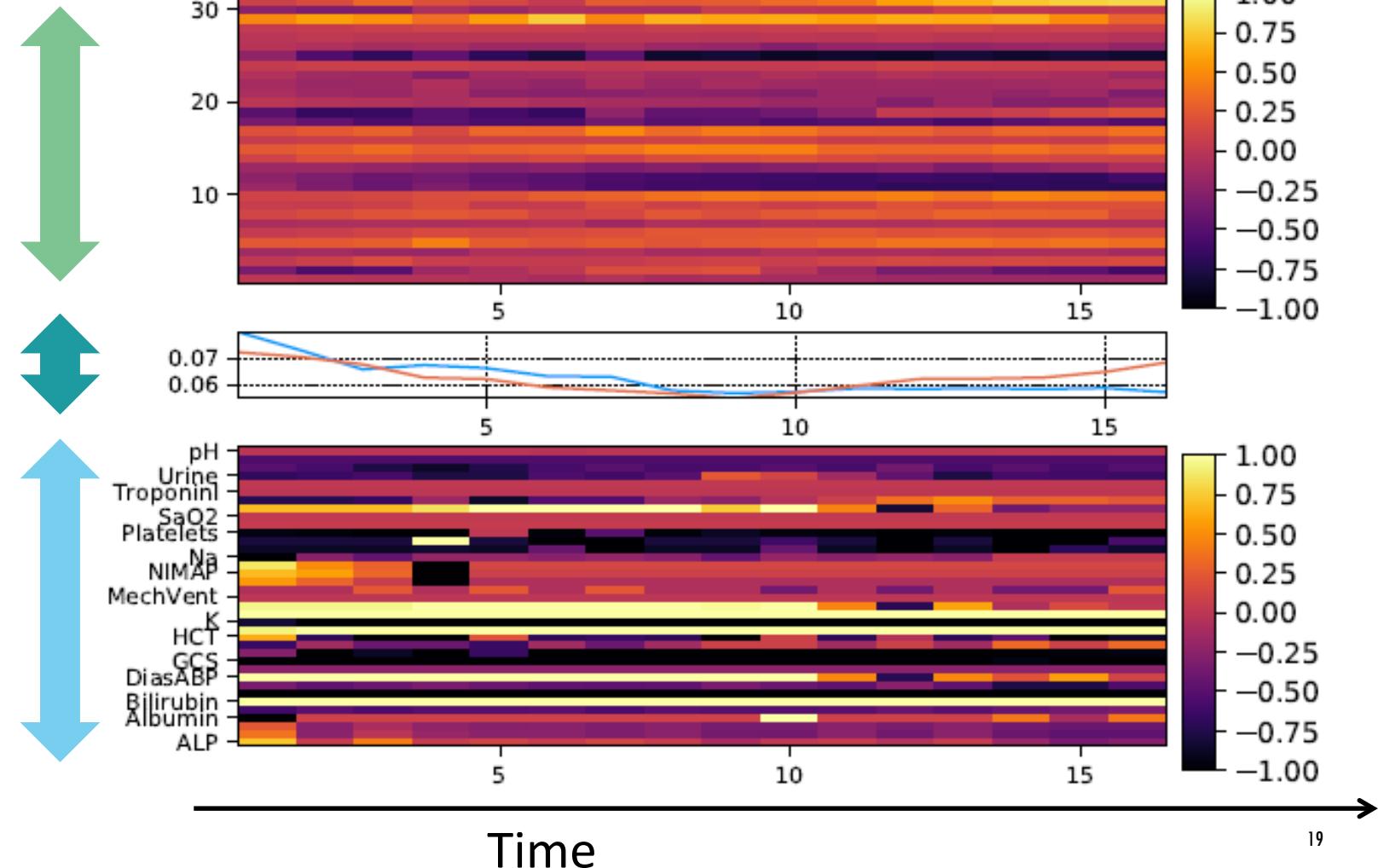
# Attend to risks in ICU

#REF: Phuoc Nguyen, Truyen Tran, Svetha Venkatesh, "Deep Learning to Attend to Risk in ICU", IJCAI'17 Workshop on Knowledge Discovery in Healthcare II: Towards Learning Healthcare Systems (KDH 2017).

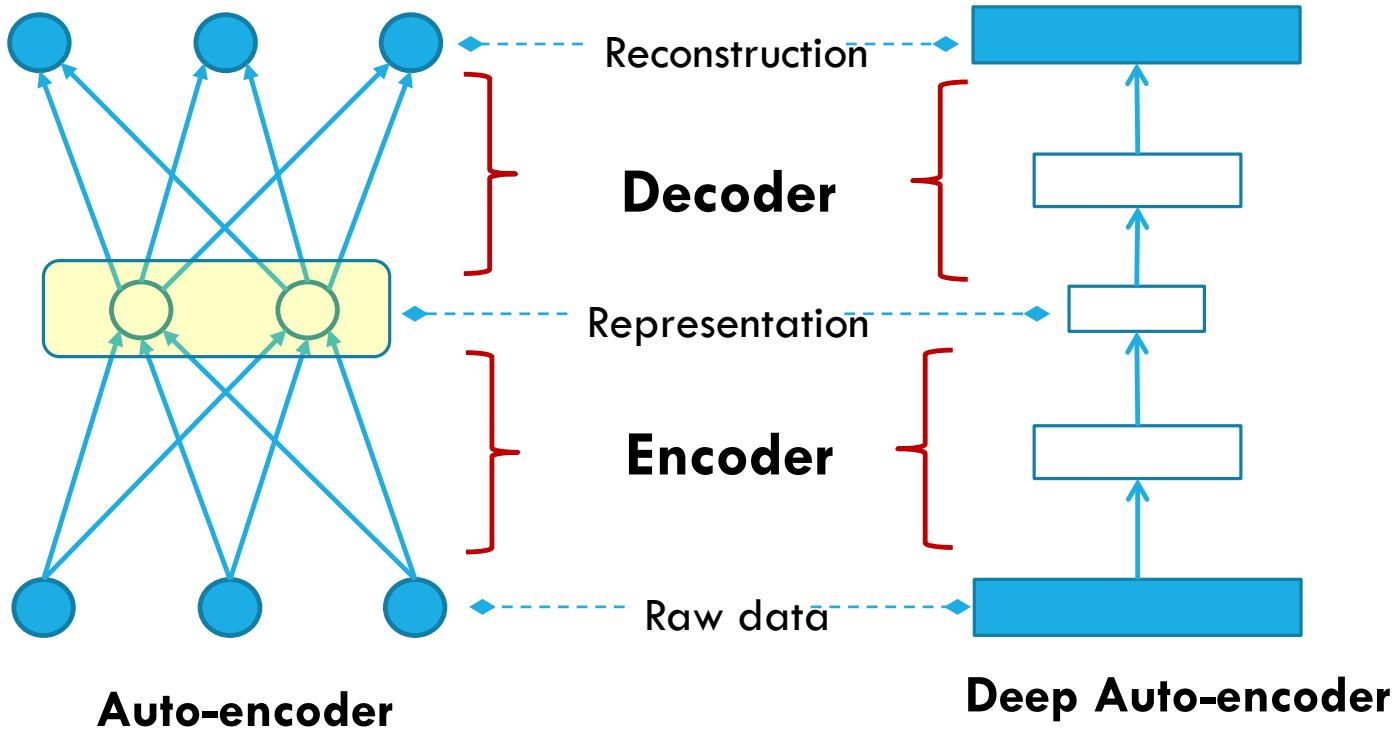
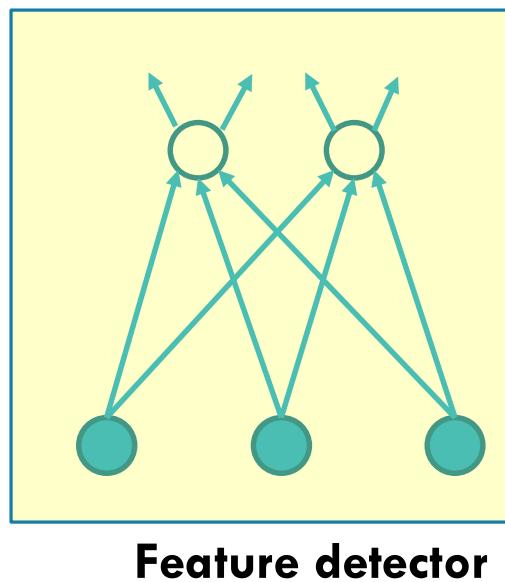
State  
transition

Attention  
probabilities

Physiological  
measures



# DeepPatient: Representing medical records with Stacked Denoising Autoencoder



#Ref: Miotto, Riccardo, et al. "Deep patient: An unsupervised representation to predict the future of patients from the electronic health records." *Scientific reports* 6 (2016): 26094.

# DeepPatient: Results on disease classification

Time Interval = 1 year (76,214 patients)			
Patient Representation	AUC-ROC	Classification Threshold = 0.6	
		Accuracy	F-Score
RawFeat	0.659	0.805	0.084
PCA	0.696	0.879	0.104
GMM	0.632	0.891	0.072
K-Means	0.672	0.887	0.093
ICA	0.695	0.882	0.101
DeepPatient	<b>0.773</b> *	<b>0.929</b> *	<b>0.181</b> *

# Trajectories modeling: Challenges & opportunities

Long-term dependencies

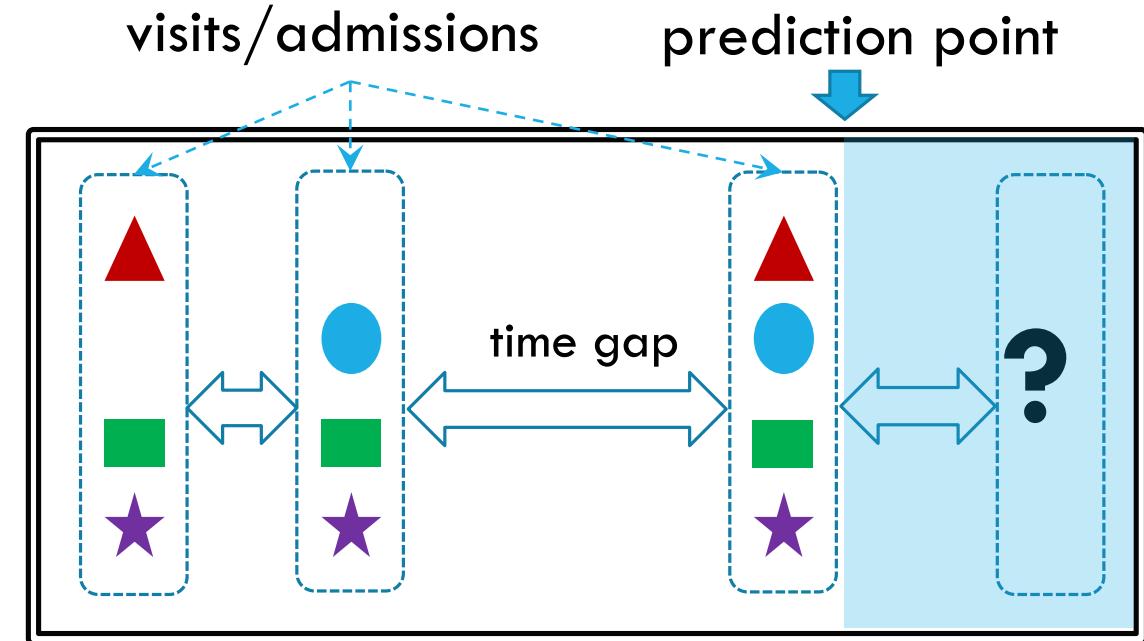
Irregular timing

Mixture of discrete codes and  
continuous measures

Complex interaction of diseases and  
care processes

Cohort of interest can be small (e.g.,  
<1K)

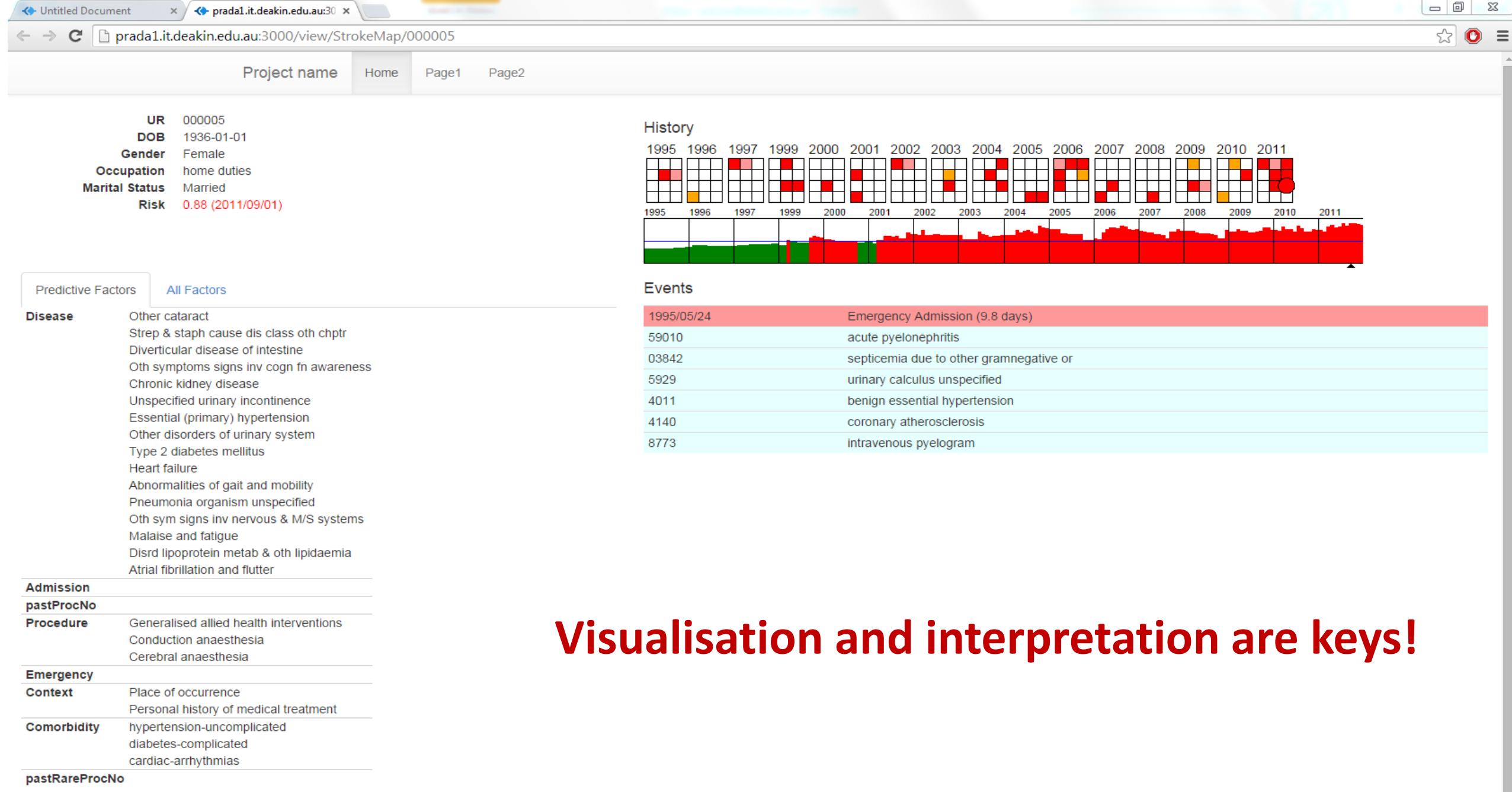
Rich domain knowledge & ontologies



Multimodalities: Text, physiological signals (e.g., EEG/ECG), images (e.g., MRI, X-ray, retina), genomics

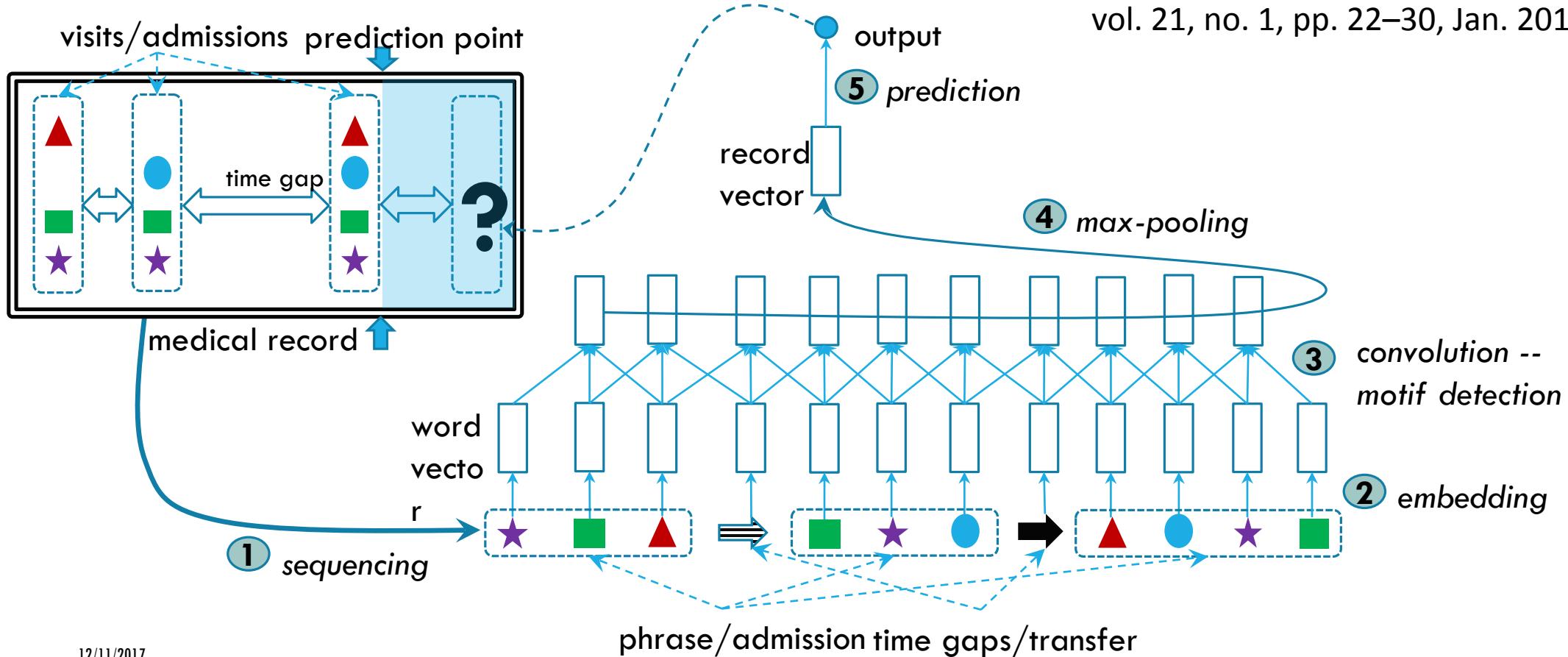
New modalities: social medial, wearable devices

**Explainability!**



# Deepr: CNN for repeated motifs and short sequences

#REF: Phuoc Nguyen et al., Deepr: A Convolutional Net for Medical Records, *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 22–30, Jan. 2017



# Deepr: Disease embedding & motifs detection

**E11 I48 I50**

Type 2 diabetes mellitus

Atrial fibrillation and flutter

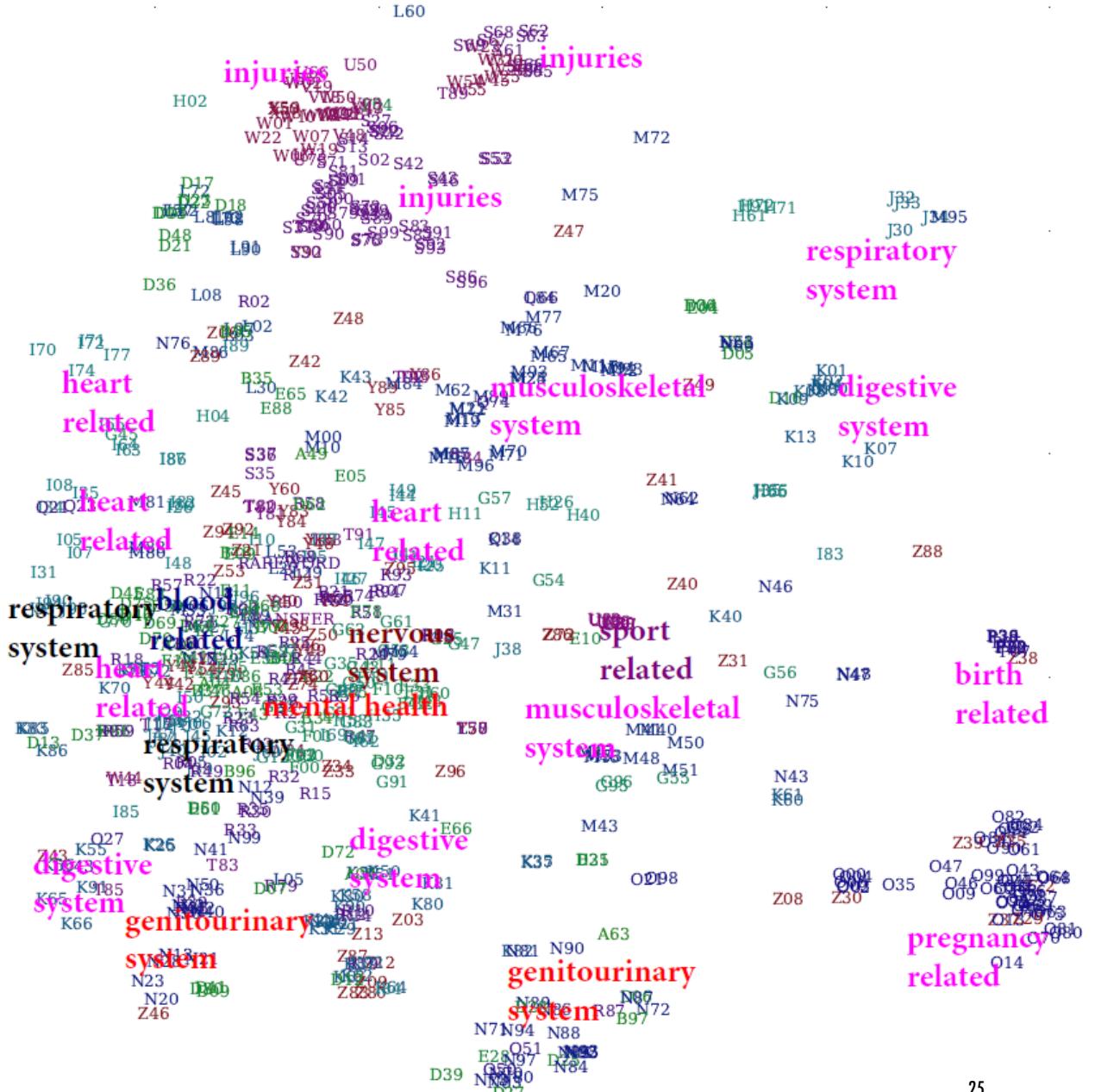
Heart failure

**E11 I50 N17**

Type 2 diabetes mellitus

Heart failure

Acute kidney failure



# DeepCare: intervened long-term memory of health

**Illness states are a dynamic memory process** → moderated by time and intervention

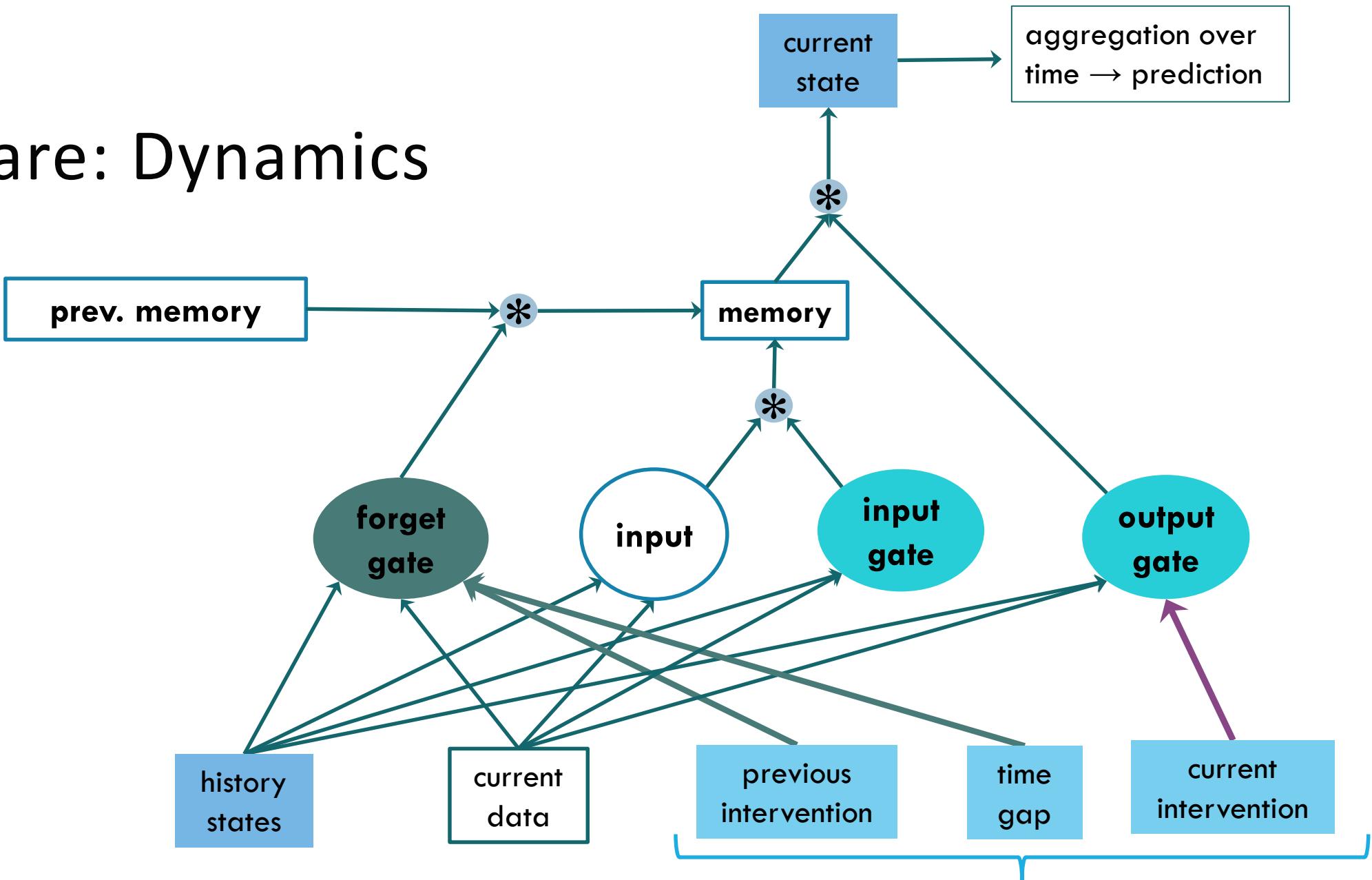
Discrete admission, diagnosis and procedure → vector embedding

Time and previous intervention → “forgetting” of illness

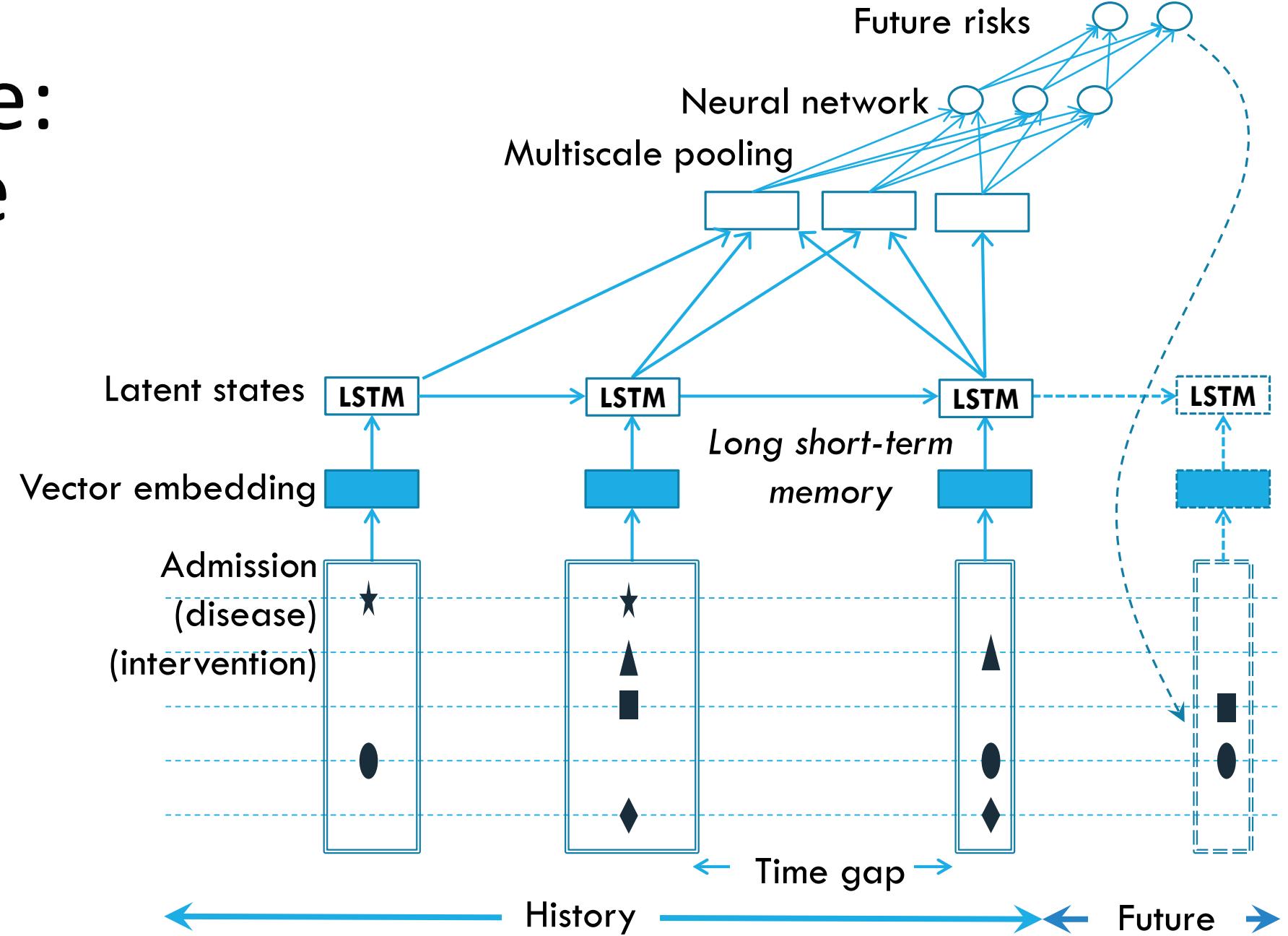
Current intervention → controlling the risk states

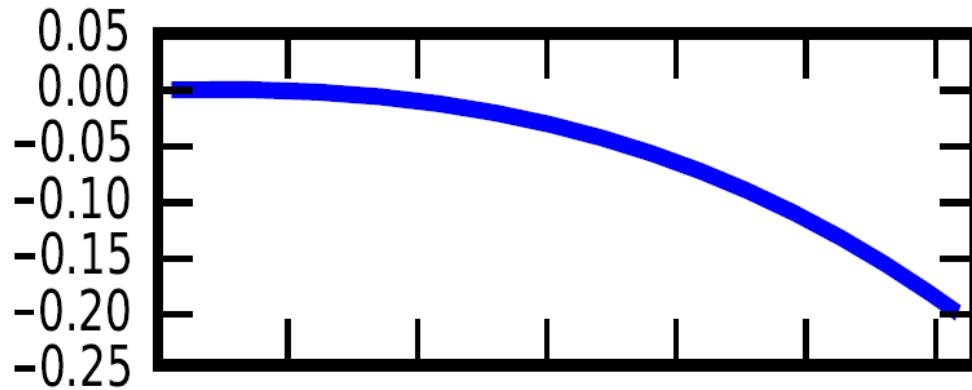
#REF: Trang Pham, et al., Predicting healthcare trajectories from medical records: A deep learning approach, *Journal of Biomedical Informatics*, April 2017, DOI: 10.1016/j.jbi.2017.04.001.

# DeepCare: Dynamics



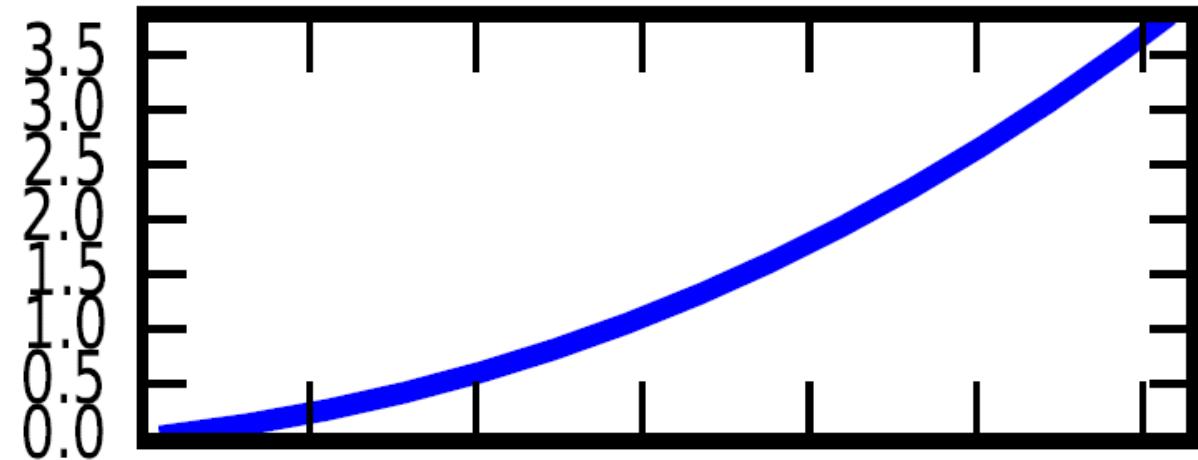
# DeepCare: Structure





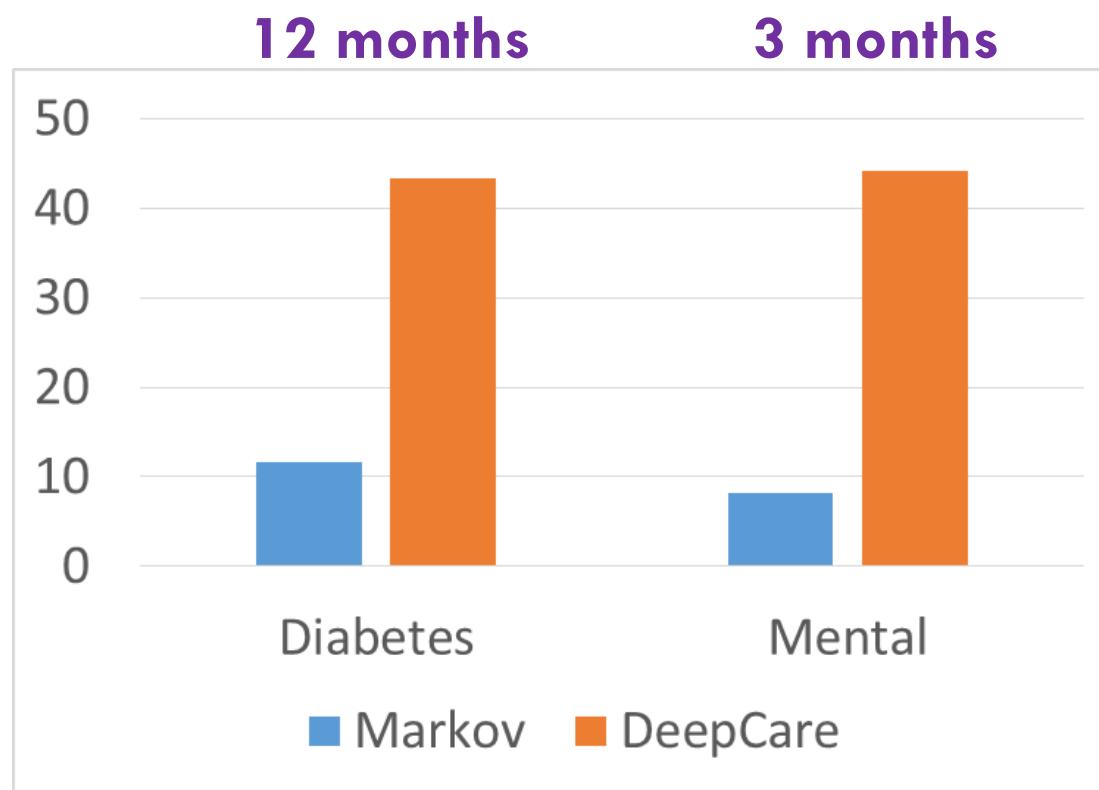
→ decreasing illness

→ Increasing illness

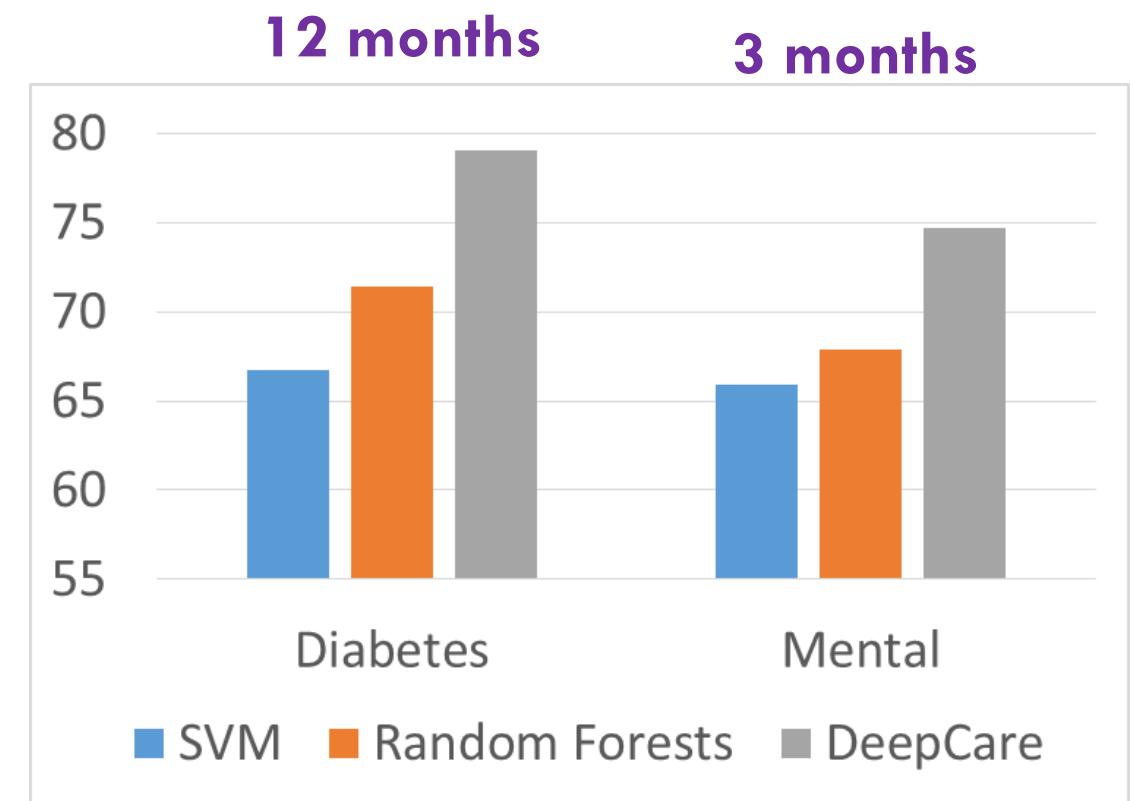


DeepCare: Two modes of forgetting as a function of time

# DeepCare: prediction results



Intervention recommendation (precision@3)



Unplanned readmission prediction (F-score)

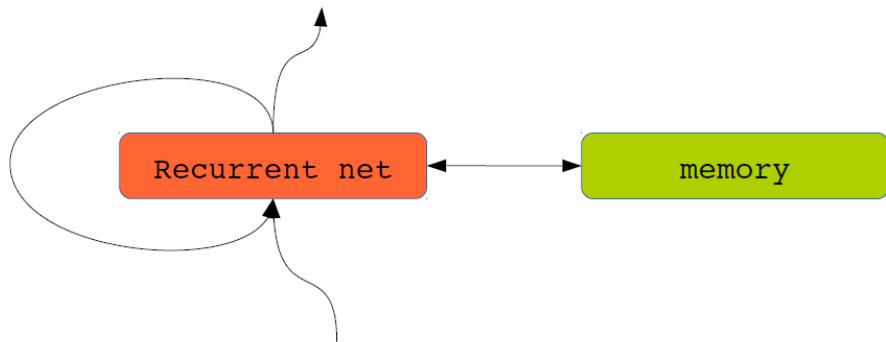
# Trajectories prediction

Generating a subset of treatments

Generating an entire health/care trajectory

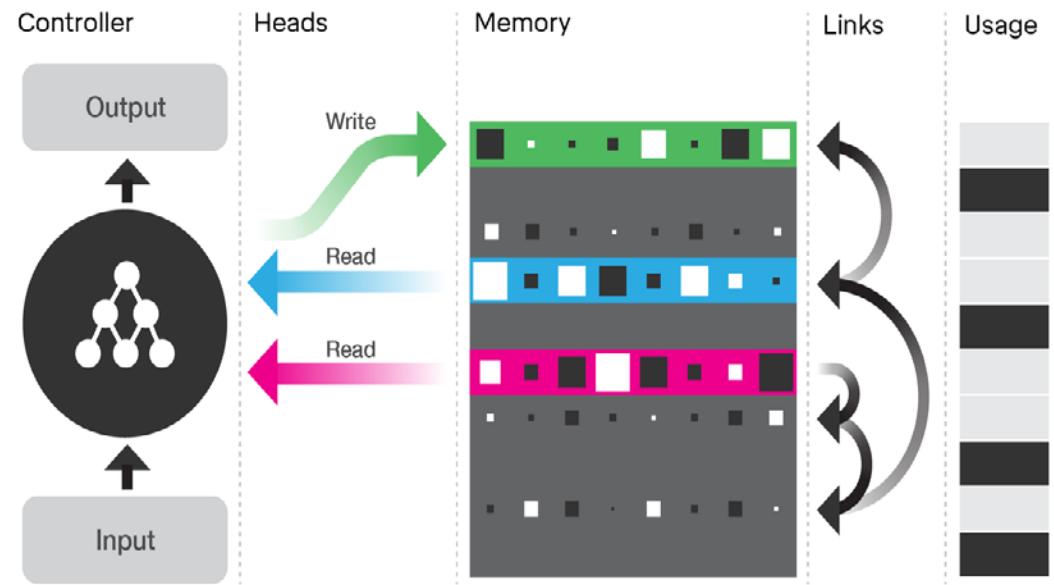
Challenges: global loss, meaningful evaluation metrics

A solution: Memory-augmented neural nets (MANN)



(LeCun, 2015)

Illustration of the DNC architecture



Source: [deepmind.com](http://deepmind.com)

#REF: Graves, Alex, et al. "Hybrid computing using a neural network with dynamic external memory." *Nature* 538.7626 (2016): 471-476.

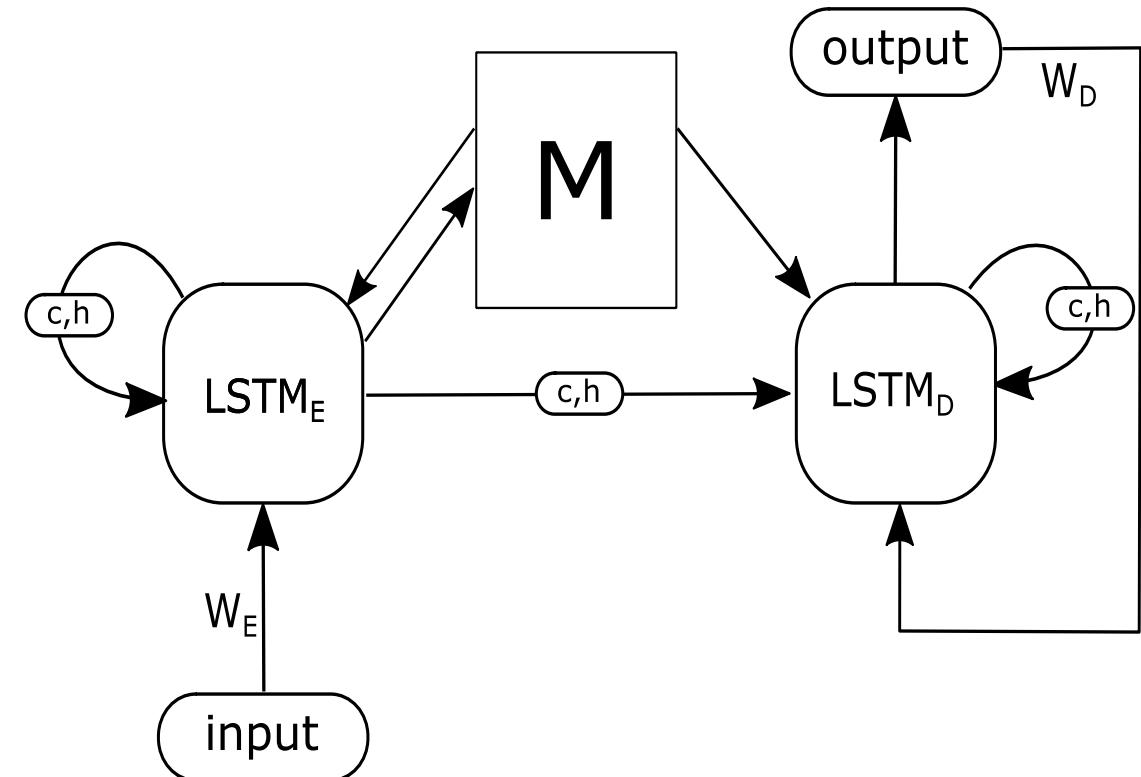
# Dual Controller Write-protected Memory-Augmented Neural Networks (DCw-MANN) (*Le et al, work in progress*)

Separate controllers for encoding and decoding

Memory is write-protected during decoding

Use memory as a replacement for:

- Attention
- Skip connection



# DCw-MANN

Model	Procedure Output		Drug Output	
	Recall	BLEU	Recall	BLEU
Logistic Regression	25.60	N/A	41.21	N/A
Random Forest	26.84	N/A	49.06	N/A
Seq2Seq	27.16	36.42	23.56	37.71
Seq2Seq with attention	27.33	36.83	24.82	39.70
DNC	28.28	37.69	58.16	61.72
DC-MANN	28.54	38.51	58.79	62.70
DCw-MANN	<b>30.36</b>	<b>39.74</b>	<b>59.76</b>	<b>63.80</b>

**Table 3.** Results on MIMIC-III dataset for procedure prediction and drug prescription. The higher the better.

# Modeling multiple disease-treatment interactions over time

Co-morbidity is the norm in modern medicine

Each hospital visit contains a set of diseases and a set of treatments

There are interactions between multi-diseases and multiple-treatments

**Algebraic view:** Health =  $RNN(\text{Illness} - \text{Intervention})$

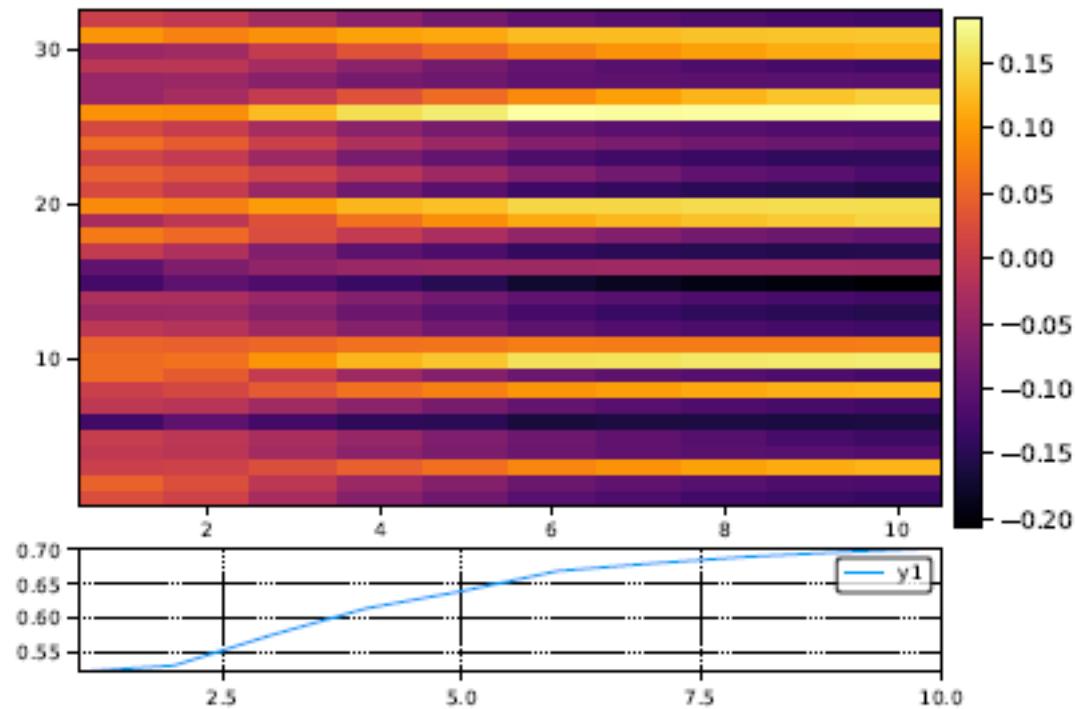
$$v_t = \rho(\Delta) \quad \text{where} \quad \Delta = d_t - p_t$$

$$f_e(S) \leftarrow \frac{\bar{e}_S}{\epsilon + \|\bar{e}_S\|} \quad \text{where} \quad \bar{e}_S = \max(0, \sum_{i \in S} e_i)$$

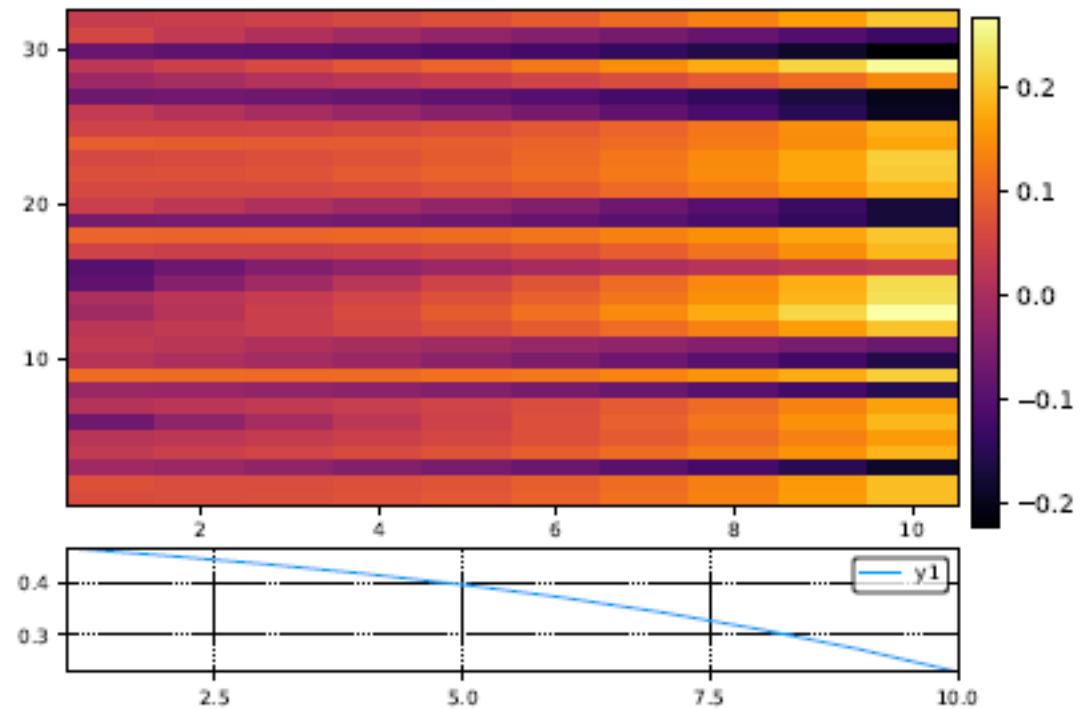
#REF: P Nguyen, T Tran, S Venkatesh, “Finding Algebraic Structure of Care in Time: A Deep Learning Approach”, NIPS Workshop on Machine Learning for Health (ML4H), 2017

# Results (AUC)

Method	Diabetes	Mental health
BoW+LR	0.673	0.705
Deepr [14]	0.680	0.714
<b>MDMTP+LTSM</b>	<b>0.718</b>	<b>0.726</b>
<b>MDMT+LSTM</b>	<b>0.701</b>	<b>0.730</b>



(a) Worsening progression ( $P = 0.70$ )



(b) Improving progression ( $P = 0.23$ )

# Big room: Towards personalized healthcare

Medical practice as recommender systems (Xavier Amatriain, healthcare Recsys Workshop, Como, 2017)

Clinical Practice Guides are not personalized

Research done on “homogeneous”, healthy subjects

It is very hard for doctors to “manually” personalize their “recommendations”

# Other refs

Choi, Edward, et al. "Doctor ai: Predicting clinical events via recurrent neural networks." *Machine Learning for Healthcare Conference*. 2016.

Lipton, Zachary C., et al. "Learning to diagnose with LSTM recurrent neural networks." *arXiv preprint arXiv:1511.03677*(2015).

Harutyunyan, Hrayr, et al. "Multitask Learning and Benchmarking with Clinical Time Series Data." *arXiv preprint arXiv:1703.07771* (2017).

Choi, Edward, et al. "RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism." *Advances in Neural Information Processing Systems*. 2016.

# Agenda

Topic 1: Introduction (20 mins)

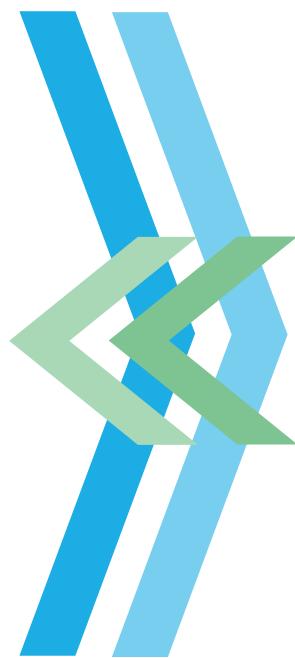
Topic 2: Brief review of deep learning (25 mins)

- Classic architectures
- Capsules
- Graphs
- Memory-augmented nets

Topic 3: Genomics (25 mins)

- Nanopore sequencing
- Genomics modelling

QA (10 mins)



**Break (20 mins)**

Topic 4: Biomedical imaging (15 mins)

- Cellular imaging
- Diagnostics imaging
- EEG/ECG

Topic 5: Healthcare (25 mins)

- Time-series of physio measures
- Trajectories prediction

Topic 6: Generative biomed (30 mins)

- Few-shot learning
- Generative models
- Drug design
- Future outlook

QA (10 mins)

# Few-shot deep learning

Lots of biomedical problems are data poor

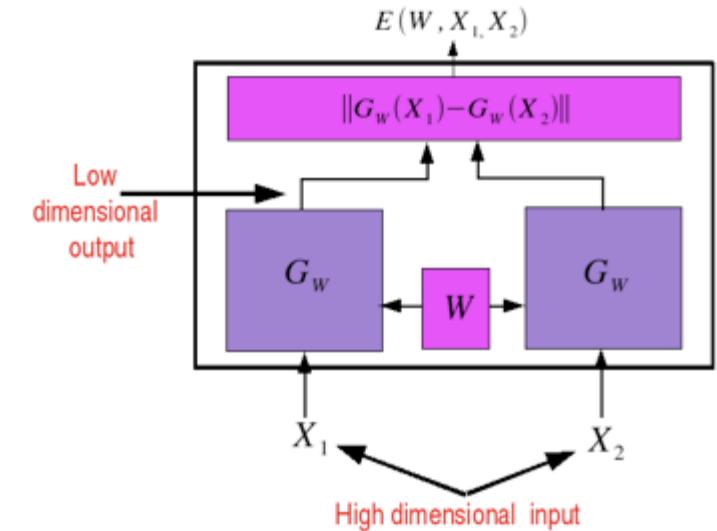
- Rare drugs
- Rare diseases

Distance metrics learning (DML) methods

- Learn to pull any pair of the similar data points, and push the dissimilar
- Well-known methods: **Siamese networks**

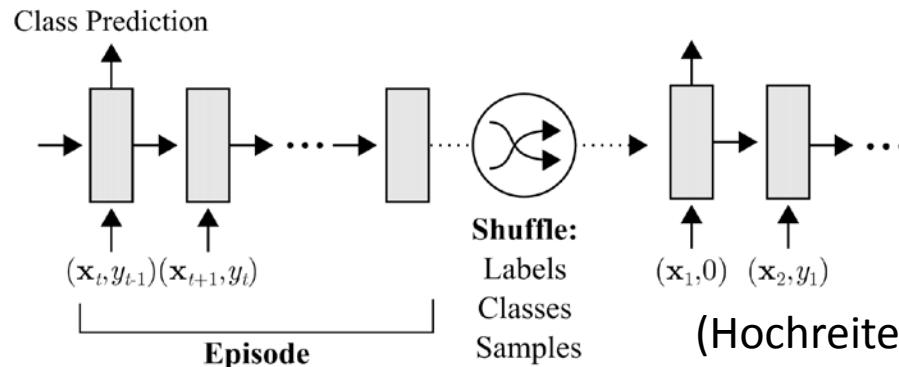
Meta-learning strategies

- Tasks are presented in sequence
- New tasks can borrow from similar prior tasks

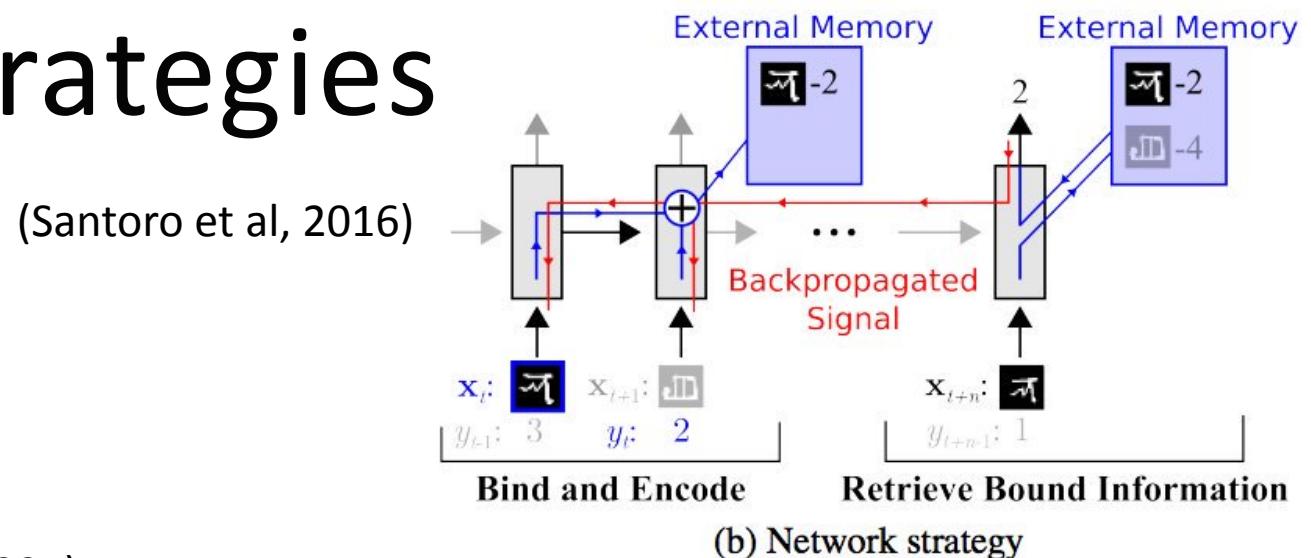


#REF: Chopra, Sumit, Raia Hadsell, and Yann LeCun. "Learning a similarity metric discriminatively, with application to face verification." *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.* Vol. 1. IEEE, 2005.

# Meta-learning strategies



(Hochreiter et al., 2001)



(b) Network strategy

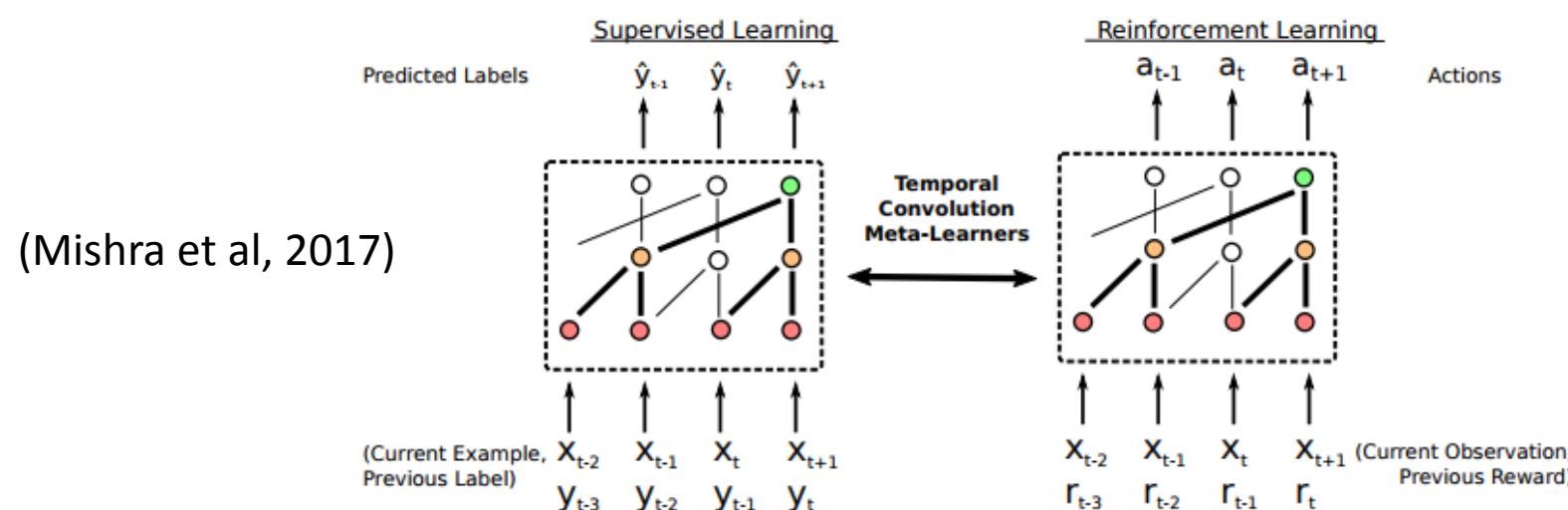


Figure 1: Overview of our temporal-convolution-based meta-learner (TCML). The same class of model architectures can be applied to both supervised and reinforcement learning.

# Generative models

## Many applications:

- Text to speech
- **Simulate data that are hard to obtain/share in real life (e.g., healthcare)**
- Generate meaningful sentences conditioned on some input (foreign language, image, video)
- Semi-supervised learning
- Planning

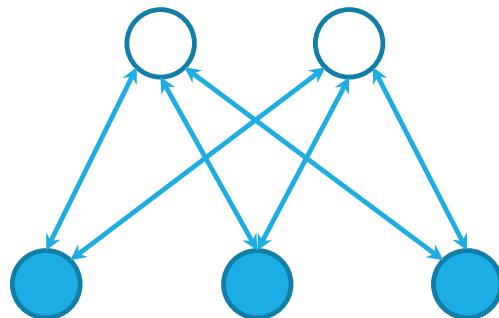
$$\mathbf{v} \sim P_{model}(\mathbf{v})$$

$$P_{model}(\mathbf{v}) \approx P_{data}(\mathbf{v})$$

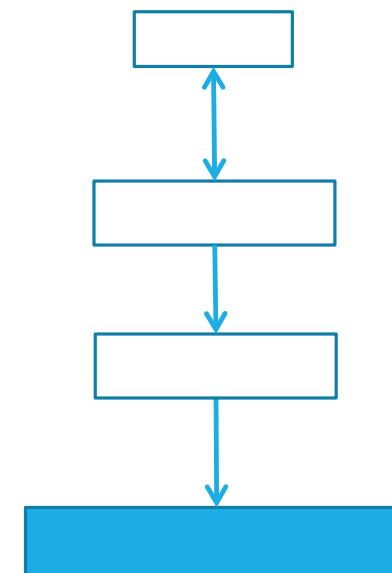
# A family: RBM → DBN → DBM

$$p(\mathbf{v}, \mathbf{h}; \psi) \propto \exp [-E(\mathbf{v}, \mathbf{h}; \psi)]$$

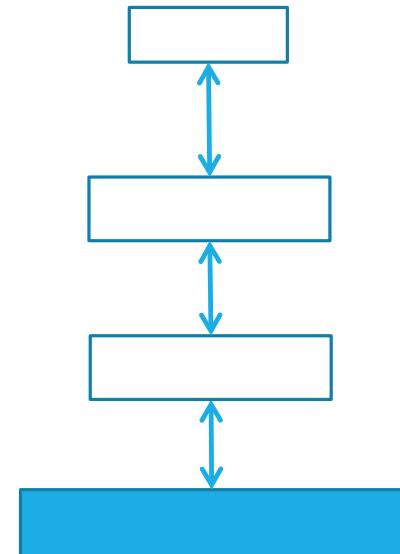
*energy*



**Restricted Boltzmann Machine**  
(~1994, 2001)



**Deep Belief Net**  
(2006)

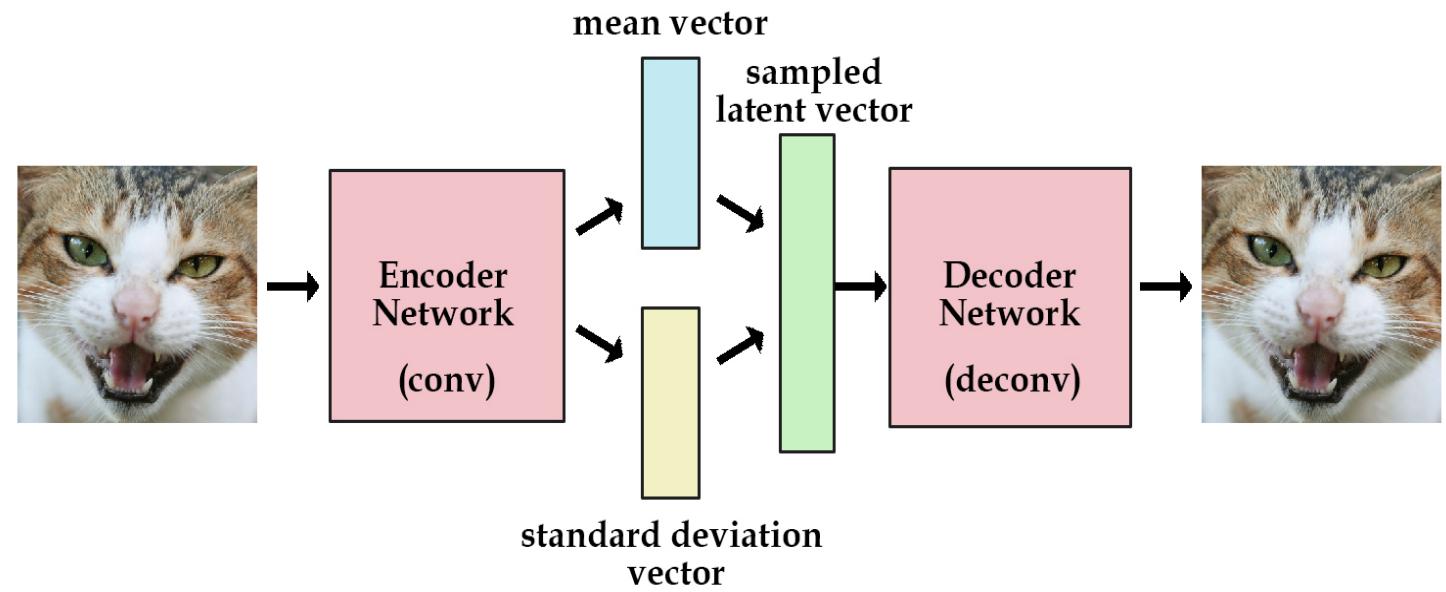
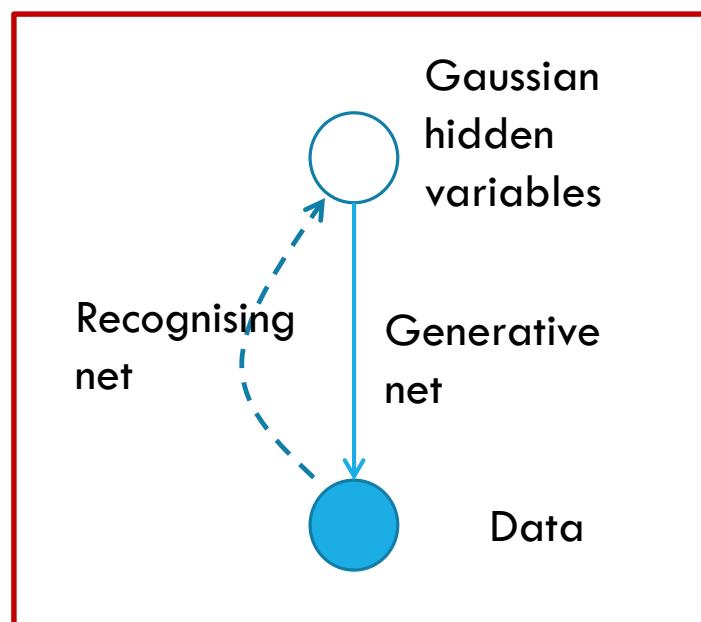


**Deep Boltzmann Machine**  
(2009)

# Variational Autoencoder

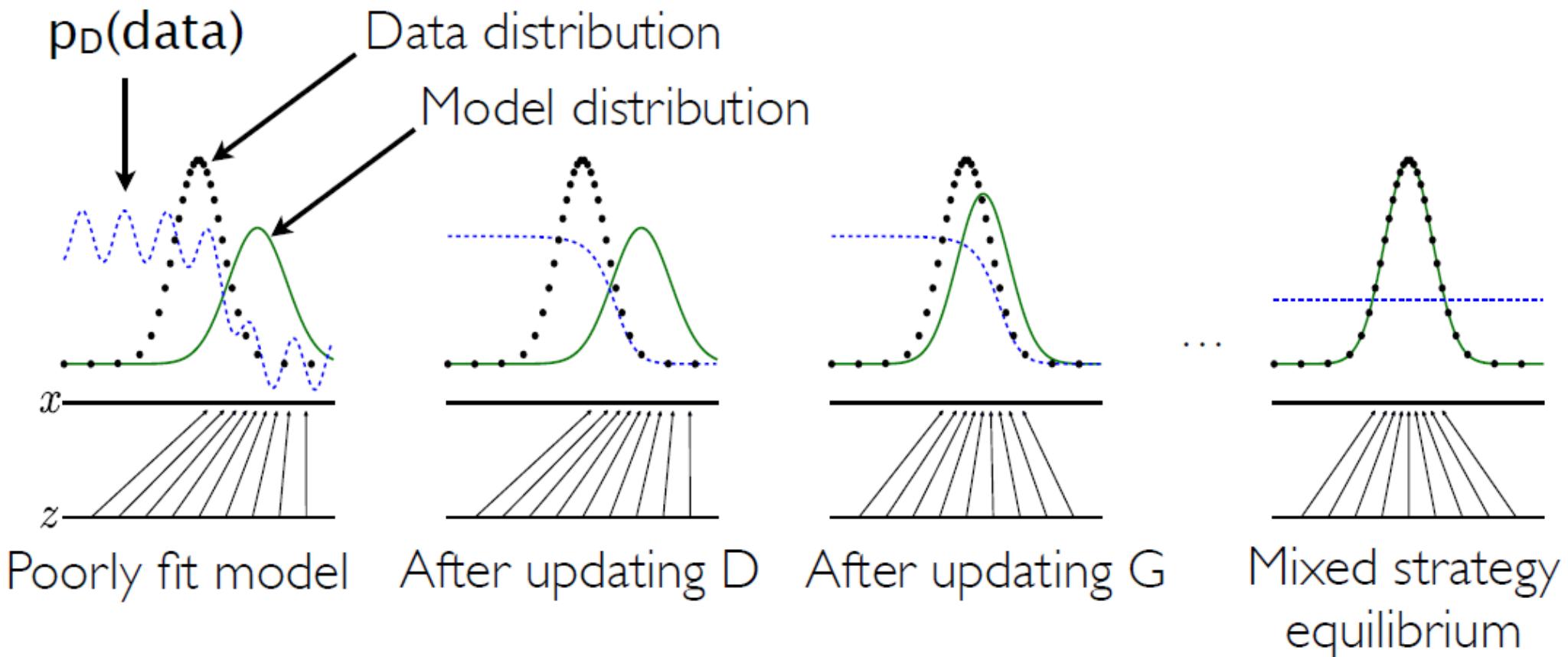
(Kingma & Welling, 2014)

Two separate processes: generative (hidden → visible) versus recognition (visible → hidden)



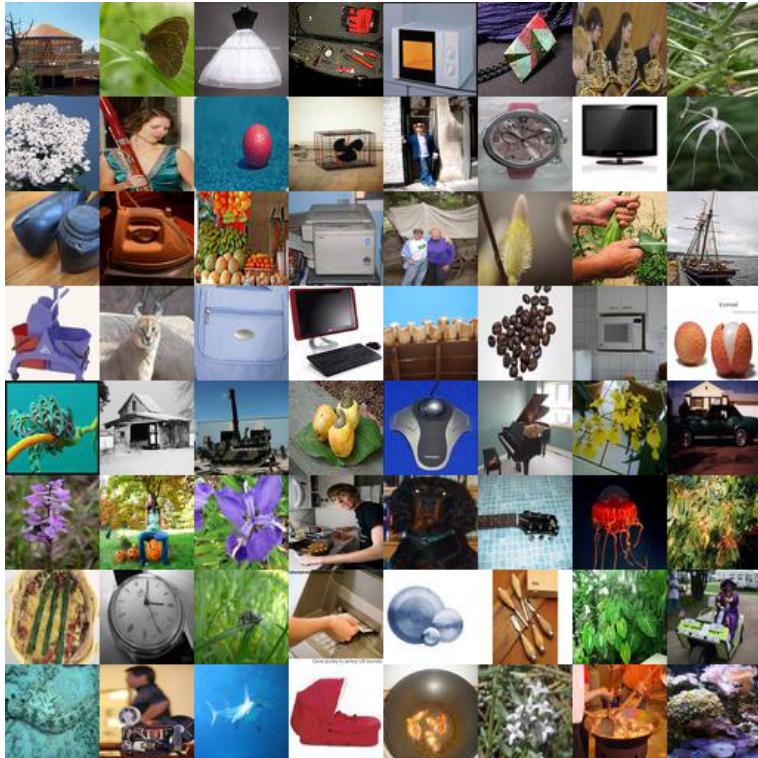
# GAN: implicit density models

(Adapted from Goodfellow's, NIPS 2014)



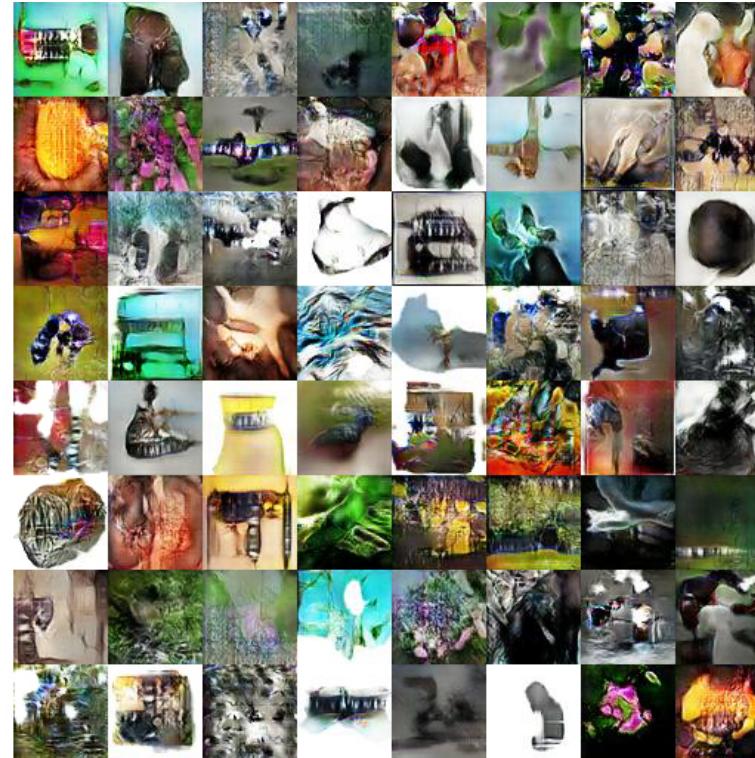
# GAN: Generated Samples

Some high quality pictures generated by GAN



Real

12/11/2017



Generated

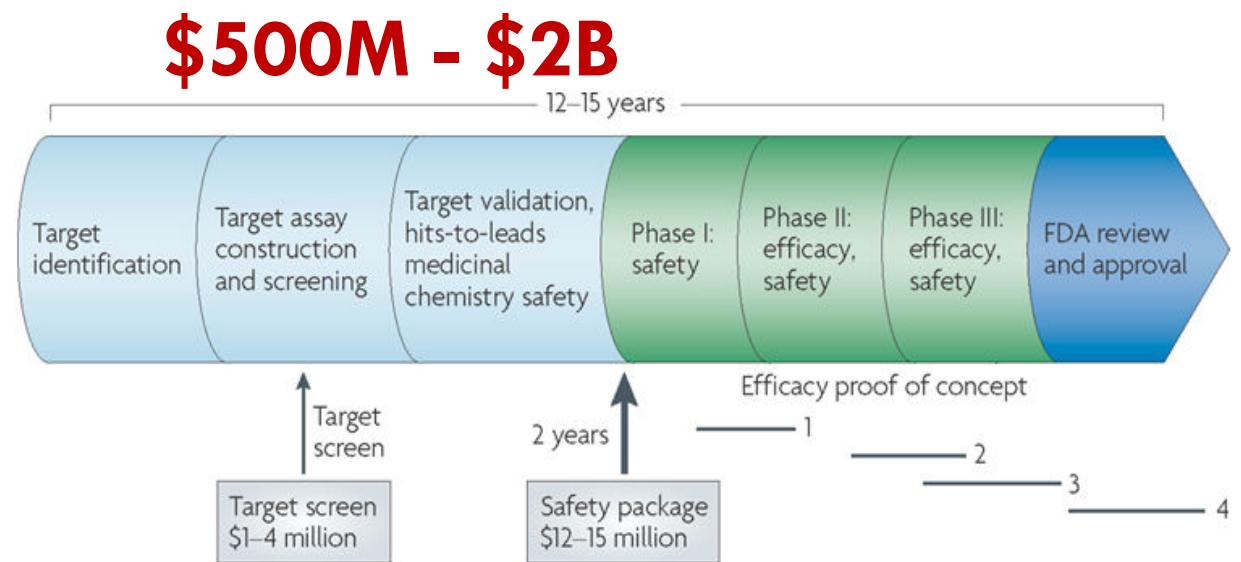
<http://kvfrans.com/generative-adversarial-networks-explained/>

# Generative deep learning for drug discovery

Predicting bioactivities from molecules

Drug representation,  
unsupervised learning from graphs

Generate from bioactivities to molecular graphs



Nature Reviews | Drug Discovery

#REF: Roses, Allen D. "Pharmacogenetics in drug discovery and development: a translational perspective." *Nature reviews Drug discovery* 7.10 (2008): 807-817.

# Combinatorial chemistry

Generate variations on a template

Returns a list of molecules from this template that

- Bind to the pocket with good pharmacodynamics?
- Have good pharmacokinetics?
- Are synthetically accessible?

#REF: Talk by Chloé-Agathe Azencott titled “Machine learning for therapeutic research”, 12/10/2017

# First step: Map molecule → drug properties (binding/acting)

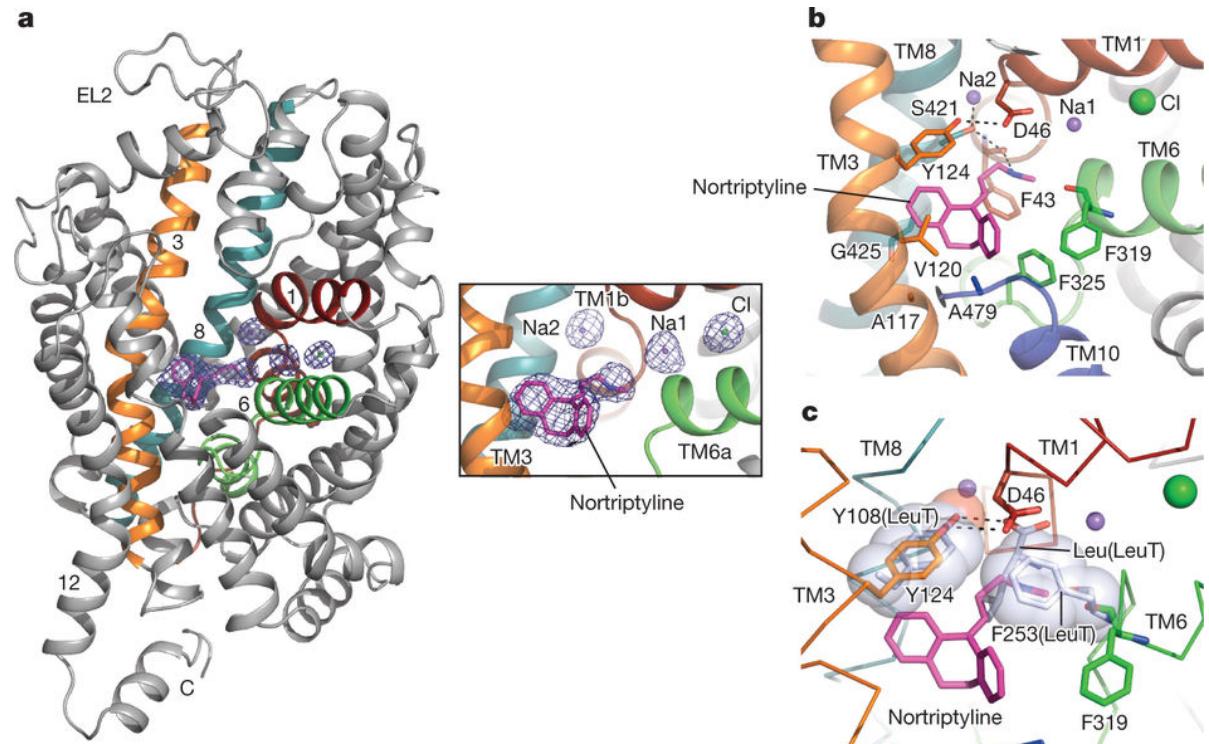
Drugs are small bio-molecules

Traditional techniques:

- Graph kernels (ML)
- Molecular fingerprints (Chemistry)

Modern techniques

- Molecule as graph: atoms as nodes, chemical bonds as edges



#REF: Penmatsa, Aravind, Kevin H. Wang, and Eric Gouaux. "X-ray structure of dopamine transporter elucidates antidepressant mechanism." *Nature* 503.7474 (2013): 85-90.

# Molecular fingerprints

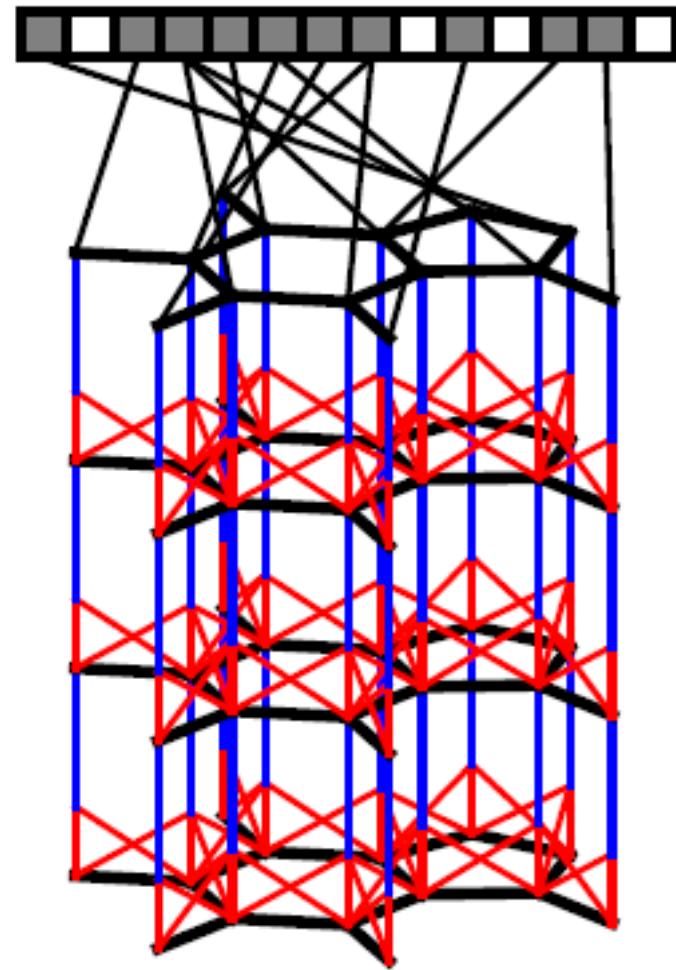
---

## Algorithm 1 Circular fingerprints.

---

```
1 : Input: molecule, radius  $R$ , fingerprint length  $S$ 
2 : Initialize: fingerprint vector  $x \leftarrow \mathbf{0}_S$ 
3 : foreach atom  $a$  in molecule
4 :    $r_a \leftarrow q(a)$  #extract initial atom features
5 : for  $L = 1$  to  $R$  #loop through layers
6 :   foreach atom  $a$  in molecule
7 :      $r_1 \dots r_N = \text{neighbors}(a)$ 
8 :      $v \leftarrow [r_a, r_1, \dots, r_N]$  #combine neighbor features
9 :      $r_a \leftarrow \text{hash}(v)$  #refine atom features
10:     $i \leftarrow \text{mod}(r_a, S)$  #convert to index
11:     $x_i \leftarrow 1$  #Write 1 (indicator) at index
12: Return: binary vector  $x$ .
```

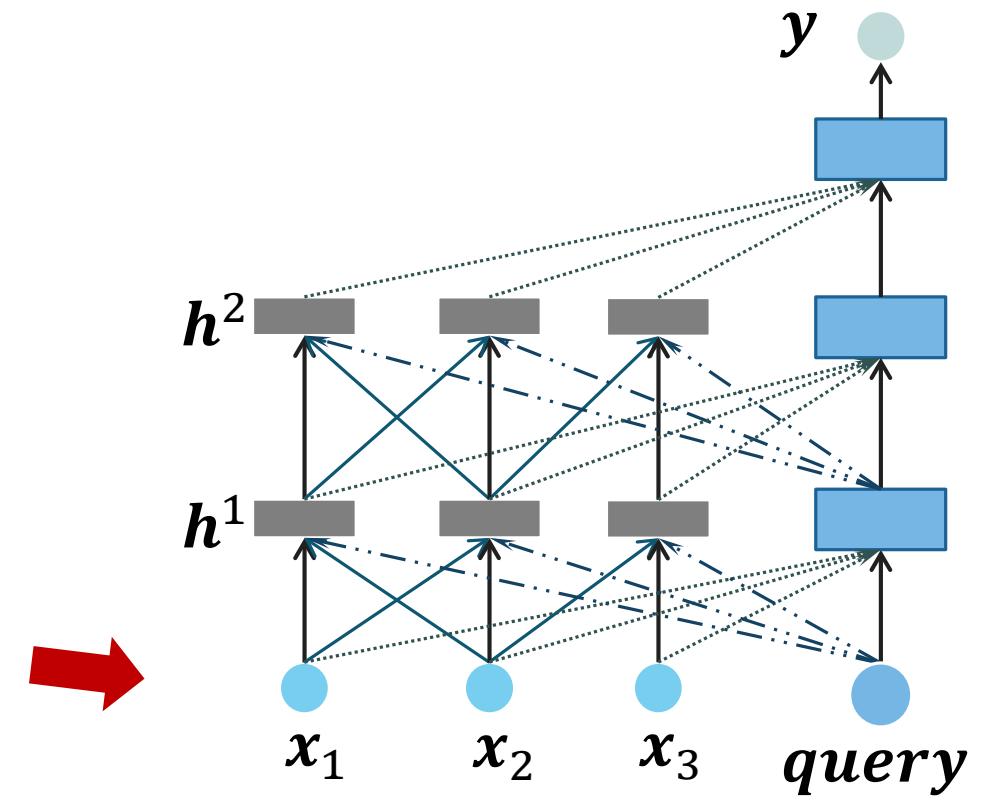
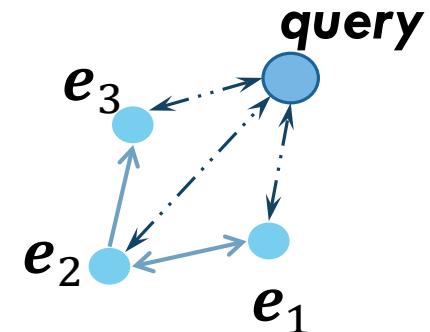
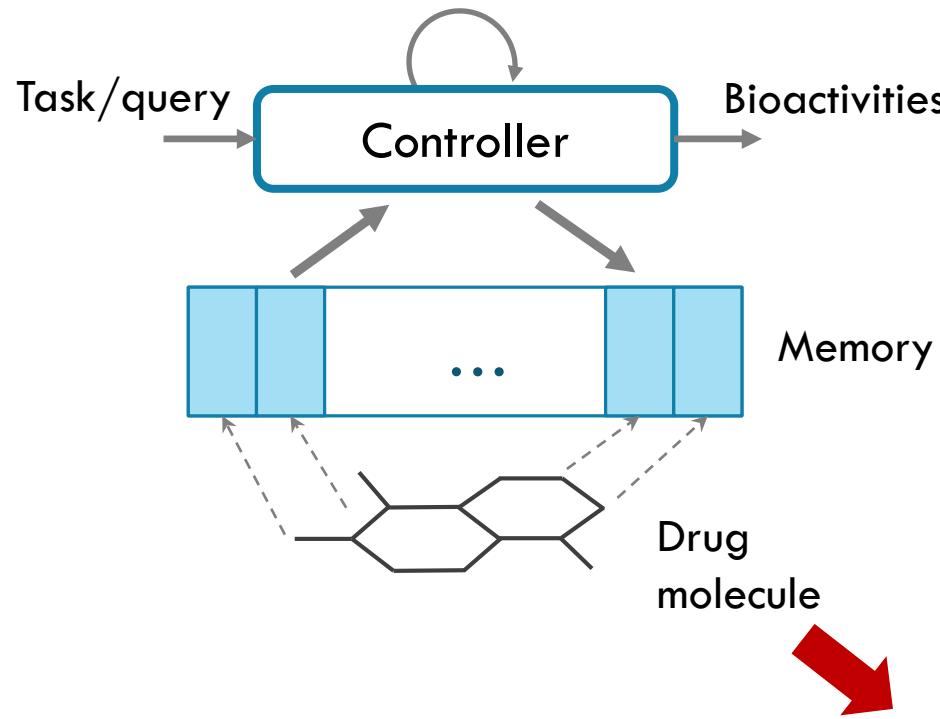
---



#REF: Duvenaud, David K., et al.  
"Convolutional networks on graphs for learning molecular fingerprints." *Advances in neural information processing systems*. 2015.

# Graph memory networks

(Pham et al, 2017, work in progress)



# Graph memory networks: Results

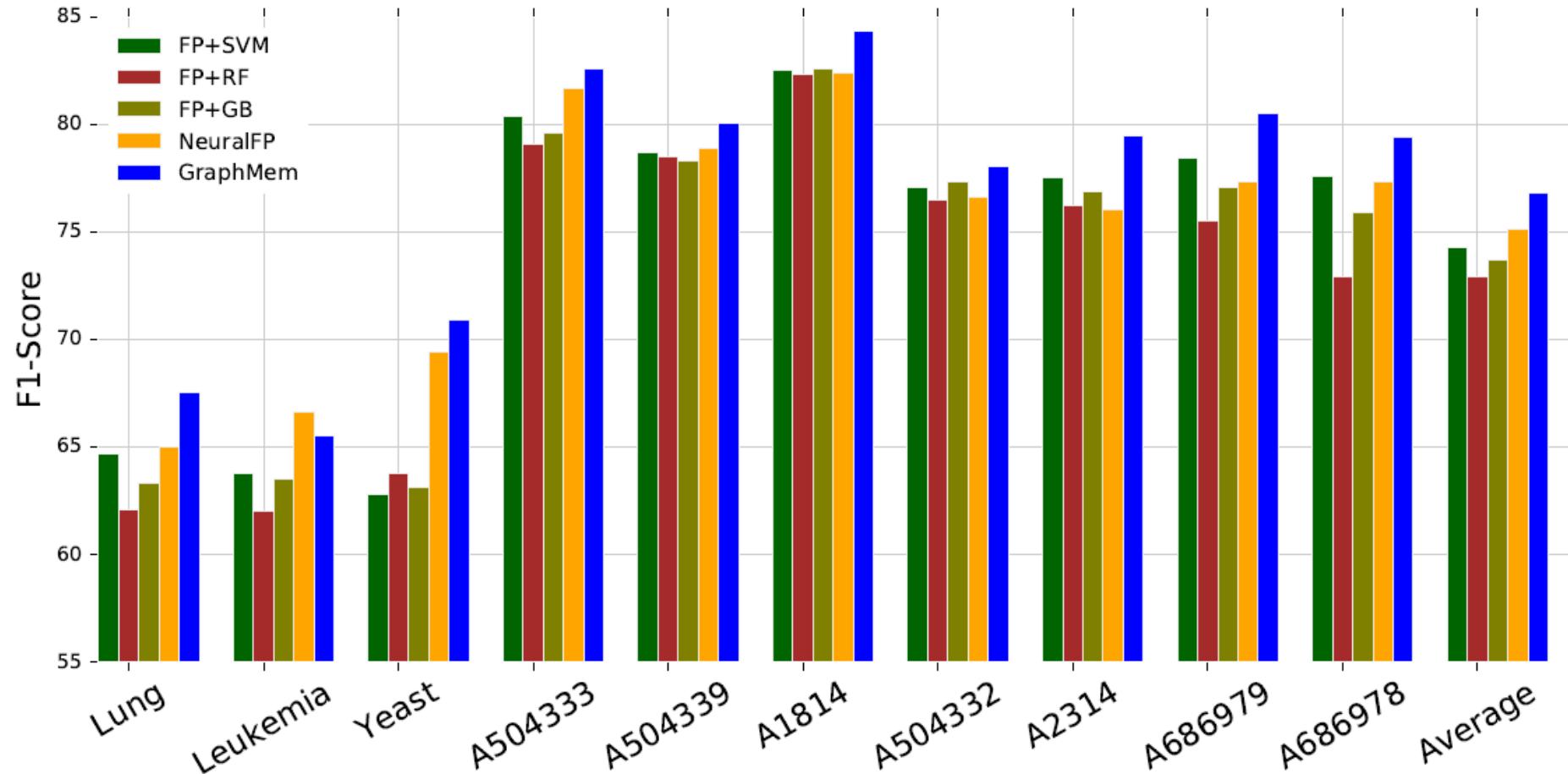
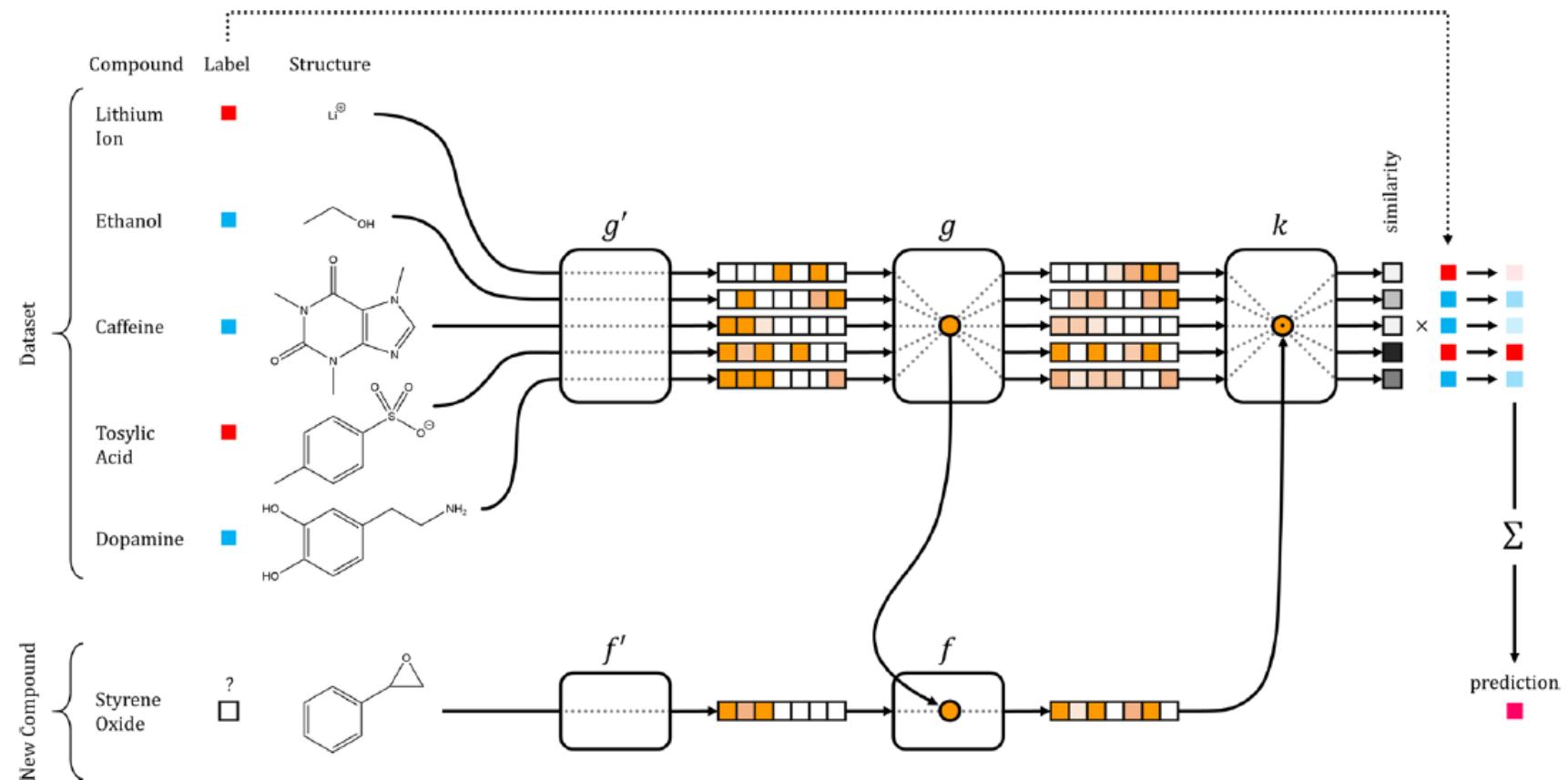


Figure 2: F1-score (%) for NCI datasets. FP = Fingerprint; RF = Random Forests; GBM = Gradient Boosting Machine. Best view in color.

# One-shot learning for drug discovery

#REF: Altae-Tran, Han, et al.  
"Low Data Drug Discovery  
with One-Shot  
Learning." *ACS central  
science* 3.4 (2017): 283-293.



# Drug design and generation

We now have methods for compute bioactivities of a drug molecule

**We need a reverse method to generate drug molecules from desirable bioactivities**

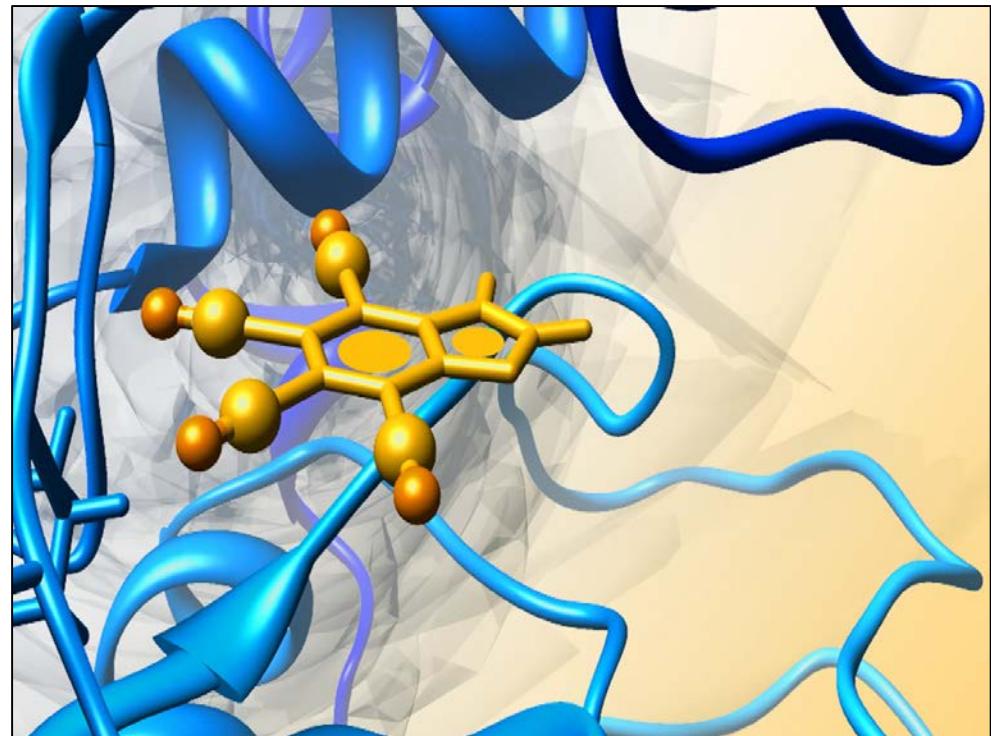
The space of drugs is estimated to be  $1e+23$  to  $1e+60$

- Only  $1e+8$  substances synthesized thus far.

It is impossible to model this space fully.

The current technologies are not ready for graph generations.

But approximate techniques do exist.



Source: pharmafactz.com

# The trick: Molecule → string

Using SMILES representation of drug, to convert a molecular graph into a string

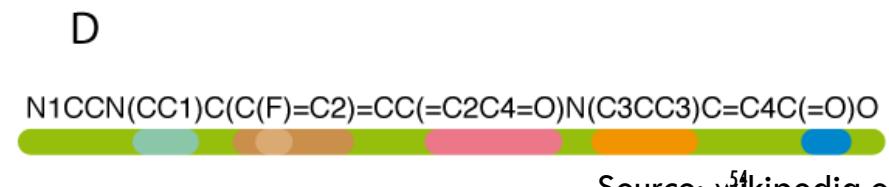
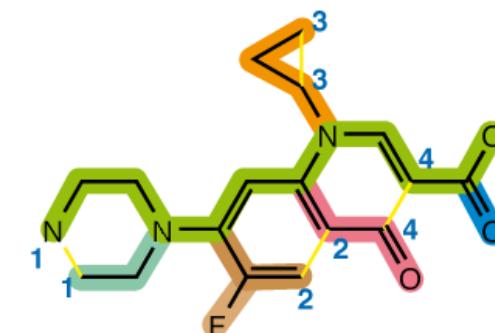
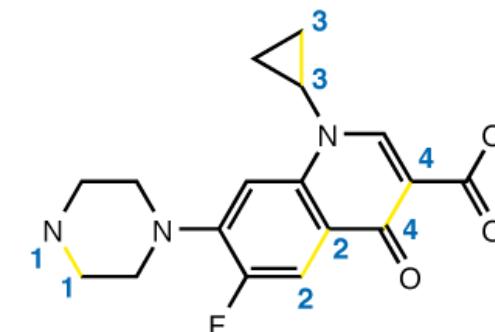
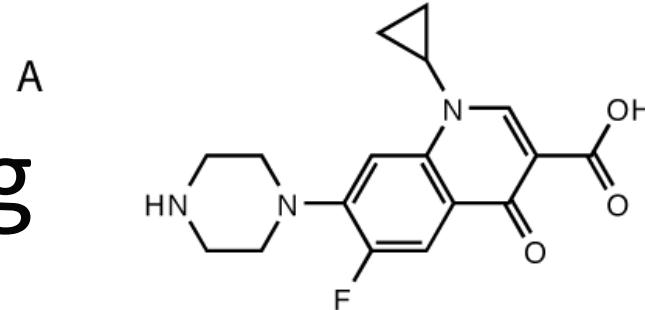
- SMILES = Simplified Molecular-Input Line-Entry System

Then using sequence-to-sequence + VAE/GAN to model the continuous space that encodes/decodes SMILES strings

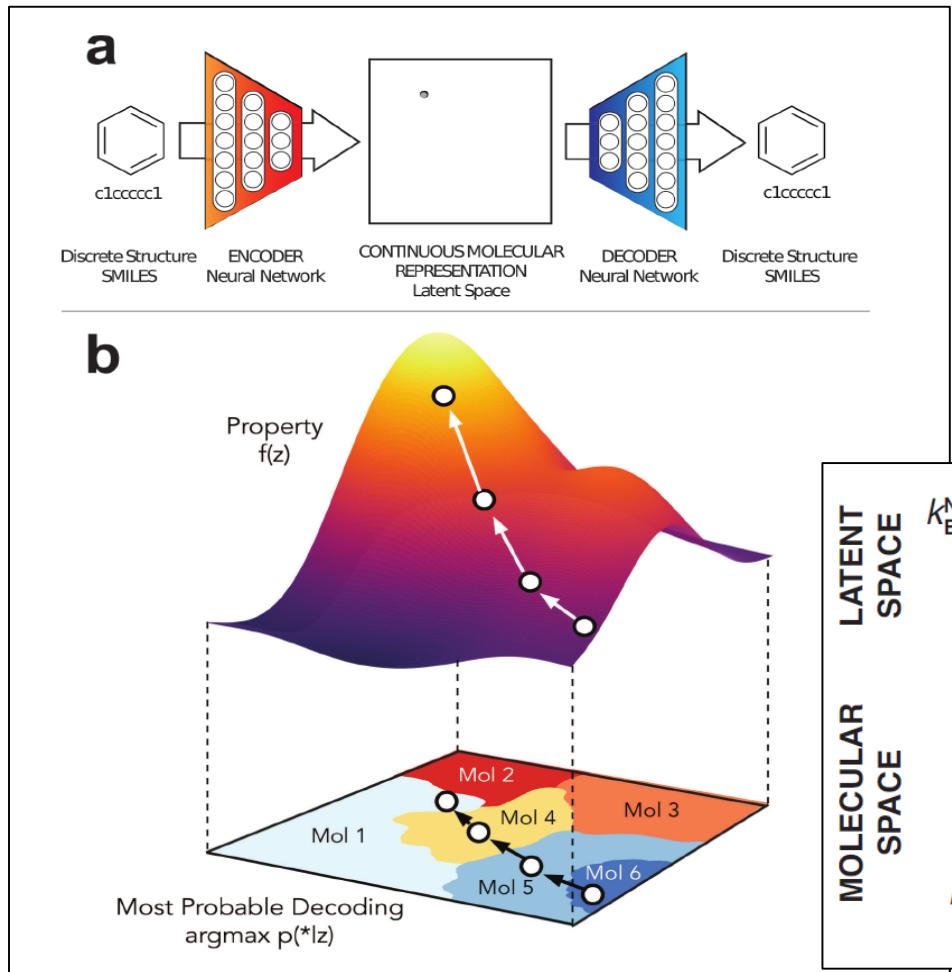
- Allow easy optimization on the continuous space
- Problem: String → graphs is not unique!

A better way is to encode/decode graph directly.

#REF: Gómez-Bombarelli, Rafael, et al. "Automatic chemical design using a data-driven continuous representation of molecules." *arXiv preprint arXiv:1610.02415* (2016).

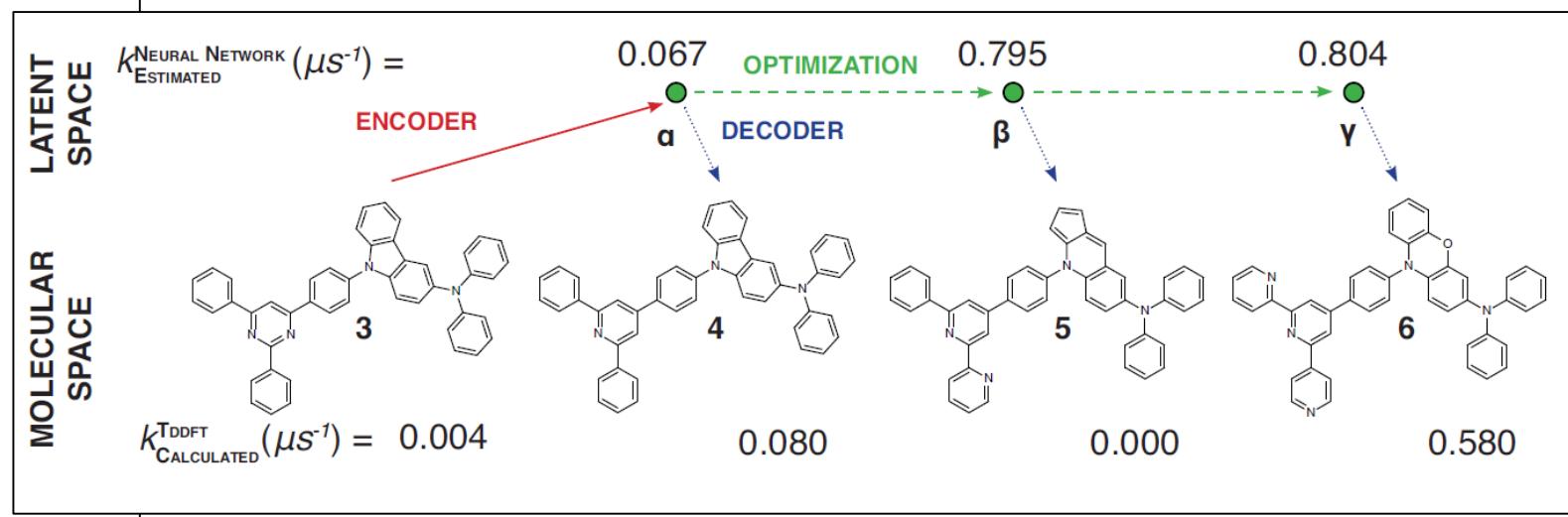


# VAE for drug space modelling



Uses VAE for sequence-to-sequence.

#REF: Bowman, Samuel R., et al. "Generating sentences from a continuous space." *arXiv preprint arXiv:1511.06349* (2015).



# More refs

Segler, Marwin HS, et al. "Generating focussed molecule libraries for drug discovery with recurrent neural networks." *arXiv preprint arXiv:1701.01329* (2017).

Kusner, Matt J., Brooks Paige, and José Miguel Hernández-Lobato. "Grammar Variational Autoencoder." *arXiv preprint arXiv:1703.01925* (2017).

Gupta, Anvita, et al. "Generative Recurrent Networks for De Novo Drug Design." *Molecular Informatics* (2017).

# Living in the future: AI for health care

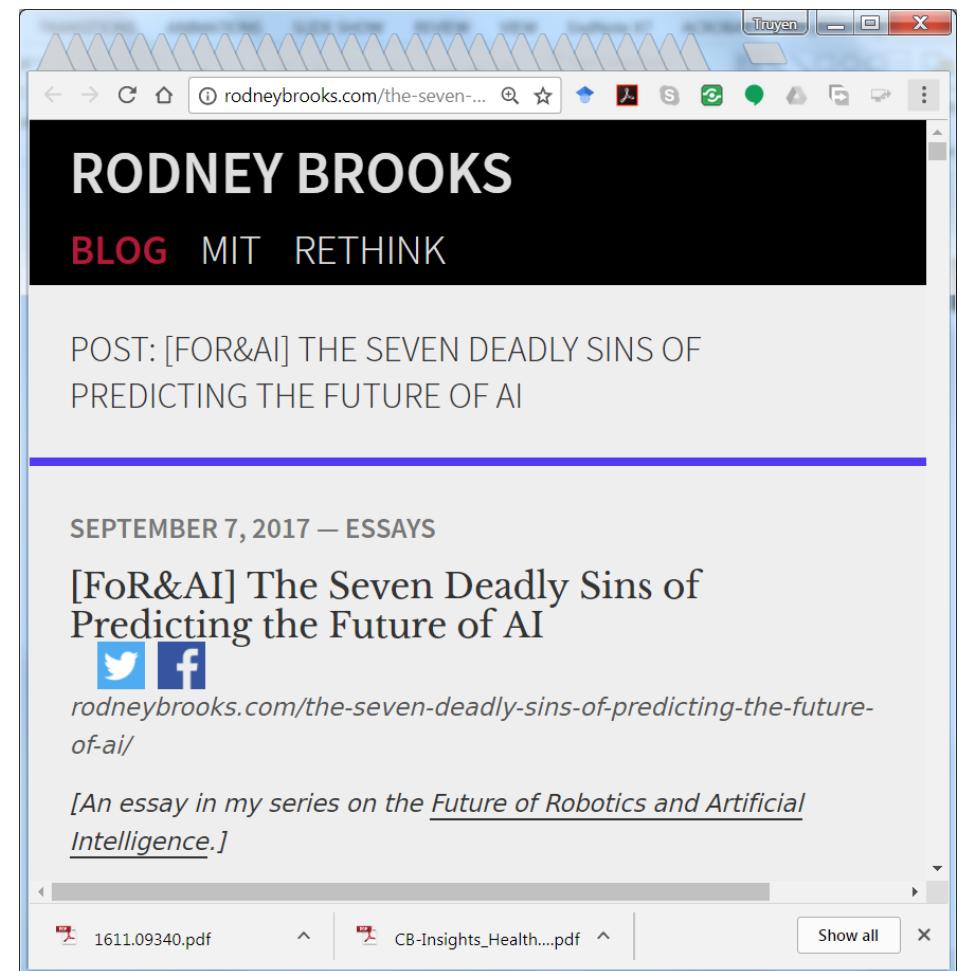
Some speculations (by me):

- <https://letdataspeak.blogspot.com.au/2017/02/living-in-future-deep-learning-for.html>

Bear in mind that anything beyond 5 years are nearly impossible to predict!

Kai-Fu Lee's vision:

- Wave 1: Internet data (→ PubMed, social media)
- **Wave 2: Business data (→EMR)**
- **Wave 3: Digitalize the physical world (→Drugs)**
- Wave 4: Full automation (→ Robot surgeons, GPs)



# Toward personalize medicine

Will this patient response to that treatment?

Can we find the best treatment for a patient?

Which biomarkers predict the patient's response?

Sound familiar to Recommender Systems (patient = user, treatment = item)?

Drug repurposing

Drug-protein interactions

Drug-drug interactions

Automating gene editing

One set of drugs per patient?

#REF: Talk by Chloé-Agathe Azencott titled “Machine learning for therapeutic research”, 12/10/2017

# Towards a dialog system → Replace GP?

## Leveraging existing knowledge

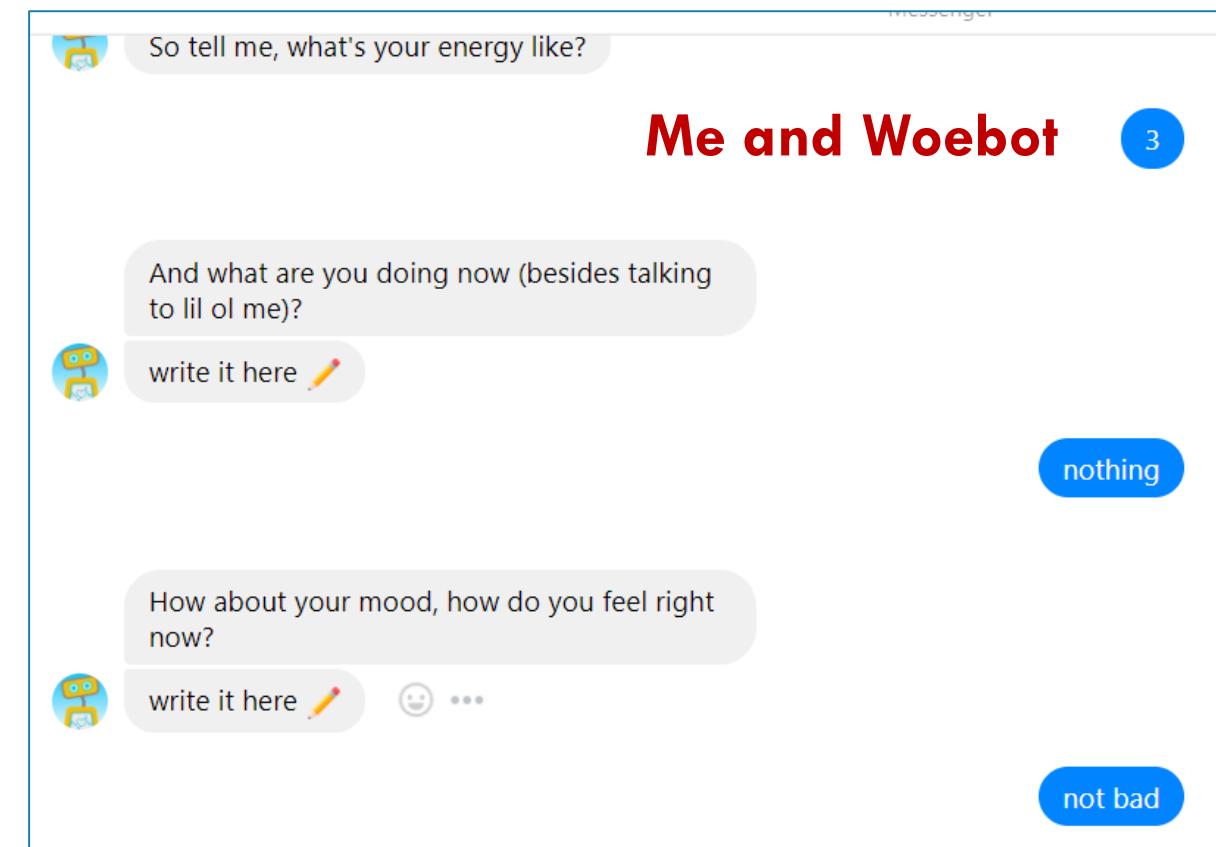
- Medical knowledge bases
- Medical texts
- Probably needs to build knowledge bases from text

## Personalizing through EMRs

- Learn from hospitals data

Ask right questions → Finding answers from databases → Generating dialog

Never ending learning (NEL).



# Other rooms for deep learning

## -omics

- Gene expression
- Proteomics

## Neuroscience

- Models for spike trains
- Deep learning for connectomics
- Neuroscience-inspired DL

## Biomedical NLP

- Classical NLP
- Social media
- Knowledge graphs

## Wearables

- Tracking the state of physical and mental health
- Lifestyle management & monitoring

## Health Insurance

- Future illness/spending prediction
- Proactive prevention programs
- **WARNING:** Working for insurance companies does raise ethical concerns!

## Nutrition

- Mobile phone vision → calories

## Explainable AI

- Seeing through the black-box, e.g., visualization, motifs
- Explainable architectures that use biological mechanisms and medical ontologies
- Dual architecture: predictor & explainer



Thank you!

We're hiring

PhD & Postdocs

*truyen.tran@deakin.edu.au*