



Discovery



Diagnosis



Prognosis



Care

Deep Learning for Biomedical Discovery and Data Mining II



Truyen Tran
Deakin University

Melbourne, June 2017



truyen.tran@deakin.edu.au



truyentran.github.io



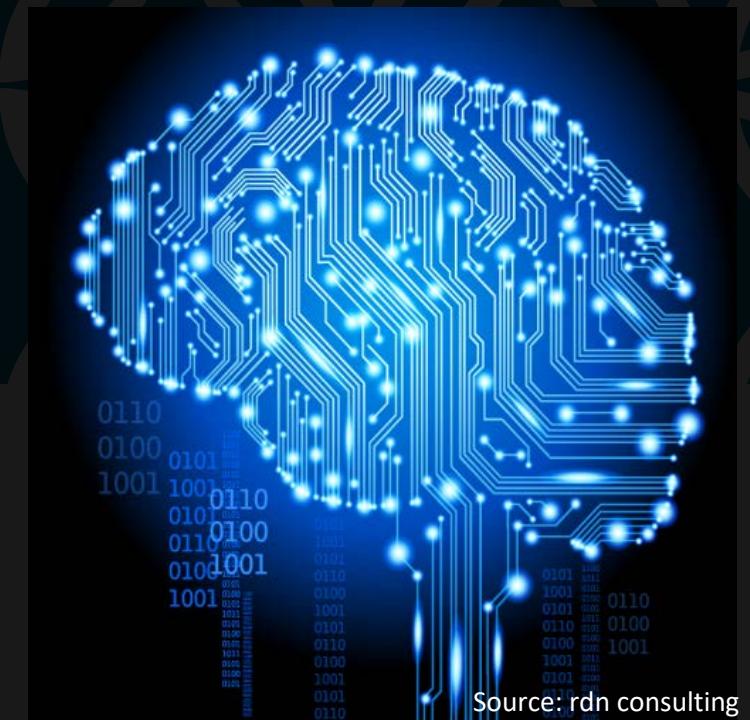
@truyenoz



letdataspeak.blogspot.com



goo.gl/3jJ100



Resources

Slides and references:

- <https://truyentran.github.io/pakdd18-tute.html>
- Shorten URL: goo.gl/UuZZJ9

Key survey paper (updated frequently):

- **Ching, Travers, et al. "Opportunities And Obstacles For Deep Learning In Biology And Medicine." *bioRxiv* (2018): 142760**

Agenda

Topic 1: Introduction (20 mins)

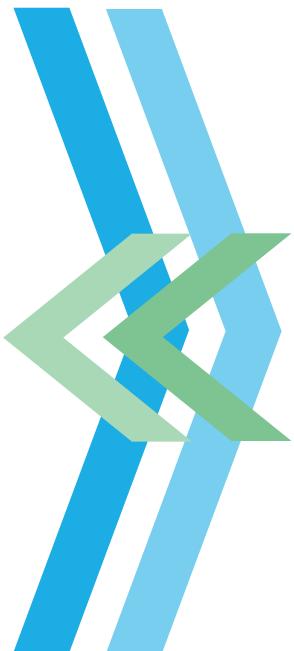
Topic 2: Brief review of deep learning (30 mins)

- Classic architectures
- Capsules & graphs
- Memory & attention

Topic 3: Genomics (30 mins)

- Nanopore sequencing
- Genomics modelling

QA (10 mins)



Break (30 mins)

Topic 4: Healthcare (40 mins)

- Time series (regular & irregular)
- EMR analysis: Trajectories prediction
- EMR analysis: Sequence generation

Topic 5: Data efficiency (40 mins)

- Few-shot learning
- Generative models
- Unsupervised learning of drugs

Topic 6: Future outlook

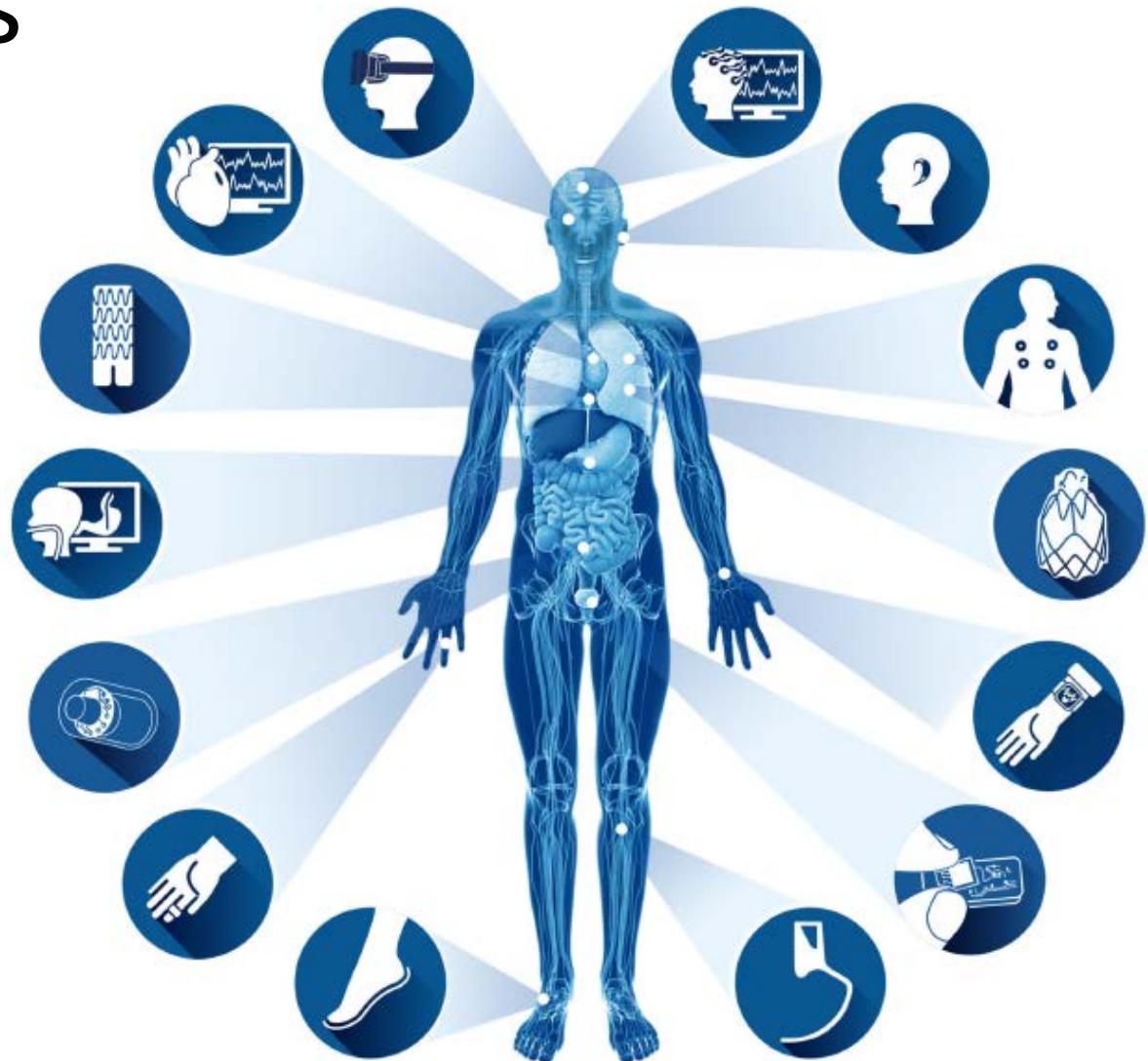
QA (10 mins)

Sensing technologies and data

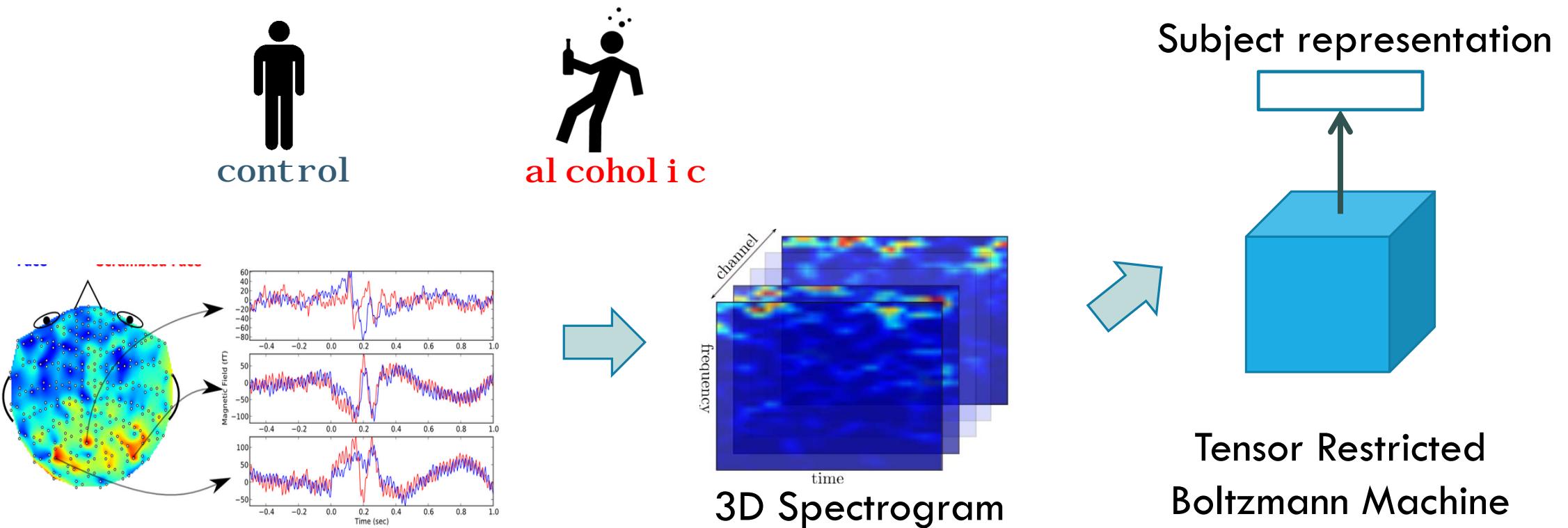
Raw signals are ideal candidates for deep learning

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.



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

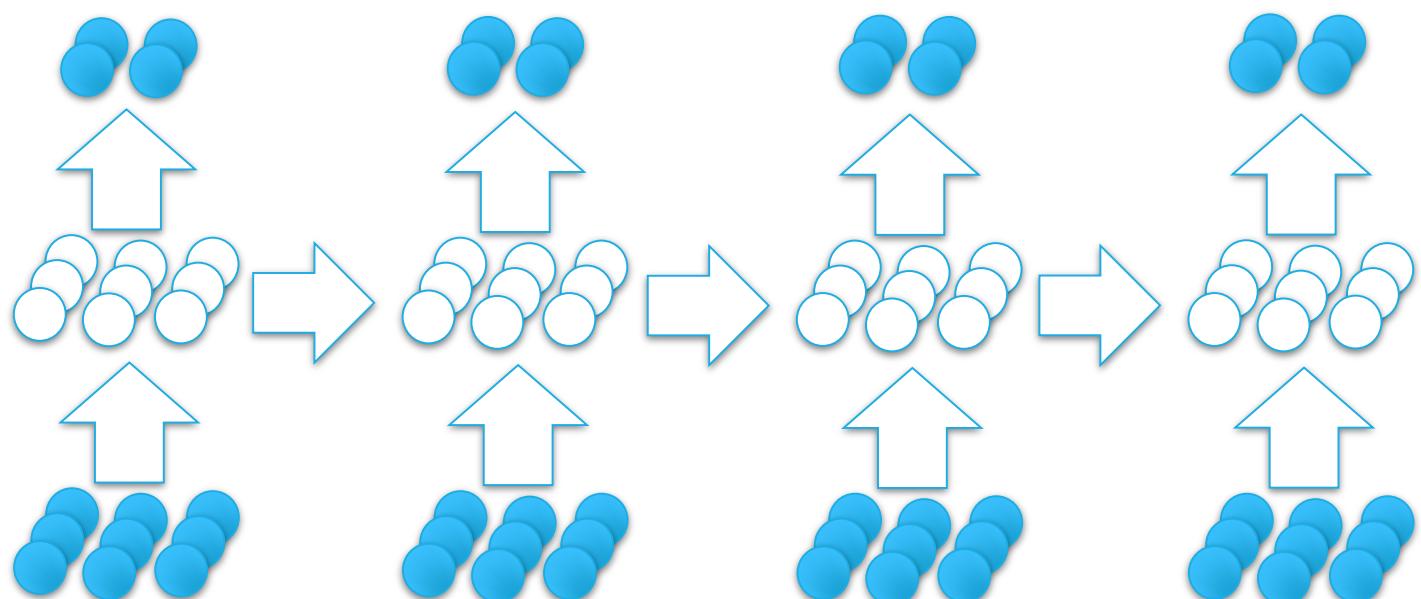
EEG segments as matrices

Temporal dynamics as recurrence

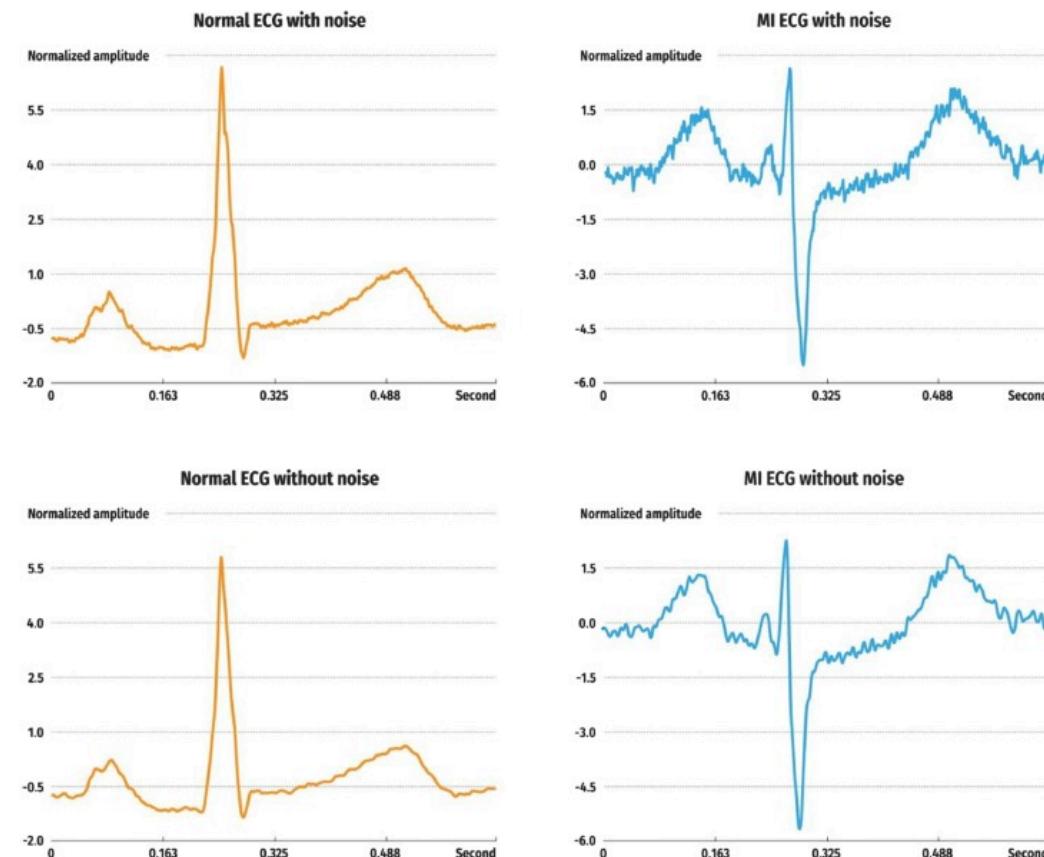
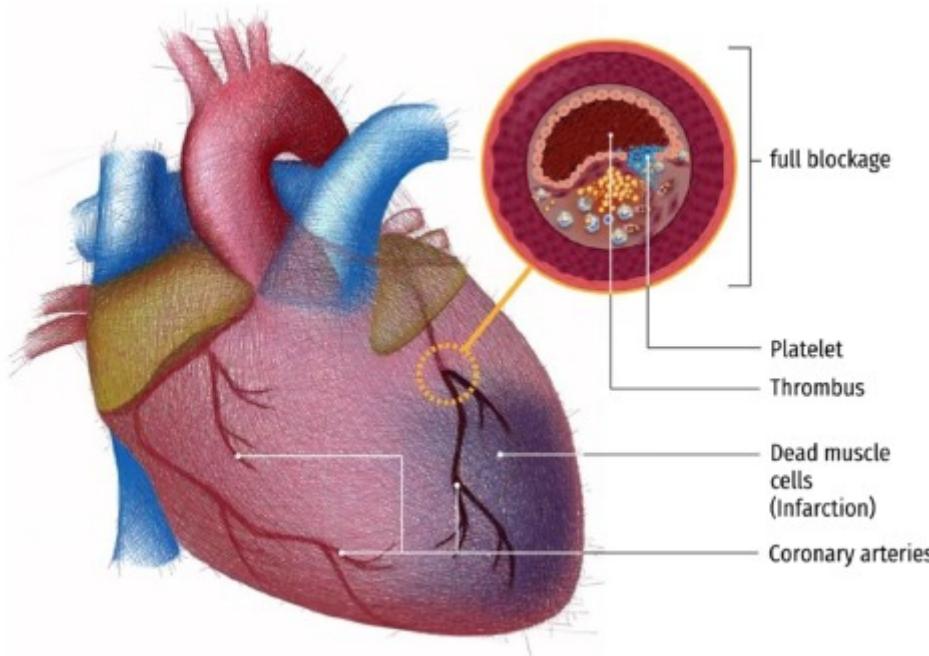
#REF: Kien Do, Truyen Tran, Svetha Venkatesh, "Learning Deep Matrix Representations", *arXiv preprint arXiv:1703.01454*

Recurrent dynamics

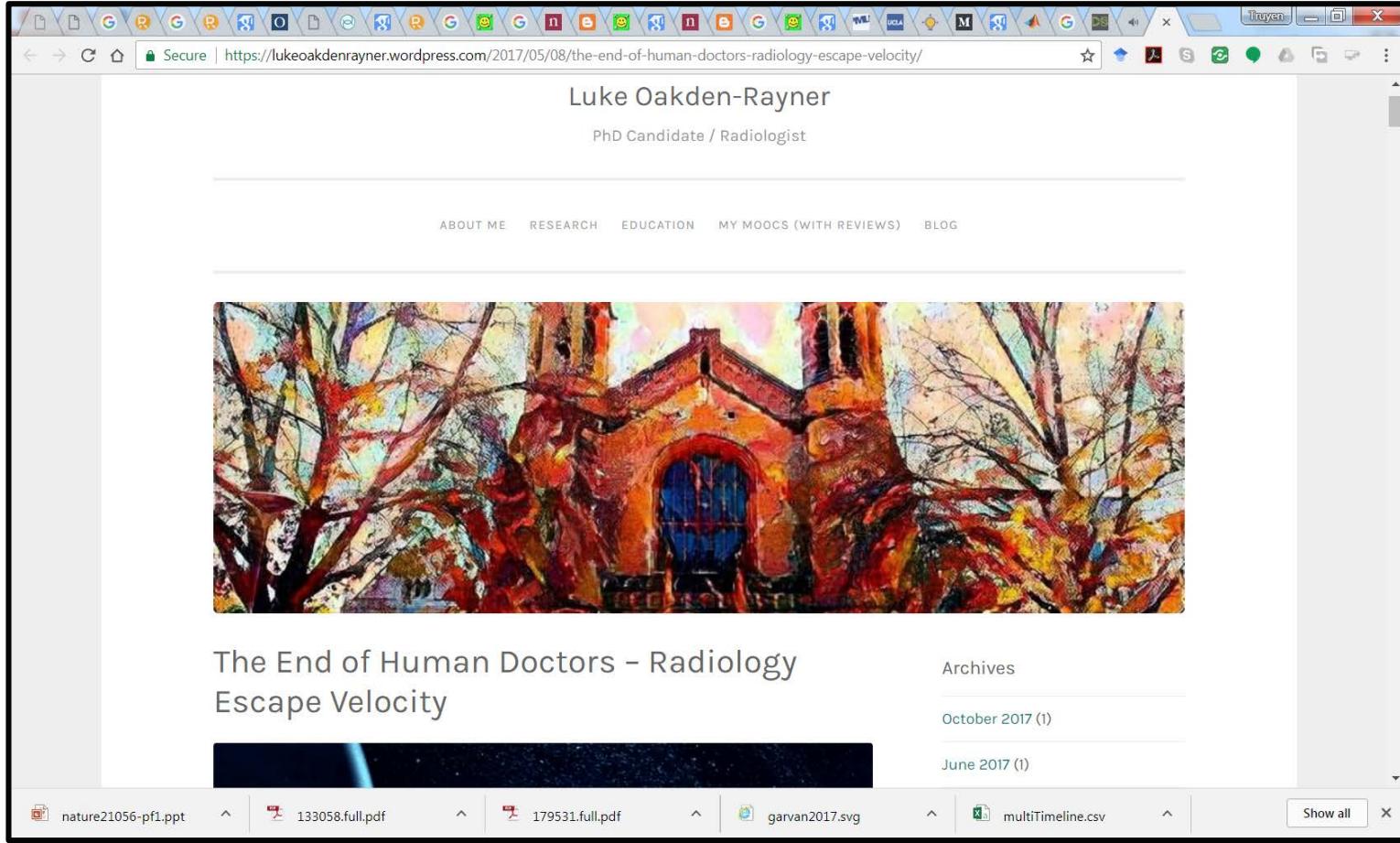
$$H_t = \sigma(U_x^\top X_t V_x + U_h^\top H_{t-1} V_h + B)$$



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)

Handling irregular time-series

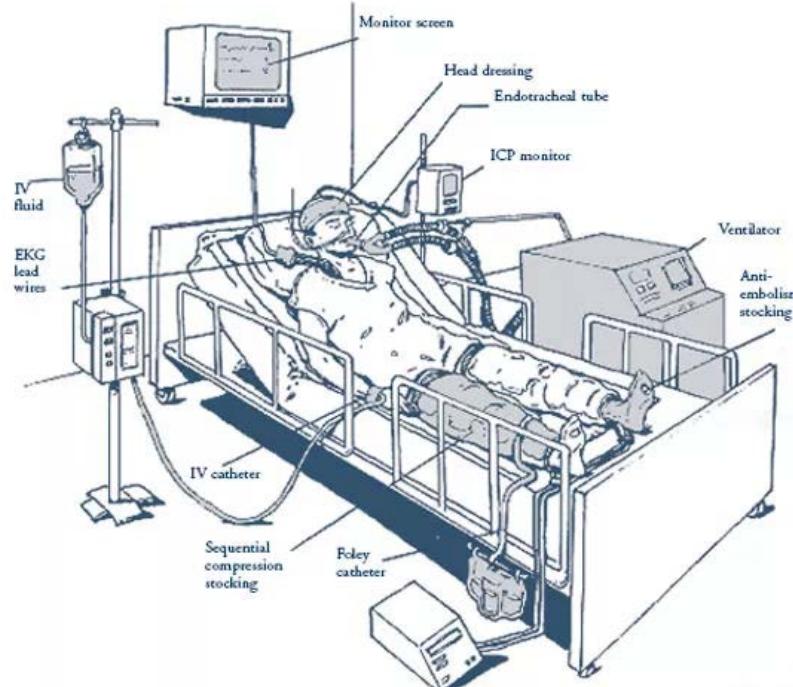
The needs

- Accuracy
- Interpretability
- As early as possible

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).



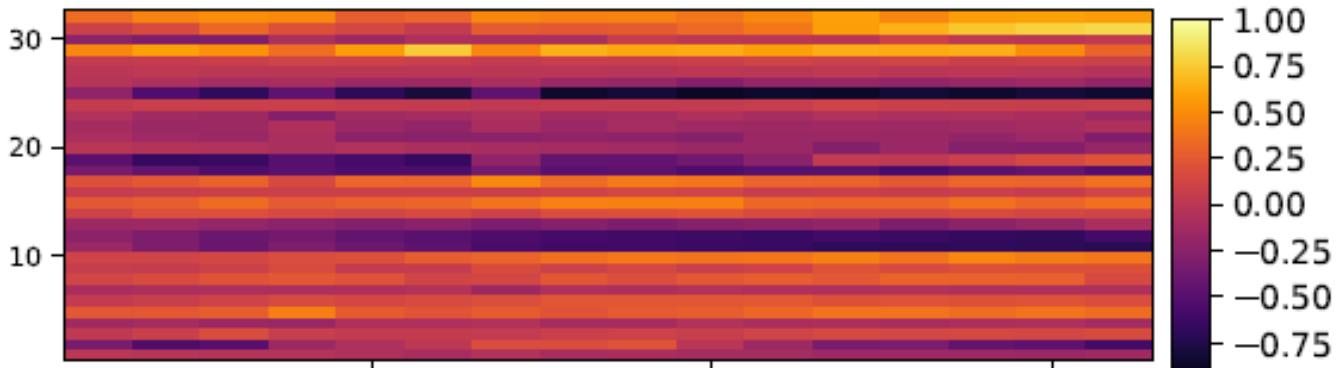
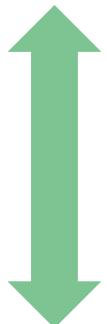
Source: healthpages.org

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**

Result: Attend to risks in ICU

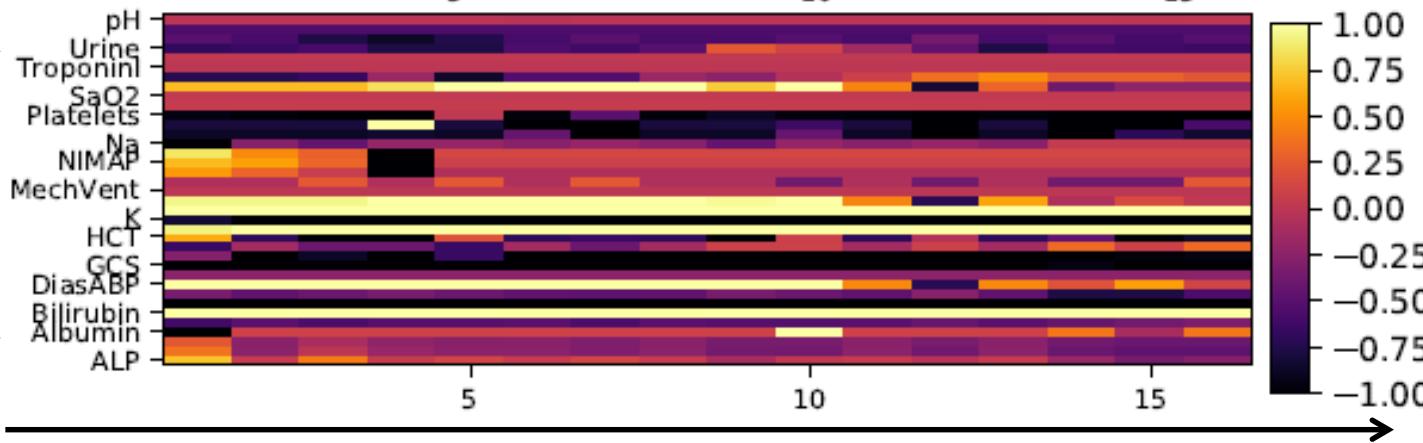
State
transition



Attention
probabilities



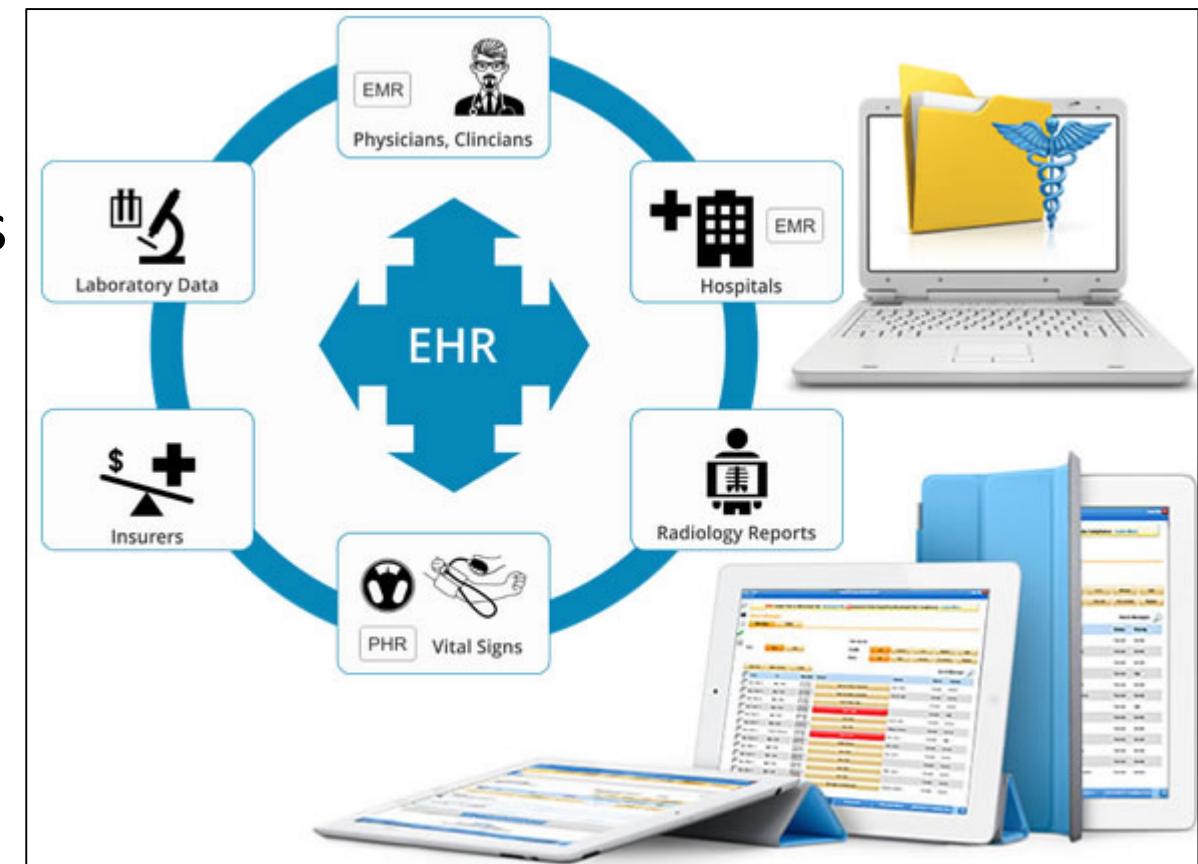
Physiological
measures



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

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

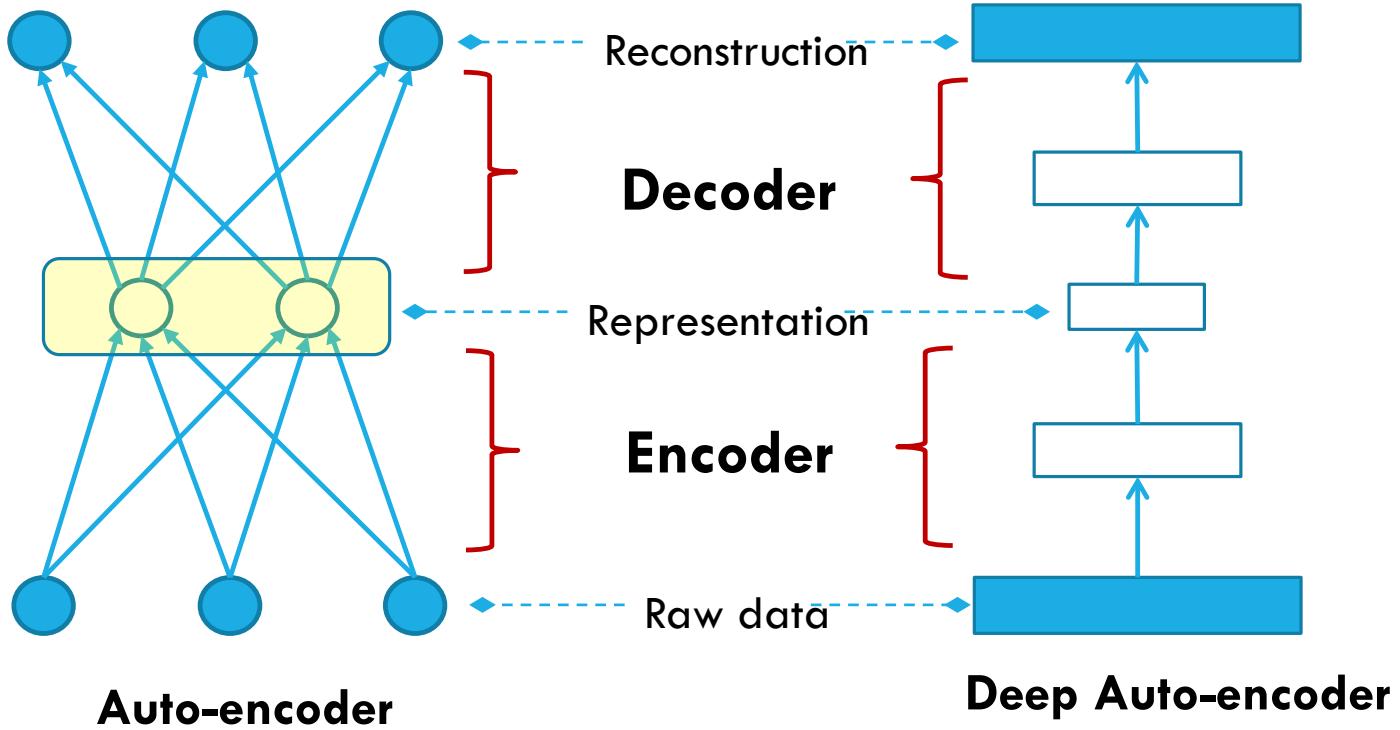
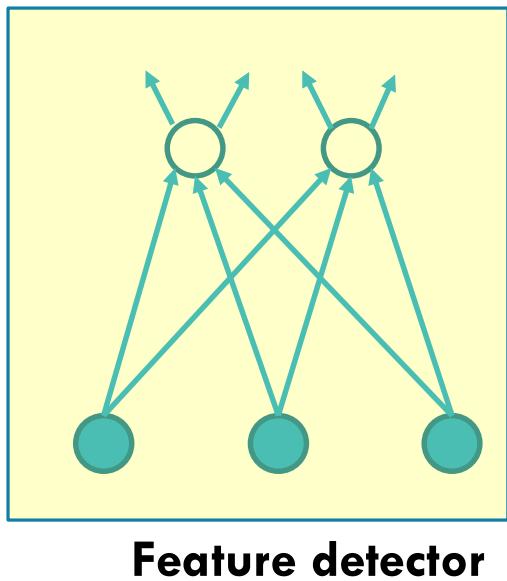
Drugs & tests

- 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

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	0.773 *	0.929 *	0.181 *

Trajectories modeling: Challenges & opportunities

Long-term dependencies

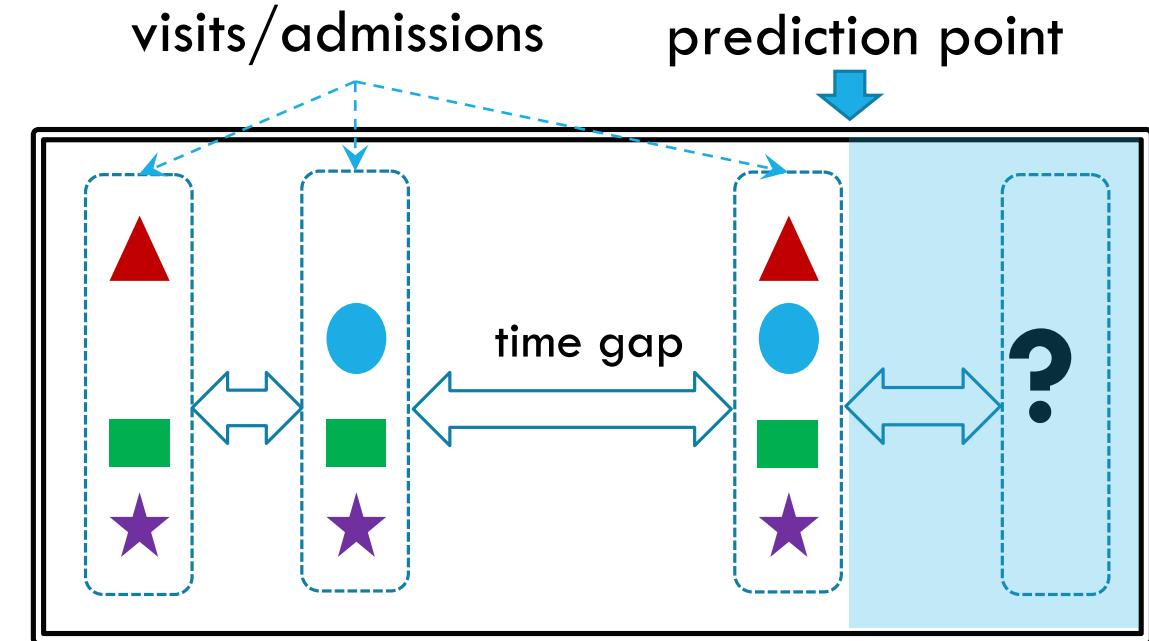
Irregular timing

Mixture of discrete codes and
continuous measures

Complex interaction of diseases and
care processes

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

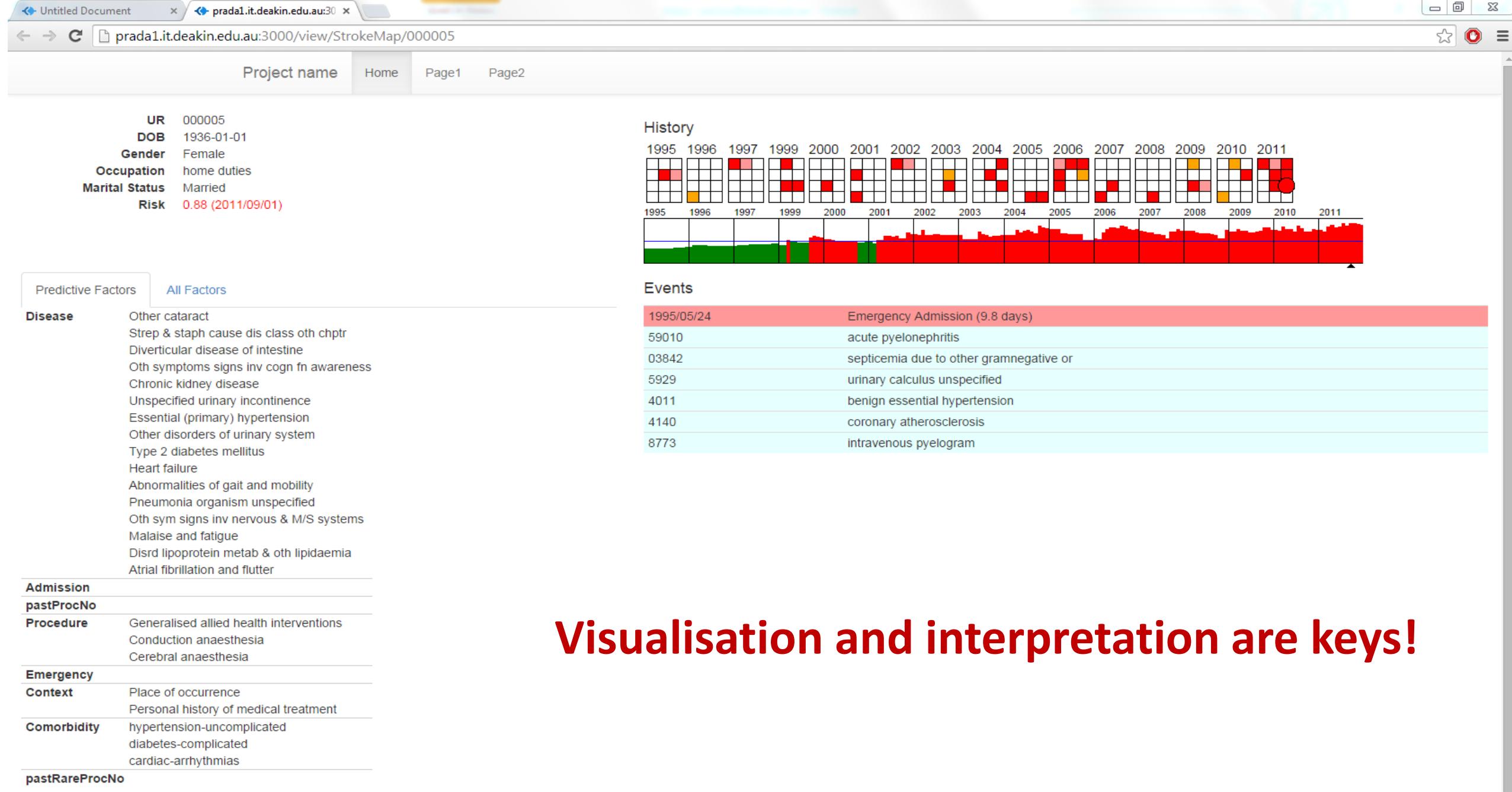
Rich domain knowledge & ontologies



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

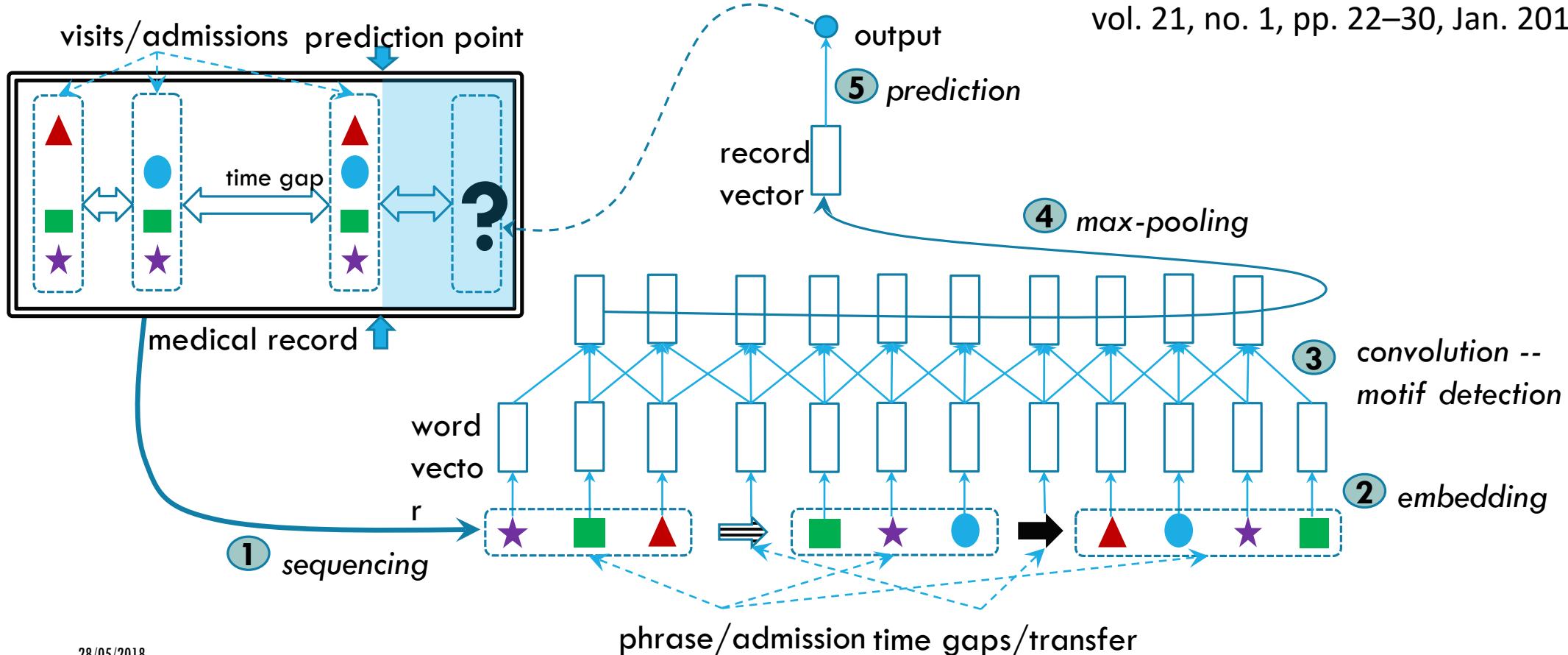
New modalities: social medial, wearable
devices

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

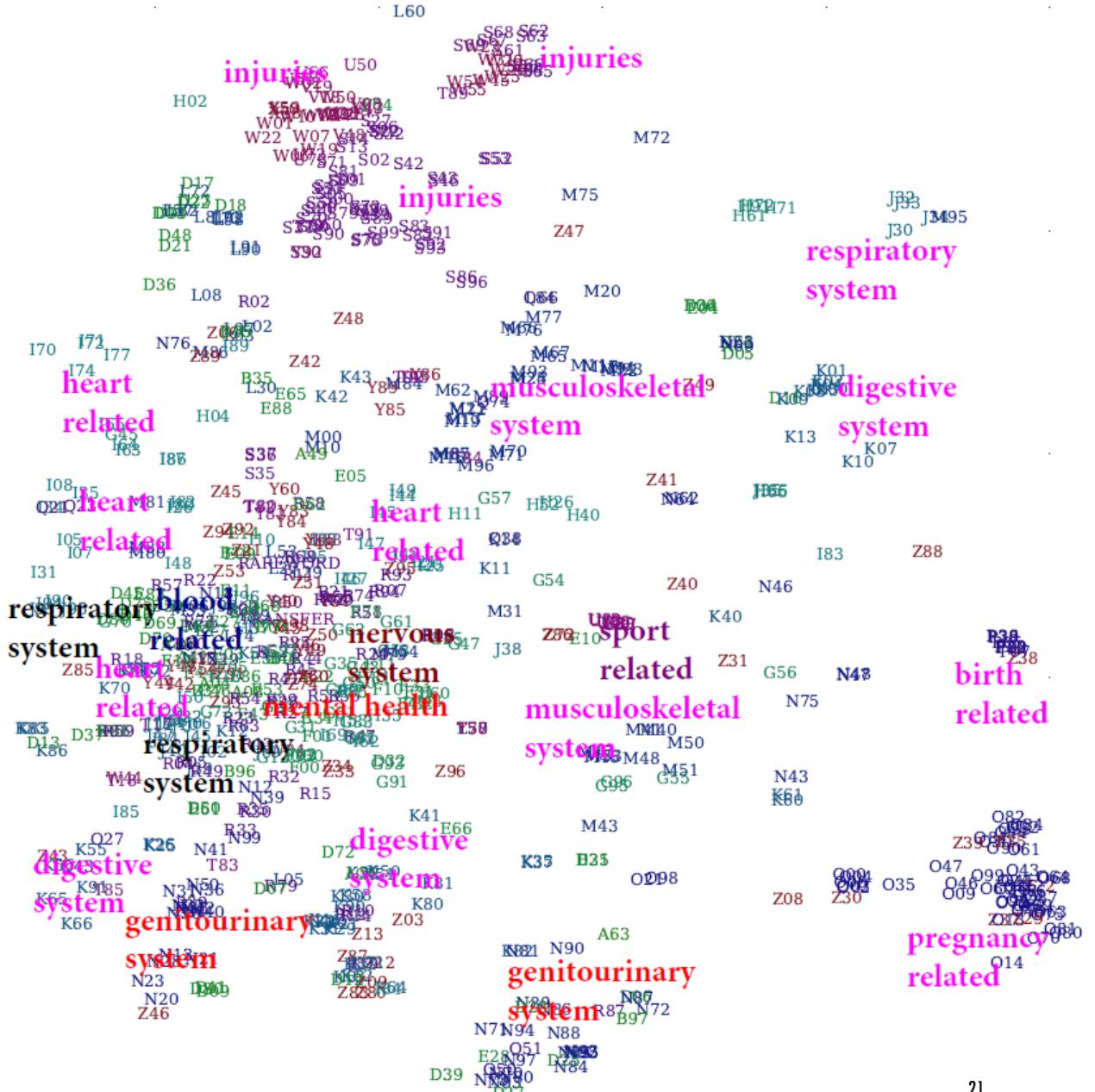
Heart failure

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

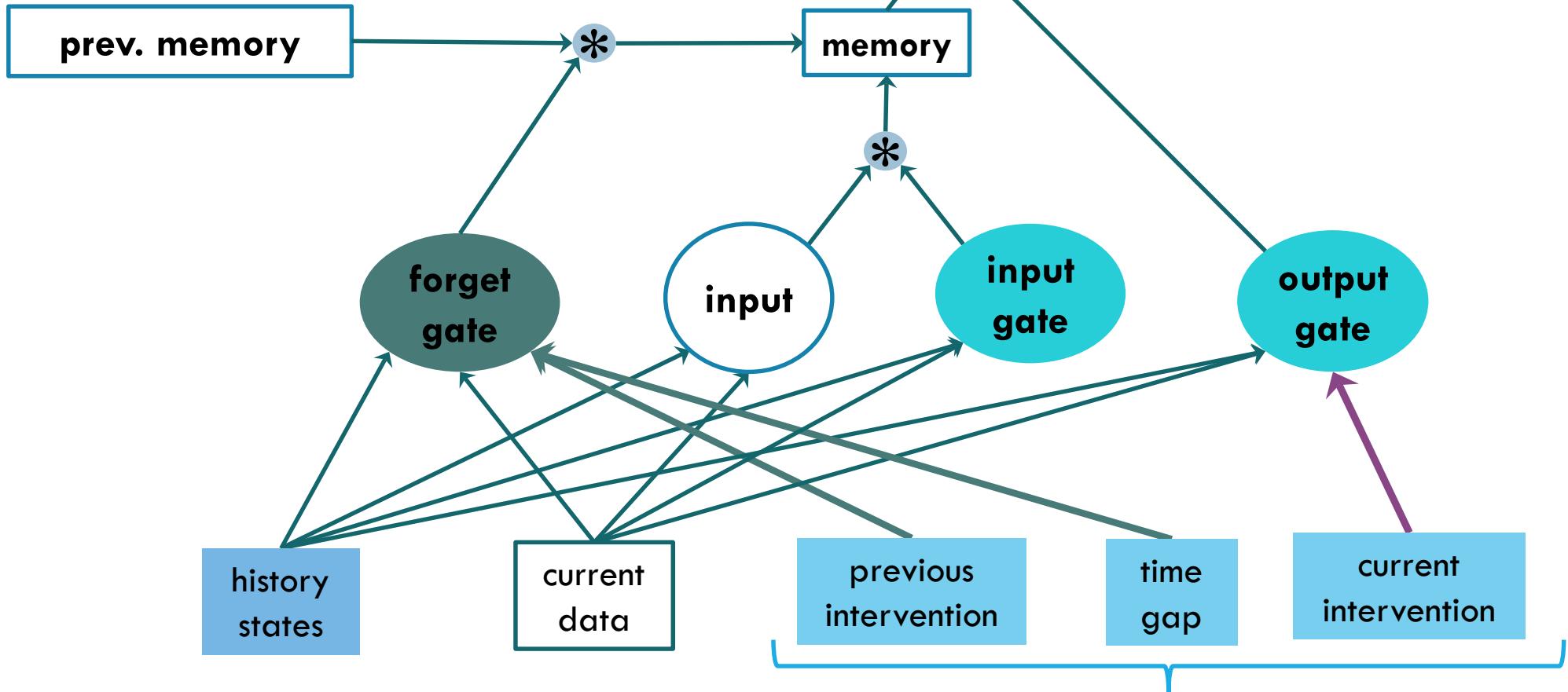
Time and previous intervention → “forgetting” of illness

Current intervention → controlling the risk states

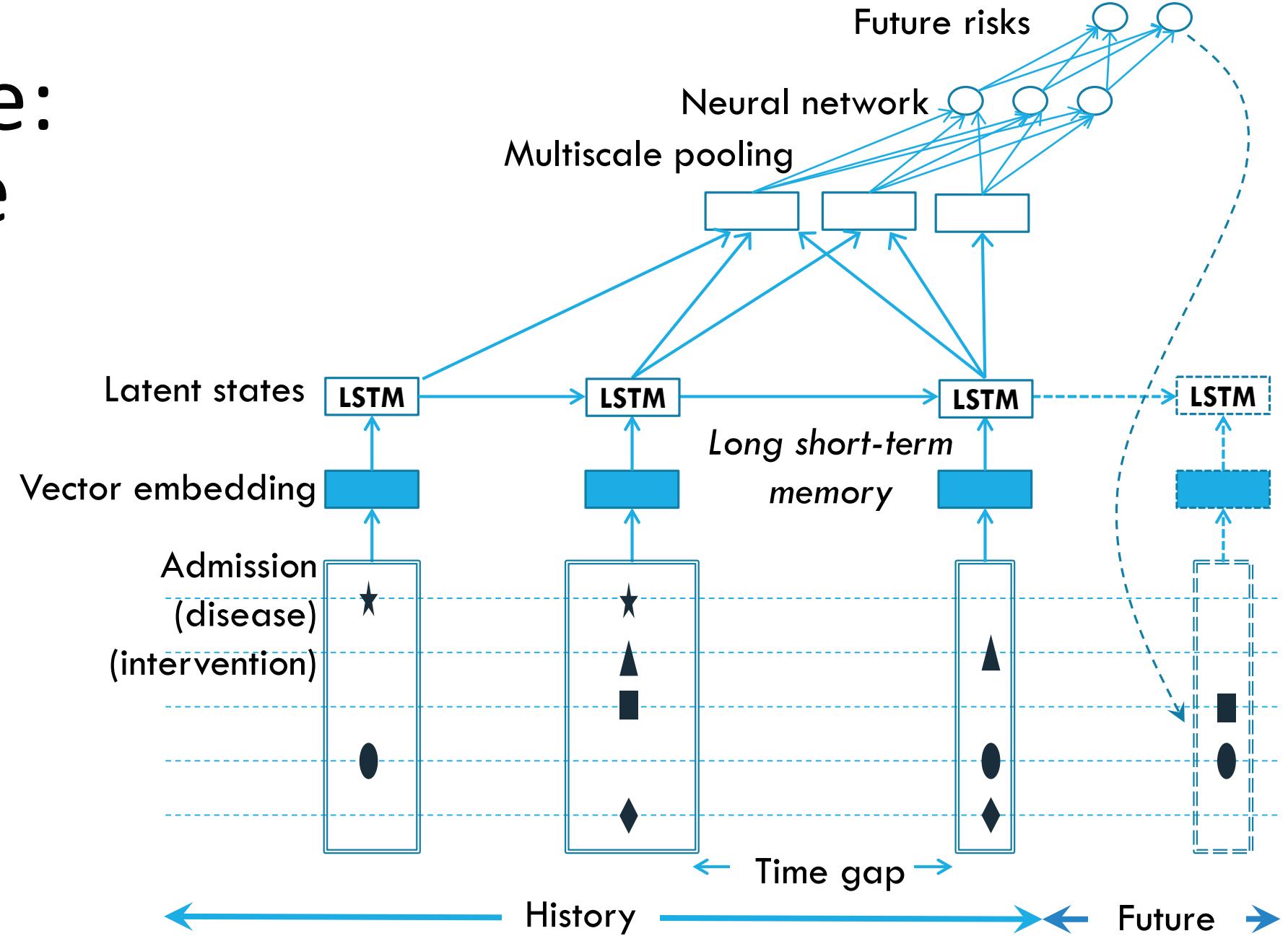
#REF: Trang Pham, et al., Predicting healthcare trajectories from medical records: A deep learning approach, *Journal of Biomedical Informatics*, April 2017, DOI: 10.1016/j.jbi.2017.04.001.

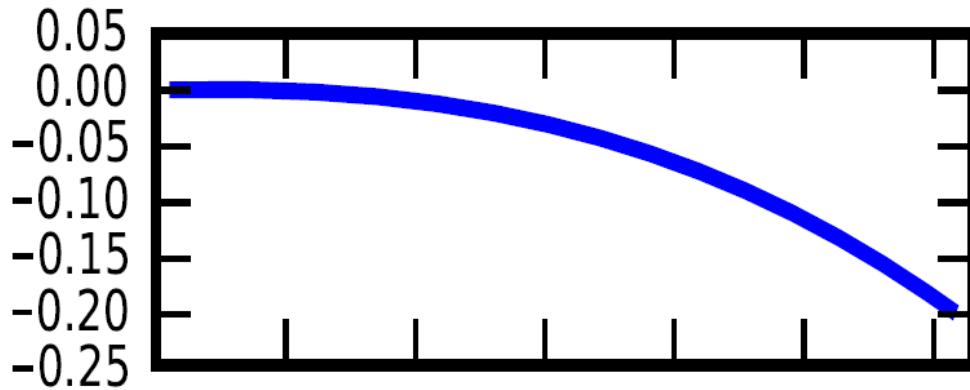
aggregation over
time → prediction

DeepCare: Dynamics



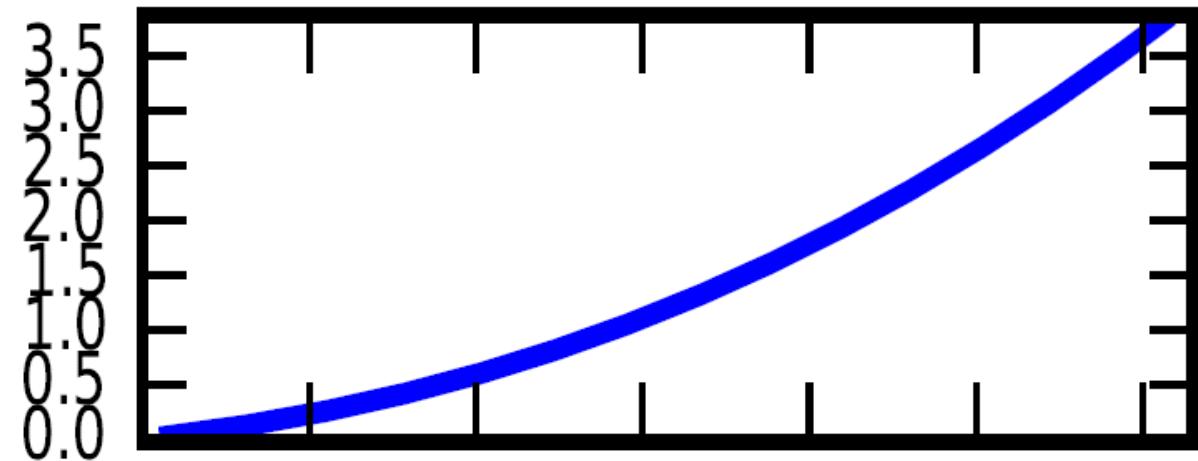
DeepCare: Structure





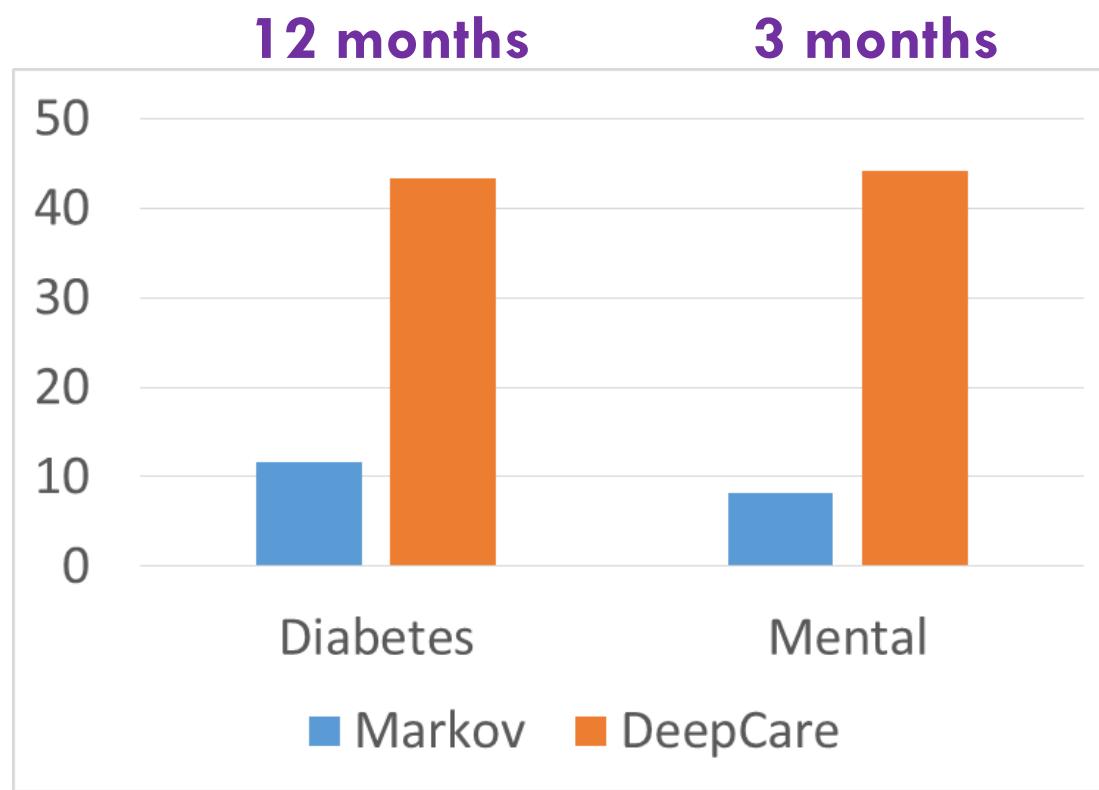
→ decreasing illness

→ Increasing illness

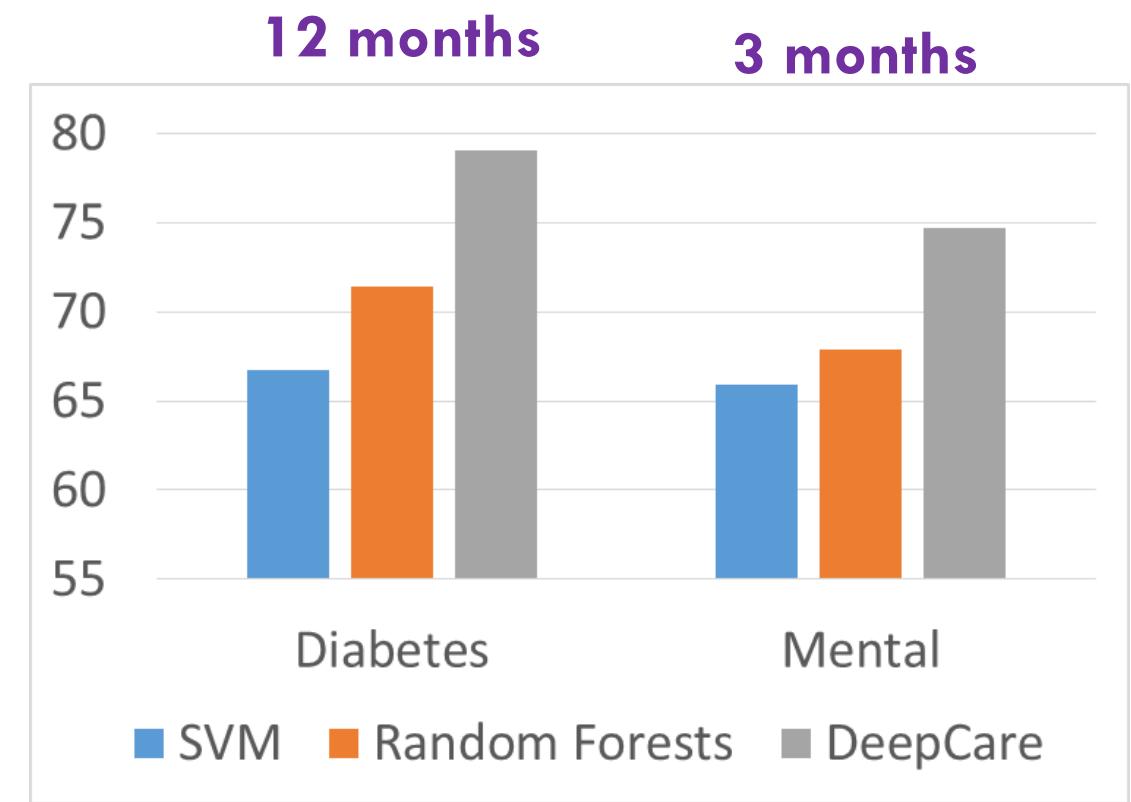


DeepCare: Two modes of forgetting as a function of time

DeepCare: prediction results



Intervention recommendation (precision@3)



Unplanned readmission prediction (F-score)

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

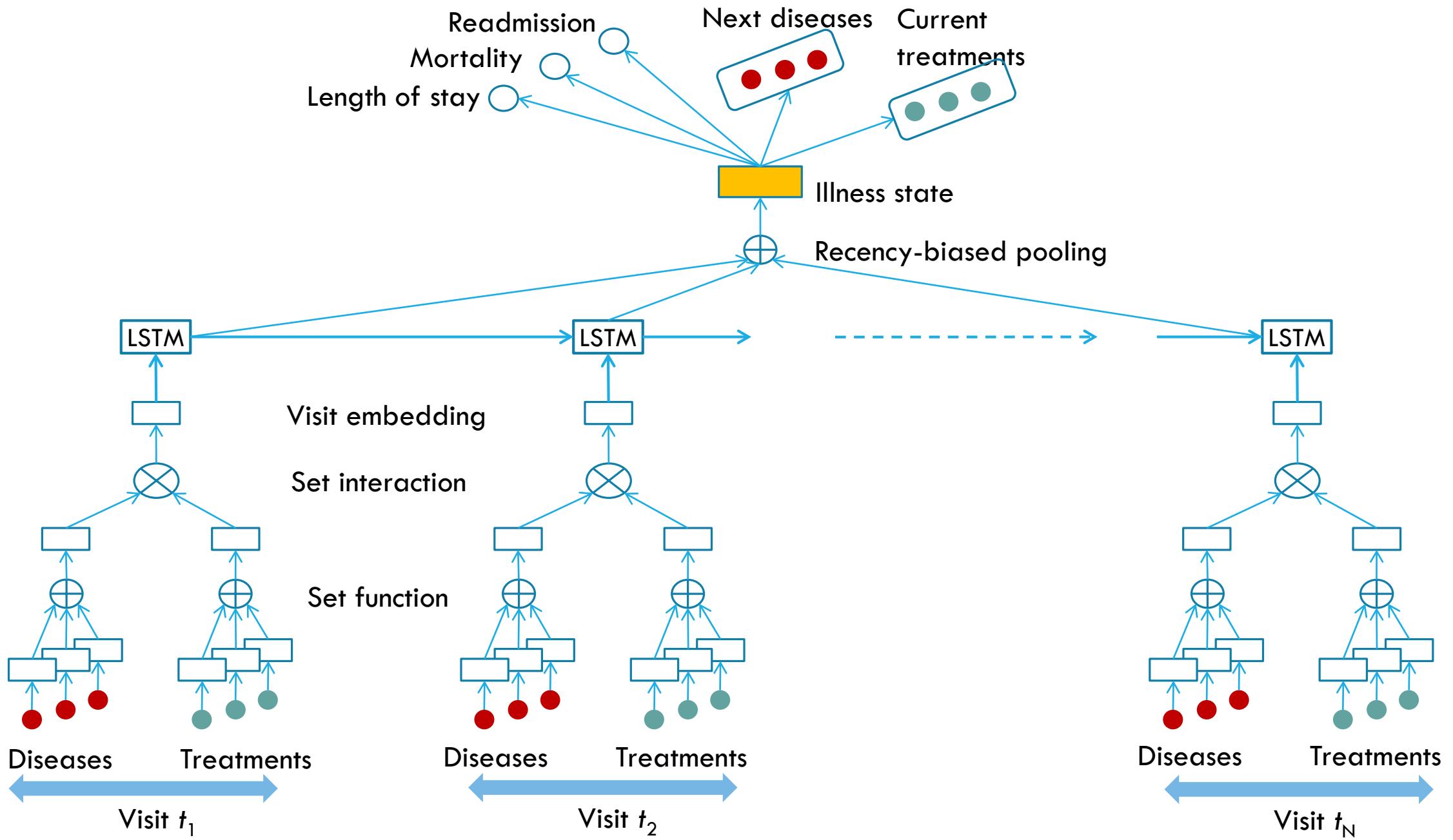
There are interactions between multi-diseases and multiple-treatments

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

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

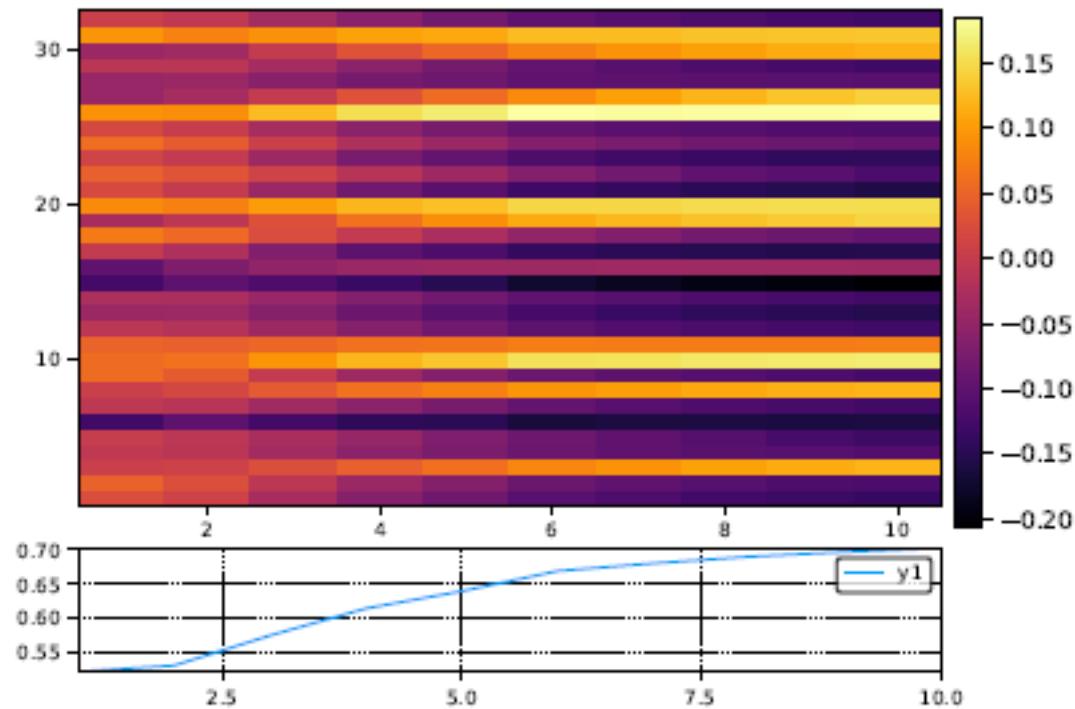
$$f_e(S) \leftarrow \frac{\bar{e}_S}{\epsilon + \|\bar{e}_S\|} \quad \text{where} \quad \bar{e}_S = \max(0, \sum_{i \in S} e_i)$$

#REF: Phuoc Nguyen, Truyen Tran, and Svetha Venkatesh. "Ressel: A Recurrent Model for Sequence of Sets with Applications to Electronic Medical Records." IJCNN (2018).

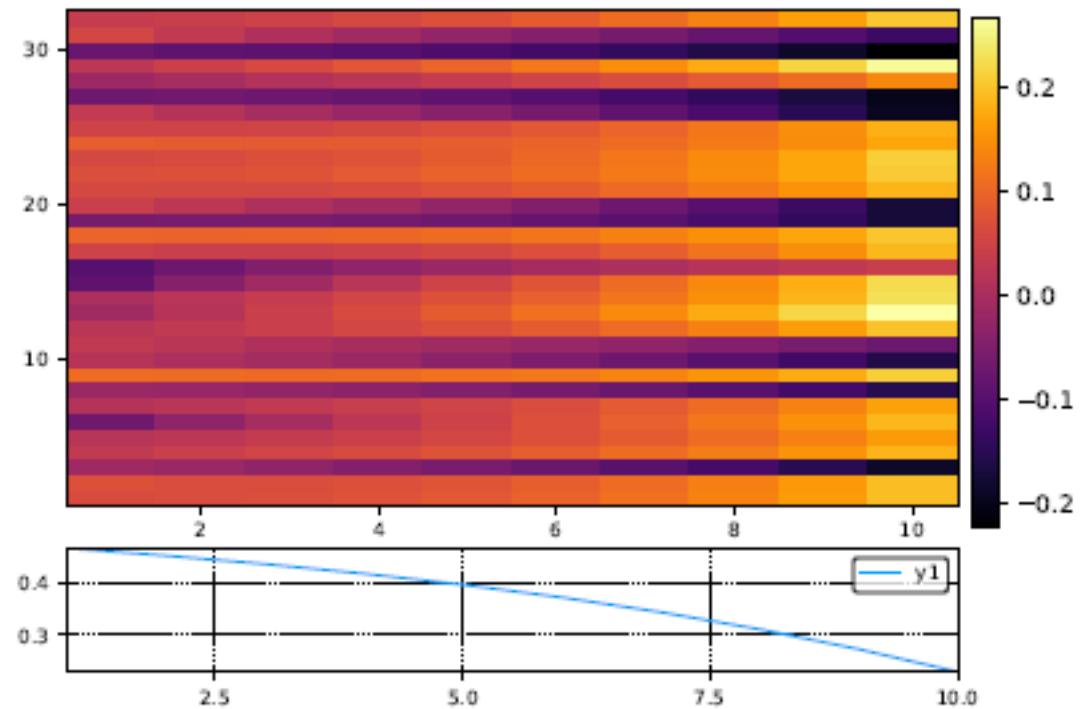


Results (AUC)

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



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



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

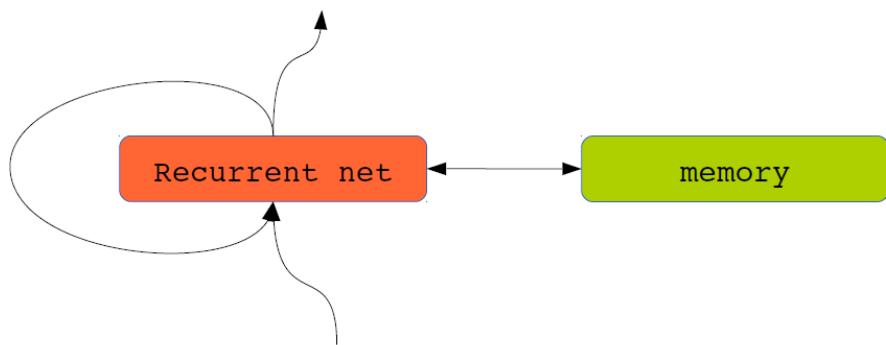
Trajectories prediction

Generating a subset of treatments

Generating an entire health/care trajectory

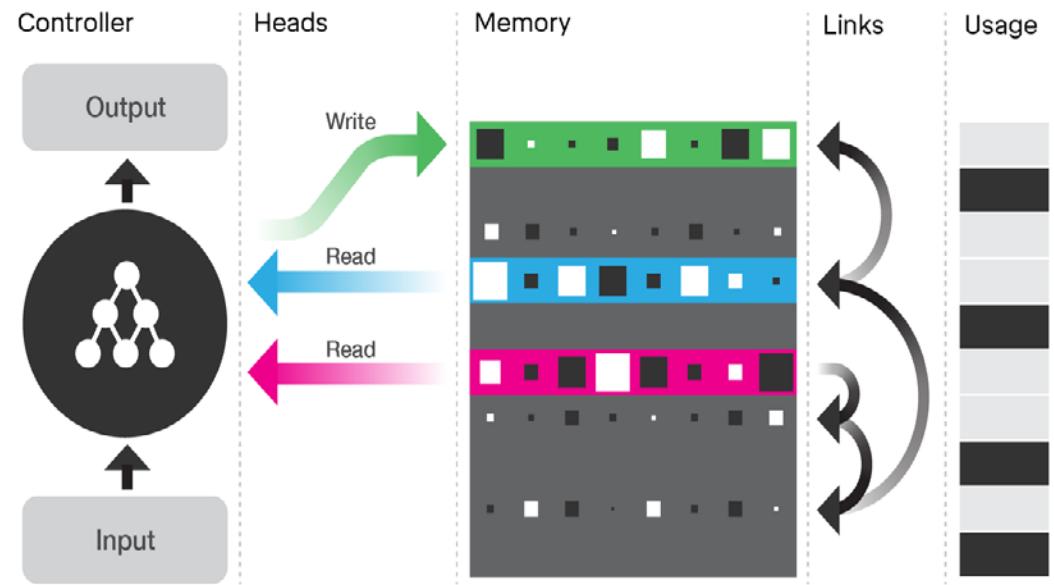
Challenges: global loss, meaningful evaluation metrics

A solution: Attention and Memory-augmented neural nets (MANN)



(LeCun, 2015)

Illustration of the DNC architecture



Source: deepmind.com

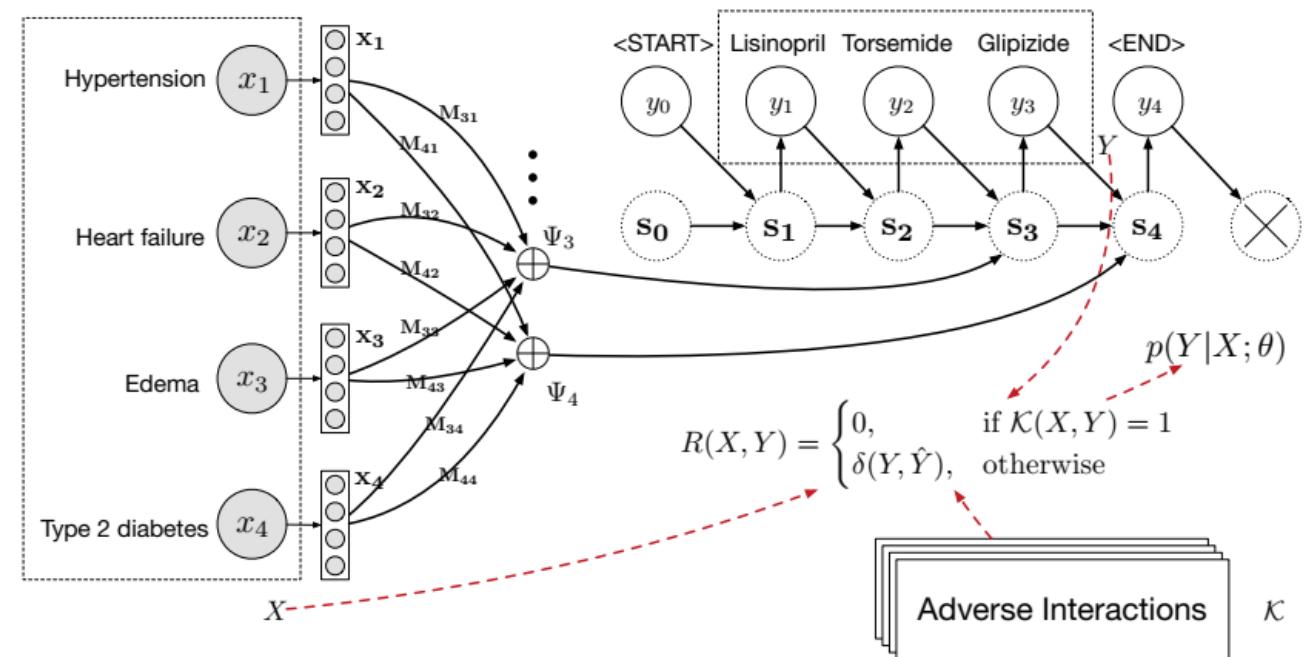
#REF: Graves, Alex, et al. "Hybrid computing using a neural network with dynamic external memory." *Nature* 538.7626 (2016): 471-476.

Treatment recommendations using attention: LEAP

Attend to the diagnoses

Generating a sequence of medicine codes

Fine tune the output with reinforcement learning: avoid adverse drug interaction (need external database adverse drug interaction pairs K)



#REF: Zhang et al., “Leap: Learning to prescribe effective and safe treatment combinations for multimorbidity”, KDD17.

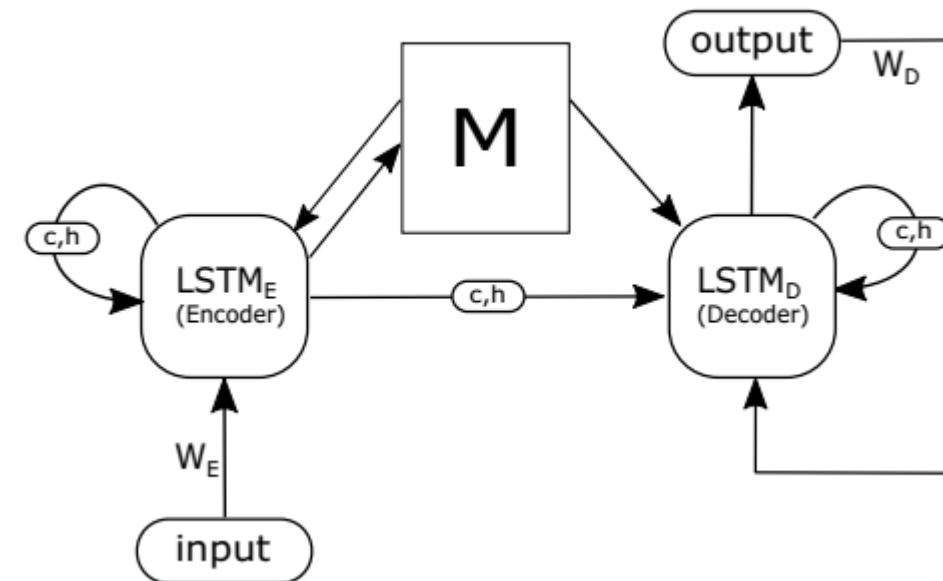
Dual Control Memory-augmented Memory Network

Diagnoses are encoded to an external memory by the encoder

The decoder reads the memory and produces a sequence of treatment codes. During decoding, the memory is write-protected (DCw-MANN)

Memory captures long-term dependencies inside and amongst admissions

Memory enables skip-connection attentions



#REF: Hung Le, Truyen Tran, and Svetha Venkatesh. “Dual Control Memory Augmented Neural Networks for Treatment Recommendations”, PAKDD18.

Treatment Prediction Results

First drug predictions: precision and Jaccard Score MIMIC3 dataset.

Models	Precision	Jaccard
Logistic Regression	0.412	0.311
Random Forest	0.491	0.405
LSTM	0.220	0.138
LSTM + attention	0.224	0.142
DNC	0.577	0.529
DCw-MANN	0.598	0.556

Top Drug predictions: Jaccard Score on top code GPI 1-3 MIMIC3 dataset.

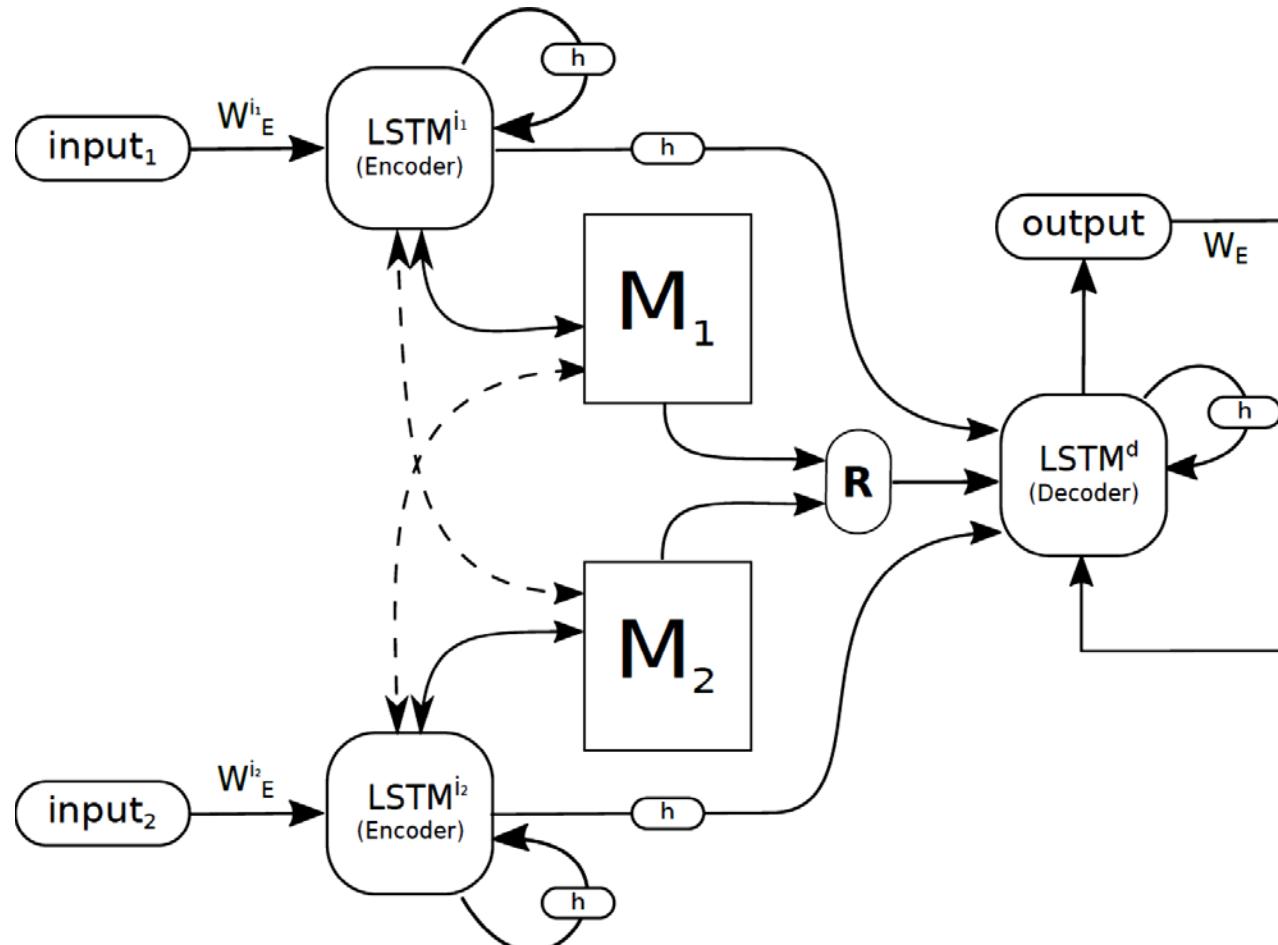
Models	Jaccard on GPI 1	Jaccard on GPI 3
Basic LEAP	0.510	0.385
LEAP + RL	0.558	0.434

Dual Memory Neural Computer (DMNC)

Reads multiple EMR channels (disease, procedure, medication)

Memory can be shared or separated.

Generate a sequence of outputs (e.g., medications recommendation, or future disease progression).



#Ref: Le, Hung, Truyen Tran, and Svetha Venkatesh. "Dual Memory Neural Computer for Asynchronous Two-view Sequential Learning." KDD18.

DMNC: drug prescription

Two views:

+ Diagnoses

+ Procedures

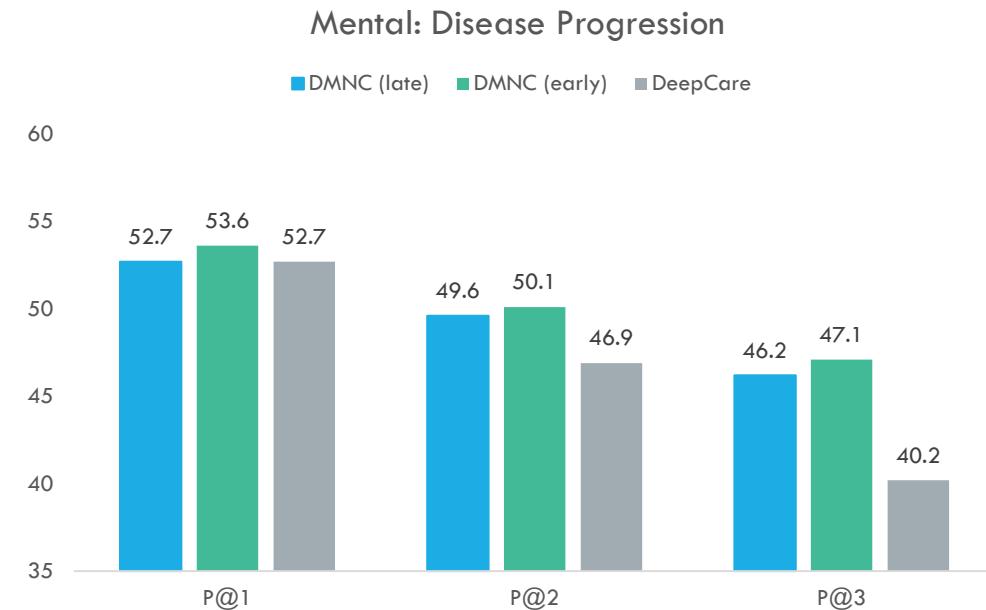
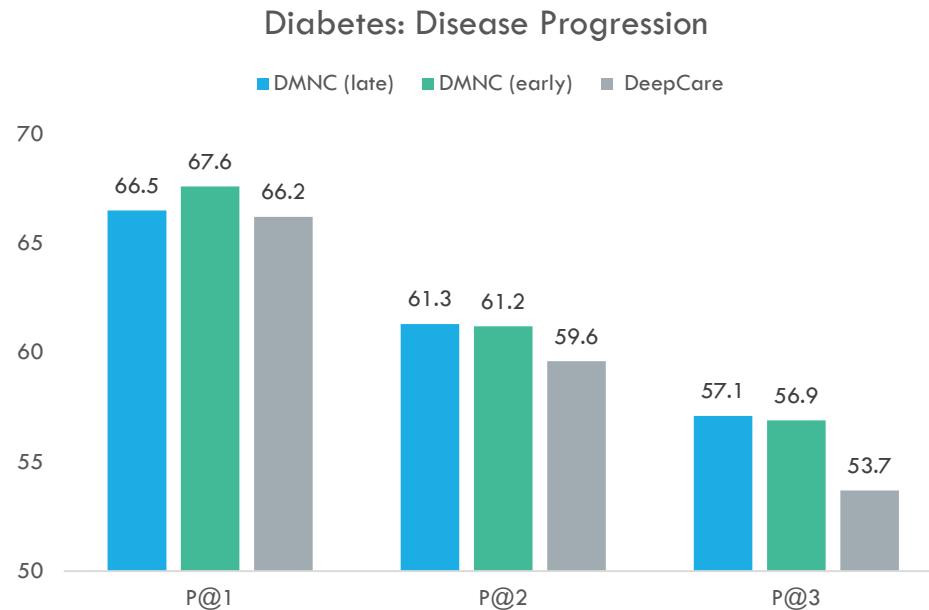
Two modes of memory:

+ Late fusion

+ Early fusion

Model	AUC	F1	P@1	P@2	P@5
Diagnosis Only					
Binary Relevance	82.6	69.1	79.9	77.1	70.3
Classifier Chains	66.8	63.8	68.3	66.8	61.1
LSTM	84.9	70.9	90.8	86.7	79.1
DNC	85.4	71.4	90.0	86.7	79.8
Procedure Only					
Binary Relevance	81.8	69.4	82.6	80.1	73.6
Classifier Chains	63.4	61.7	83.7	80.3	71.9
LSTM	83.9	70.8	88.1	86.0	78.4
DNC	83.2	70.4	88.4	85.8	78.7
Diagnosis and procedure					
Binary Relevance	84.1	70.3	81.0	78.2	72.3
Classifier Chains	64.6	63.0	84.6	81.5	74.2
LSTM	85.8	72.1	91.6	86.8	80.5
DNC	86.4	72.4	90.9	87.4	80.6
Dual LSTM	85.4	71.4	90.6	87.1	80.5
WLAS	86.6	72.5	91.9	88.1	80.9
<i>DMNC_l</i>	87.4	73.2	92.4	88.9	82.6
<i>DMNC_e</i>	87.6	73.4	92.1	89.9	82.5

DMNC: disease progression



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”

Agenda

Topic 1: Introduction (20 mins)

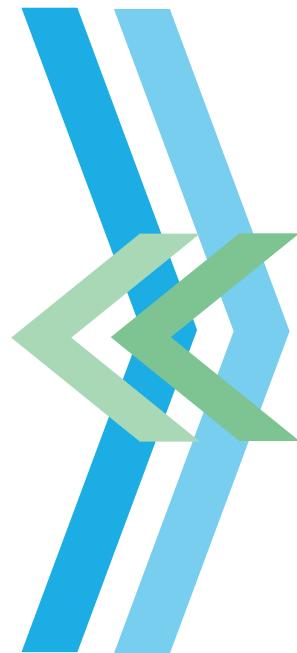
Topic 2: Brief review of deep learning (30 mins)

- Classic architectures
- Capsules & graphs
- Memory & attention

Topic 3: Genomics (30 mins)

- Nanopore sequencing
- Genomics modelling

QA (10 mins)



Break (30 mins)

Topic 4: Healthcare (40 mins)

- Time series (regular & irregular)
- EMR analysis: Trajectories prediction
- EMR analysis: Sequence generation

Topic 5: Data efficiency (40 mins)

- Few-shot learning
- Generative models
- Unsupervised learning of drugs

Topic 6: Future outlook

QA (10 mins)

Few-shot deep learning

Lots of biomedical problems are data poor

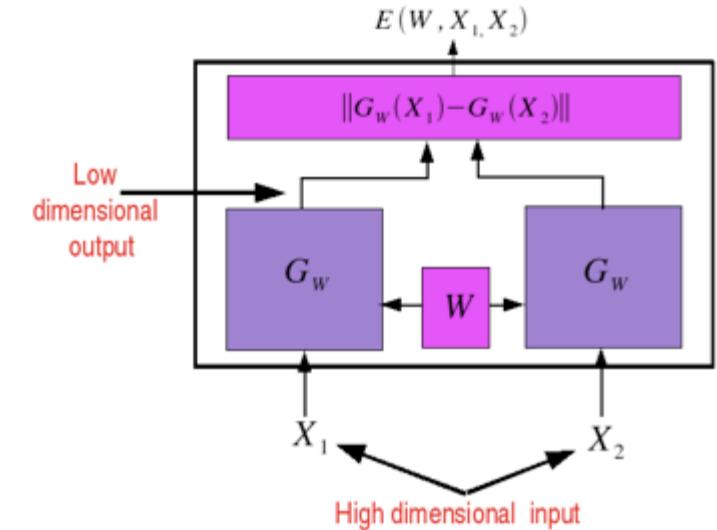
- Rare drugs
- Rare diseases

Distance metrics learning (DML) methods

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

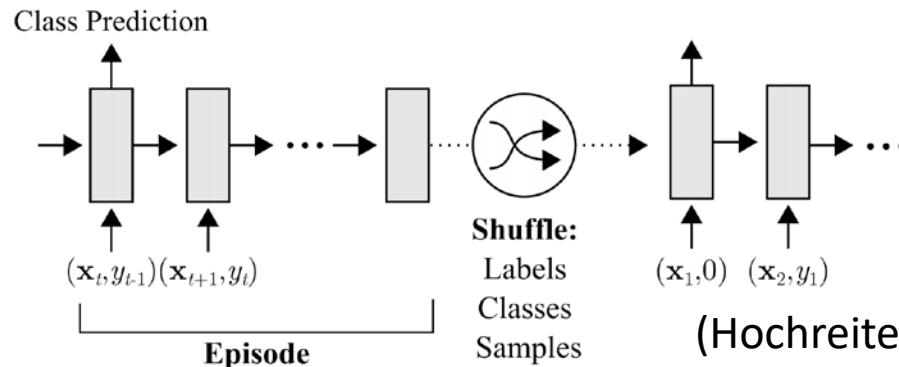
Meta-learning strategies

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



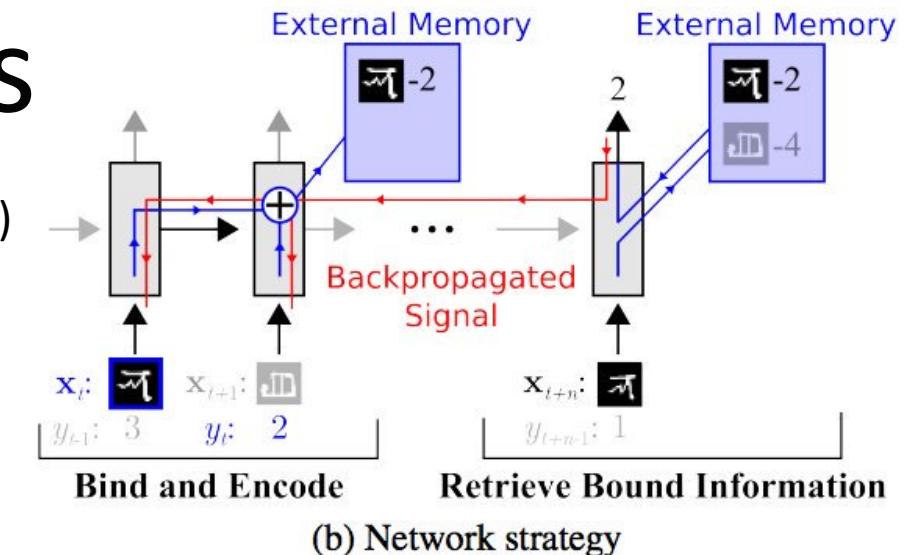
#REF: Chopra, Sumit, Raia Hadsell, and Yann LeCun. "Learning a similarity metric discriminatively, with application to face verification." *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.* Vol. 1. IEEE, 2005.

Meta-learning strategies



(Hochreiter et al., 2001)

(Santoro et al, 2016)



(Mishra et al, 2017)

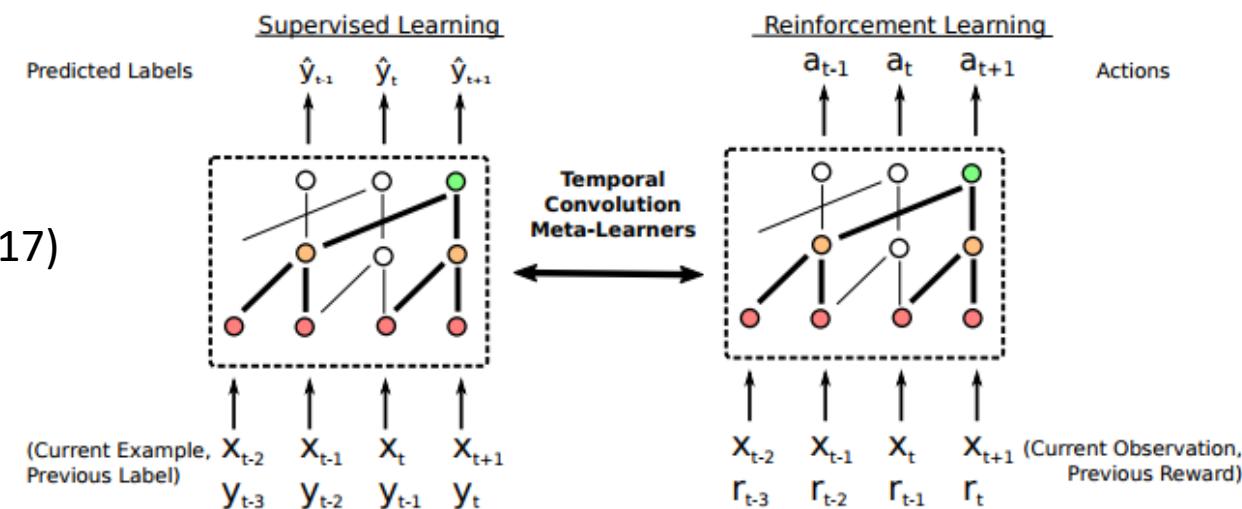


Figure 1: Overview of our temporal-convolution-based meta-learner (TCML). The same class of model architectures can be applied to both supervised and reinforcement learning.

Generative models

Many applications:

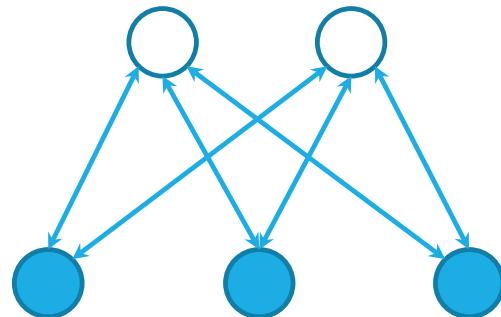
- Text to speech
- **Simulate data that are hard to obtain/share in real life (e.g., healthcare)**
- Generate meaningful sentences conditioned on some input (foreign language, image, video)
- Semi-supervised learning
- Planning

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

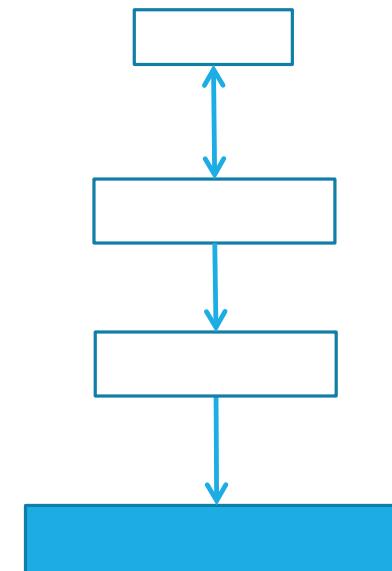
A family: RBM → DBN → DBM

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

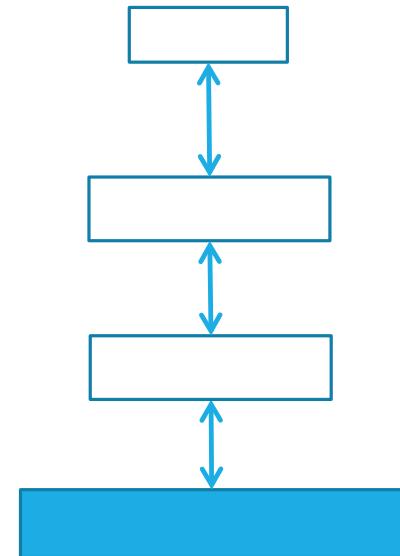
energy



Restricted Boltzmann Machine
(~1994, 2001)



Deep Belief Net
(2006)

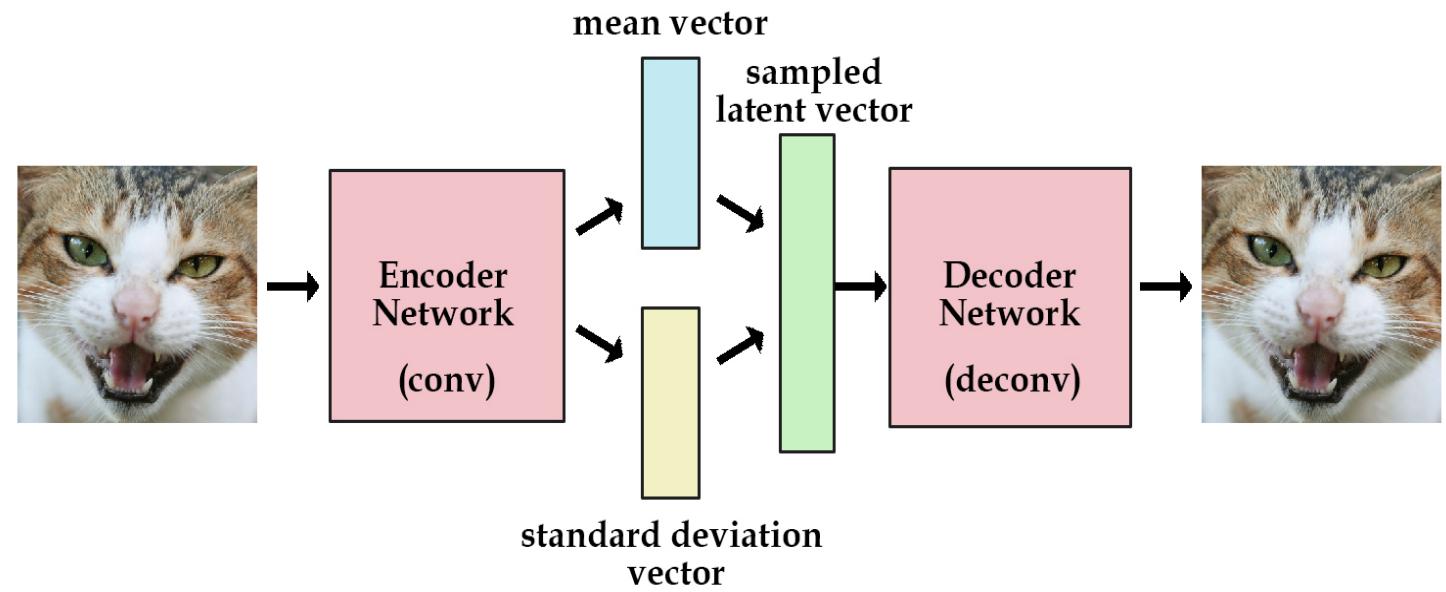
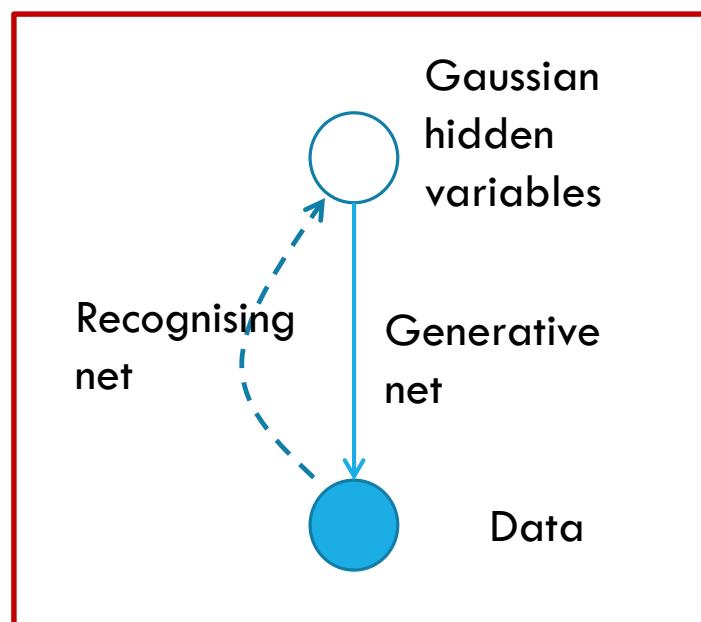


Deep Boltzmann Machine
(2009)

Variational Autoencoder

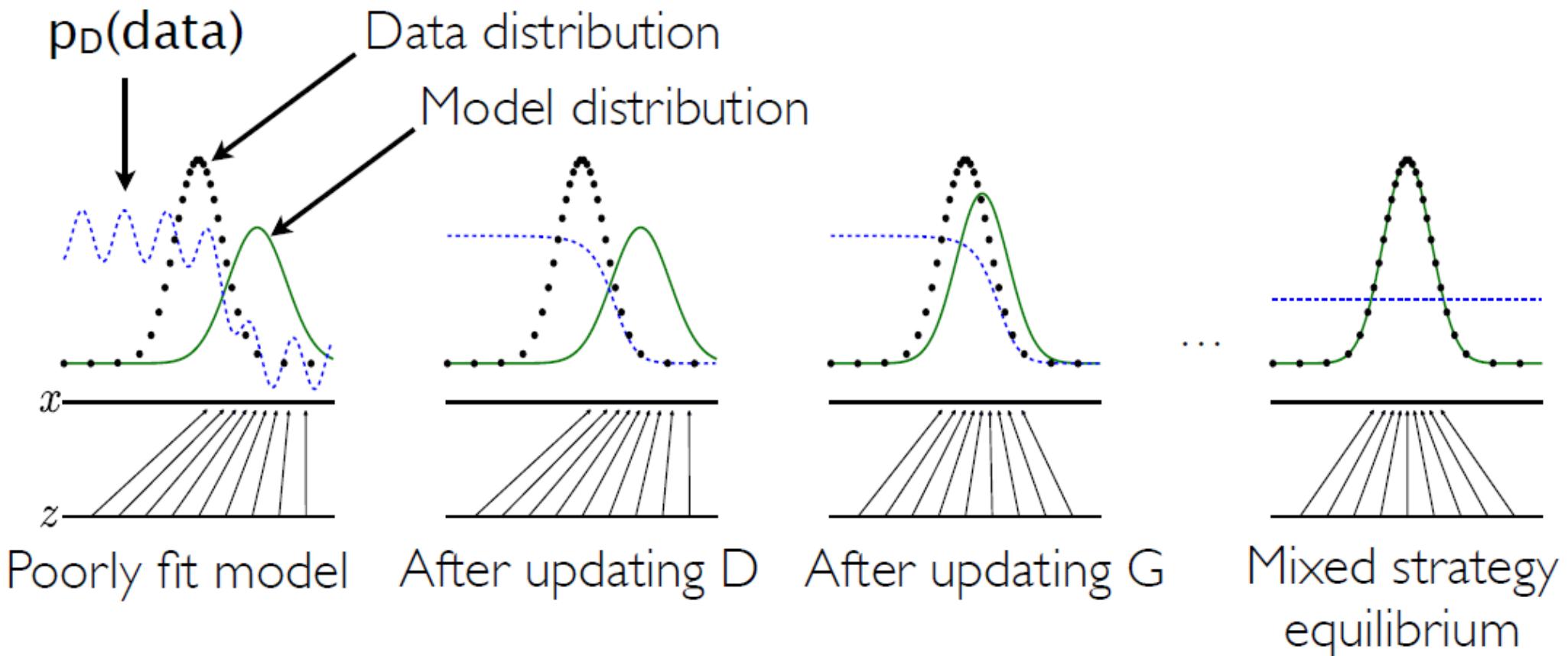
(Kingma & Welling, 2014)

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



GAN: implicit density models

(Adapted from Goodfellow's, NIPS 2014)



Progressive GAN: Generated images



female1.png



female2.png



female3.png



female4.png



female6.png



male1.png



male2.png



male3.png

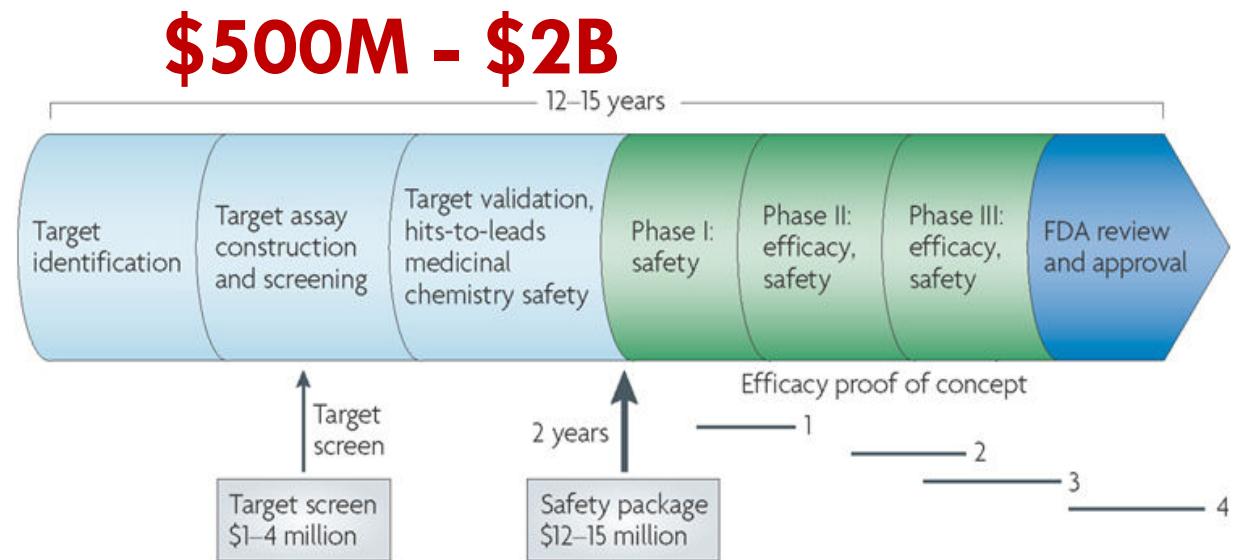
Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*.

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)

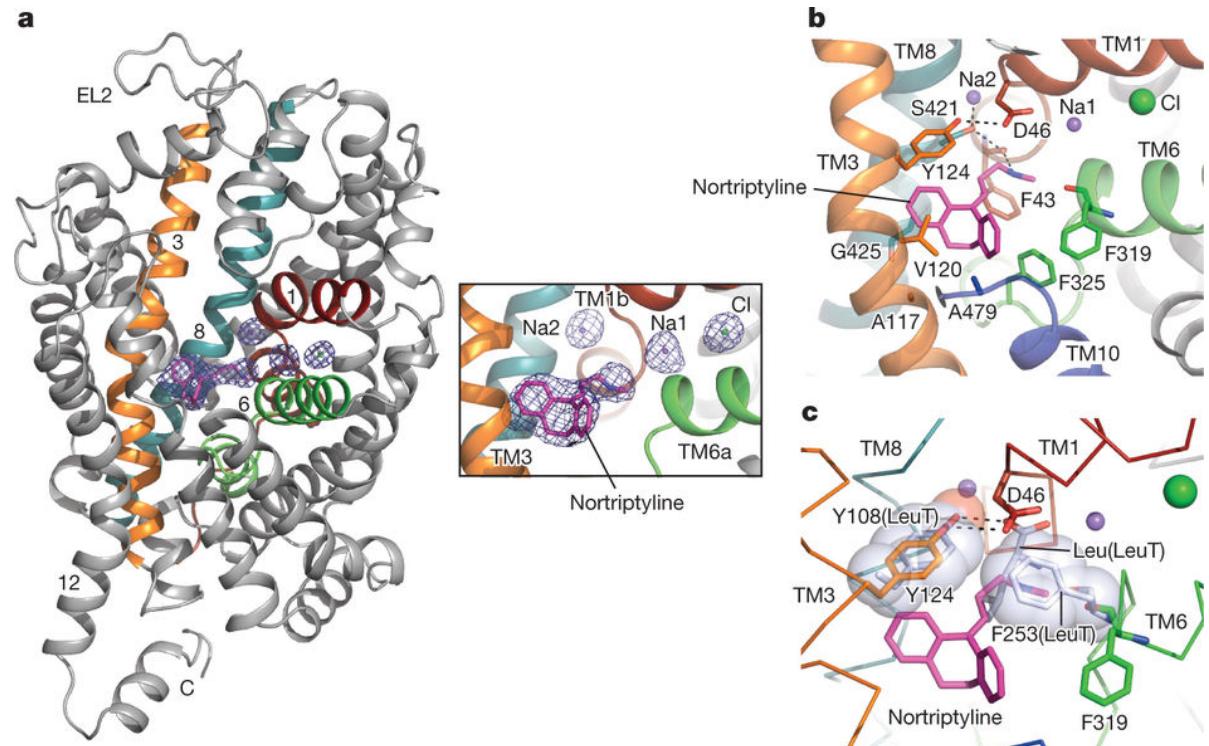
Drugs are small bio-molecules

Traditional techniques:

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

Modern techniques

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

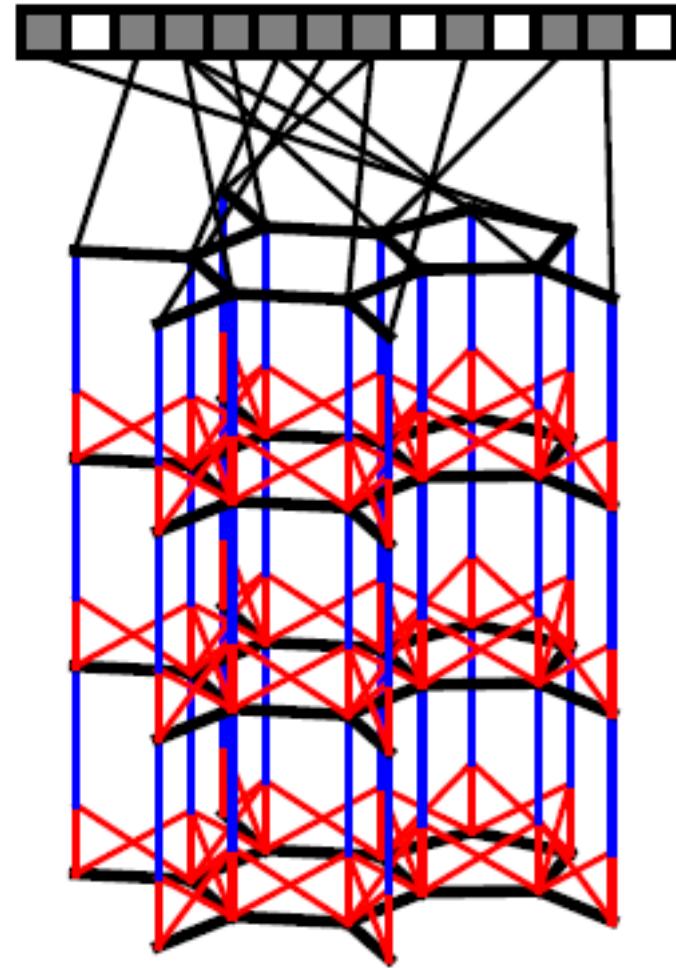


#REF: Penmatsa, Aravind, Kevin H. Wang, and Eric Gouaux. "X-ray structure of dopamine transporter elucidates antidepressant mechanism." *Nature* 503.7474 (2013): 85-90.

Molecular fingerprints

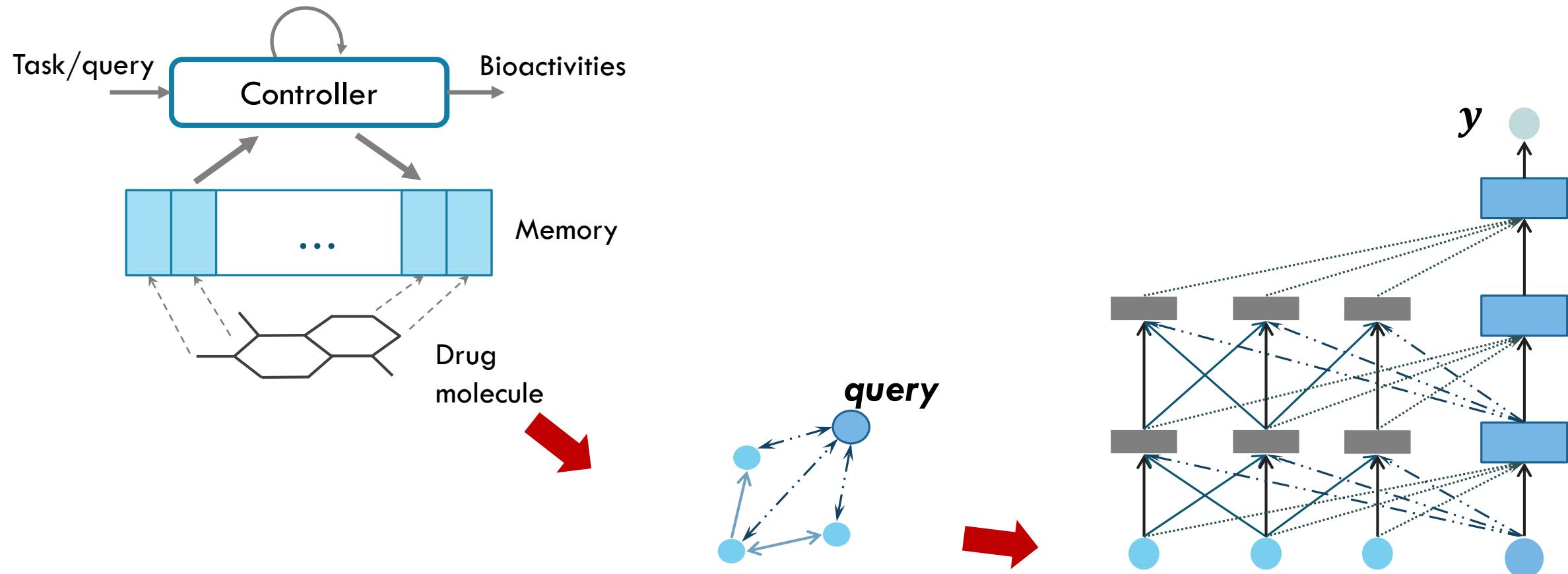
Algorithm 1 Circular fingerprints.

```
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



#Ref: Pham, Trang, Truyen Tran, and Svetha Venkatesh. "Graph Memory Networks for Molecular Activity Prediction." ICPR'18.

Graph memory networks: Results

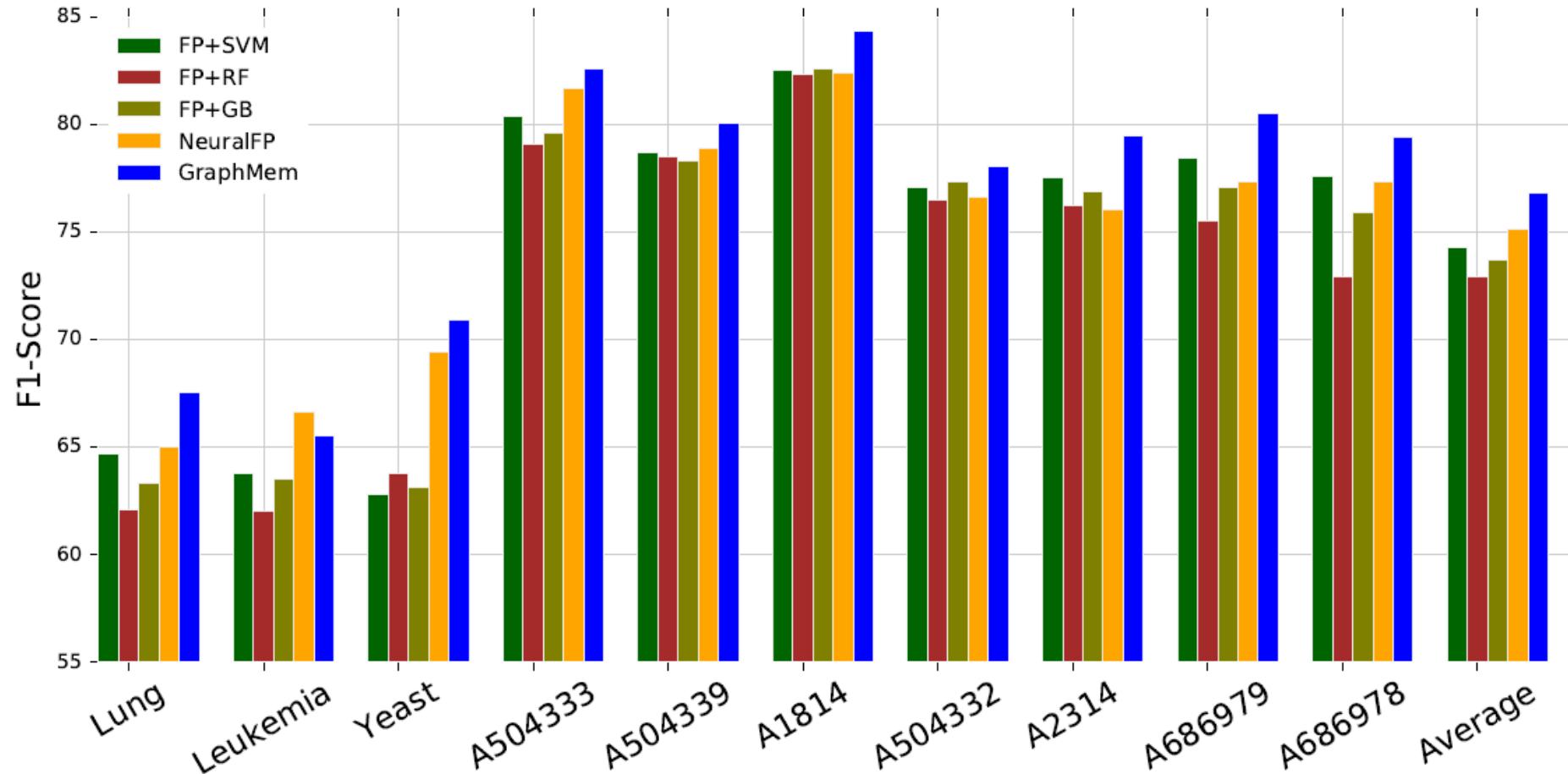
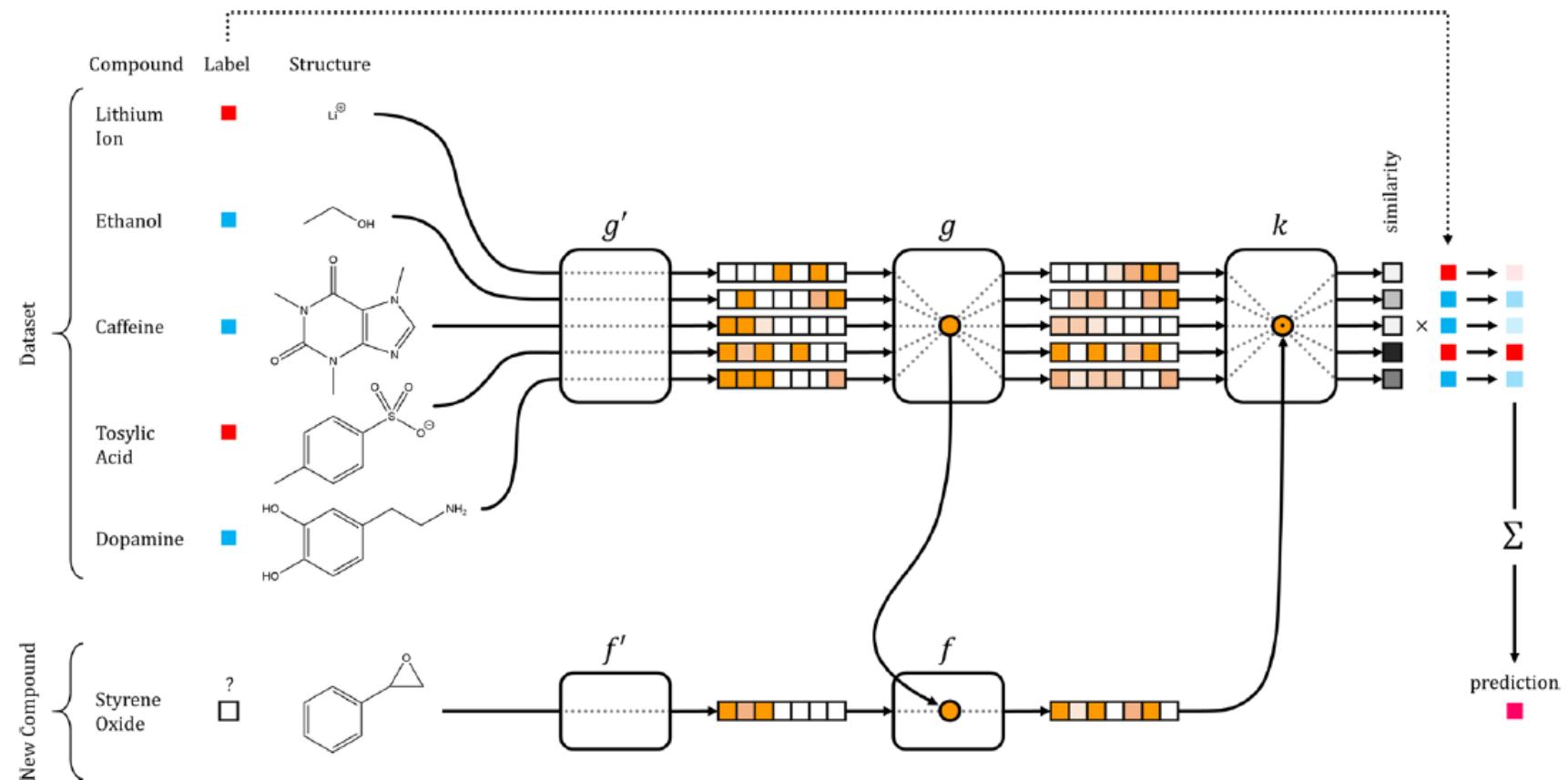


Figure 2: F1-score (%) for NCI datasets. FP = Fingerprint; RF = Random Forests; GBM = Gradient Boosting Machine. Best view in color.

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

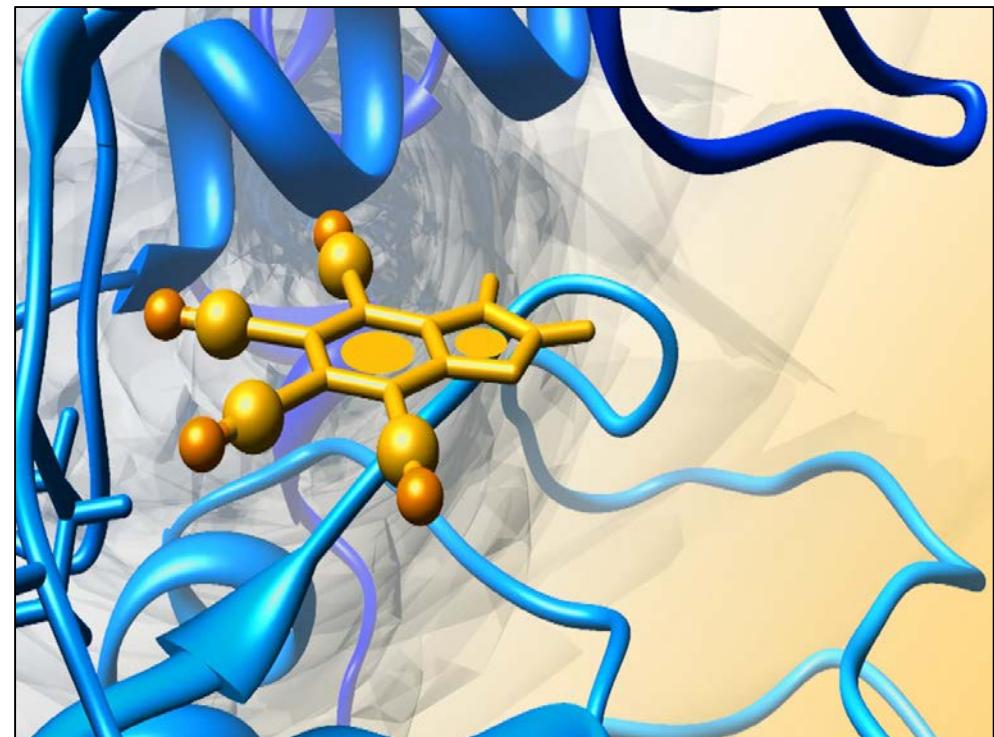
The space of drugs is estimated to be $1e+23$ to $1e+60$

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

It is impossible to model this space fully.

The current technologies are not mature for graph generations.

But approximate techniques do exist.



Source: pharmafactz.com

Generating molecular graphs

No regular, fixed-size structures

Graphs are ***permutation invariant***:

- #permutations are exponential function of #nodes
- The probability of a generated graph G need to be marginalized over all possible permutations

Aims for:

- **Diversity** of generated graphs
- **Smoothness** of latent space

Approaches

Representing graph as string,
and use sequence VAEs or
GANs.

Graph VAE & GAN

- Model nodes & interactions
- Model cliques

Sequences

- Iterative methods

Reinforcement learning

- Discrete objectives

Any combination of these +
memory.

Molecule → string

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

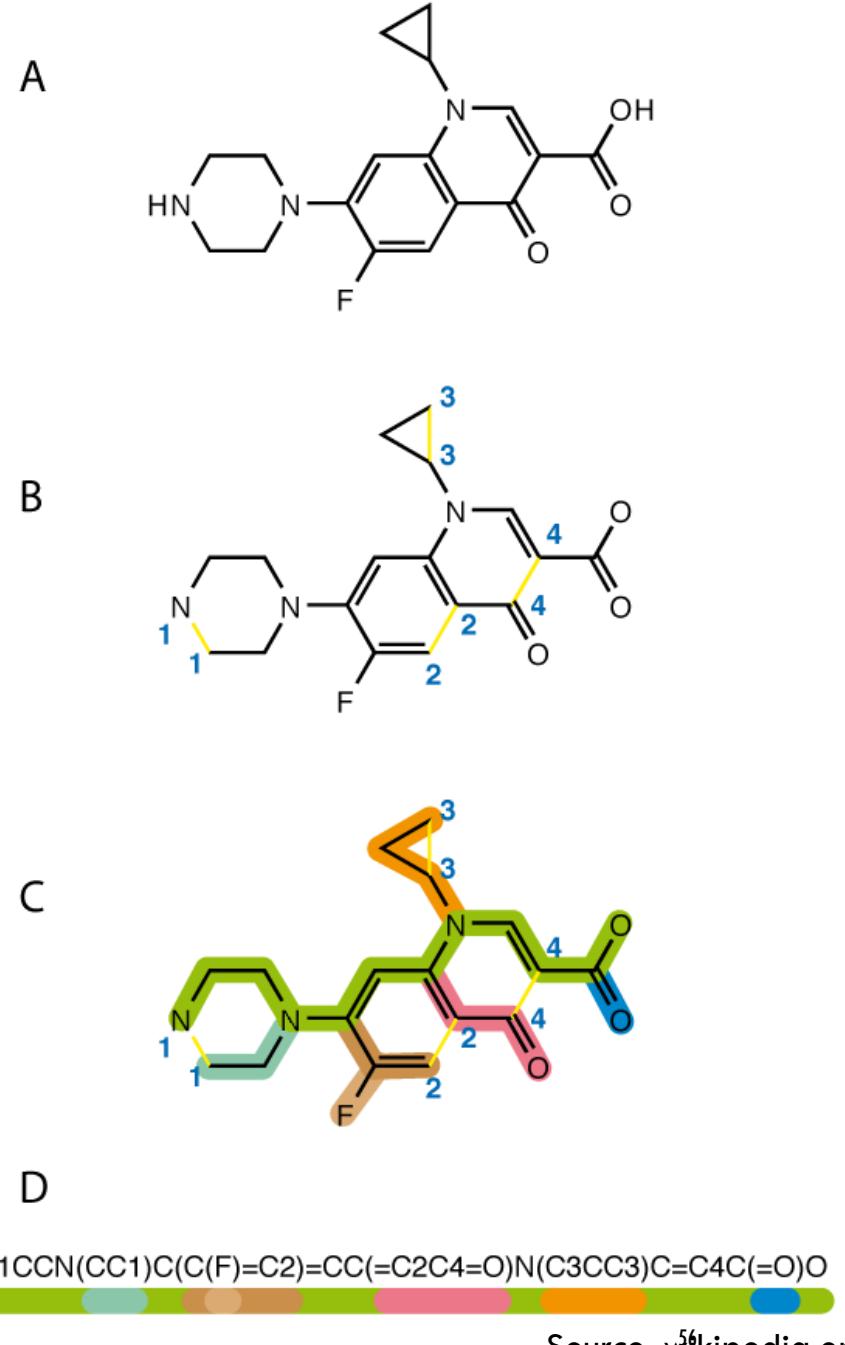
- SMILES = Simplified Molecular-Input Line-Entry System

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

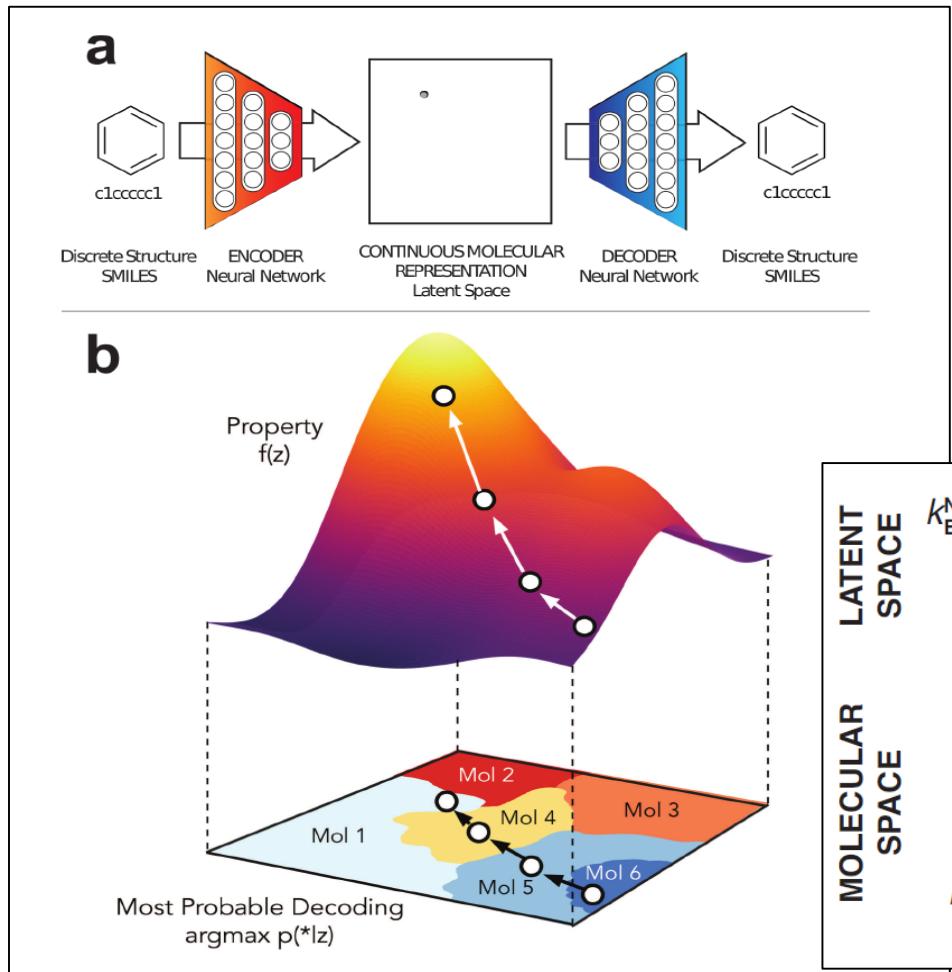
- Allow easy optimization on the continuous space
- Problem: String → graphs is not unique!

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).

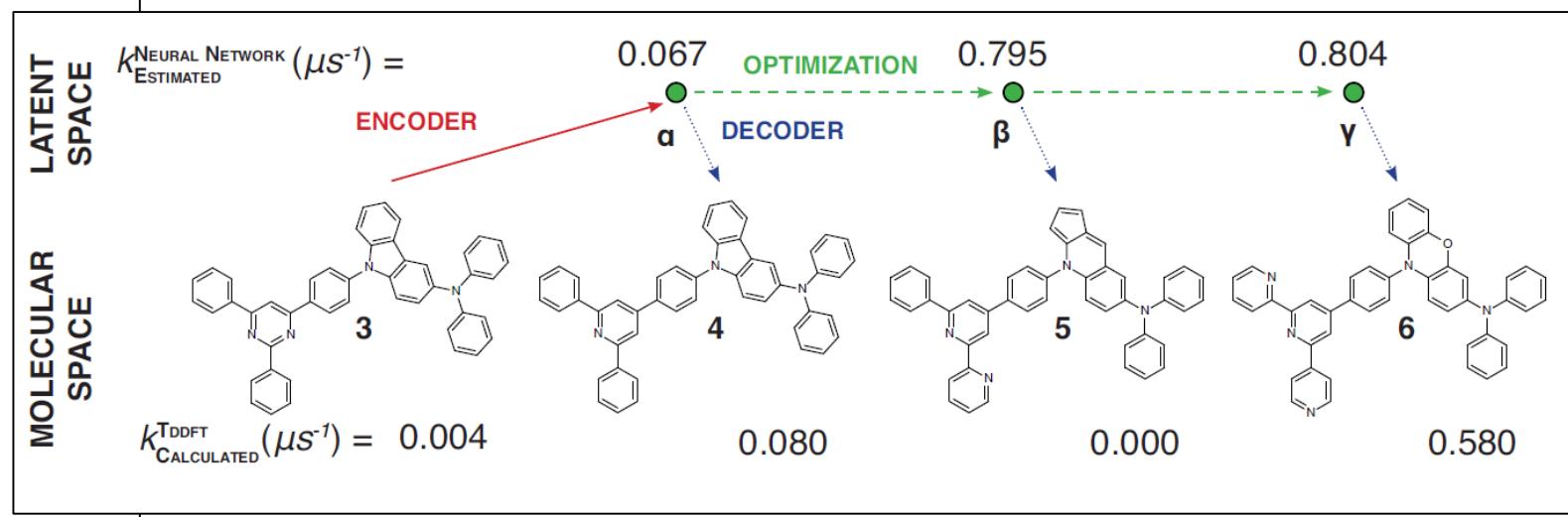


VAE for drug space modelling



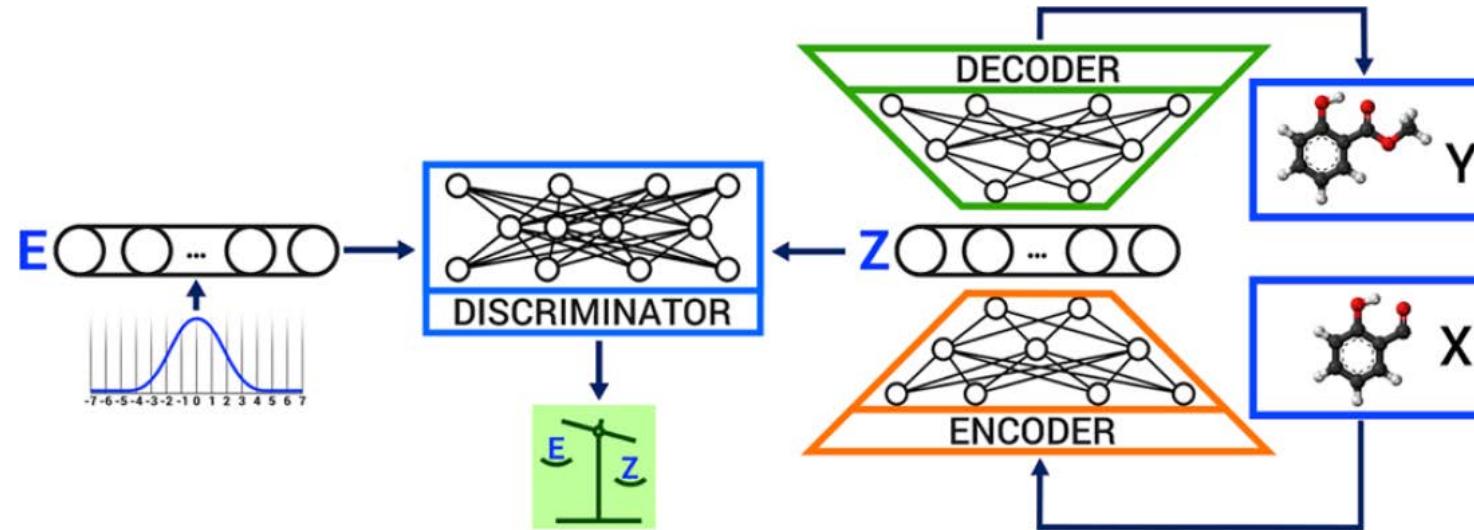
Uses VAE for sequence-to-sequence.

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



Adversarial Autoencoders

Kadurin et al. Molecular pharmaceutics 2017



Input of the encoder : the fingerprint of a molecule

The decoder outputs the predicted fingerprint .

The generative model generates a vector E, which is then discriminated from the latent vector of the real molecule by the discriminator.

GraphVAE

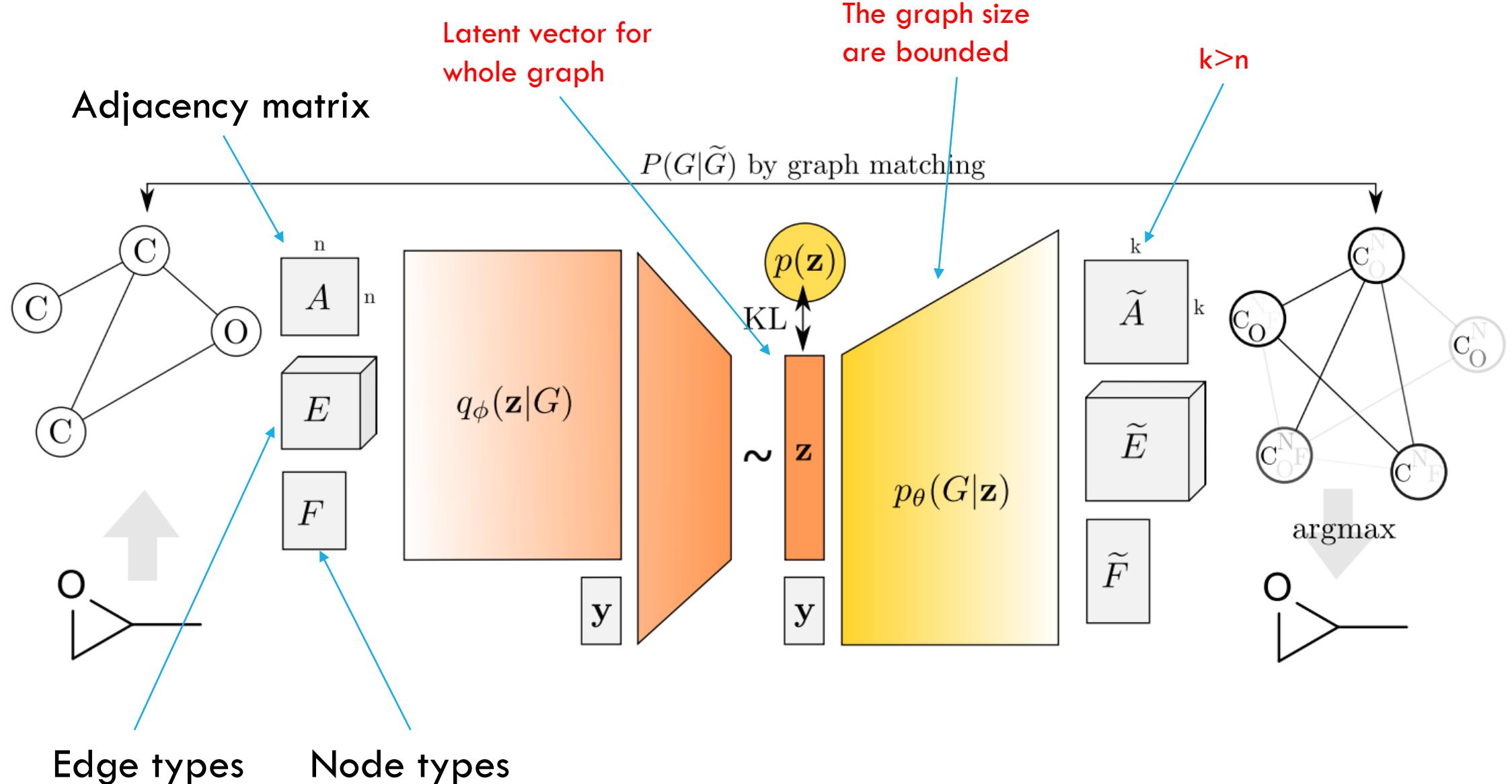
Handle irregular structures?: Predict the whole adjacency matrix, node types and edge types

Variable size graph?: Yes, but bounded by the size of the largest graph in training data.

Handle permutation invariance?: Matching every pair of nodes in 2 graphs

Diversity? To some extent. | Smoothness? No clue.

#REF: Simonovsky, M., & Komodakis, N. (2018). GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders. *arXiv preprint arXiv:1802.03480*.



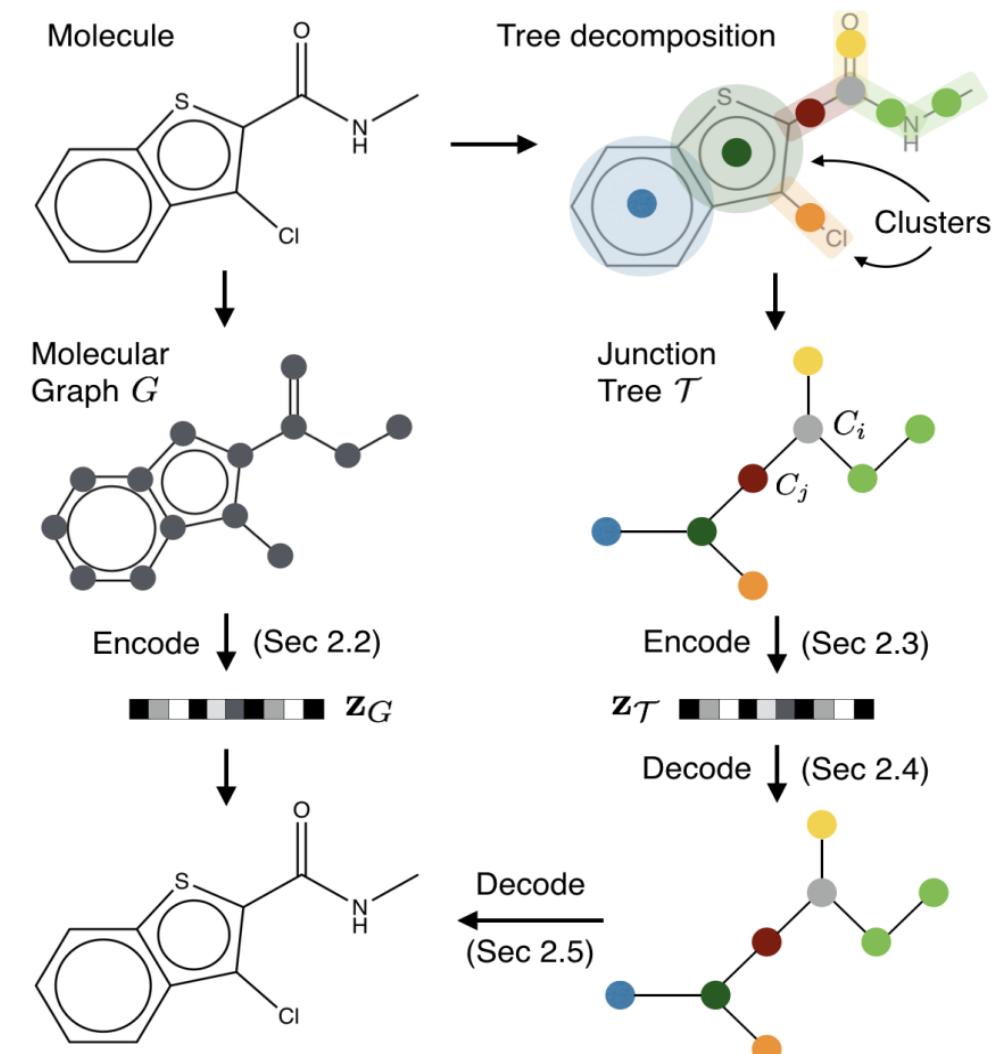
#REF: Simonovsky, M., & Komodakis, N. (2018). GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders. *arXiv preprint arXiv:1802.03480*.

Junction tree VAE

Cluster vocab:

- rings
- bonds
- atoms

Jin, W., Barzilay, R., & Jaakkola, T. (2018). Junction Tree Variational Autoencoder for Molecular Graph Generation. *arXiv preprint arXiv:1802.04364*.



Algorithm 2 Tree decomposition of molecule $G = (V, E)$

$V_1 \leftarrow$ the set of bonds $(u, v) \in E$ that do not belong to any rings.

$V_2 \leftarrow$ the set of simple rings of G .

for r_1, r_2 **in** V_2 **do**

Merge rings r_1, r_2 into one ring if they share more than two atoms (bridged rings).

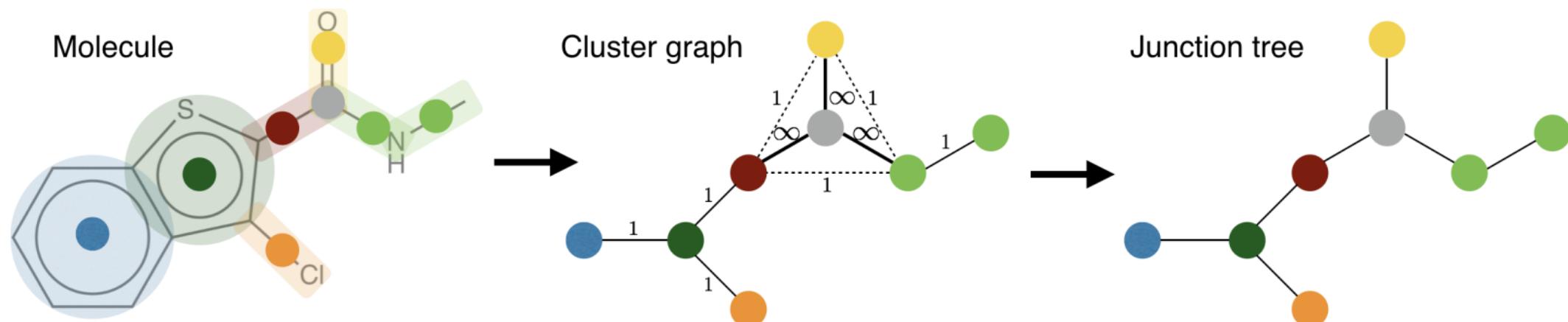
end for

$V_0 \leftarrow$ atoms being the intersection of three or more clusters in $V_1 \cup V_2$.

$\mathcal{V} \leftarrow V_0 \cup V_1 \cup V_2$

$\mathcal{E} \leftarrow \{(i, j, c) \in \mathcal{V} \times \mathcal{V} \times \mathbb{R} \mid |i \cap j| > 0\}$. Set $c = \infty$ if $i \in V_0$ or $j \in V_0$, and $c = 1$ otherwise.

Return The maximum spanning tree over cluster graph $(\mathcal{V}, \mathcal{E})$.



Jin, W., Barzilay, R., & Jaakkola, T. (2018). Junction Tree Variational Autoencoder for Molecular Graph Generation. *arXiv preprint arXiv:1802.04364*.

Agenda

Topic 1: Introduction (20 mins)

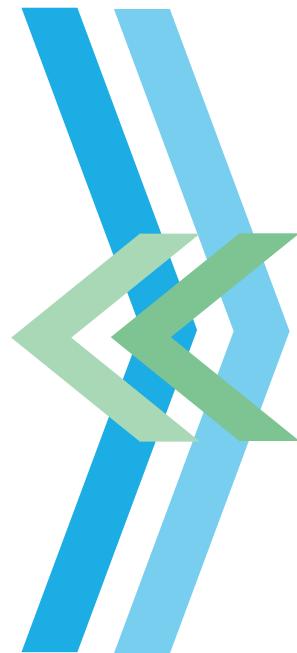
Topic 2: Brief review of deep learning (30 mins)

- Classic architectures
- Capsules & graphs
- Memory & attention

Topic 3: Genomics (30 mins)

- Nanopore sequencing
- Genomics modelling

QA (10 mins)



Break (30 mins)

Topic 4: Healthcare (40 mins)

- Time series (regular & irregular)
- EMR analysis: Trajectories prediction
- EMR analysis: Sequence generation

Topic 5: Data efficiency (40 mins)

- Few-shot learning
- Generative models
- Unsupervised learning of drugs

Topic 6: Future outlook

QA (10 mins)

Living in the future: AI for health care

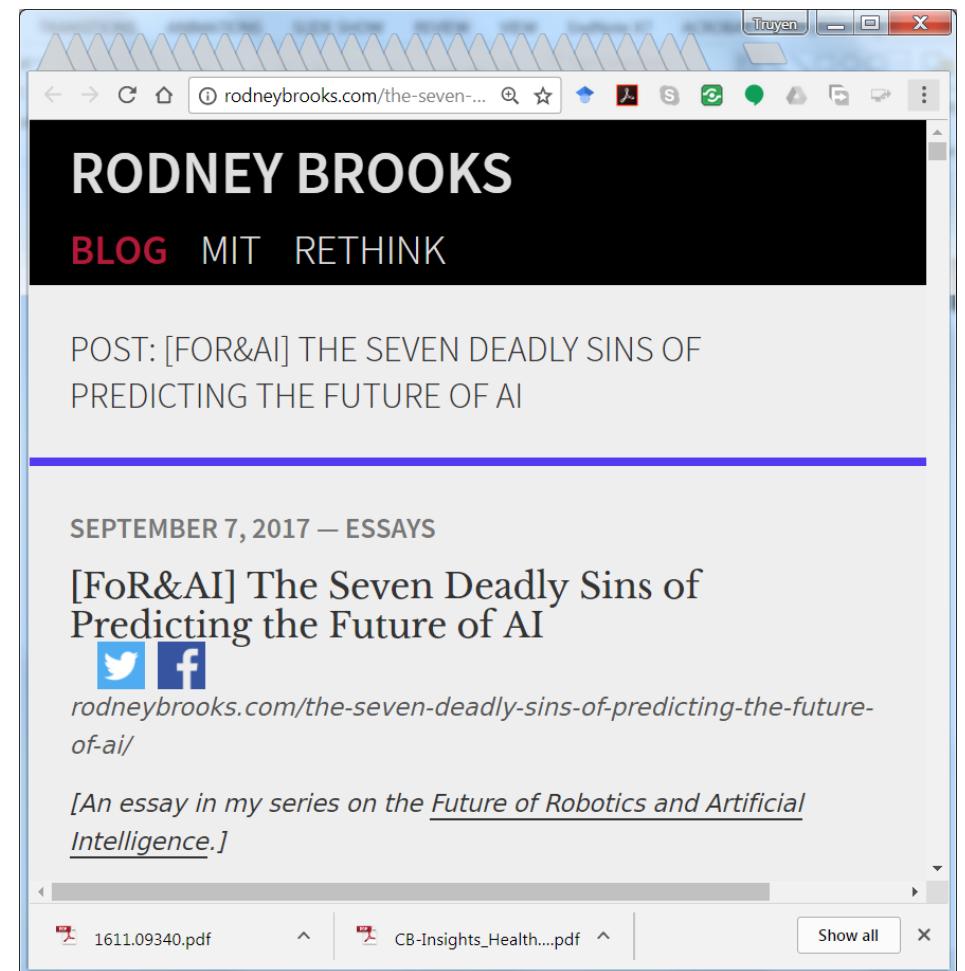
Some speculations (by me):

- <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 personalized medicine

Will this patient respond 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)?

Can we synthesize a drug for the patient **on-demand?**

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

Towards a dialog system → Replace GP?

Leveraging existing knowledge

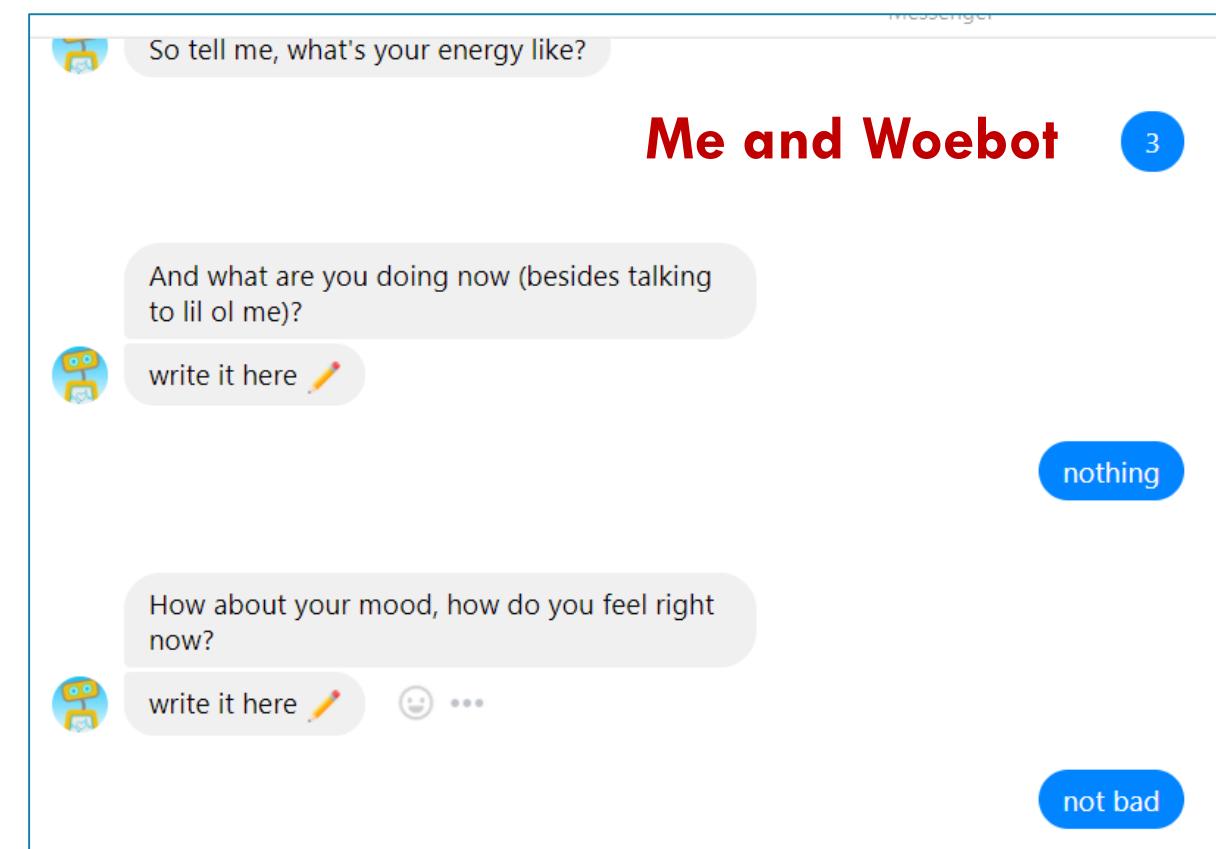
- Medical knowledge bases
- Medical texts
- Probably needs to build knowledge bases from text

Personalizing through EMRs

- Learn from hospitals data

Ask right questions → Finding answers from databases → Generating dialog

Never ending learning (NEL).





Thank you!

We're hiring

PhD & Postdocs

truyen.tran@deakin.edu.au

The team that helps with slides



References

- 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
- Altae-Tran, Han, et al. "Low Data Drug Discovery with One-Shot Learning." *ACS central science* 3.4 (2017): 283-293.
- Angermueller, Christof, et al. "Deep learning for computational biology." *Molecular systems biology* 12.7 (2016): 878.
- Bowman, Samuel R., et al. "Generating sentences from a continuous space." *arXiv preprint arXiv:1511.06349* (2015).
- Choi, Edward, et al. "Doctor ai: Predicting clinical events via recurrent neural networks." *Machine Learning for Healthcare Conference*. 2016.
- Choi, Edward, et al. "RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism." *Advances in Neural Information Processing Systems*. 2016.
- Duvenaud, David K., et al. "Convolutional networks on graphs for learning molecular fingerprints." *Advances in neural information processing systems*. 2015.
- Gupta, Anvita, et al. "Generative Recurrent Networks for De Novo Drug Design." *Molecular Informatics* (2017).
- Gómez-Bombarelli, Rafael, et al. "Automatic chemical design using a data-driven continuous representation of molecules." *arXiv preprint arXiv:1610.02415* (2016).
- Harutyunyan, Hrayr, et al. "Multitask Learning and Benchmarking with Clinical Time Series Data." *arXiv preprint arXiv:1703.07771* (2017).
- Hung Le, Truyen Tran, and Svetha Venkatesh. "Dual Memory Neural Computer for Asynchronous Two-view Sequential Learning." *KDD18*.

References (2)

Hung Le, Truyen Tran, and Svetha Venkatesh. "Dual Control Memory Augmented Neural Networks for Treatment Recommendations", *PAKDD18*.

Jin, W., Barzilay, R., & Jaakkola, T. (2018). Junction Tree Variational Autoencoder for Molecular Graph Generation. *arXiv preprint arXiv:1802.04364*.

Kadurin, A., Aliper, A., Kazennov, A., Mamoshina, P., Vanhaelen, Q., Khrabrov, K., & Zhavoronkov, A. (2017). The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget*, 8(7), 10883.

Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A., & Zhavoronkov, A. (2017). druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. *Molecular pharmaceutics*, 14(9), 3098-3104.

Kien Do, et al. "Attentional Multilabel Learning over Graphs-A message passing approach." *arXiv preprint arXiv:1804.00293*(2018).

Kien Do, Truyen Tran, Svetha Venkatesh, "Learning Deep Matrix Representations" *arXiv preprint arXiv:1703.01454*

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

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

Litjens, Geert, et al. "A survey on deep learning in medical image analysis." *arXiv preprint arXiv:1702.05747* (2017). 308 papers surveyed, 242 published in 2016 or Jan of 2017.

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.

References (3)

- Penmatsa, Aravind, Kevin H. Wang, and Eric Gouaux. "X-ray structure of dopamine transporter elucidates antidepressant mechanism." *Nature* 503.7474 (2013): 85-90.
- Pham, Trang, Truyen Tran, and Svetha Venkatesh. "Graph Memory Networks for Molecular Activity Prediction." *ICPR'18*.
- 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
- Phuoc Nguyen, Truyen Tran, and Svetha Venkatesh. 'Resset: A Recurrent Model for Sequence of Sets with Applications to Electronic Medical Records.' *IJCNN* (2018).
- 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)*
- Roses, Allen D. "Pharmacogenetics in drug discovery and development: a translational perspective." *Nature reviews Drug discovery* 7.10 (2008): 807-817.
- Segler, Marwin HS, et al. "Generating focussed molecule libraries for drug discovery with recurrent neural networks." *arXiv preprint arXiv:1701.01329* (2017).
- Simonovsky, M., & Komodakis, N. (2018). GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders. *arXiv preprint arXiv:1802.03480*.
- 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
- Tu D. Nguyen, Truyen Tran, D. Phung, and S. Venkatesh, "Tensor-variate Restricted Boltzmann Machines", *AAAI 2015*
- Zhang et al., "Leap: Learning to prescribe effective and safe treatment combinations for multimorbidity", *KDD17*.