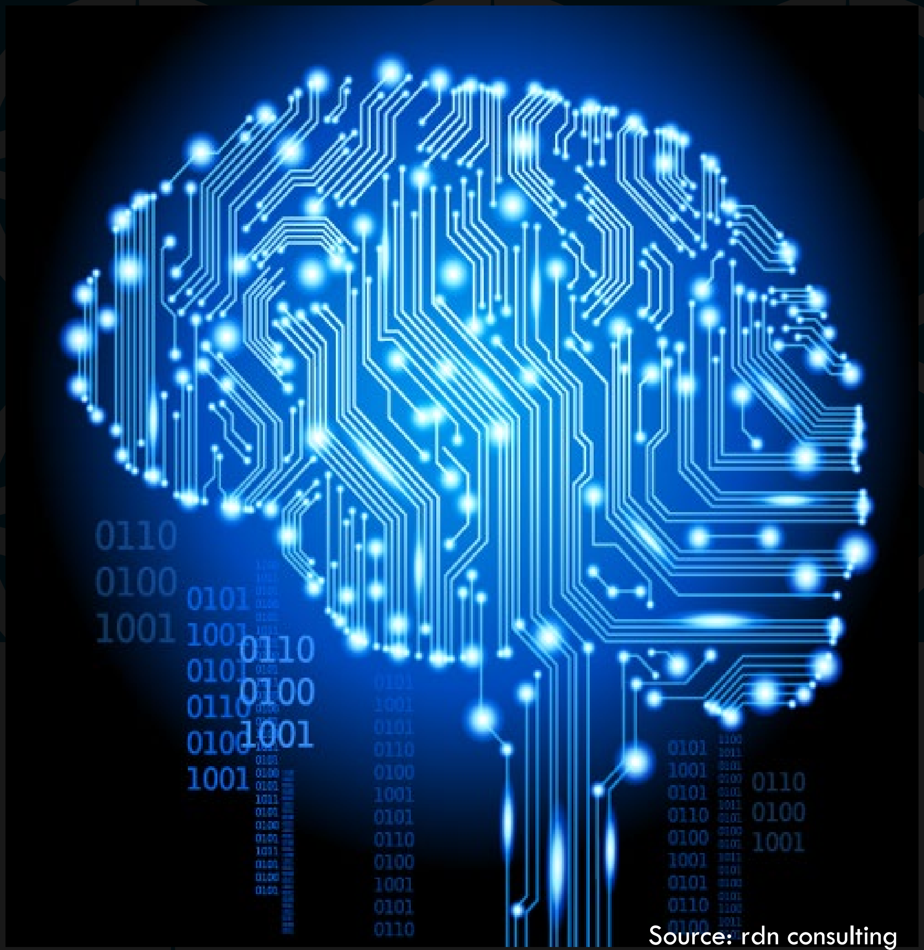
DEEP NEURAL NETS FOR HEALTHCARE

Truyen Tran
Deakin University

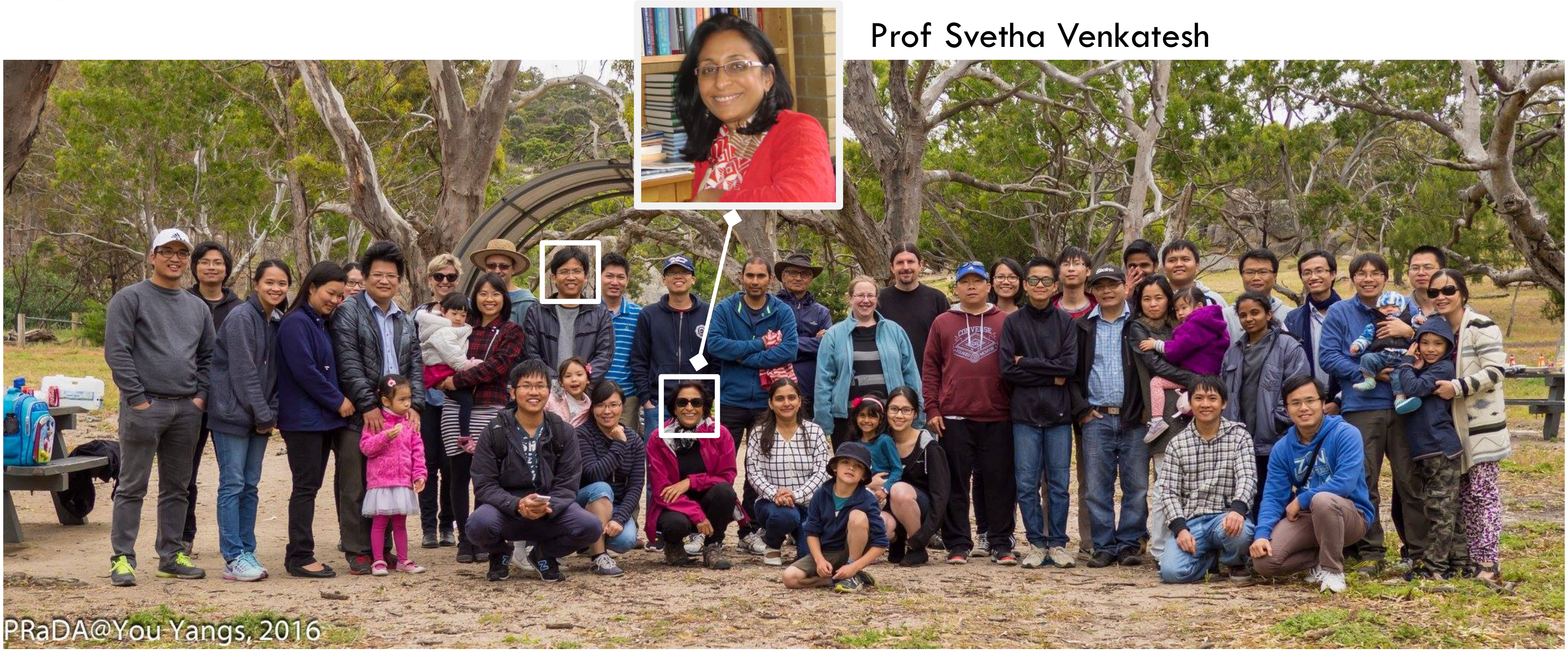
truyen.tran@deakin.edu.au
truyentran.github.io
tranthetruyen
letdataspeak.blogspot.com
@truyenoz
goo.gl/3jJ1O0


Source: rdn consulting

Seattle, Feb 24th 2017

PRADA @ DEAKIN, MELBOURNE, AUSTRALIA

Prof Svetha Venkatesh



AGENDA

Introduction

- Our engagement in health
- Deep learning

Discovery

- Stable discovery of risk factors with Autoencoder
- Deepr - Discovery of predictive EMR motifs using CNN

Diagnosis

- EEG-based diagnosis with CNN + matrix-LSTM

Prognosis

- DeepCare - Health trajectory modelling
- Symbolic ICU - a symbolic representation of ICU time-series + deep nets

SOLVING HEALTH PROBLEMS IS VERY REWARDING

Barwon Health

Royal North Shore Hospital

Epworth HealthCare

Cerebral Palsy ALLIANCE

Formerly known as The Spastic Centre

Black Dog Institute

GARVAN INSTITUTE

Partnership

TOBY Autism Therapy

iHosp™

Startups

FFN,1986

Yann LeCun

CNN, 1988

Geoff Hinton

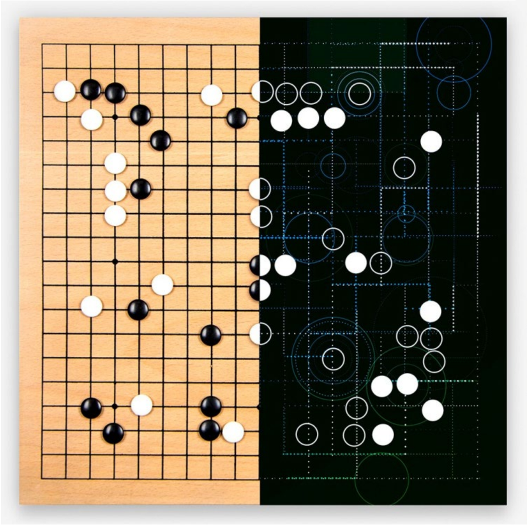
DBN,2006

Jurgen Schmidhuber

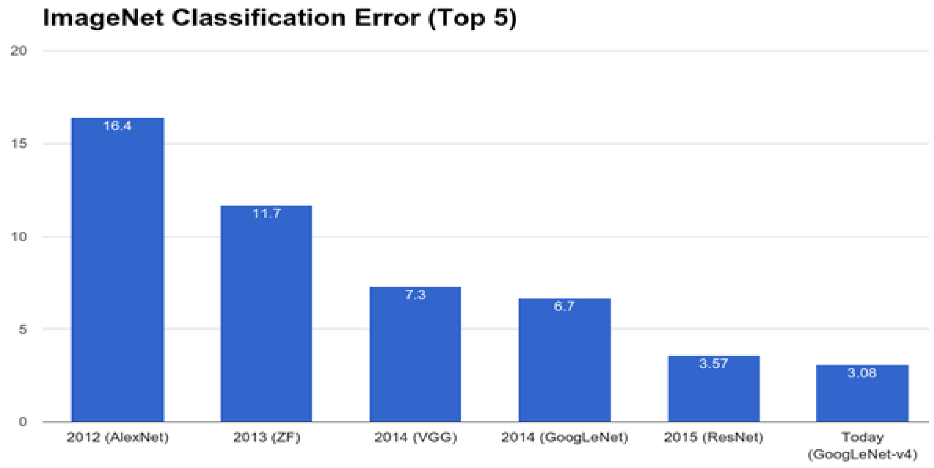
LSTM, 1997

AlexNet, 2012

AlphaGo



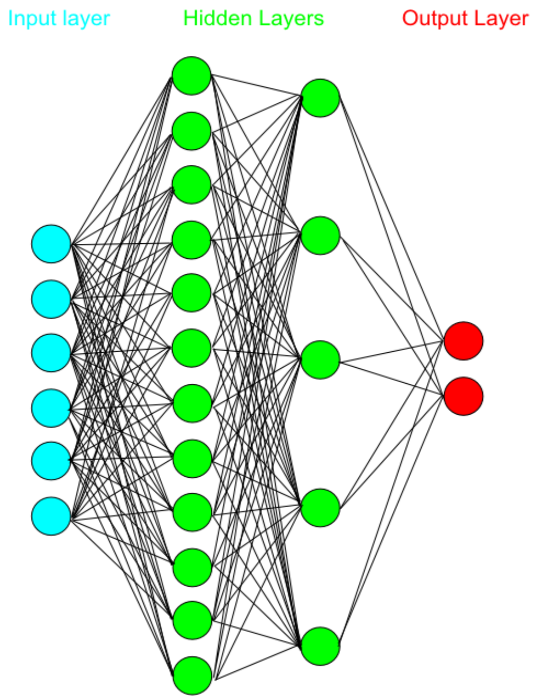
2016-2017



http://redcatlabs.com/2016-07-30_FifthElephant-DeepLearning-Workshop/#/

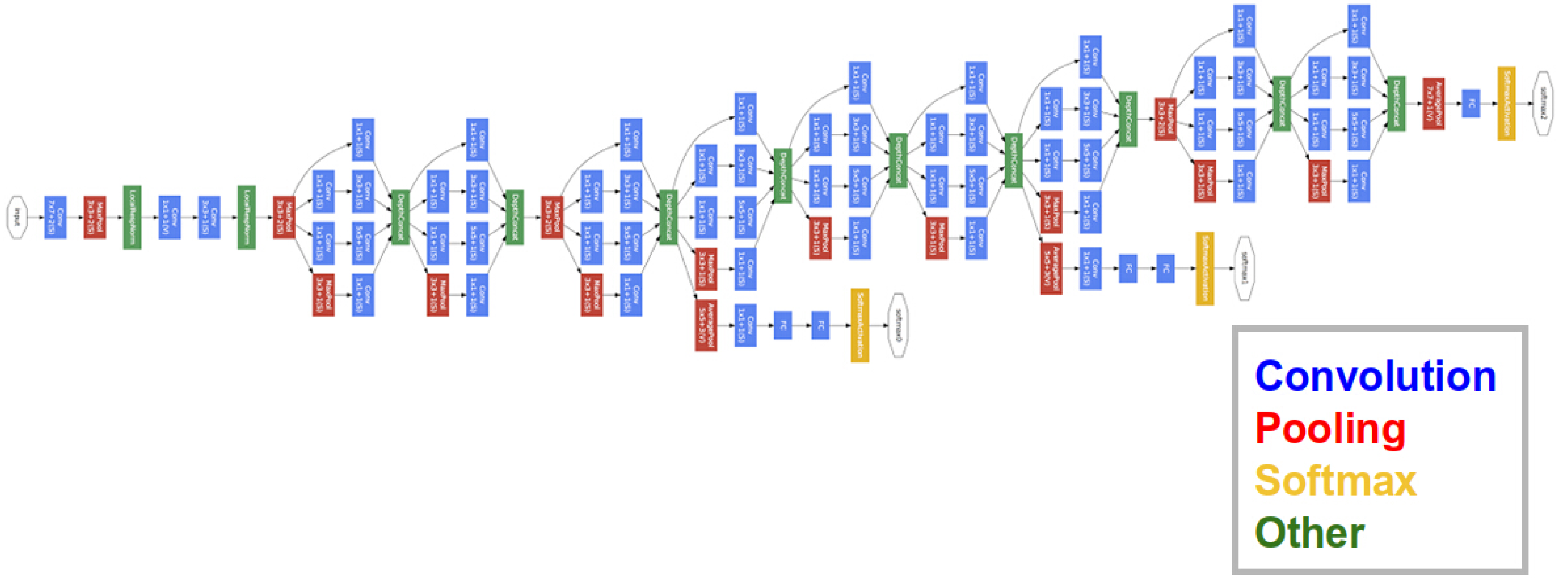
DEEP LEARNING IS NEURAL NETS, BUT MUCH HAS CHANGED

1986



http://blog.refu.co/wp-content/uploads/2009/05/mlp.png

2016



THE LEARNING IS ALSO CHANGING

Supervised learning

(mostly machine)

A →

Will be quickly solved for "easy" problems (Andrew Ng)

Unsupervised learning

...nan)

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

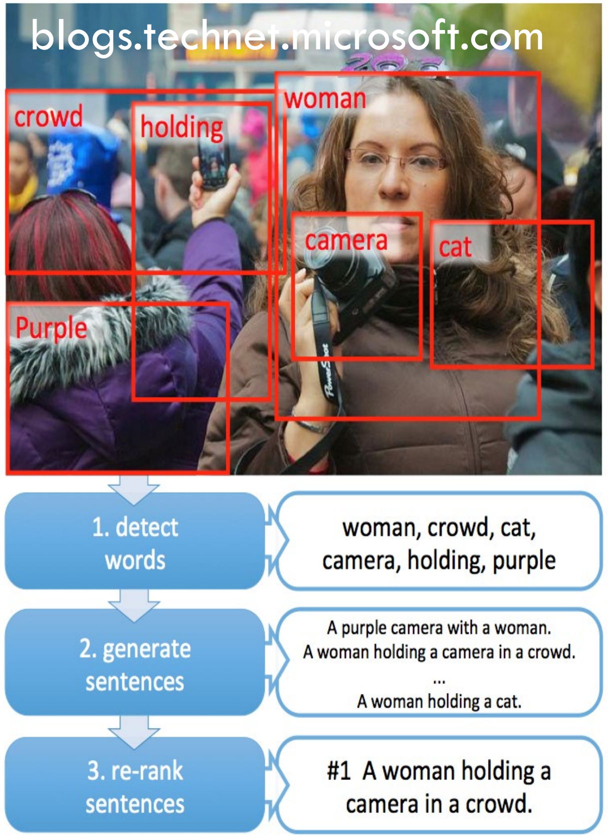
$$\ldots\mathbf{v}) \approx P_{data}(\mathbf{v})$$

Anywhere in between: semi-supervised learning, reinforcement learning, lifelong learning.

DEEP LEARNING IN COGNITIVE DOMAINS

Google Translate

Where human can recognise, act or answer accurately within seconds



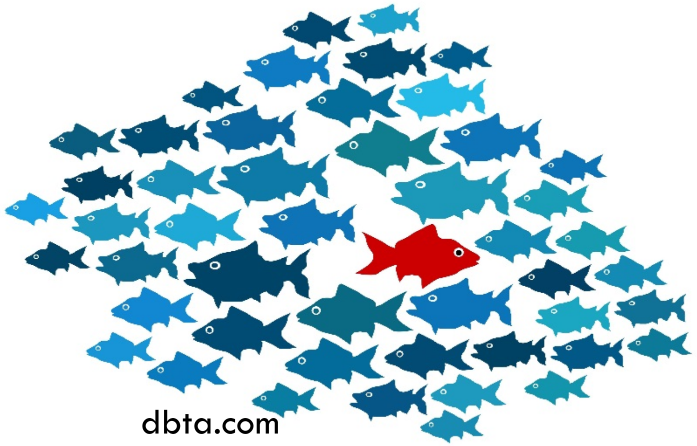
DEEP LEARNING IN NON-COGNITIVE DOMAINS

- **Where humans need extensive training to do well**
- Domains that demand transparency & interpretability.

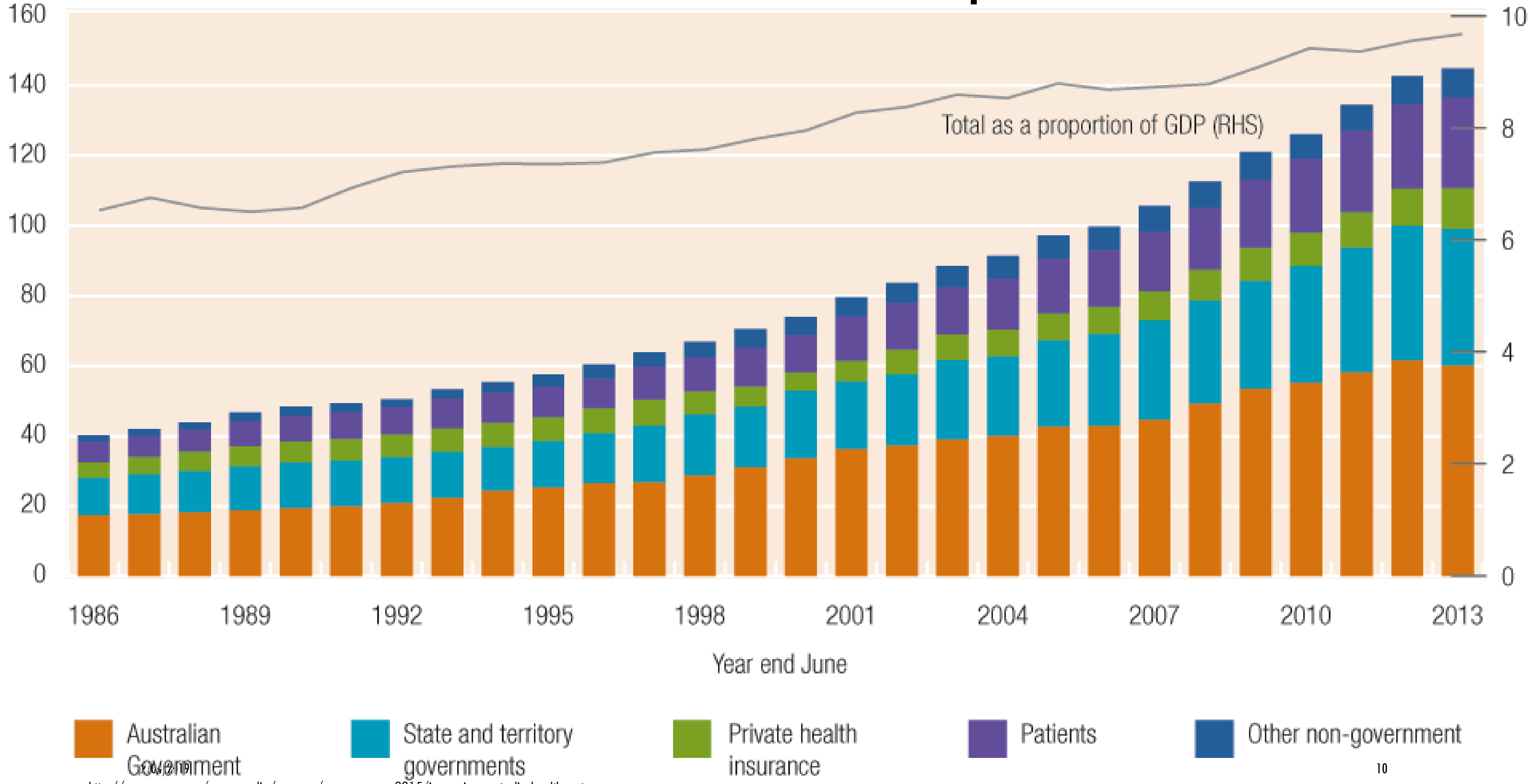
… <u>**healthcare**</u>

… security

… genetics, foods, water …

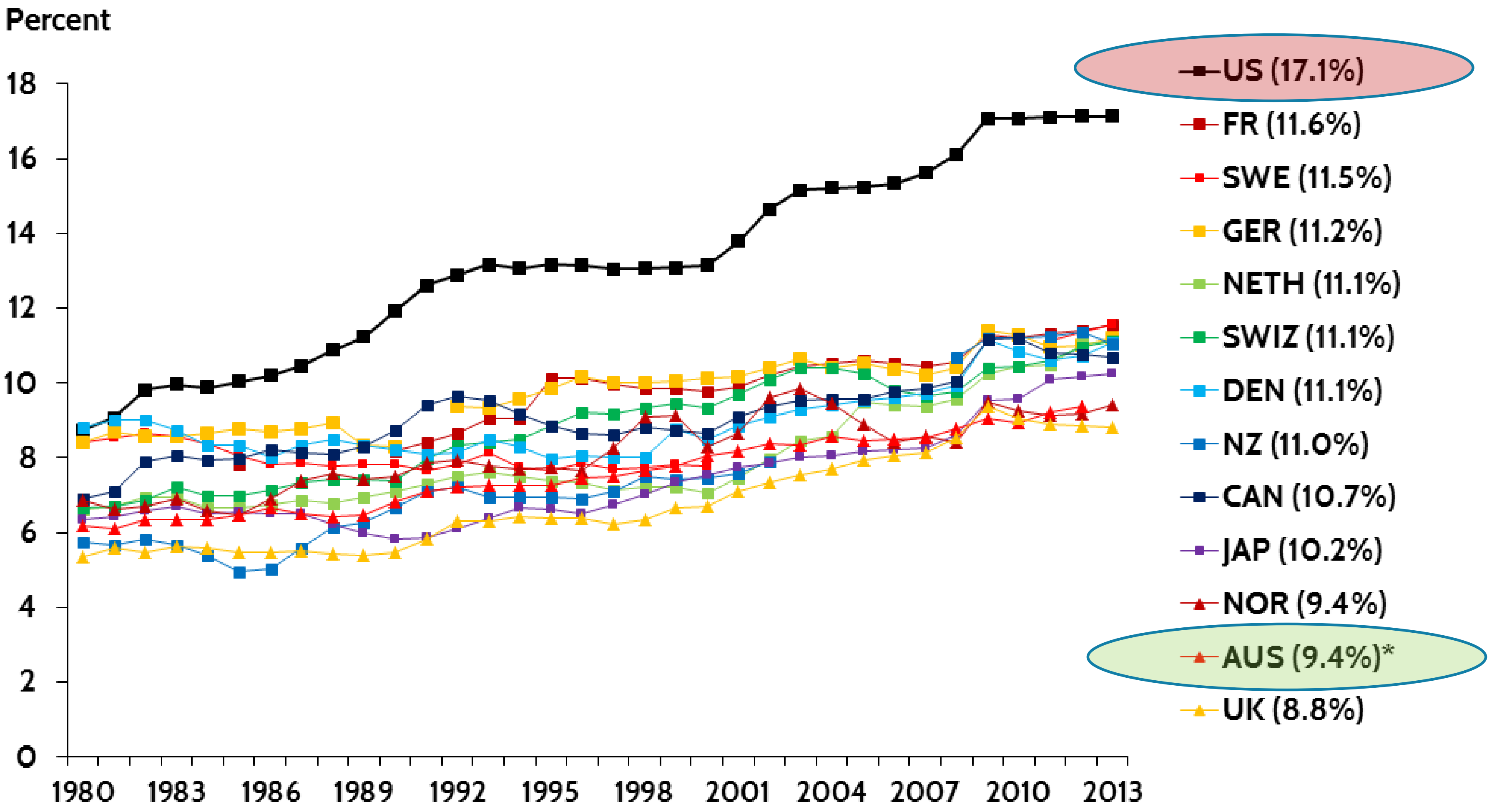


Australian health expenditure



http://www.pc.gov.au/news-media/pc-news/pc-news-may-2015/improving-australia-health-system

Exhibit 1. Health Care Spending as a Percentage of GDP, 1980–2013



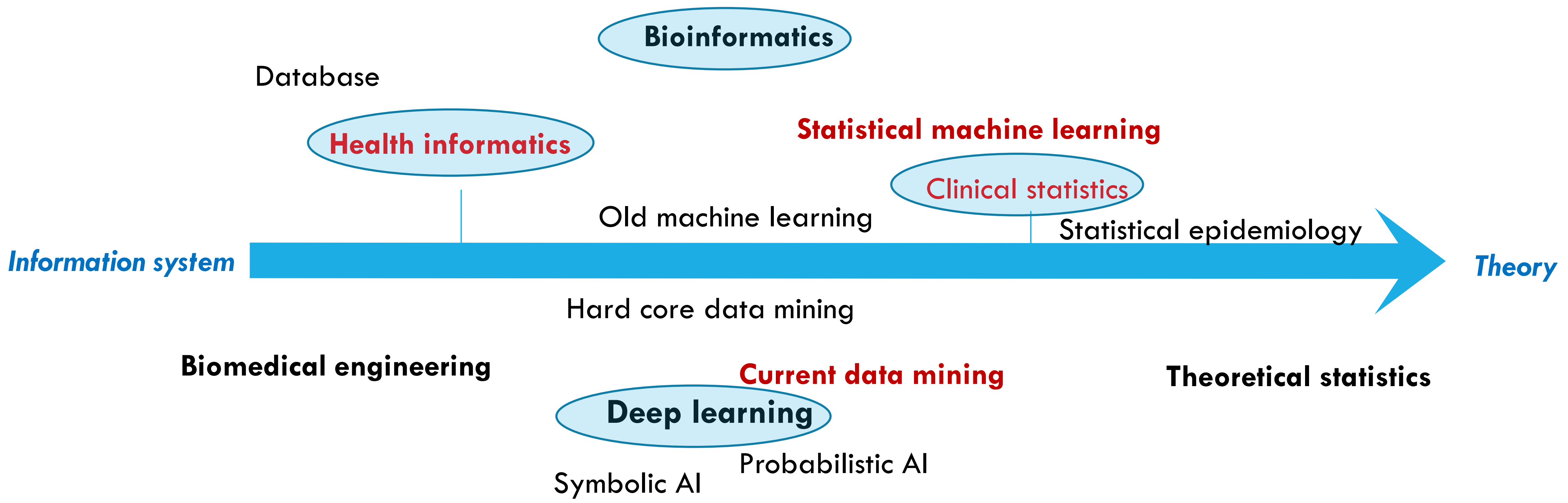
* 2012.

Notes: GDP refers to gross domestic product. Dutch and Swiss data are for current spending only, and exclude spending on capital formation of health care providers.

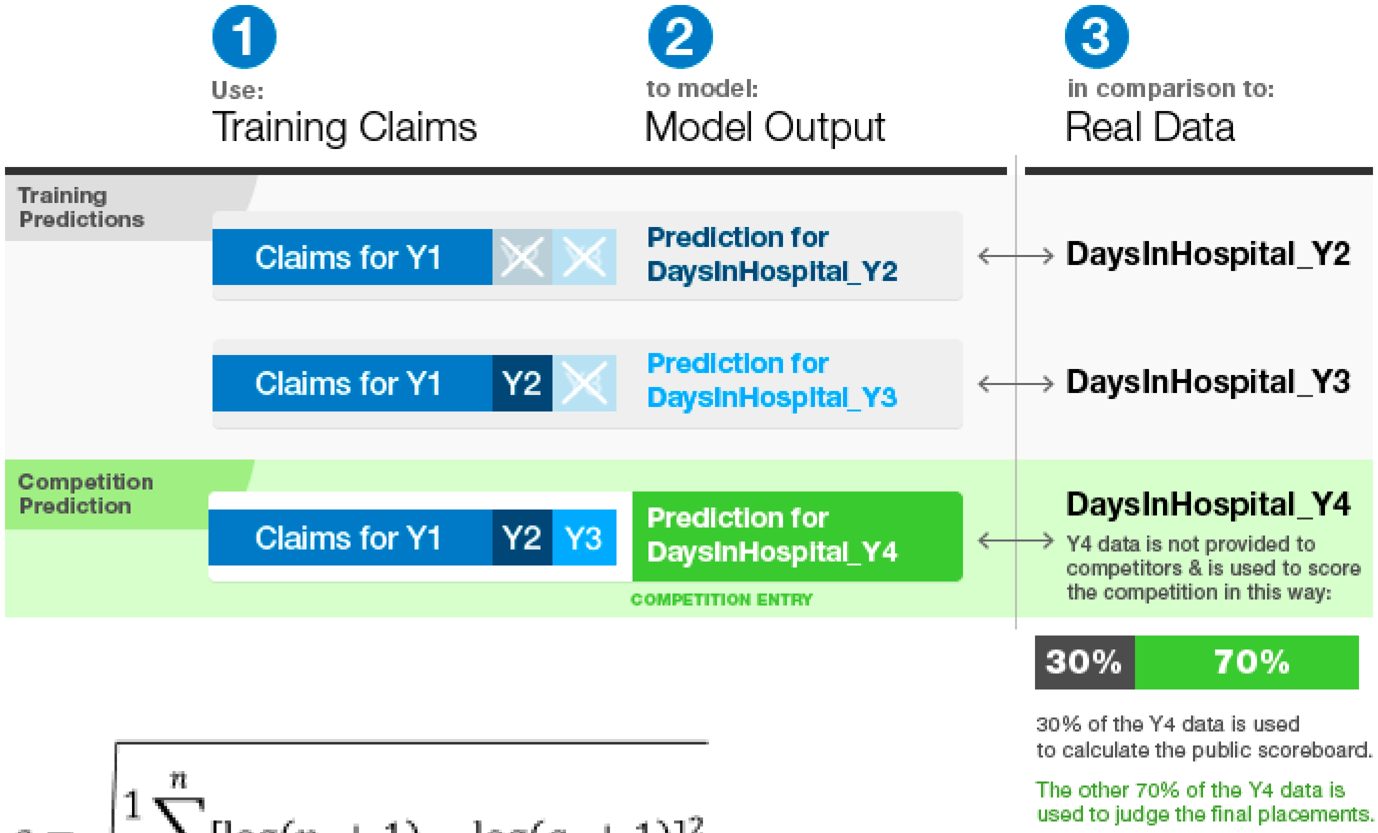
Source: OECD Health Data 2015.

http://www.commonwealthfund.org/publications/issue-briefs/2015/oct/us-health-care-from-a-global-perspective

HEALTHCARE ENGAGEMENT: SPEAK THEIR LANGUAGE(S)



HERITAGE HEALTH PRIZE ($3M, 2012-2013)



$$\varepsilon = \sqrt{\frac{1}{n}\sum_{i}^{n}[\log(p_i + 1) - \log(a_i + 1)]^2}$$

HERITAGE PROVIDER NETWORK HEALTH PRIZE

Truyen Tran | In the News | Judging Panel | Visit HPN

Dashboard

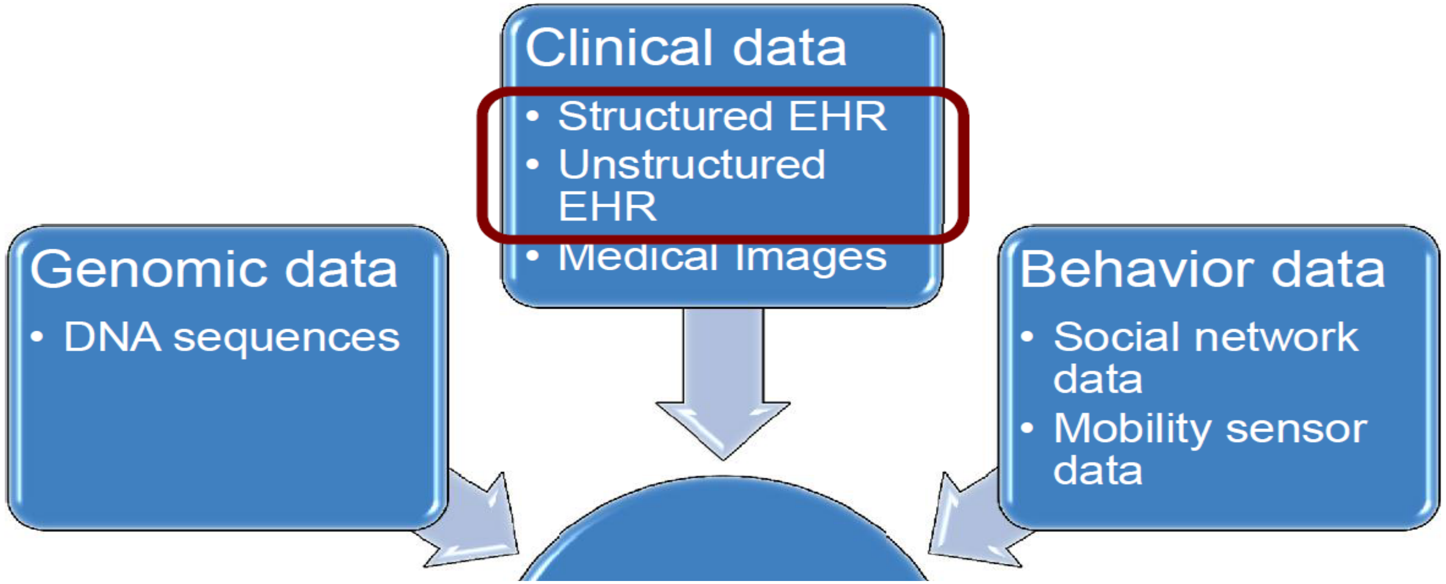
Leaderboard - Heritage Health Prize

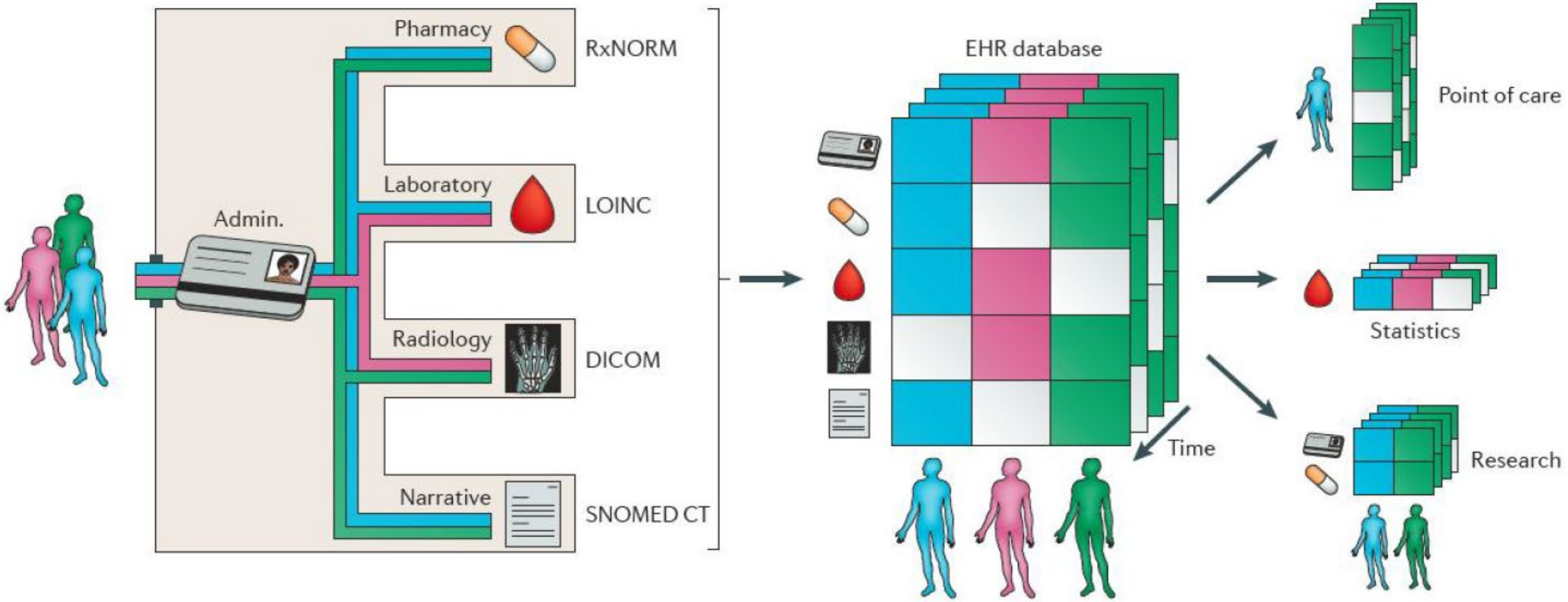
This competition has completed. This leaderbo

See someone using multiple accounts? Let us know.

- Heavy feature engineering
- Feature conjunction
- Gradient boosting
- No medical knowledge

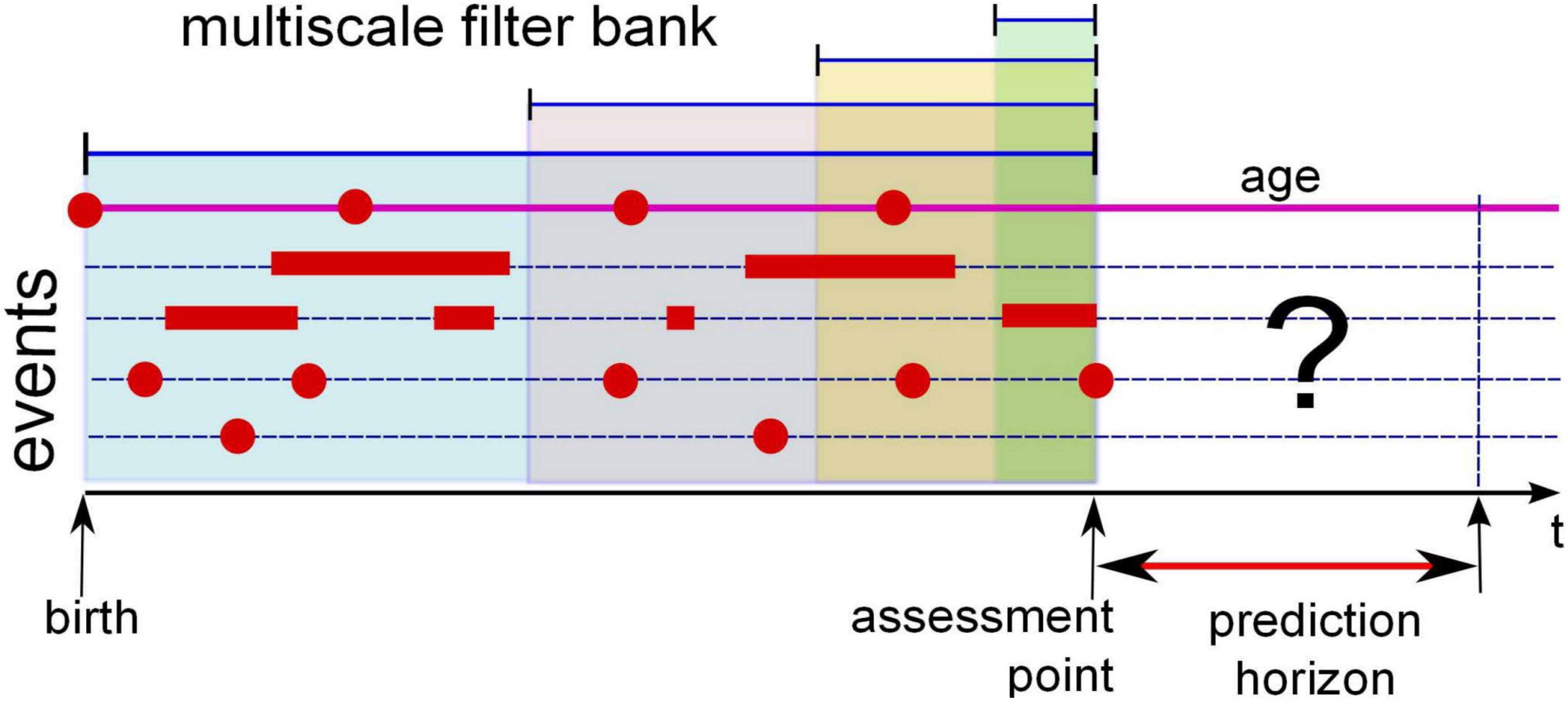
#	Δ1w	Team Name *in the money			t Submission UTC (Best – Last Submission)
1	-	POWERDOT *			, 04 Apr 2013 05:12:00 (-12.3d)
2	↑60	EXL Analytics			, 04 Apr 2013 00:06:09 (-3.4d)
3	↑15	J.A. Guerrero			, 04 Apr 2013 06:03:09
47	↓4	Midnight Run			i, 15 Feb 2013 02:18:14 (-194.5d)
48	↓4	PookyPANTS	0.467387	6	Fri, 03 Feb 2012 21:30:44
49	↑31	**Vietlabs**	**0.467543**	**8**	**Thu, 28 Mar 2013 22:36:51**
50	↓5	jsf	0.467545	18	Wed, 03 Apr 2013 17:31:42 (-118d)





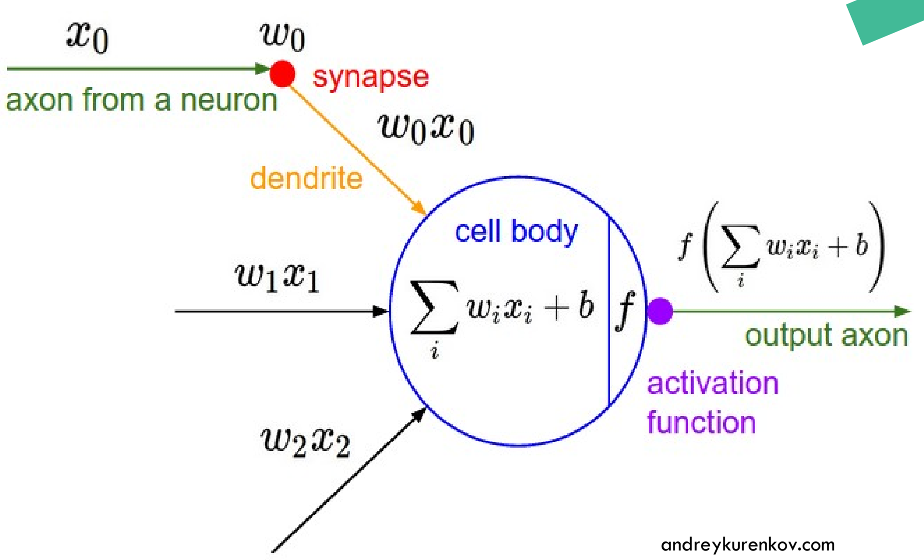
Source: Sun & Reddy, Big Data Analytics for Healthcare, Tutorial at SDM'13

FEATURE ENGINEERING (2012-2014)

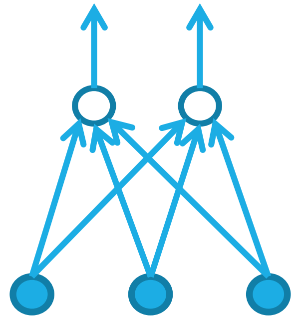


ENTER DEEP LEARNING AS FEATURE LEARNING

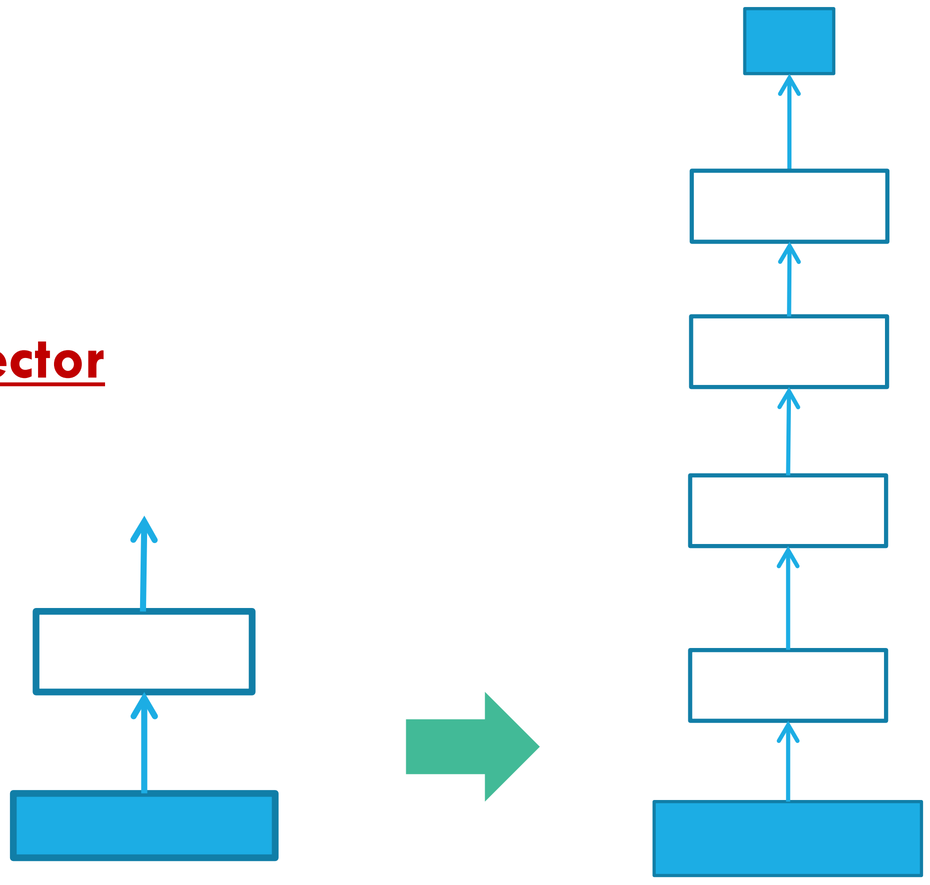
Integrate-and-fire neuron



andreykurenkov.com

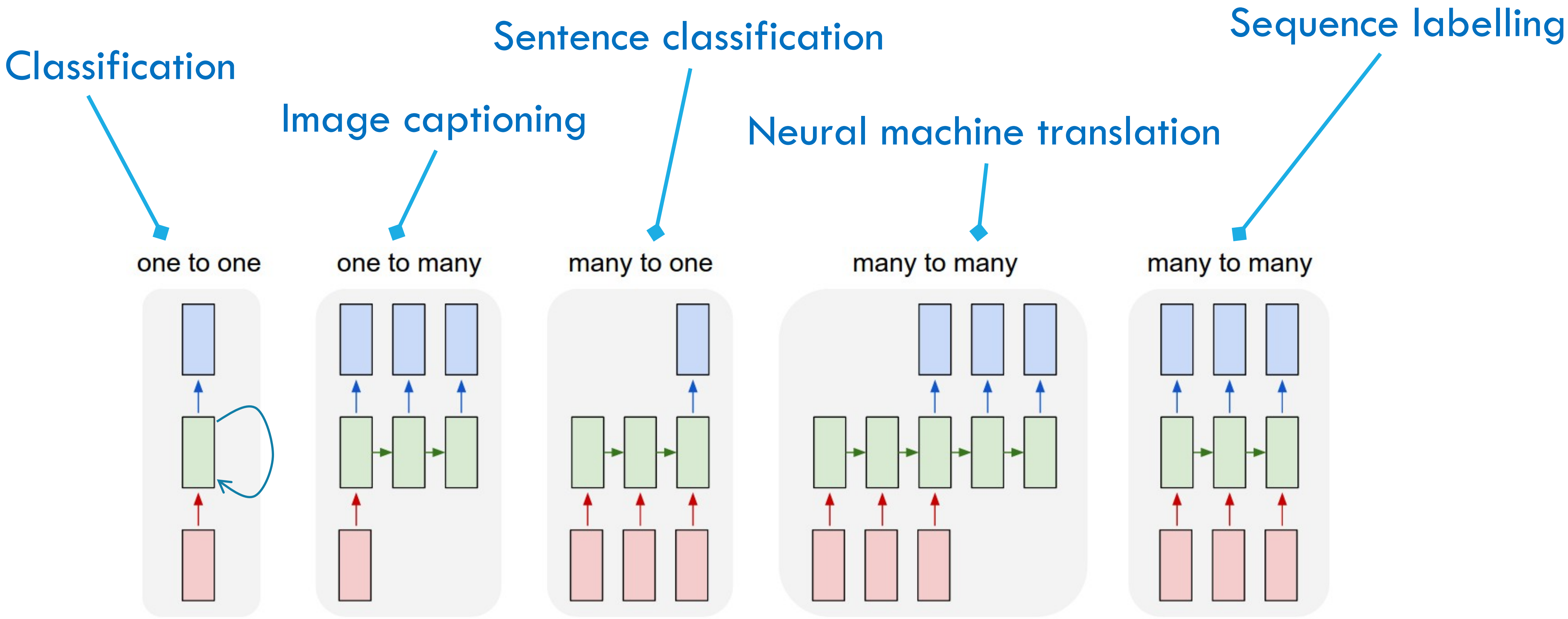


Feature detector



Block representation

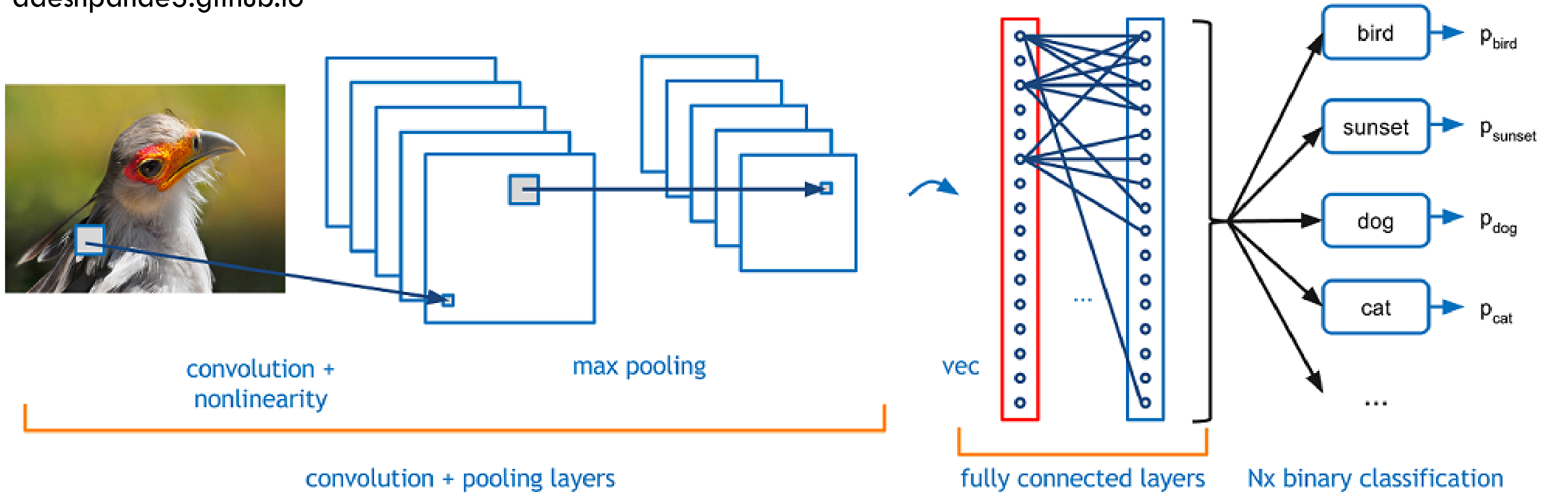
ARCHITECTURE ENGINEERING: RECURRENT NEURAL NETWORKS



Source: http://karpathy.github.io/assets/rnn/diags.jpeg

ARCHITECTURE ENGINEERING: CNN IS (CONVOLUTION → POOLING) REPEATED

adeshpande3.github.io



Design parameters:

- Padding
- Stride
- #Filters (maps)
- Filter size
- Pooling size
- #Layers
- Activation function

can be repeated N times - depth

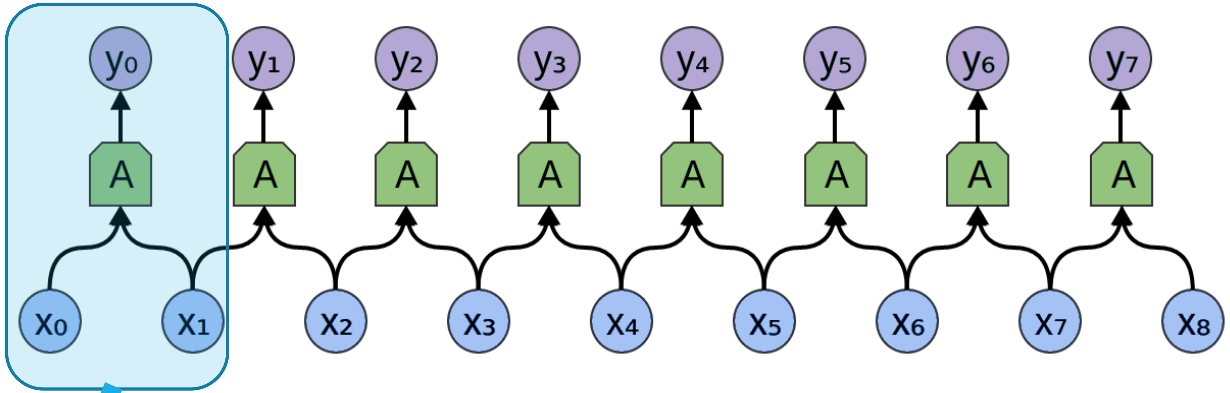
$F(x) =$ NeuralNet(Pooling(Rectifier(Conv(x))))

classifier max/mean nonlinearity feature detector

LEARNABLE CONVOLUTION AS MOTIFS DETECTOR

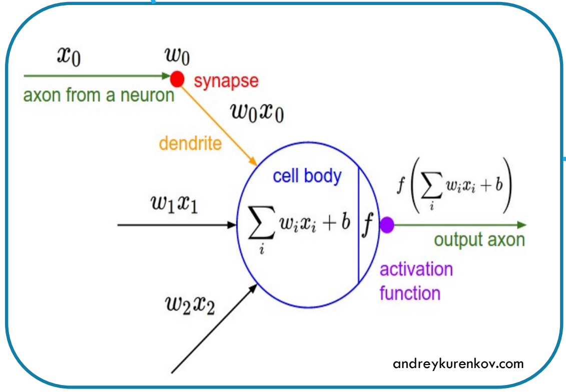
Learnable kernels

$$y_i = \sum_c \boxed{K(c)} x_{i+c}$$

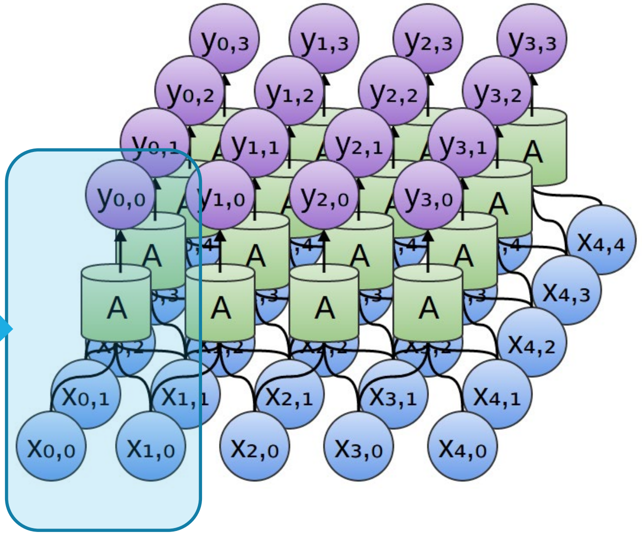


http://colah.github.io/posts/2015-09-NN-Types-FP/

$$y_{ij} = \sum_{c,d} \boxed{K(c,d)} x_{i+c,j+d}$$



Feature detector, often many



HOW DOES AI WORK FOR HEALTH?

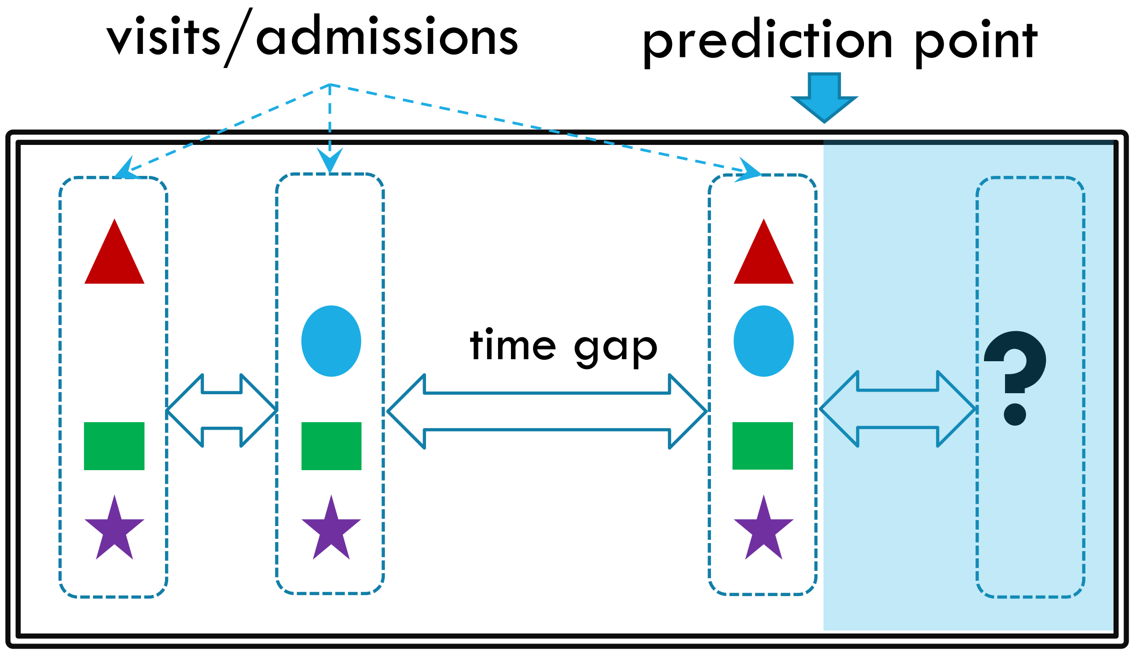
Discovery

Diagnosis

Prognosis

http://hubpages.com/education/Top-Medical-Inventions-of-The-1950s, http://www.ctrr.net/journal/

PREDICTIVE HEALTH USING ELECTRONIC MEDICAL RECORDS (EMR)

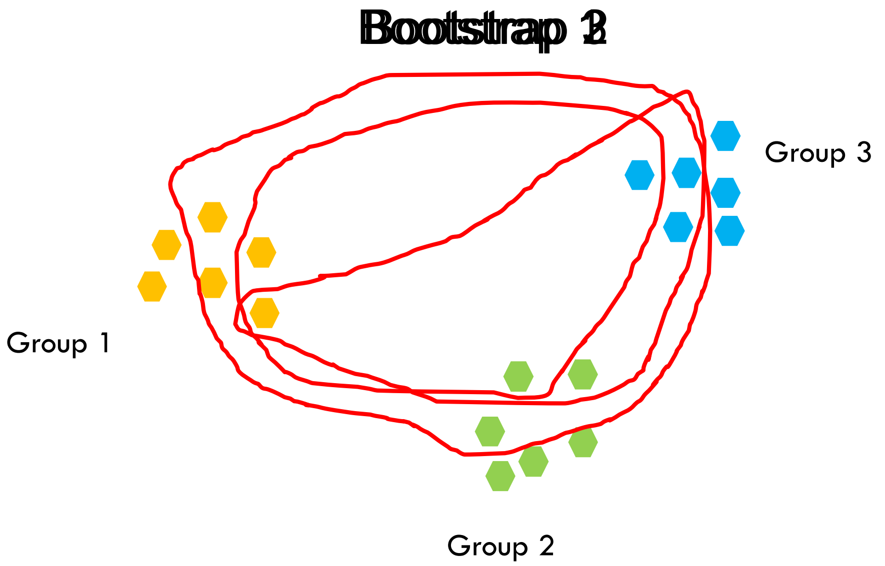


- Time-stamped
- Coded data: diagnosis, procedure & medication
- Numerical measurements
- Signals & imaging
- Text not considered, but in principle can be embedded into vector (LSTM/GRU, para2vec, word2vec)

DISCOVERY OF STABLE RISK FACTORS
(S. GOPAKUMAR ET AL, ADMD'16)

In medicine, transparent models are results. In ML, it is performance.

- Decision trees
- Linear models (sometimes with integer coefficients)

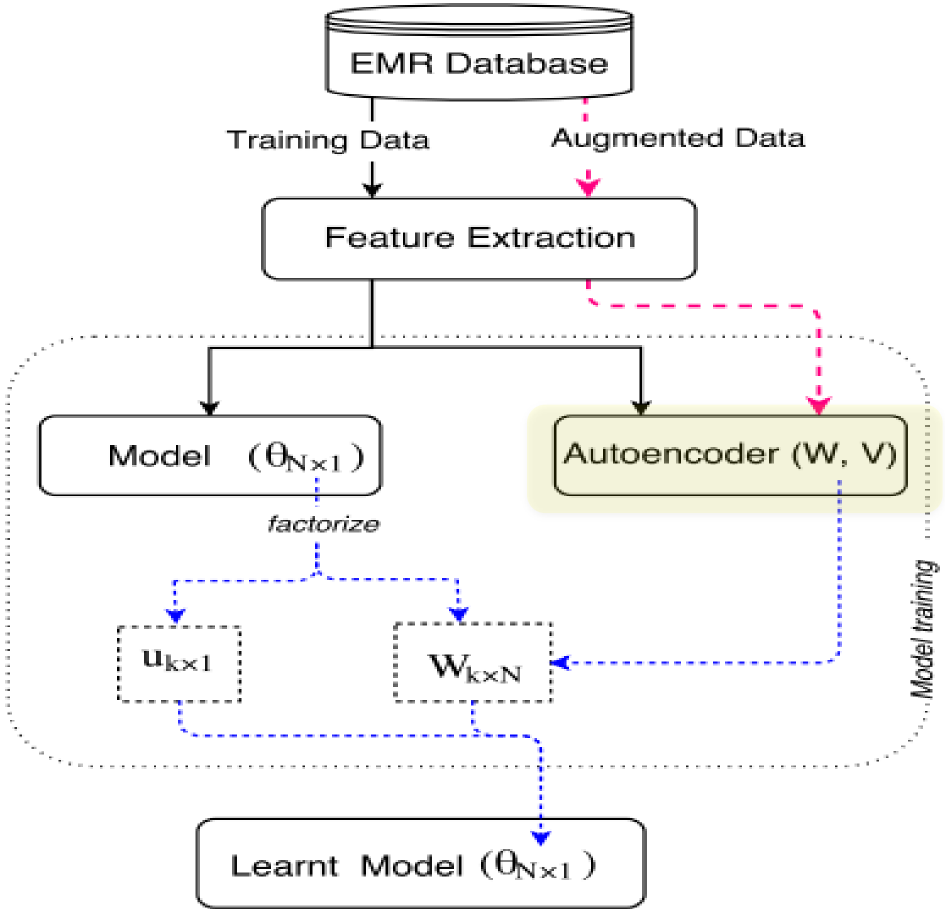


Modern healthcare data is high-dimensional and correlated, redundant.

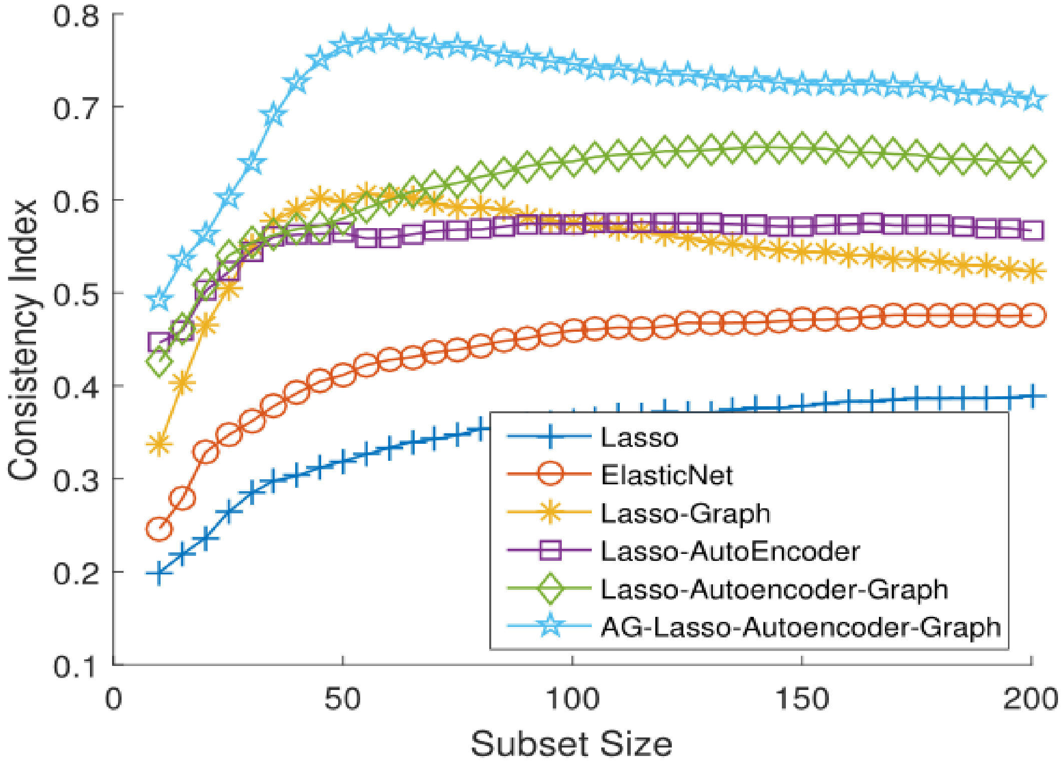
Automatic feature selection, e.g., lasso, in such data causes **model instability**

We can't ship different models from time to time!

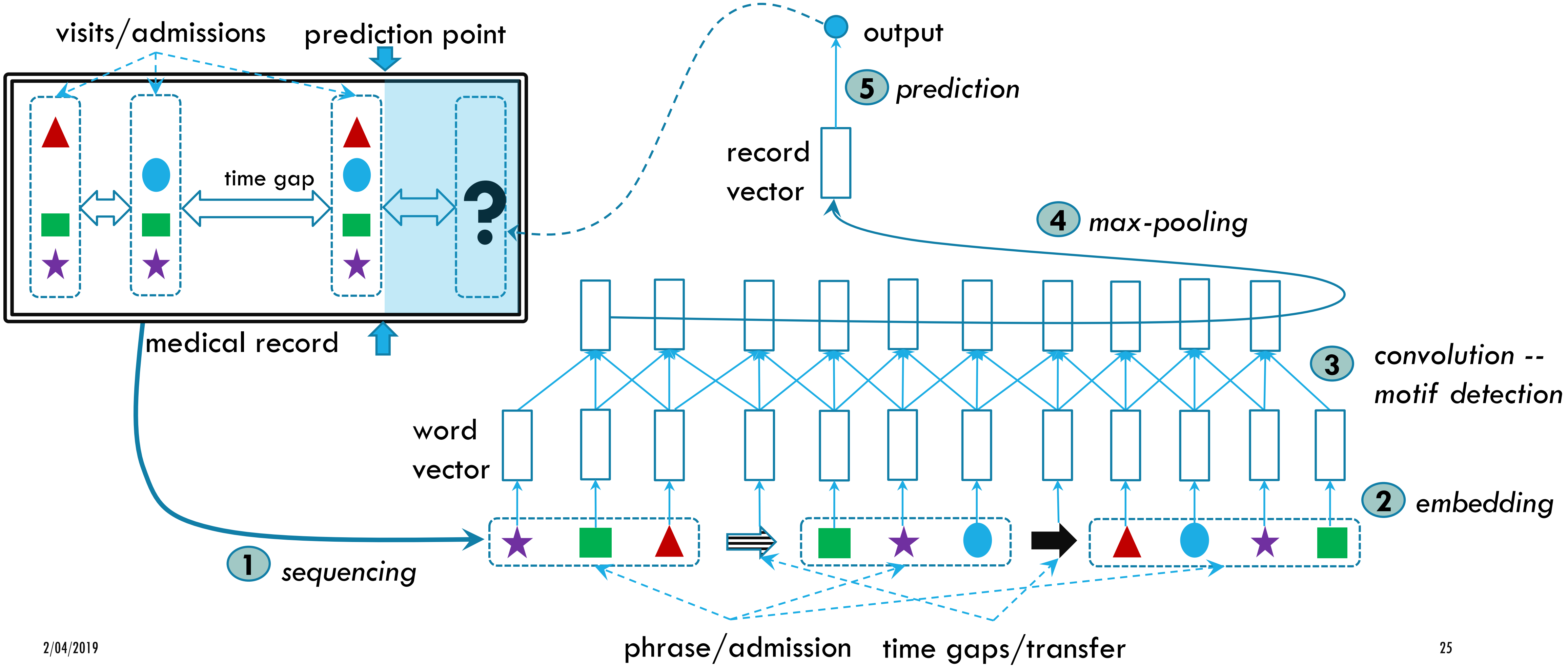
AUTO-ENCODER AS STABILIZING AGENT



Feature subset stability



DISCOVERY OF CARE MOTIFS VIA **DEEPR**
(PHUOC NGUYEN ET AL, IEEE J-BHI 2017)



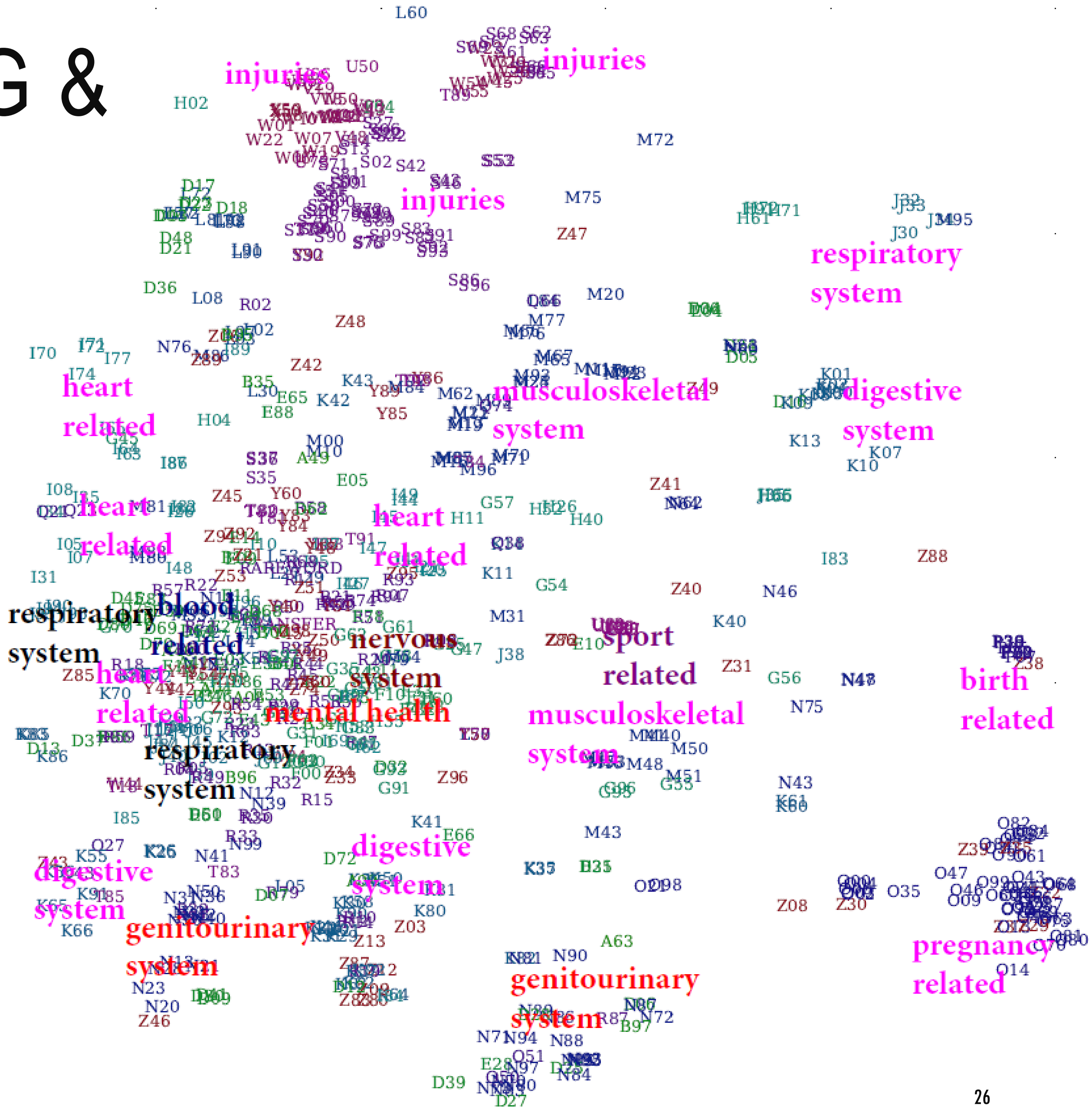
DISEASE EMBEDDING & MOTIFS DETECTION

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



HOW DOES AI WORK FOR HEALTH?

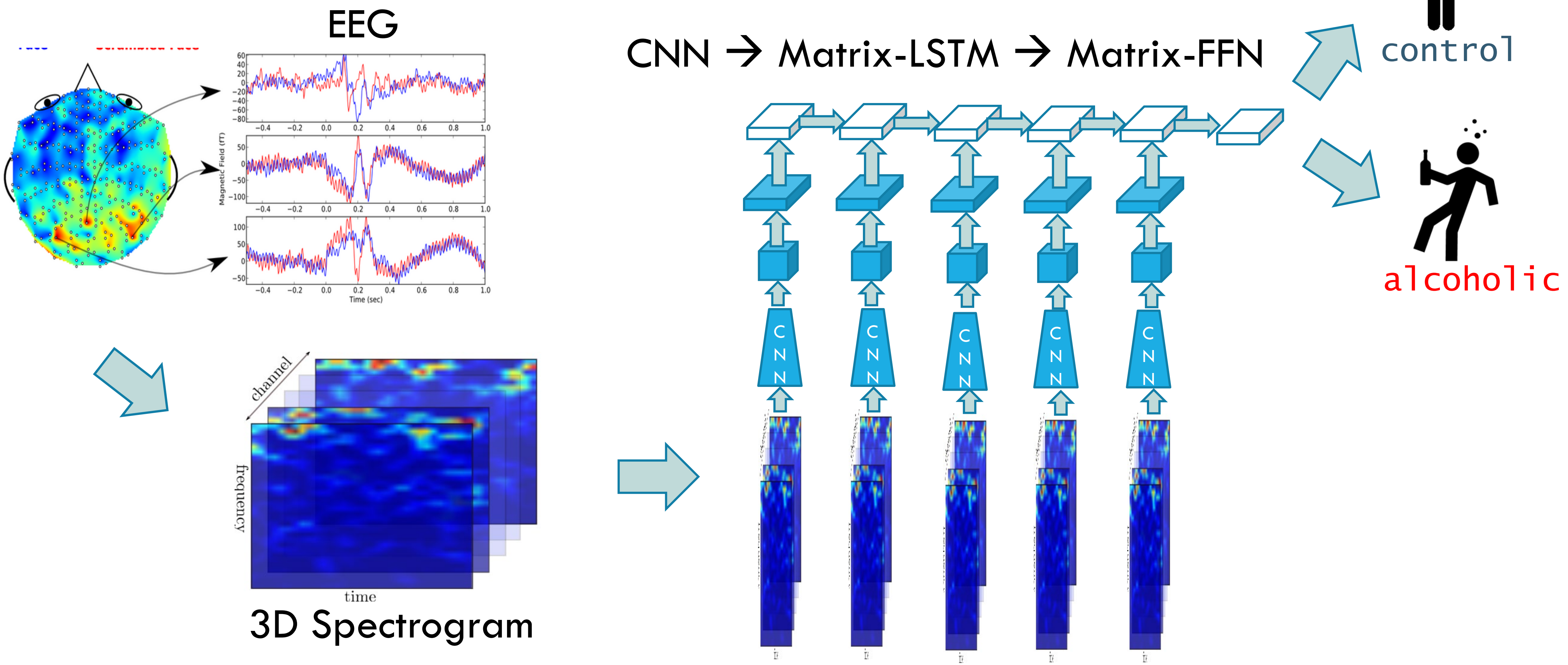
Discovery

Diagnosis

Prognosis

http://hubpages.com/education/Top-Medical-Inventions-of-The-1950s, http://www.ctrr.net/journal/

DIAGNOSIS OF ALCOHOLIC



MATRIX-LSTM
(KIEN DO, ET AL., 2017)

$$\mathrm{mat}(X, H; \boldsymbol{\theta}) \quad := \quad U_x^{\intercal} X V_x + U_h^{\intercal} H V_h + B$$

$$\text{Gates}\ \left\{\begin{aligned} I_t &= \sigma(\mathrm{mat}(X_t, H_{t-1}; \boldsymbol{\theta}_i)) \\ F_t &= \sigma(\mathrm{mat}(X_t, H_{t-1}; \boldsymbol{\theta}_f)) \\ O_t &= \sigma(\mathrm{mat}(X_t, H_{t-1}; \boldsymbol{\theta}_o)) \end{aligned}\right.$$

$$\text{Memory}\ \left\{\begin{aligned} \hat{C}_t &= \tanh(\mathrm{mat}(X_t, H_{t-1}; \boldsymbol{\theta}_c)) \\ C_t &= F_t \odot C_{t-1} + I_t \odot \hat{C}_t \end{aligned}\right.$$

$$\text{Output}\ \left\{ H_t = O_t \odot \tanh(C_t) \right.$$

RESULTS ON WITHIN-SUBJECT TEST TRIALS

Model	*# Params*	*Err (%)*
vec-LSTM (1)	1,844,201	5.29
mat-LSTM (2)	160,601	1.71
CNN-g + vec-LSTM (3)	1,435,729	1.90
CNN-m + vec-LSTM (4)	2,266,829	2.63
CNN-s + mat-LSTM (5)	200,729	4.12
CNN-m + mat-LSTM (6)	**248,029**	**1.44**

HOW DOES AI WORK FOR HEALTH?

Discovery

Diagnosis

Prognosis

http://hubpages.com/education/Top-Medical-Inventions-of-The-1950s, http://www.ctrr.net/journal/

heart failure

diabetes

mental health

COPD

heart attack

cancers

preterm

RISK PREDICTION (PROGNOSIS)

suicide attempts

side effects

toxicity

death

quality-of-life

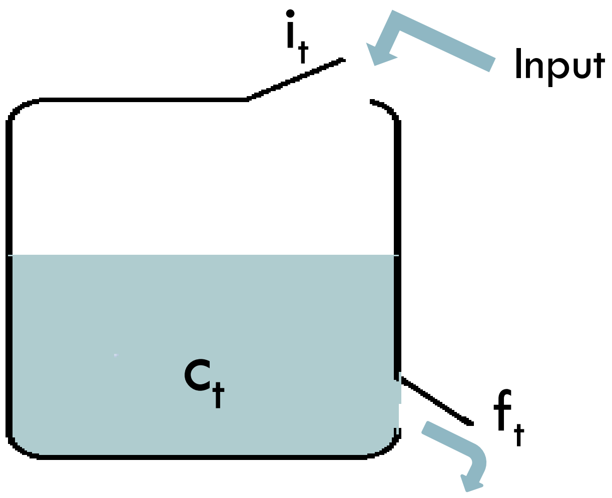
readmission

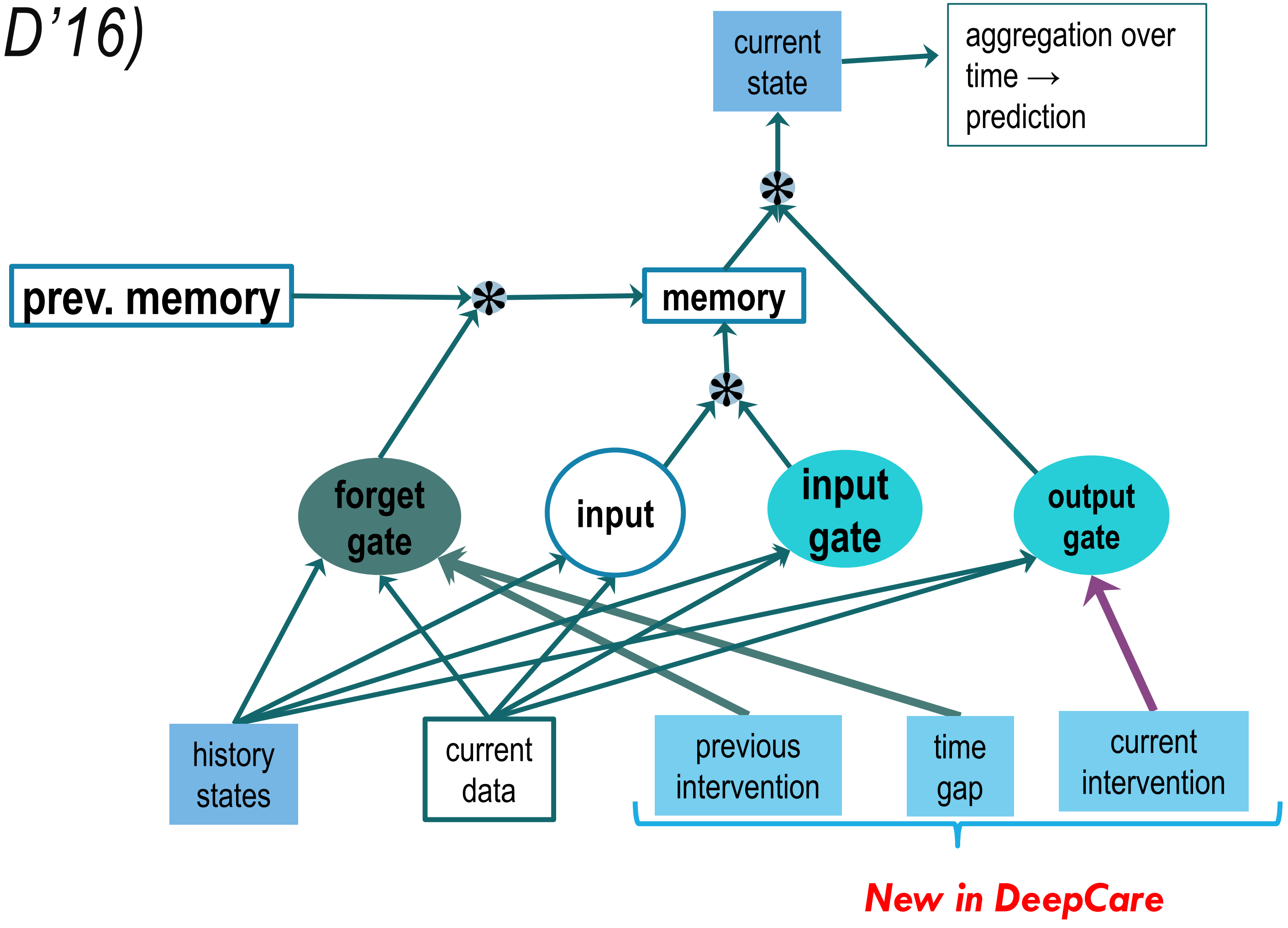
stress

progression to advanced stages

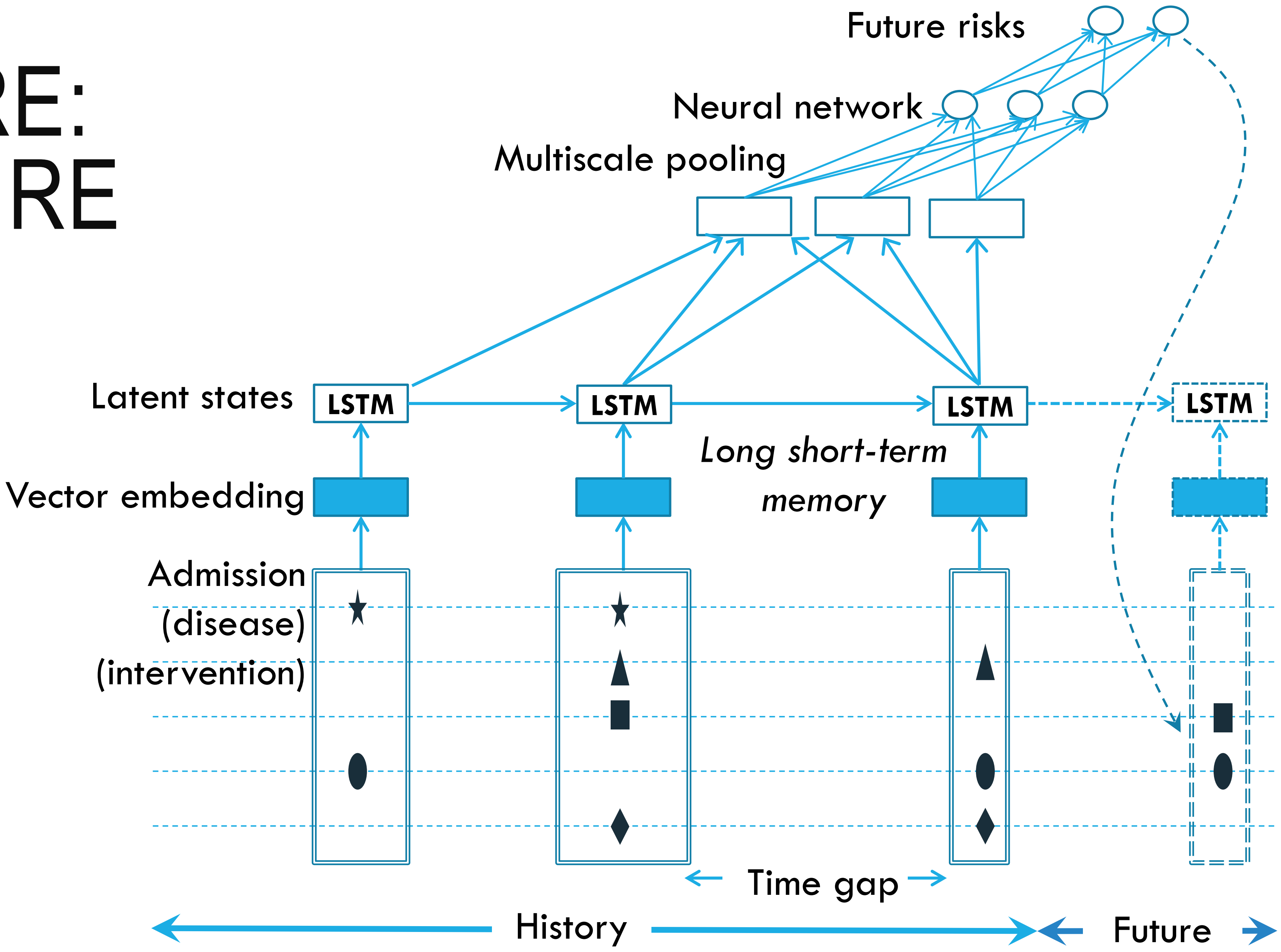
length-of-stay

DEEPCARE: INTERVENED LONG-TERM MEMORY OF HEALTH
(TRANG PHAM ET AL, PAKDD'16)

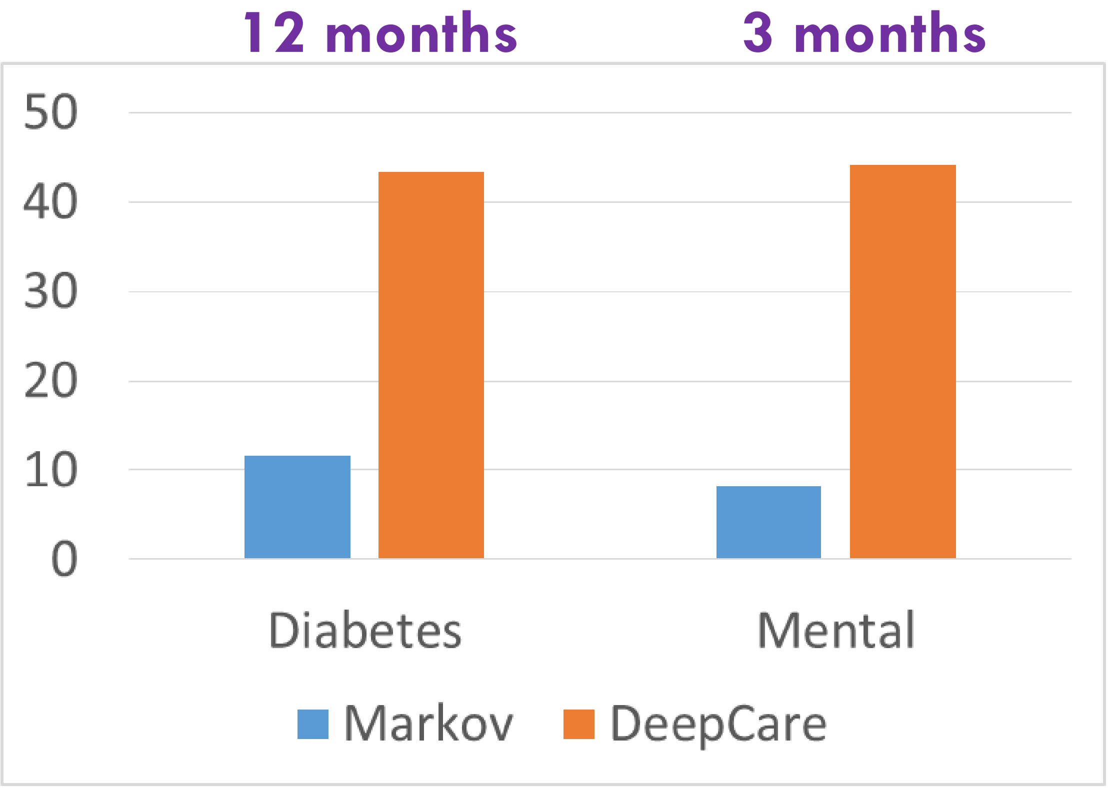




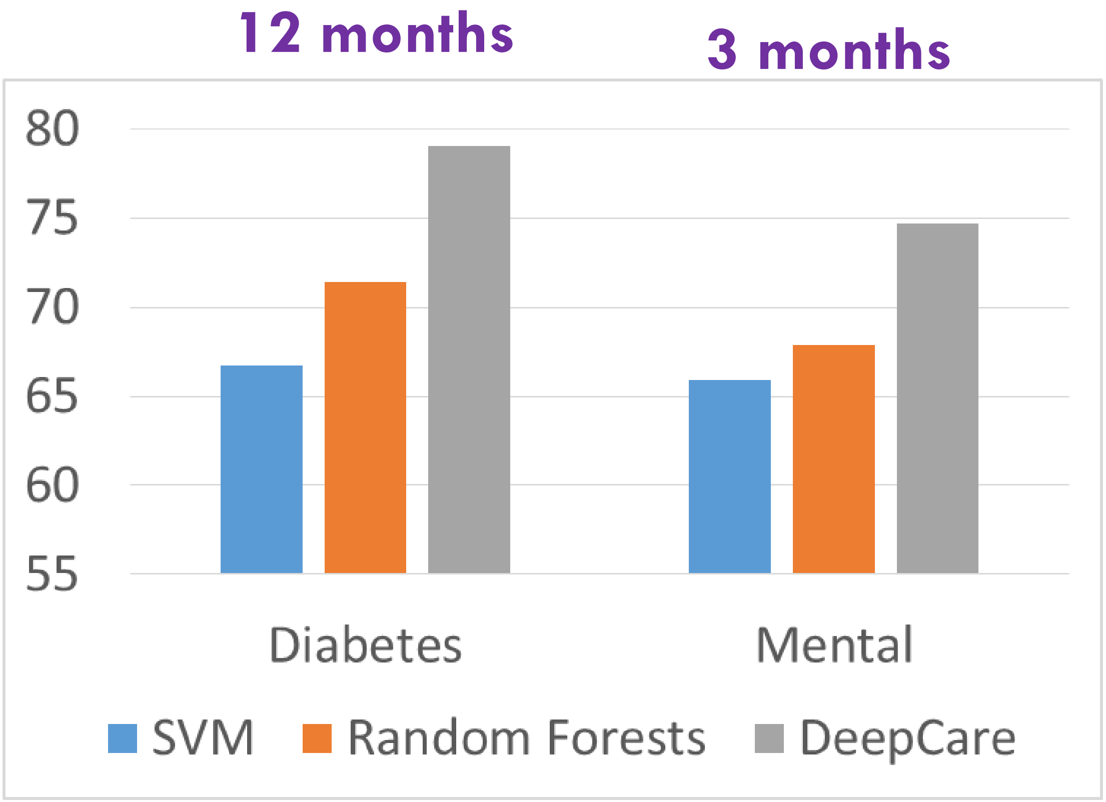
DEEPCARE: STRUCTURE



DEEPCARE: PREDICTION RESULTS



Intervention recommendation (precision@3)



Unplanned readmission prediction (F-score)

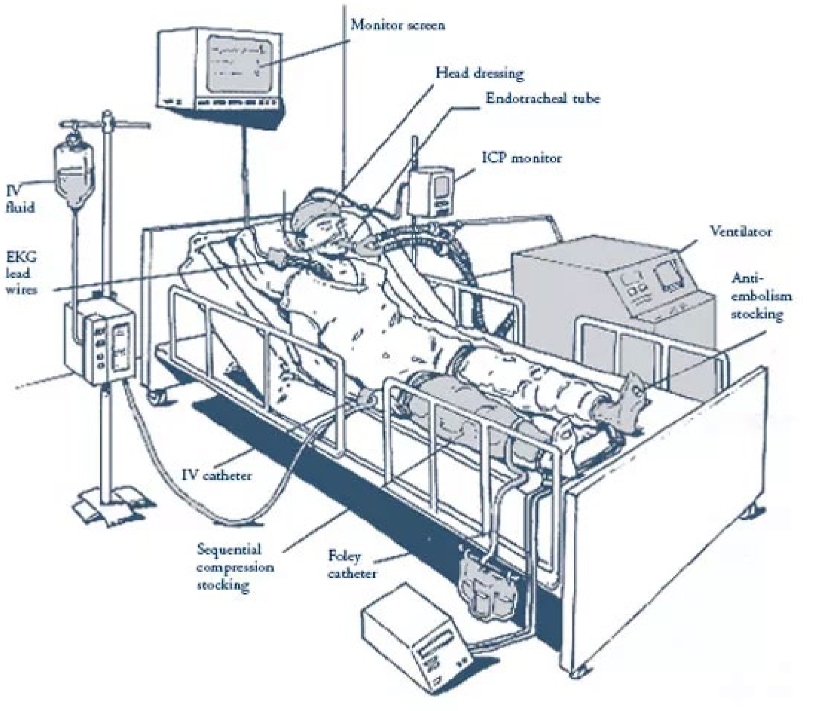
DEEPICU: MORTALITY PREDICTION IN ICU
(PHUOC NGUYEN ET AL, 2017)

Existing methods: Handcoded features, LSTM with missingness and time-gap as input.

New method: **Deepic**

Steps:

- Measurement quantization
- Time gap quantization
- Sequencing words into sentence
- CNN+LSTM+more



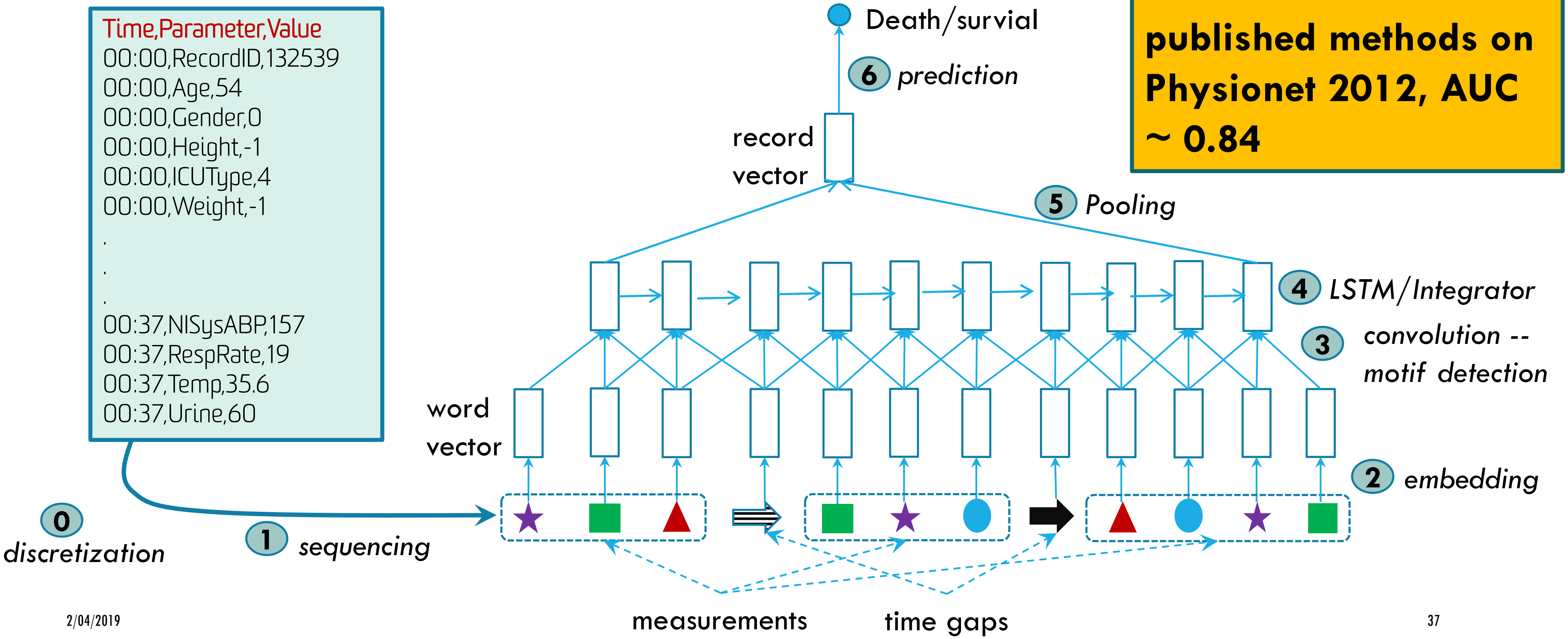
http://www.healthpages.org/brain-injury/brain-injury-intensive-care-unit-icu/

```
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**

DEEPICU: SYMBOLIC & TIME GAP REPRESENTATION OF DATA

Results ~ best published methods on Physionet 2012, AUC ~ 0.84



REFERENCES

- Matrix-Centric Neural Nets, Kien Do, Truyen Tran, Svetha Venkatesh, *In submission*.
- Stabilizing Linear Prediction Models using Autoencoder, Shivapratap Gopakumara, Truyen Tran, Dinh Phung, Svetha Venkatesh, *International Conference on Advanced Data Mining and Applications* (ADMA 2016).
- Deepr: A Convolutional Net for Medical Records, Phuoc Nguyen, Truyen Tran, Nilmini Wickramasinghe, Svetha Venkatesh, *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 22–30, Jan. 2017, Doi: 10.1109/JBHI.2016.2633963
- DeepCare: A Deep Dynamic Memory Model for Predictive Medicine, Trang Pham, Truyen Tran, Dinh Phung, Svetha Venkatesh, *PAKDD'16*, Auckland, NZ, April 2016.
- Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (e-NRBM), Truyen Tran, Tu D. Nguyen, D. Phung, and S. Venkatesh, *Journal of Biomedical Informatics*, 2015,
- Tensor-variate Restricted Boltzmann Machines, Tu D. Nguyen, Truyen Tran, D. Phung, and S. Venkatesh, *AAAI* 2015.
- A framework for feature extraction from hospital medical data with applications in risk prediction, T Tran, W Luo, D Phung, S Gupta, S Rana, RL Kennedy, A Larkins, *BMC bioinformatics* 15 (1), 425, 2014
- Latent patient profile modelling and applications with Mixed-Variate Restricted Boltzmann Machine, Tu D. Nguyen, Truyen Tran, D. Phung, and S. Venkatesh, In *Proc. of 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining* (PAKDD'13), Gold Coast, Australia, April 2013.
- An evaluation of randomized machine learning methods for redundant data: Predicting short and medium-term suicide risk from administrative records and risk assessments, T Nguyen, T Tran, S Gopakumar, D Phung, S Venkatesh, *arXiv* arXiv:1605.01116

Thank you!

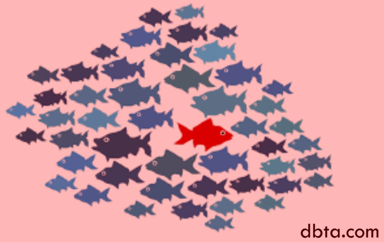
http://ahsanqawl.com/2015/10/qa/

BONUS: HOW DOES AI WORK FOR HEALTH?

Discovery

Diagnosis

Prognosis



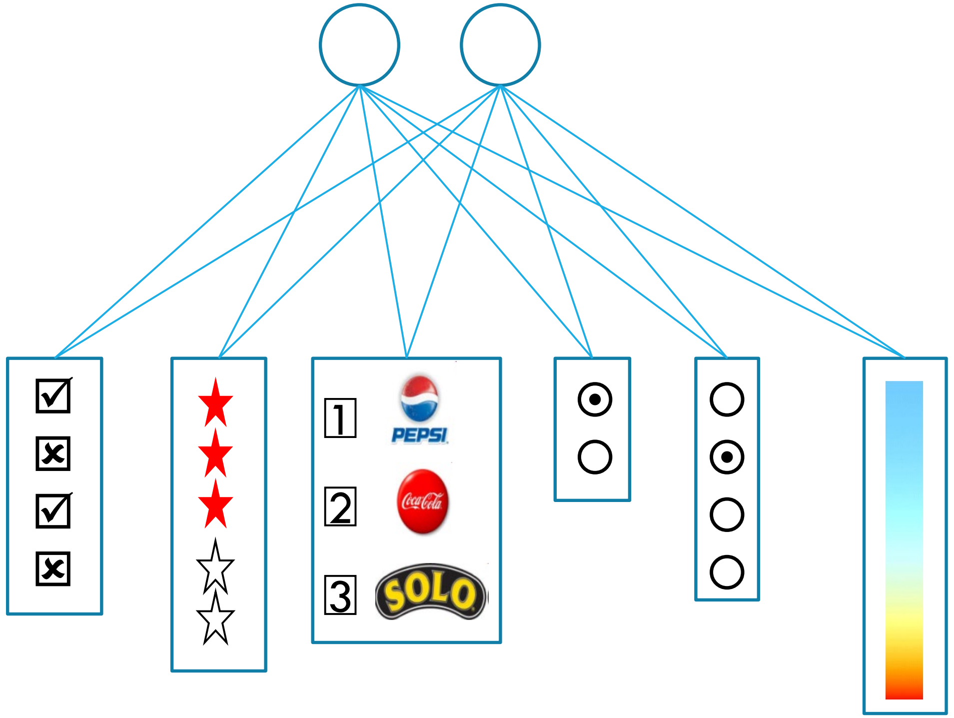
Outliers

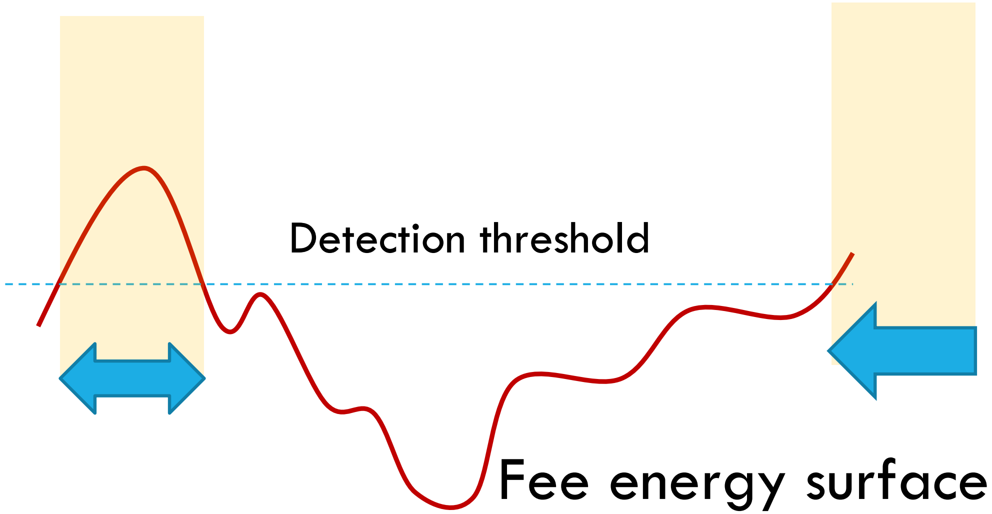
http://hubpages.com/education/Top-Medical-Inventions-of-The-1950s, http://www.ctrr.net/journal/

MIXED DATA ANOMALY DETECTION

	A	B	C	D	E	F	G	H	I	J
1	Age	Sex	Chest pain type	Resting blood pressure	Serum cholestoral (mg/dl)	Fasting blood sugar > 120 mg/dl ?	Resting electrocardiographic result	Maximum heart rate achieved	Exercise induced angina	oldpeak = ST depression induced by exercise relative to rest
2	70	male	asymptomatic (4)	130.0	322.0	no	2	109.0	no	2.4
3	67	female	non-anginal pain (3)	115.0	564.0	no	2	160.0	no	1.6
4	57	male	atypical angina (2)	124.0	261.0	no	0	141.0	no	0.3
5	64	male	asymptomatic (4)	128.0	263.0	no	0	105.0	yes	0.2
6	74	female	atypical angina (2)	120.0	269.0	no	2	121.0	yes	0.2
7	65	male	asymptomatic (4)	120.0	177.0	no	0	140.0	no	0.4
8	56	male	non-anginal pain (3)	130.0	256.0	yes	2	142.0	yes	0.6
9	59	male	asymptomatic (4)	110.0	239.0	no	2	142.0	yes	1.2
10	60	male	asymptomatic (4)	140.0	293.0	no	2	170.0	no	1.2
11	63	female	asymptomatic (4)	150.0	407.0	no	2	154.0	no	4.0
12	59	male	asymptomatic (4)	135.0	234.0	no	0	161.0	no	0.5
13	53	male	asymptomatic (4)	142.0	226.0	no	2	111.0	yes	0.0
14	44	male	non-anginal pain (3)	140.0	235.0	no	2	180.0	no	0.0
15	61	male	typical angina (1)	134.0	234.0	no	0	145.0	no	2.6
16	57	female	asymptomatic (4)	128.0	303.0	no	2	159.0	no	0.0
17	71	female	asymptomatic (4)	112.0	149.0	no	0	125.0	no	1.6
18	46	male	asymptomatic (4)	140.0	311.0	no	0	120.0	yes	1.8
19	53	male	asymptomatic (4)	140.0	203.0	yes	2	155.0	yes	3.1
20	64	male	typical angina (1)	110.0	211.0	no	2	144.0	yes	1.8
21	40	male	typical angina (1)	140.0	199.0	no	0	178.0	yes	1.4
22	67	male	asymptomatic (4)	120.0	229.0	no	2	129.0	yes	2.6

MIXED-VARIATE RBM
(TRAN ET AL, 2011 & DO ET AL, 2016)





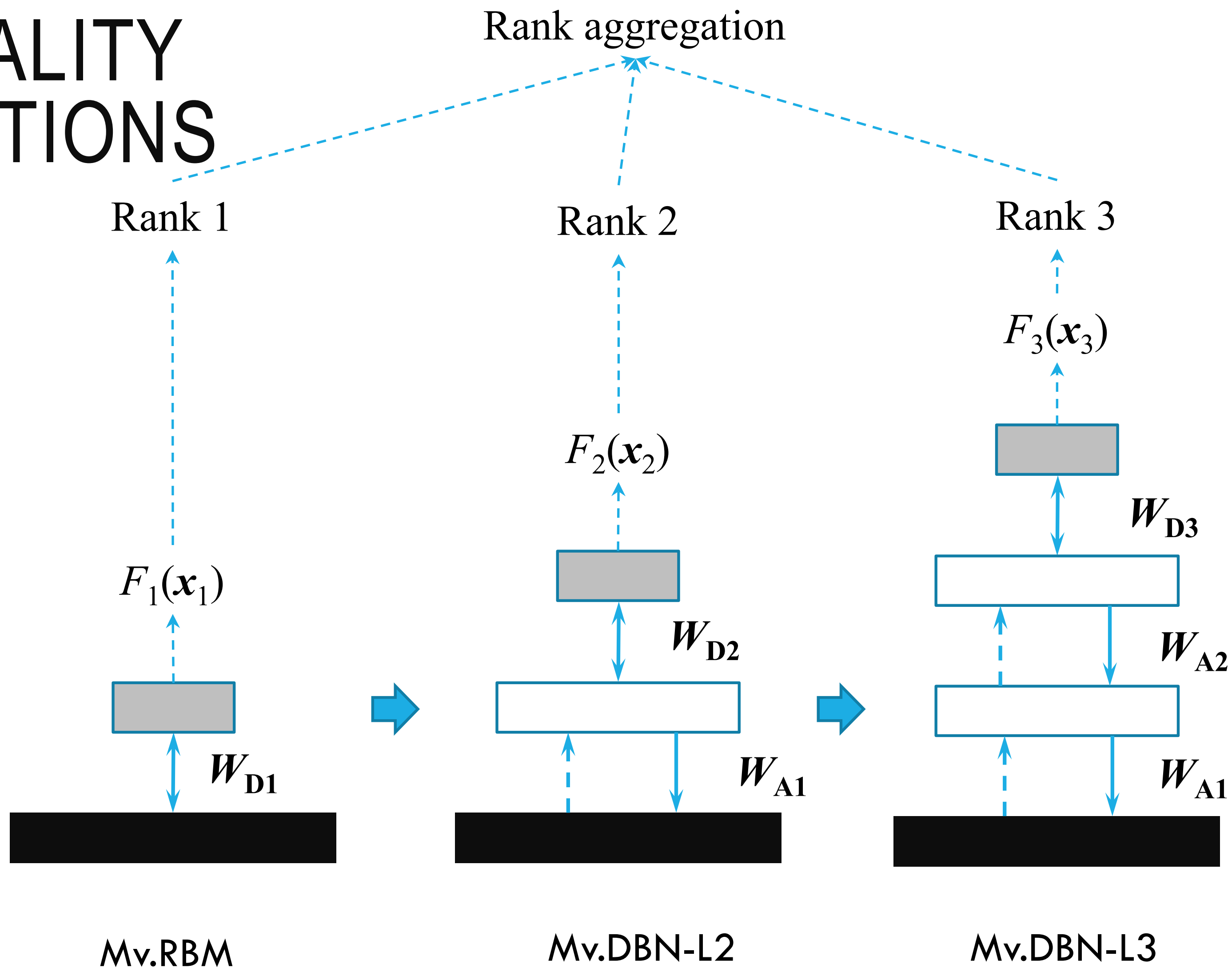
$$F(\boldsymbol{x}) = -\sum_i \left(a_i x_i + \sum_k \log(1 + \exp(x_i W_{ik} + b_k)) \right)$$

DETECTION RESULTS

Dataset	Single type			mixed-type			
	GMM	OCSVM	PPCA	BMM	ODMAD	GLM-t	Mv.RBM
KDD99-10	0.42	0.54	0.55	–	–	–	**0.71**
Australian Credit	0.74	0.84	0.38	**0.972**	0.942	–	0.90
German Credit	0.86	0.86	0.02	0.934	0.810	–	**0.95**
Heart	0.89	0.76	0.64	0.872	0.630	0.72	**0.94**
Thoracic Surgery	0.71	0.71	0.70	**0.939**	0.879	–	0.90
Auto MPG	**1.00**	**1.00**	0.67	0.625	0.575	0.64	**1.00**
Contraceptive	0.62	0.84	0.02	0.673	0.523	–	**0.91**
Average	*0.75*	*0.79*	*0.43*	*0.84*	*0.73*	*0.68*	***0.91***

MIXMAD: ABNORMALITY ACROSS ABSTRACTIONS

$$\bar{r}_i(p) = \left(\sum_{l=1}^{L} r_{li}^{p}\right)^{1/p}$$



RESULTS

	KDD	AuCredit	GeCredit	Heart	ThSurgery	AMPG	Contra.
BMM [9, 12]	–	0.97	0.93	0.87	0.94	0.62	0.67
ODMAD [12, 20]	–	0.94	0.81	0.63	0.88	0.57	0.52
GLM-t [12, 22]	–	–	–	0.72	–	0.64	–
Mv.RBM [12]	0.71	0.90	0.95	0.94	0.90	**1.00**	0.91
MIXMAD-L2p0.5	0.72	0.93	**0.97**	0.94	**0.97**	**1.00**	**0.95**
MIXMAD-L2p1	0.72	0.93	0.95	0.94	**0.97**	**1.00**	**0.95**
MIXMAD-L2p2	0.69	0.93	**0.97**	0.94	**0.97**	**1.00**	**0.95**
MIXMAD-L2p∞	0.69	0.73	**0.97**	**1.00**	**0.97**	**1.00**	**0.95**
MIXMAD-L3p0.5	**0.73**	**0.98**	**0.97**	0.94	**0.97**	0.70	**0.95**
MIXMAD-L3p1	0.72	**0.98**	**0.97**	0.94	**0.97**	0.70	**0.95**
MIXMAD-L3p2	0.71	**0.98**	**0.97**	0.94	**0.97**	0.70	**0.95**
MIXMAD-L3p∞	0.50	0.78	**0.97**	0.94	**0.97**	0.57	**0.95**