AI in Medicine

Working Group Discussion Session, APMSS 2019

Chair: 謝德威 Alexander Te-Wei Shieh

Agenda

Introduction

1. Machine Learning and Information Retrieval
2. Computer Vision / Natural Language Processing
3. Reinforcement Learning and Unsupervised Learning
4. Current AI applications in medicine
5. Some Challenges in medical AI

Discussion

(0) Self-Introduction

謝德威 Alexander Te-Wei Shieh

NTU Medicine, 3rd year

NTU Computer Science and Information Engineering (double major)

My interests are: Natural Language Processing (and related ML techniques), Tech Startups.

1. Cheminformatics / Drug Development (Computational Molecular Design and Metabolomics Lab)
2. Natural Language Processing (Machine Intelligence and Understanding Lab)

(0) Self-Introduction

---

1. Name
2. School / n-th year
3. Why you choose this topic?
4. (Any previous experience related to AI)

Memoji by Apple

(0) AI in Medicine

This session will be extremely fast-paced.
If you have any questions please shout out!

Today our discussion is much more limited:

Artificial General Intelligence (Strong AI)

Artificial Narrow Intelligence **(Weak AI)**

- **Machine Learning**
  - **Deep Learning**
  - Probabilistic Models, SVMs, Random Forests, ...
- Robotics, Search, Genetic Algorithms, ...

(1) Machine Learning and Information Retrieval

Information Retrieval: Vector-Space Model

How did PubMed find your article?

Indexing, Retrieval Model, Query-Expansion, Re-Rank (Relevance Feedback)

Document Vector

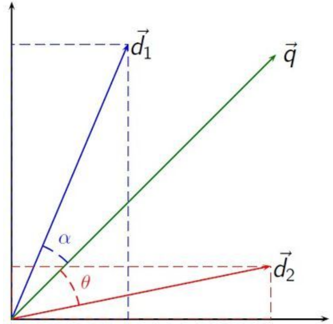
$$\mathbf{v}_d = [w_{1,d}, w_{2,d}, \ldots, w_{N,d}]^T$$

$$w_{t,d} = \mathrm{tf}_{t,d} \cdot \log \frac{|D|}{|\{d' \in D \,|\, t \in d'\}|}$$

TF-IDF Weighting

Cosine Similarity
between Document Vector and Query Vector

$$\cos(d_j, q) = \frac{\mathbf{d_j} \cdot \mathbf{q}}{\|\mathbf{d_j}\| \, \|\mathbf{q}\|} = \frac{\sum_{i=1}^{N} w_{i,j} w_{i,q}}{\sqrt{\sum_{i=1}^{N} w_{i,j}^2} \sqrt{\sum_{i=1}^{N} w_{i,q}^2}}$$



TREC precision medicine / clinical support track: http://www.trec-cds.org/2018.html

(1) Machine Learning and Information Retrieval

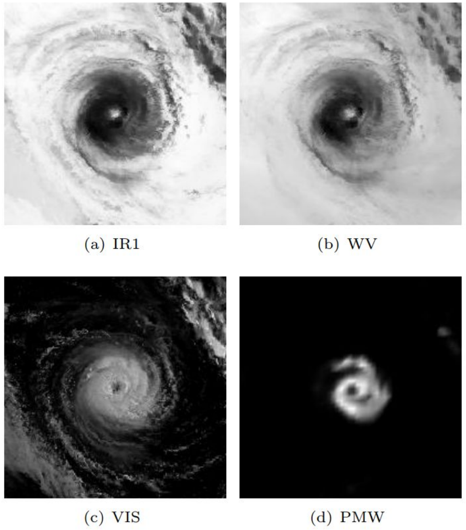
Why machine learning?

Machine Learning Problems

1. **Supervised** Learning [Labeled]
    a. Regression: [Compound, Activity]
    b. Classification: [Image, Benign / Malignant], [Text, ICD Codes]
2. **Unsupervised** Learning [Unlabeled]
    a. Representation Learning: Word2Vec, Autoencoders, Language Models
    b. Clustering
3. **Reinforcement** Learning [Reward]

(1) Machine Learning and Information Retrieval

Supervised Learning: Regression (Convolutional Neural Networks)



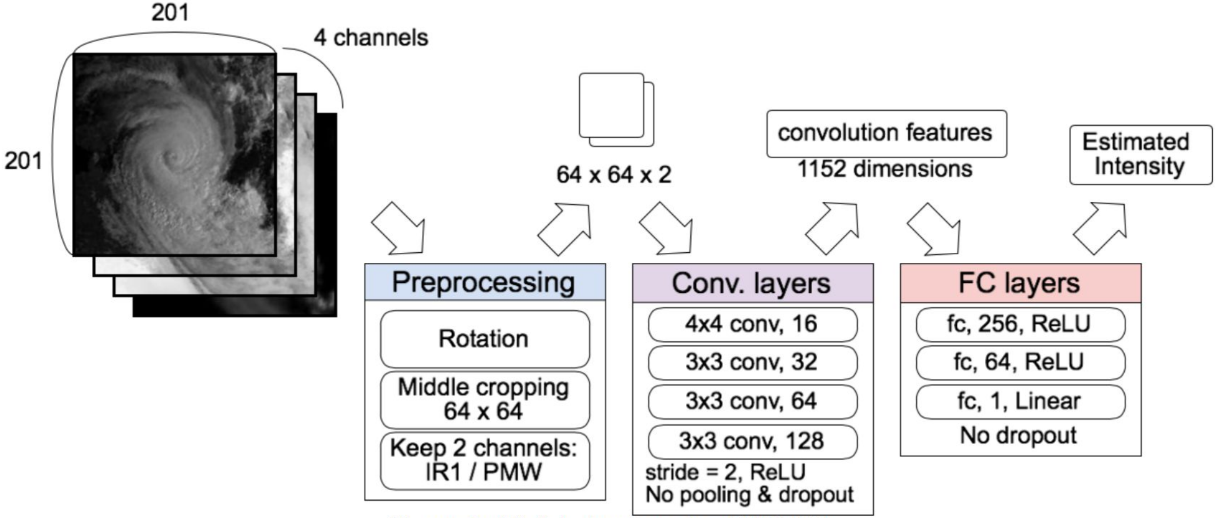


Figure 2: Model structure for CNN-TC

Chen et al. **Rotation-blended CNNs** on a New Open Dataset for **Tropical Cyclone Image-to-intensity Regression**. KDD 2018.

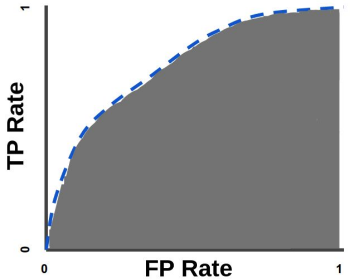
(1) Machine Learning and Information Retrieval

Evaluation for Regression Problems

$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \quad \text{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

- Root Mean Squared Error (**RMSE**), Mean Average Error (**MAE**)

Evaluation for Classification Problems



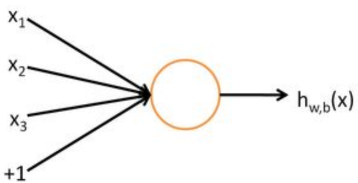
- **Accuracy** (*(TP+TN)/(P+N)*)
- **Sensitivity** (=Recall/**TPR**, *TP/P*) / **Specificity** (=**TNR**, *TN/N*)
- **Precision** (=**PPV**, *TP/(TP+FP)*) / **Recall** / **F1** (*2PPV×TPR/(PPV+TPR)*)
- **ROC curve** (TPR vs TNR at different thresholds) **/** Area Under ROC curve (**AUC**)

Others: MAP, BLEU & ROUGE, Perpelexity, Human-evaluation, Task-specific scores

(1) Machine Learning and Information Retrieval

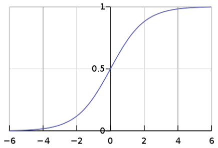
From Logistic Regression to **Deep Learning**

Logistic Regression

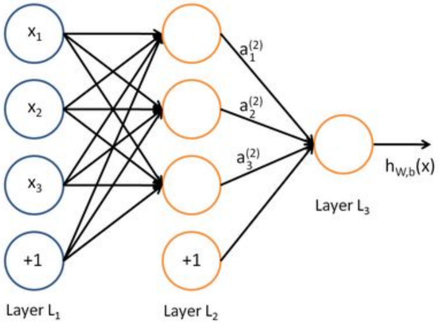


$$h_{W,b}(x) = f(W^T x) = f(\Sigma_{i=1}^{3} W_i x_i + b)$$

$$f(z) = \frac{1}{1+\exp(-z)}.$$



Multi-Layer Perceptron



$$
\begin{aligned}
a_1^{(2)} &= f(W_{11}^{(1)} x_1 + W_{12}^{(1)} x_2 + W_{13}^{(1)} x_3 + b_1^{(1)}) \\
a_2^{(2)} &= f(W_{21}^{(1)} x_1 + W_{22}^{(1)} x_2 + W_{23}^{(1)} x_3 + b_2^{(1)}) \\
a_3^{(2)} &= f(W_{31}^{(1)} x_1 + W_{32}^{(1)} x_2 + W_{33}^{(1)} x_3 + b_3^{(1)}) \\
h_{W,b}(x) = a_1^{(3)} &= f(W_{11}^{(2)} a_1^{(2)} + W_{12}^{(2)} a_2^{(2)} + W_{13}^{(2)} a_3^{(2)} + b_1^{(2)})
\end{aligned}
$$

$$
\begin{aligned}
z^{(2)} &= W^{(1)} x + b^{(1)} \\
a^{(2)} &= f(z^{(2)}) \\
z^{(3)} &= W^{(2)} a^{(2)} + b^{(2)} \\
h_{W,b}(x) = a^{(3)} &= f(z^{(3)})
\end{aligned}
$$

Matrix Form

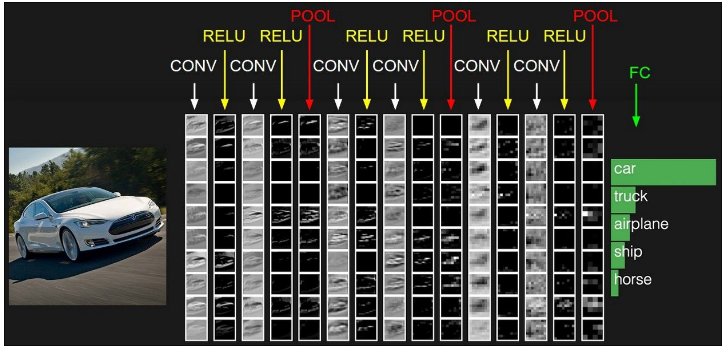
Use gradient-based optimization (SGD, Adam, …)
to get good parameters (weights).

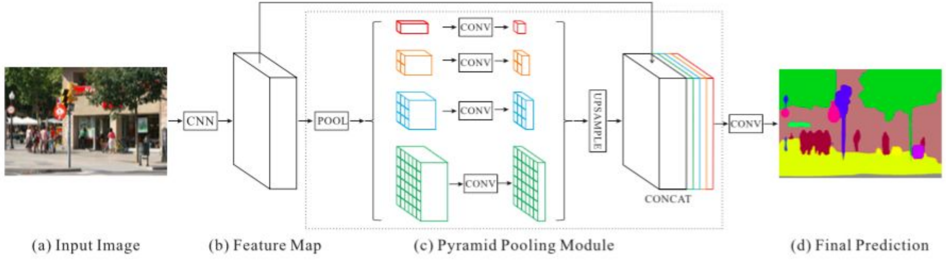
From: http://ufldl.stanford.edu/wiki/index.php/Neural_Networks

(2) CV and NLP A deep learning point of view

Convolutional Neural Networks (CNN)

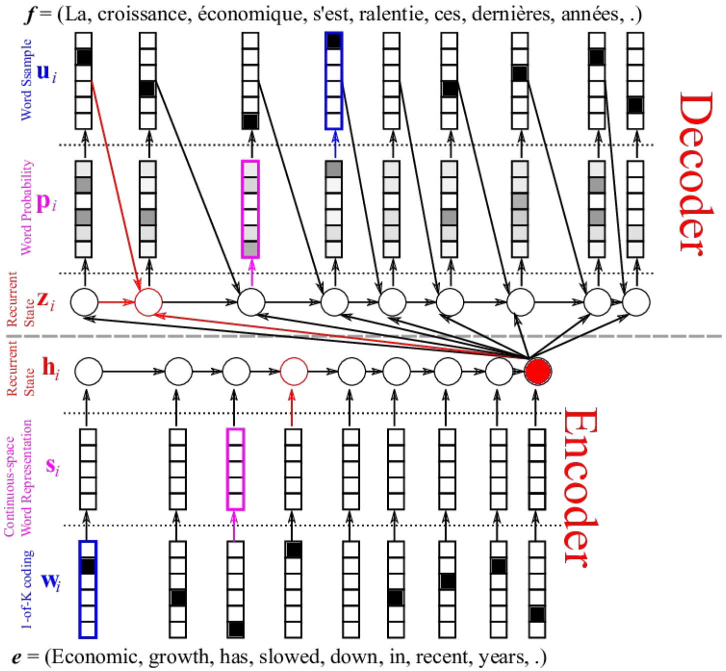
- http://cs231n.github.io/convolutional-networks/
- http://blog.qure.ai/notes/semantic-segmentation-deep-learning-review





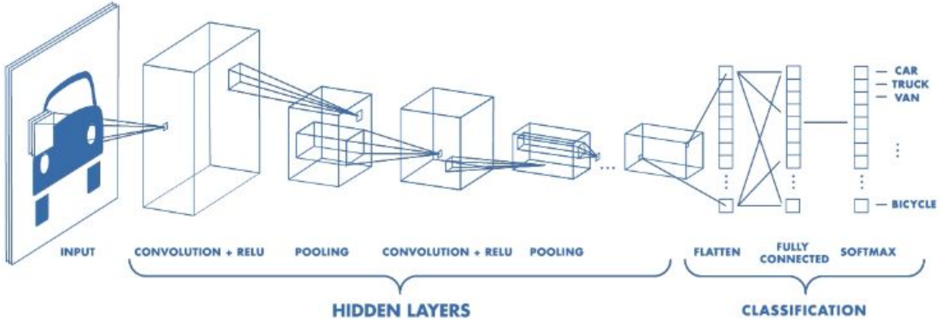
Seq2Seq with Recurrent Neural Networks (RNN)

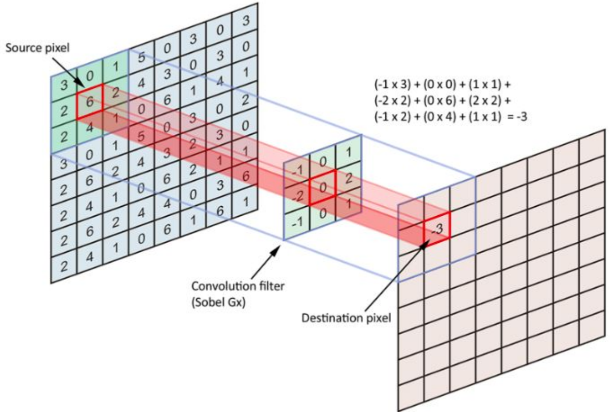
https://devblogs.nvidia.com/introduction-neural-machine-translation-gpus-part-2/

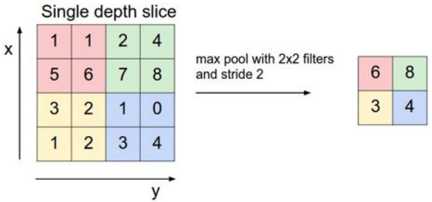


(2) CV and NLP A deep learning point of view

Convolutional Neural Network (CNN)



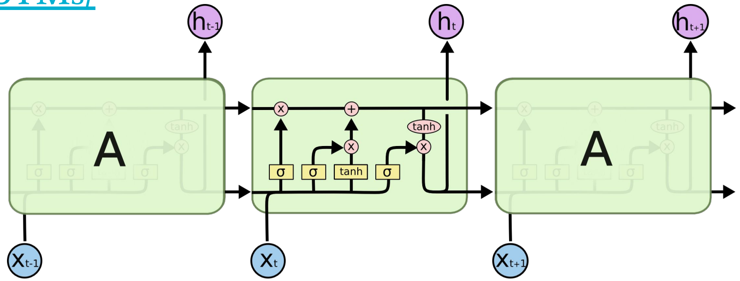


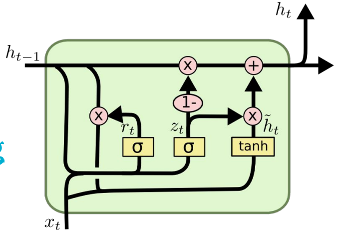


https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050

Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM) Units or Gated Recurrent Units (GRU)

http://colah.github.io/posts/2015-08-Understanding-LSTMs/





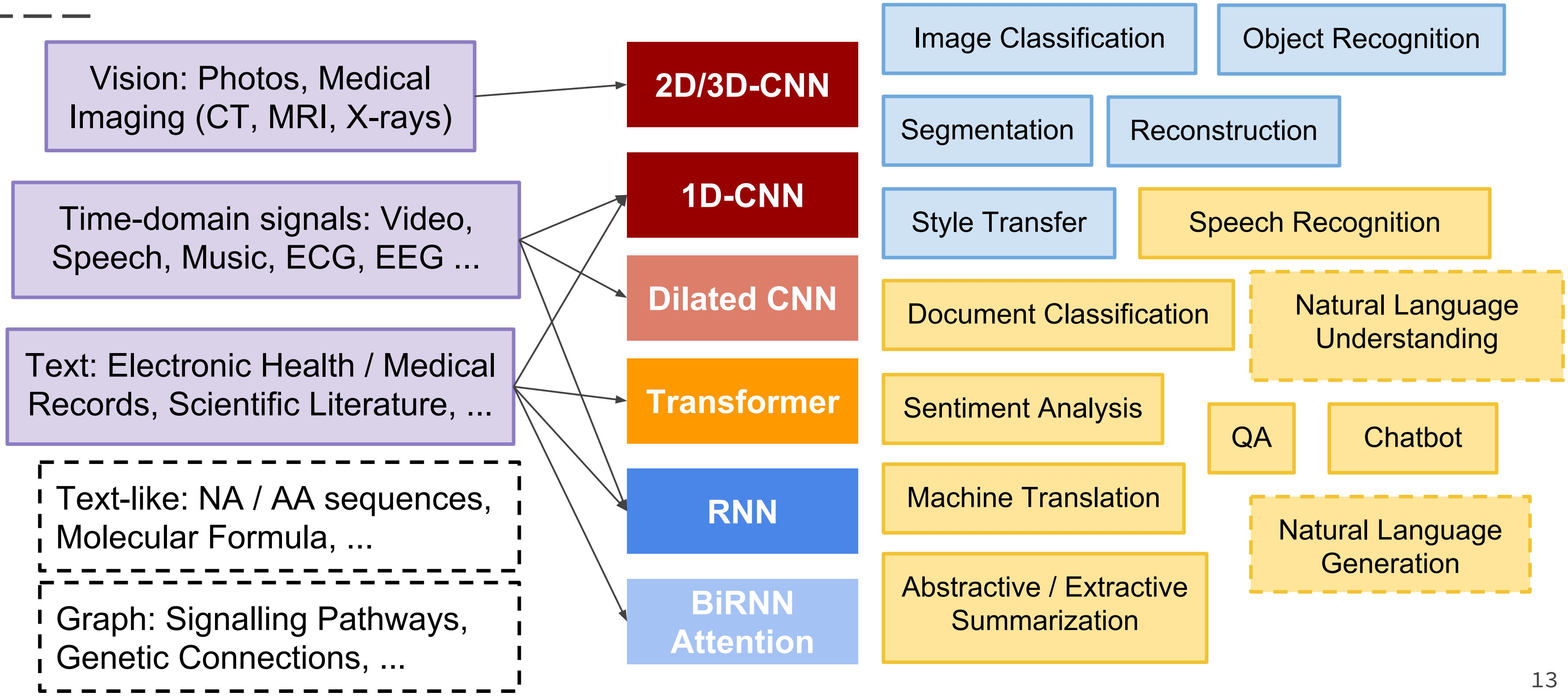
$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

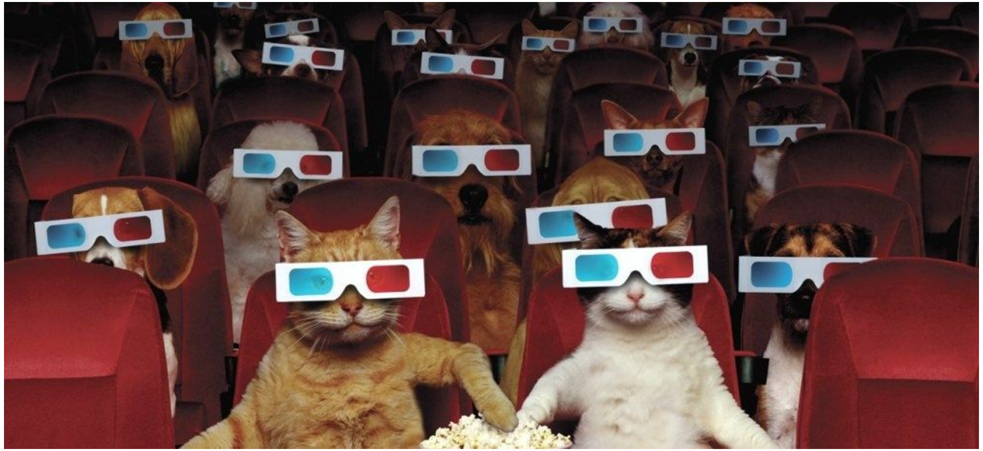
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

(2) CV and NLP A deep learning point of view



(2) CV and NLP

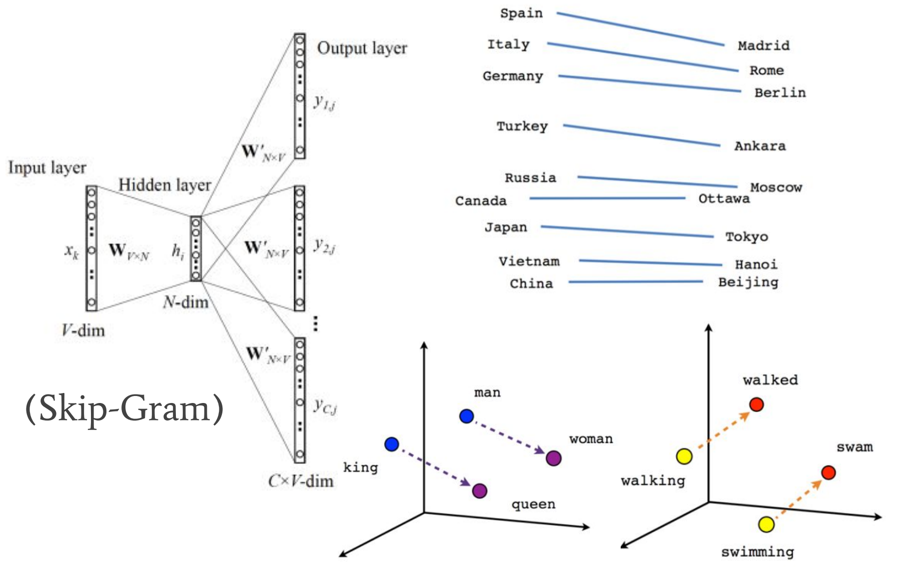
Demos

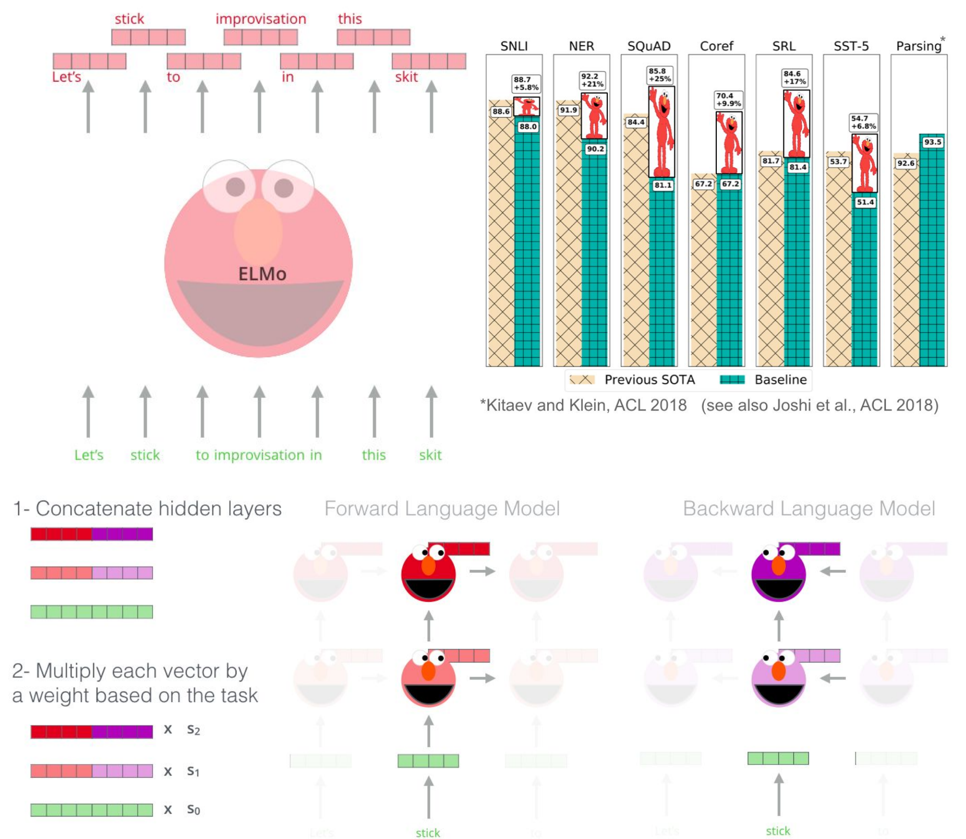


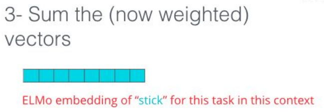
- Multimodal Unsupervised Image-to-image Translation, https://youtu.be/ab64TWzWn40
- A Style-Based Generator Architecture for Generative Adversarial Networks (GAN), https://youtu.be/kSLJriaOumA
- Image Inpainting for Irregular Holes Using Partial Convolutions, https://youtu.be/gg0F5JjKmhA
- Google's AI Assistant Can Now Make Real Phone Calls (Task-Oriented Chatbot, Speech Recognition, Natural Language Understanding, Text-to-Speech), https://youtu.be/JvbHu_bVa_g

(3) RL and Unsupervised

Unsupervised Learning: Word2Vec vs Elmo





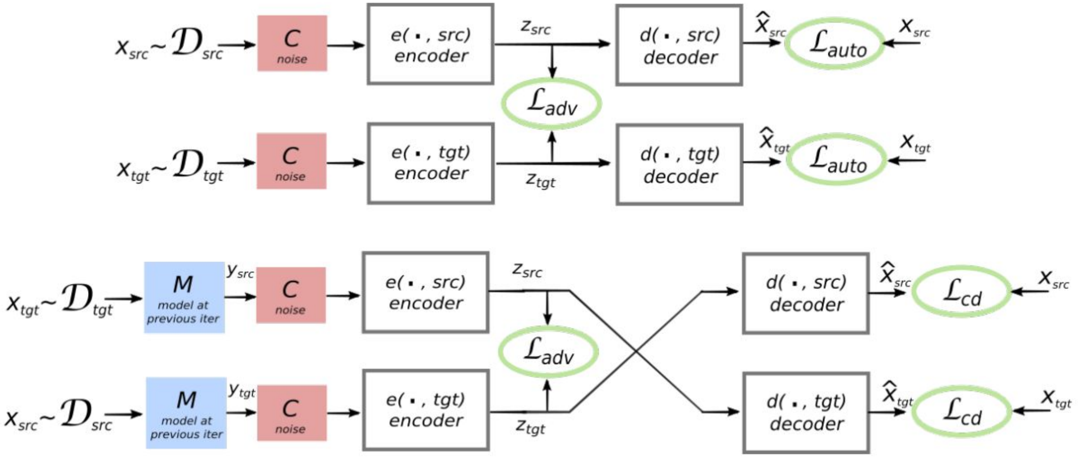


- http://jalammar.github.io/illustrated-bert/
- http://ruder.io/nlp-imagenet/

(3) RL and Unsupervised

Unsupervised Learning: Generative Adverserial Networks

- Huang et al. [Multimodal Unsupervised Image-to-Image Translation](). ECCV 2018.
- Lample et al. [Unsupervised Machine Translation Using Monolingual Corpora Only](). ICLR 2018.



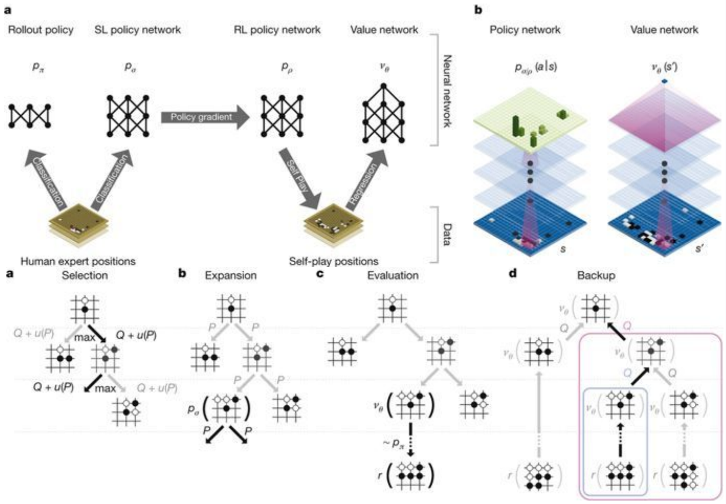
(a) edges → shoes
(b) big cats → house cats

(3) RL and Unsupervised

Reinforcement Learning: Atari DQN and AlphaGo

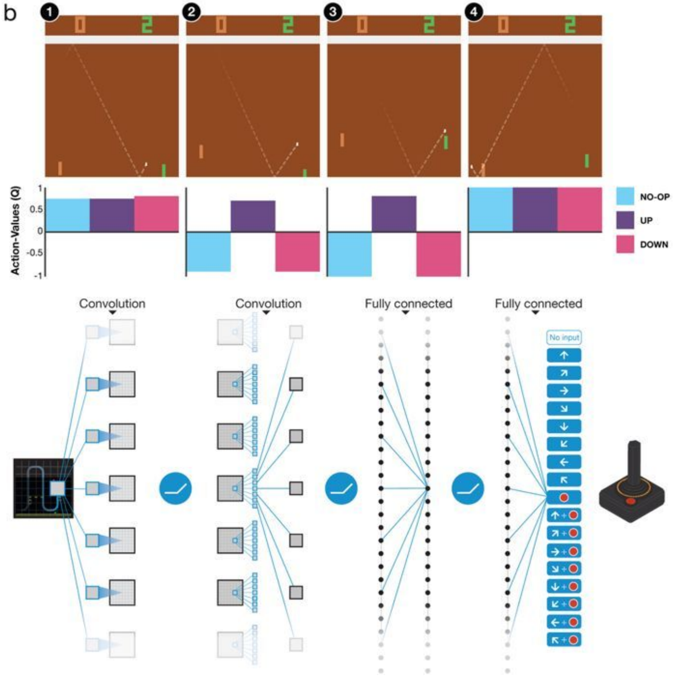
[AlphaGo](#)

Policy / Value Network
Monte-Carlo Tree Search



[Atari DQN](#)

Deep Q-Learning (From CNN to Action)
Master 49 Games



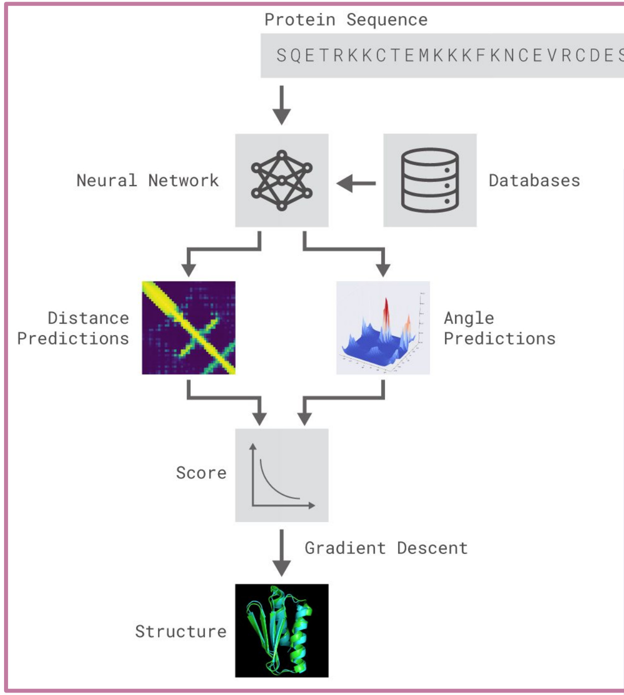
- Mnih et al. Human-level control through deep reinforcement learning. Nature 2015.
- Silver et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016.
- https://deepmind.com/blog/alphazero-shedding-new-light-grand-games-chess-shogi-and-go/

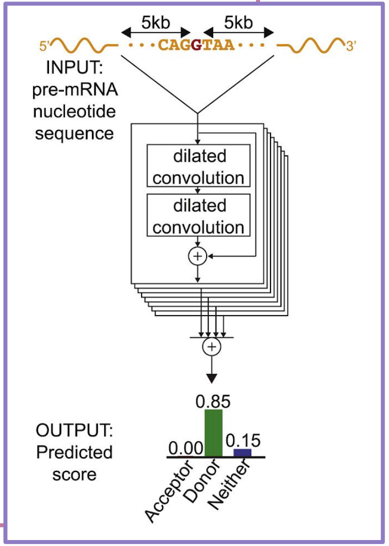
(4) Current AI Applications in Medicine

Structural / Molecular Biology

Predict **Protein Folding**: AlphaFold (DeepMind).

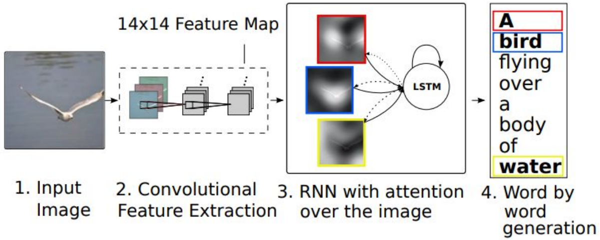
Predict **Pre-mRNA Splicing**: Jaganathan et al. Predicting Splicing from Primary Sequence with Deep Learning. Cell 2019.



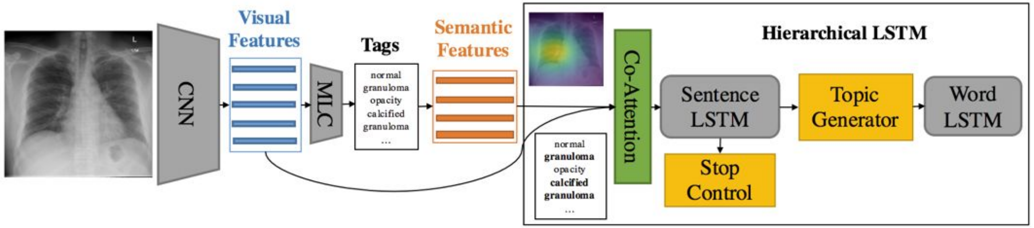


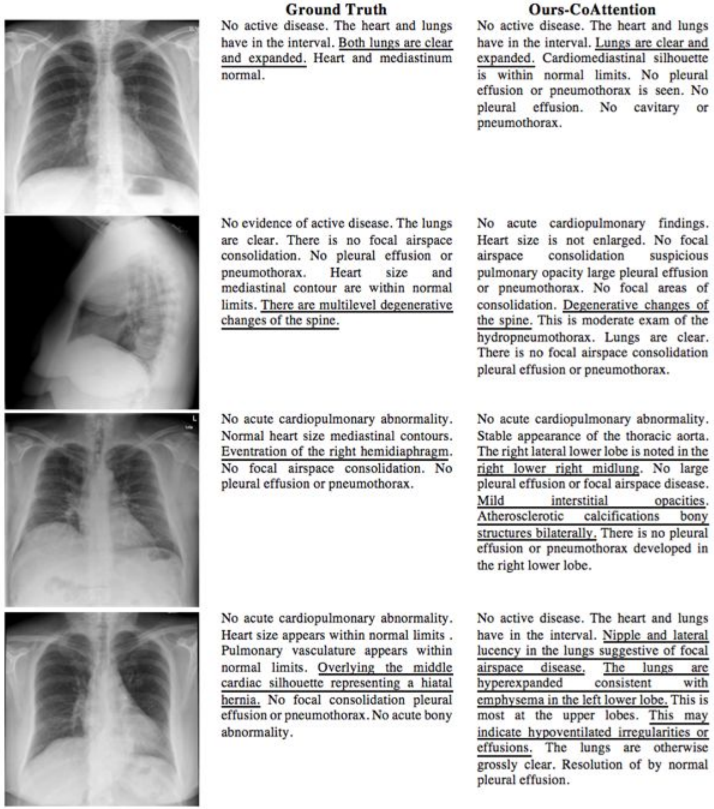
(4) Current AI Applications in Medicine

Xu et al. [Show, Attend and Tell: Neural Image Caption Generation with Visual Attention](). ICML 2015.

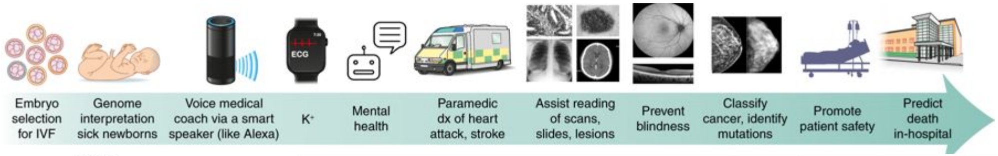


Jing et al. [On the Automatic Generation of Medical Imaging Reports](). ACL 2018.

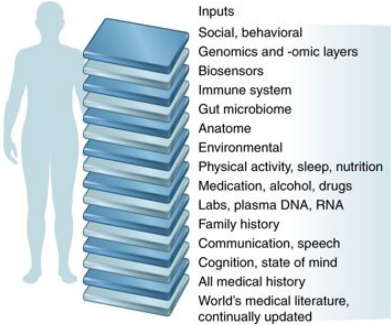




(4) Current AI Applications in Medicine



Medical Diagnosis, Prognosis



Specialty	Images	Publication
Radiology/neurology	CT head, acute neurological events	Titano et al. [27]
	CT head for brain hemorrhage	Arbabshirani et al.[19]
	CT head for trauma	Chilamkurthy et al.[20]
	CXR for metastatic lung nodules	Nam et al.[8]
	CXR for multiple findings	Singh et al.[7]
	Mammography for breast density	Lehman et al.[26]
	Wrist X-ray*	Lindsey et al.[9]
Pathology	Breast cancer	Ehteshami Bejnordi et al.[41]
	Lung cancer (+ driver mutation)	Coudray et al.[33]
	Brain tumors (+ methylation)	Capper et al.[45]
	Breast cancer metastases*	Steiner et al.[35]
	Breast cancer metastases	Liu et al.[34]
Dermatology	Skin cancers	Esteva et al.[47]
	Melanoma	Haenssle et al.[48]
	Skin lesions	Han et al.[49]
Ophthalmology	Diabetic retinopathy	Gulshan et al.[51]
	Diabetic retinopathy*	Abramoff et al.[31]
	Diabetic retinopathy*	Kanagasingam et al.[32]
	Congenital cataracts	Long et al.[38]
	Retinal diseases (OCT)	De Fauw et al.[56]
	Macular degeneration	Burlina et al.[52]
	Retinopathy of prematurity	Brown et al.[60]
	AMD and diabetic retinopathy	Kermany et al.[53]
Gastroenterology	Polyps at colonoscopy*	Mori et al.[36]
	Polyps at colonoscopy	Wang et al.[37]
Cardiology	Echocardiography	Madani et al.[23]
	Echocardiography	Zhang et al.[24]

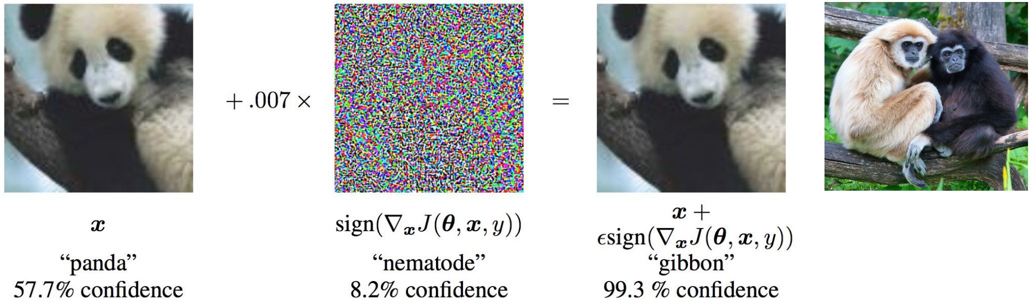
Prediction	n	AUC	Publication (Reference number)
In-hospital mortality, unplanned readmission, prolonged LOS, final discharge diagnosis	216,221	0.93* 0.75+ 0.85#	Rajkomar et al.[96]
All-cause 3–12 month mortality	221,284	0.93^	Avati et al.[91]
Readmission	1,068	0.78	Shameer et al.[106]
Sepsis	230,936	0.67	Horng et al.[102]
Septic shock	16,234	0.83	Henry et al.[103]
Severe sepsis	203,000	0.85@	Culliton et al.[104]
Clostridium difficile infection	256,732	0.82++	Oh et al.[93]
Developing diseases	704,587	range	Miotto et al.[97]
Diagnosis	18,590	0.96	Yang et al.[90]
Dementia	76,367	0.91	Cleret de Langavant et al.[92]
Alzheimer's Disease (+ amyloid imaging)	273	0.91	Mathotaarachchi et al.[98]
Mortality after cancer chemotherapy	26,946	0.94	Elfiky et al.[95]
Disease onset for 133 conditions	298,000	range	Razavian et al.[105]
Suicide	5,543	0.84	Walsh et al.[86]
Delirium	18,223	0.68	Wong et al.[100]

Company	FDA Approval	Indication
Apple	September 2018	Atrial fibrillation detection
Aidoc	August 2018	CT brain bleed diagnosis
iCAD	August 2018	Breast density via mammography
Zebra Medical	July 2018	Coronary calcium scoring
Bay Labs	June 2018	Echocardiogram EF determination
Neural Analytics	May 2018	Device for paramedic stroke diagnosis
IDx	April 2018	Diabetic retinopathy diagnosis
Icometrix	April 2018	MRI brain interpretation
Imagen	March 2018	X-ray wrist fracture diagnosis
Viz.ai	February 2018	CT stroke diagnosis
Arterys	February 2018	Liver and lung cancer (MRI, CT) diagnosis
MaxQ-AI	January 2018	CT brain bleed diagnosis
Alivecor	November 2017	Atrial fibrillation detection via Apple Watch
Arterys	January 2017	MRI heart interpretation

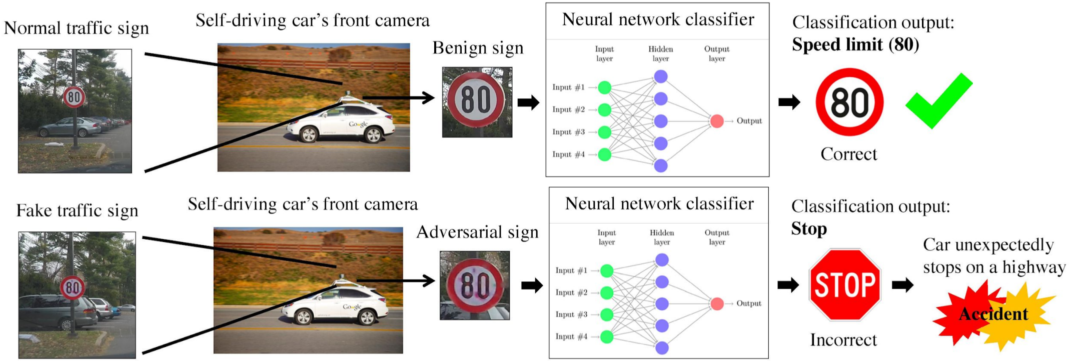
Recommended Reading: Topol. [High-performance medicine: the convergence of human and artificial intelligence](). Nature Medicine 2019.

(5) Some Challenges

- **Adverserial Examples**: Sitawarin et al. DARTS: Deceiving Autonomous Cars with Toxic Signs. ACM CCS 2018.
- The Telegraph: IBM Watson AI criticised after giving 'unsafe' cancer treatment advice.
- State-of-the-art AI systems could be **fragile**, subject to malicious attacks
- Goals for Future AI
  - **Robustness**
  - **Fairness**
  - **Interpretability**
  - **Data Efficiency**
  - **Model Complexity**



https://pytorch.org/tutorials/beginner/fgsm_tutorial.html



Key Takeaways

1. CNNs and RNNs are prevalent methods in deep learning
2. Many medical AI applications borrowed ideas for CV and NLP.
3. Use the data to find appropriate algorithms! (but not vice versa)
4. Solving a problem do not always require advanced AI / deep learning or even machine learning.
5. Use domain knowledge to establish good inductive bias.
6. Will doctors be replaced by AI? Try to build an AI and lets see!

Learn More

Machine Learning DIY: Scikit-Learn, Tensorflow(Keras), PyTorch, Google Cloud ML

Google: **Machine Learning Crash Course** https://developers.google.com/machine-learning/crash-course/

MLDS lectures by Prof. Hung-Yi Lee @ NTU EE (in Chinese)
http://speech.ee.ntu.edu.tw/~tlkagk/courses_MLDS18.html

Machine Learning for Health Workshop, NIPS 2018 https://ml4health.github.io/2018/pages/papers.html

AI for Social Good Workshop, NIPS 2018 https://aiforsocialgood.github.io/2018/

Stanford ML Group https://stanfordmlgroup.github.io/

Follow some works by Google Brain, DeepMind, Facebook AI Research, NVIDIA Research, …

Discussion

---

1. Ding et al. A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain. Radiology 2018.

2. Hannun et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine 2019. project site

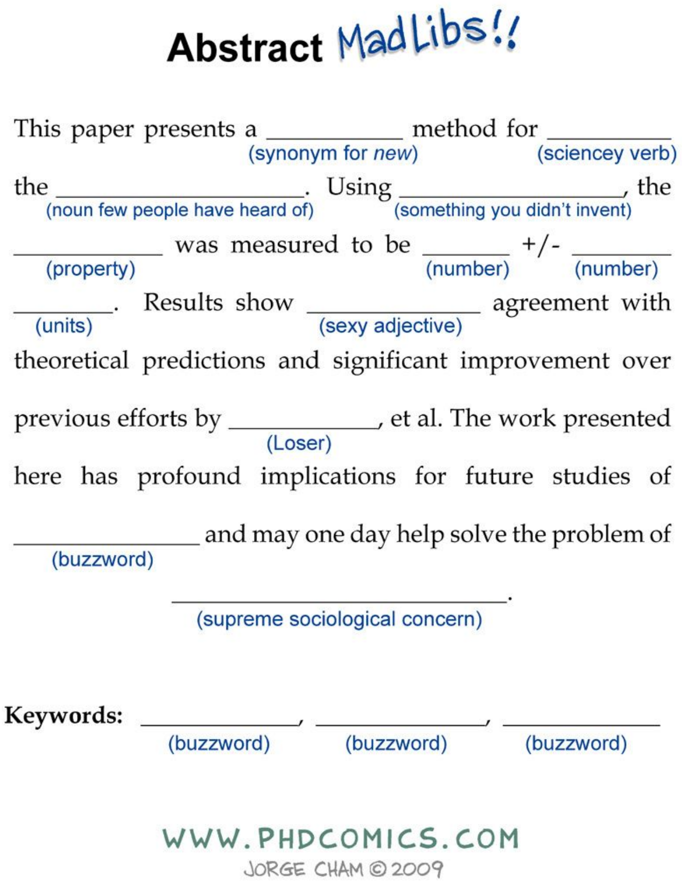
3. Ribeiro et al. Automatic Diagnosis of Short-Duration 12-Lead ECG using a Deep Convolutional Network. ML4H Workshop, NIPS 2018.

4. Komorowski et al. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. Nature Medicine 2018. related work 1, related work 2

Discussion

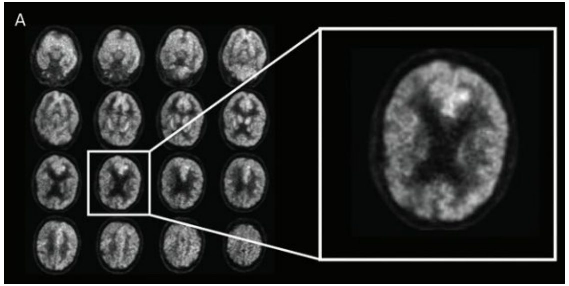
__Paper Presentation: 3 min, be concise !!!__

1. What's the **objective** of this AI application?
2. Describe the **dataset**, including size, dimensions, endpoints, etc.
3. Describe the authors' **approach**: which algorithm did they use and why?
4. Describe the **evaluation** method and results.
5. (Opt) What's the significance of this work?
6. (Opt) What are the limitations discussed by the authors?



PET Scans for Alzheimers

- Mild Cognitive Impairment (MCI) to Alzheimer's Disease (AD)
- Alzheimer’s Disease Neuroimaging Initiative (ADNI) (2109 imaging studies / 1002 patients) and independent test set (40 imaging studies / 40 patients), Size: 512×512
- Model: Inception V3 CNN



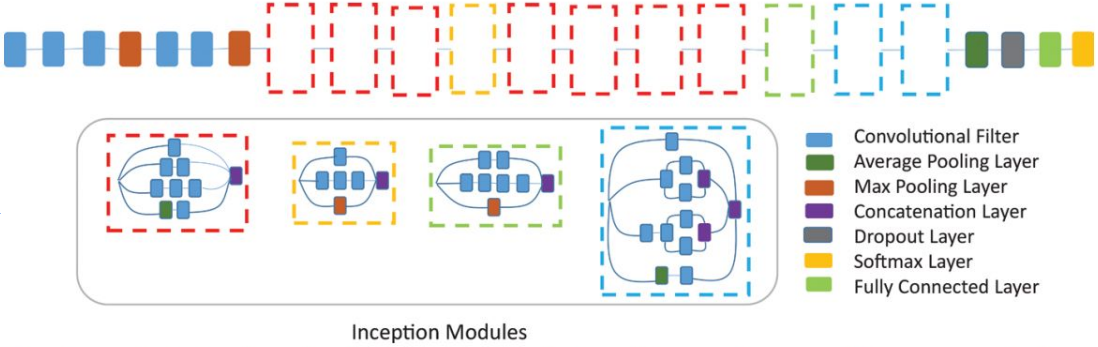


Figure 3: Convolutional neural network architecture, Inception v3, used in this study. Inception v3 network stacks 11 inception modules where each module consists of pooling layers and convolutional filters with rectified linear units as activation function. The input of the model is two-dimensional images of 16 horizontal sections of the brain placed on 4 × 4 grids as produced by the preprocessing step. Three fully connected layers of size 1024, 512, and 3 are added to the final concatenation layer. A dropout with rate of 0.6 is applied before the fully connected layers as means of regularization. The model is pretrained on ImageNet dataset and further fine-tuned with a batch size of 8 and learning rate of 0.0001.

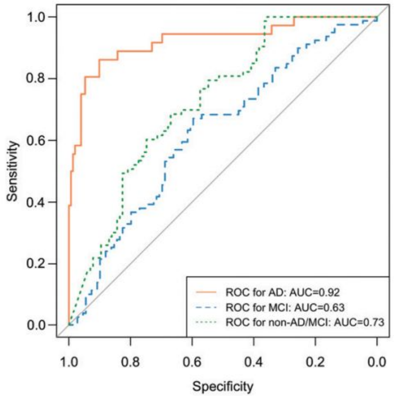
PET Scans for Alzheimers

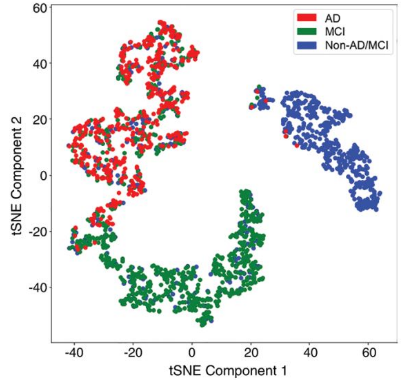
- AUC = 0.98 [0.94, 1.00], specificity = 82% when sensitivity = 100%
- (average of 75.8 months prior to the final diagnosis.)

Table 2: Performance Comparison of Deep Learning Algorithm and Radiology Readers

Parameter	Sensitivity (%)*	Specificity (%)*	Precision (%)*	F1 Score (%)	No. of Imaging Studies
Deep learning model on 10% ADNI set					
AD	81 (29/36)	94 (143/152)	76 (29/38)	78	36
MCI	54 (43/79)	68 (74/109)	55 (43/78)	55	79
Non-AD/MCI	59 (43/73)	75 (86/115)	60 (43/72)	59	73
Deep learning model on independent test set					
AD	100 (7/7)†	82 (27/33)	54 (7/13)	70†	7
MCI	43 (3/7)†	58 (19/33)	18 (3/17)†	25†	7
Non-AD/MCI	35 (9/26)	93 (13/14)†	90 (9/10)†	50	26
Radiology readers on independent test set					
AD	57 (4/7)	91 (30/33)	57 (4/7)	57	7
MCI	14 (1/7)	76 (25/33)	11 (1/9)	13	7
Non-AD/MCI	77 (20/26)	71 (10/14)	83 (20/24)	80	26

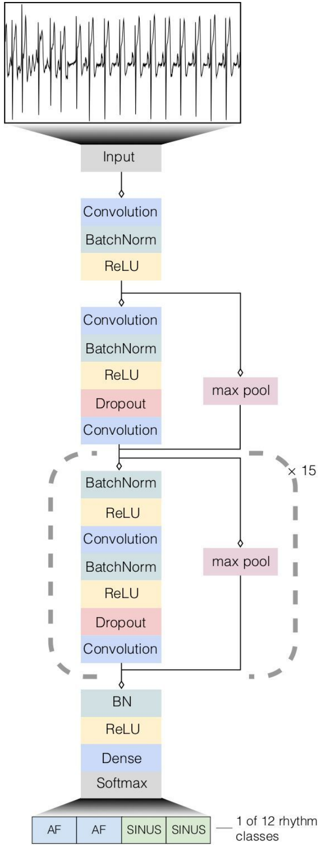
Note.–Unless otherwise indicated, data are averages ± standard deviation. ADNI = Alzheimer's Disease Neuroimaging Initiative, AD = Alzheimer disease, MCI = mild cognitive impairment, Non-AD/MCI = neither Alzheimer disease nor mild cognitive impairment.
* Numbers in parentheses are raw data used to calculate the percentage.
† Numbers indicate higher performance from deep learning algorithm compared with reader performance on independent test set.

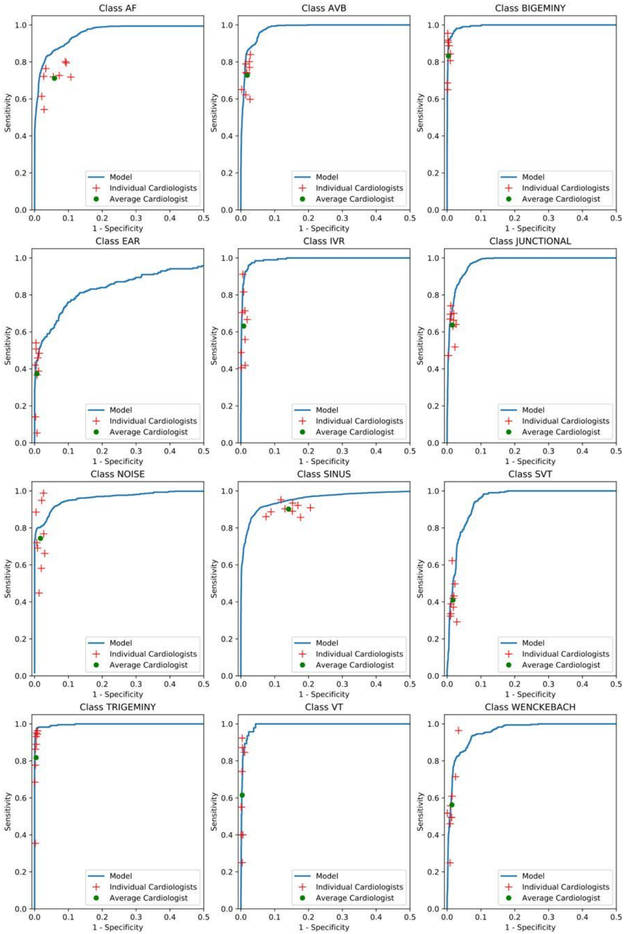




ECG with CNN

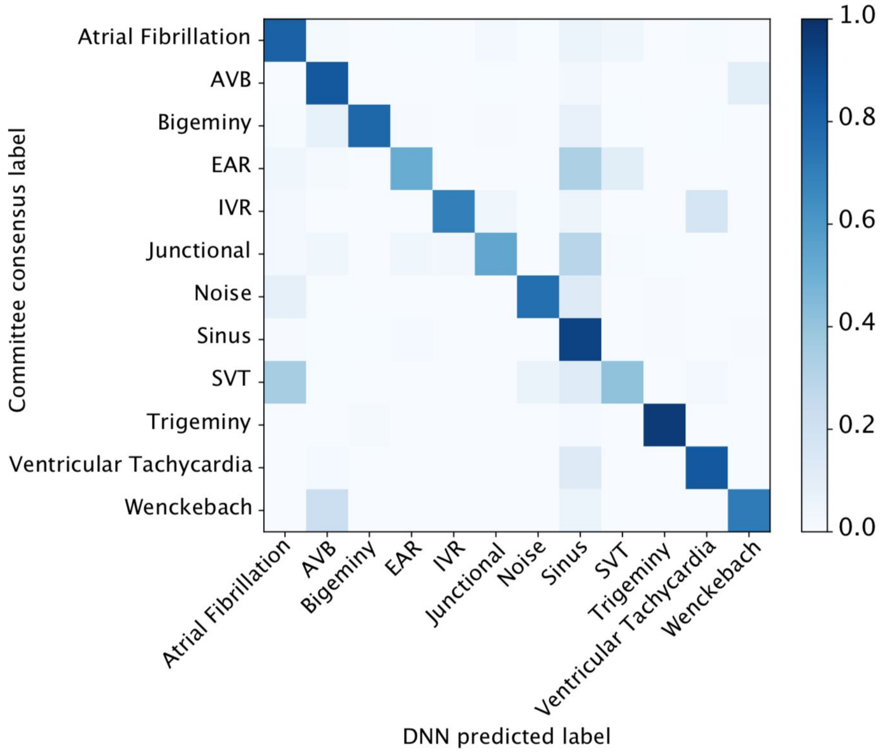
- Single-lead (Lead II), raw ECG (200 Hz), prediction every 256 input samples.
- Training: 91,232 ECG records from 53,549 patients
- Validation: 328 ECG records collected from 328 unique patients
- Model: 34 layer ResNet





ECG with CNN

- Tachyarrythmia
  - Atrial (**SVT**): PSVT(AVNRT, WPW/AVRT), Atrial Flutter, **AFib**
  - Ventricular: **VT**, **IVR**, VFib
  - **Junctional**: JET
- Bradyarrythmia
  - **AVB**: first, second(**type 1 (Mobitz I/Wenckebach**), type 2 (Mobitz II/Hay)), third
- Dysarrhythmia
  - VPC: **Bigeminy**, **Trigeminy**
  - APC (**EAR**?)



"The two confusion matrices exhibit a similar pattern, highlighting those rhythm classes that were generally more problematic to classify (that is, supraventricular tachycardia (SVT) versus atrial fibrillation, junctional versus sinus rhythm, and EAR versus sinus rhythm)."

Sepsis Treatment with RL

MIMIC-III (17,083 admissions) for model development (80% training, 20% validation), eRI (79,073 admissions) for testing. Patients that fit sepsis-3 criteria.

- **State**: 4 hr/step, 72hr/patient
- **Action**: 25 actions
- **Reward**: 90-day mortality (+100/-100)
- Evaluation: HCOPE(WIS)

MDP solved by **policy-iteration**

$$\pi(s) := \arg\max_a \left\{ \sum_{s'} P(s'|s,a) \left(R(s'|s,a) + \gamma V(s') \right) \right\}$$

$$V(s) := \sum_{s'} P_{\pi(s)}(s,s') \left(R_{\pi(s)}(s,s') + \gamma V(s') \right)$$

State

(Discretized using k-means)

Category	Items	Type	Available in MIMIC-III	Available in eRI
Demographics	Age	Cont.	+	+
	Gender	Binary	+	+
	Weight	Cont.	+	+
	Readmission to intensive care	Binary	+	+
	Elixhauser score (premorbid status)	Cont.	+	-
Vital signs	Modified SOFA*	Cont.	+	+
	SIRS	Cont.	+	+
	Glasgow coma scale	Cont.	+	+
	Heart rate, systolic, mean and diastolic	Cont.	+	+
	blood pressure, shock index	Cont.	+	+
	Respiratory rate, SpO_2	Cont.	+	+
	Temperature			
Lab values	Potassium, sodium, chloride	Cont.	+	+
	Glucose, BUN, creatinine	Cont.	+	+
	Magnesium, calcium, ionized calcium,	Cont.	+	+
	carbon dioxide	Cont.	+	+
	SGOT, SGPT, total bilirubin, albumin	Cont.	+	+
	Hemoglobin	Cont.	+	+
	White blood cells count, platelets	Cont.	+	+
	count, PTT, PT, INR			
	pH, PaO_2, $PaCO_2$, base excess,			
	bicarbonate, lactate, PaO_2/FiO_2 ratio			
Ventilation parameters	Mechanical ventilation	Binary	+	+
	FiO_2	Cont.	+	+
Medications and fluid balance	Current IV fluid intake over 4h	Cont.	+	+
	Maximum dose of vasopressor over 4h	Cont.	+	+
	Urine output over 4h	Cont.	+	+
	Cumulated fluid balance since admission (includes preadmission data when available)	Cont.	+	+
Outcome	Hospital mortality	Binary	+	+
	90-day mortality	Binary	+	-

Reward

Supplementary Table 2. Description of the variables included in the datasets. Cont.: continuous; INR: International Normalized Ratio; * Modified SOFA: SOFA based on values in the current 4h time step; PEEP: Positive End Expiratory Pressure; PT: Prothrombin Time; PTT: Partial Thromboplastin Time; SIRS: Systemic Inflammatory Response Syndrome; Shock index: systolic blood pressure/heart rate.

Actions

Discretized action	IV fluids (mL/4 hours)		Vasopressors (mcg/kg/min)	
	Range	Median dose	Range	Median dose
1	0	0	0	0
2	]0-50]	30	]0-0.08]	0.04
3	]50-180]	85	]0.08-0.22]	0.13
4	]180-530]	320	]0.22-0.45]	0.27
5	>530	946	>0.45	0.68

Discussion

Design a medical AI application

1. What's the **objective** of this AI application?
2. Where is the **data**?
3. How to collect the desired data?
4. What kind of **model** will you choose and why?
5. What is the **evaluation criteria** for your model?

Thank you!
Any Questions?

If you have further inquiries, please email me:
b05401009@ntu.edu.tw