이화여자대학교
EWHA WOMANS UNIVERSITY

CDW 연구 활용2 : CDW, CDM, 병원의무기록실 청구데이터를 활용한 다기관 machine learning 연구

이화여대 의과대학 환경의학교실 김이준

EWHA,
THE FUTURE
WE CREATE

목차

Machine learning이란?

왜 의학에 인공지능이 도입되어야 하는가?

✓ 인간과 비슷하거나 인간보다 뛰어난 **진단** 능력

- 폐암, 결핵 진단
- 병리 소견 판독

✓ 인간보다 뛰어난 **사전 예측** 능력

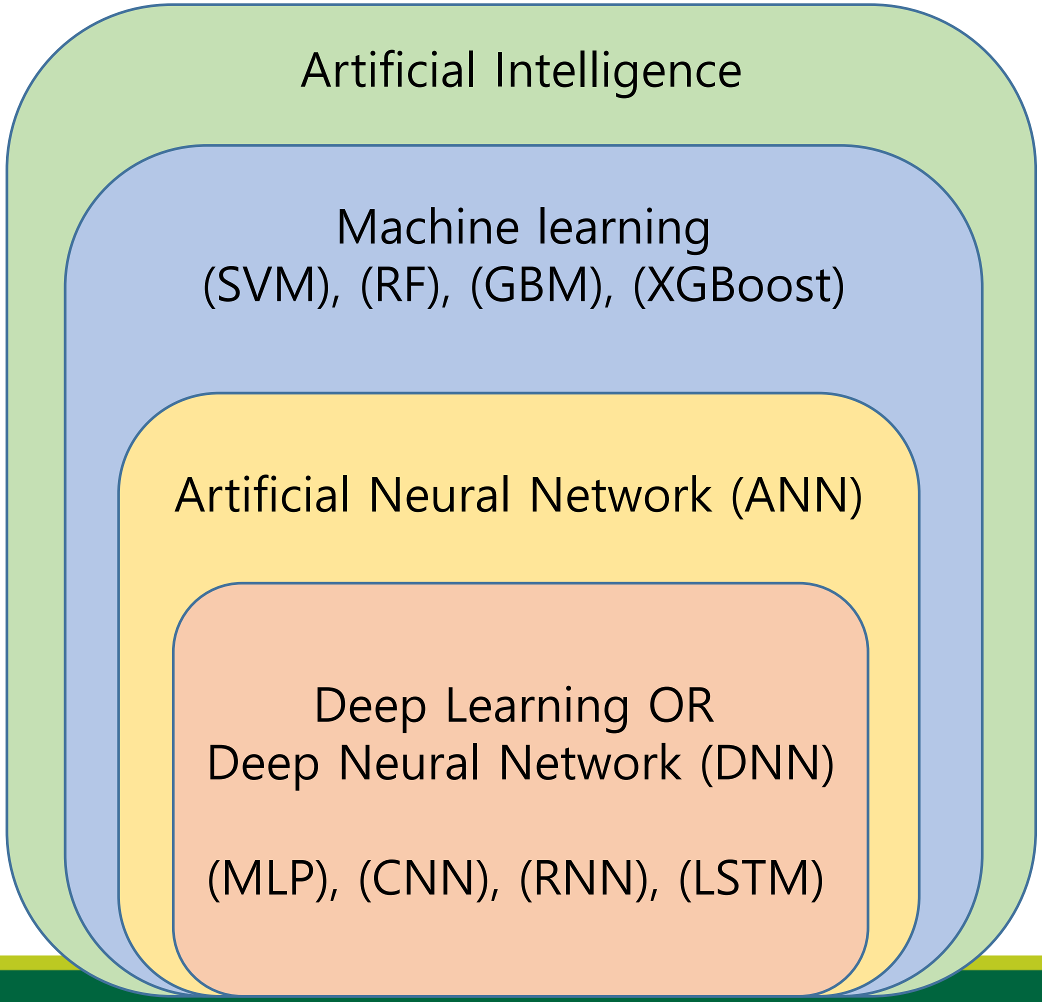
- 예후 위험 신호 탐지
- 암의 재발 위험 탐지

✓ 새로운 **창조** 능력

- Resolution이 부족한 병리 영상의 고화질화
- 가상환자 데이터 생성
- 새로운 promoter의 sequencing 생성
- Coding, Paper writing… (ChatGPT)

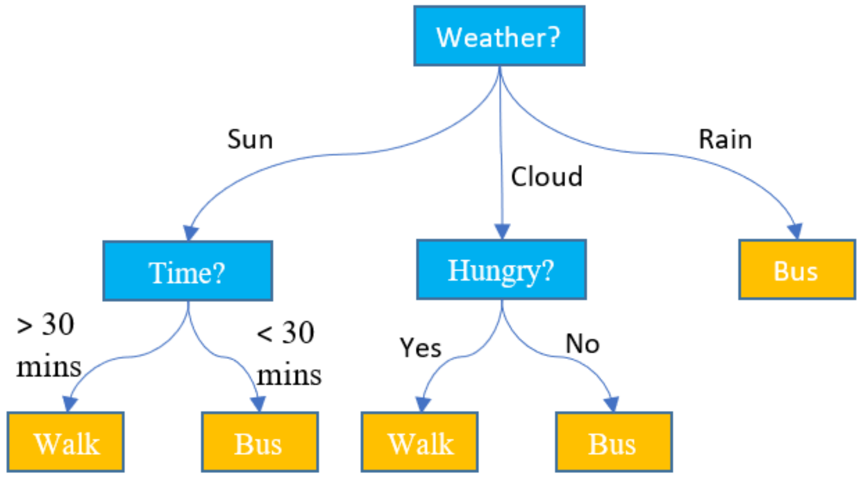
ANN, DNN, CNN, RNN???

- ✓ ANN: Artificial neural network
- ✓ DNN: Deep neural network
- ✓ CNN: Convolutional neural network (합성곱신경망)
- ✓ RNN: Recurrent Neural network (순환신경망)



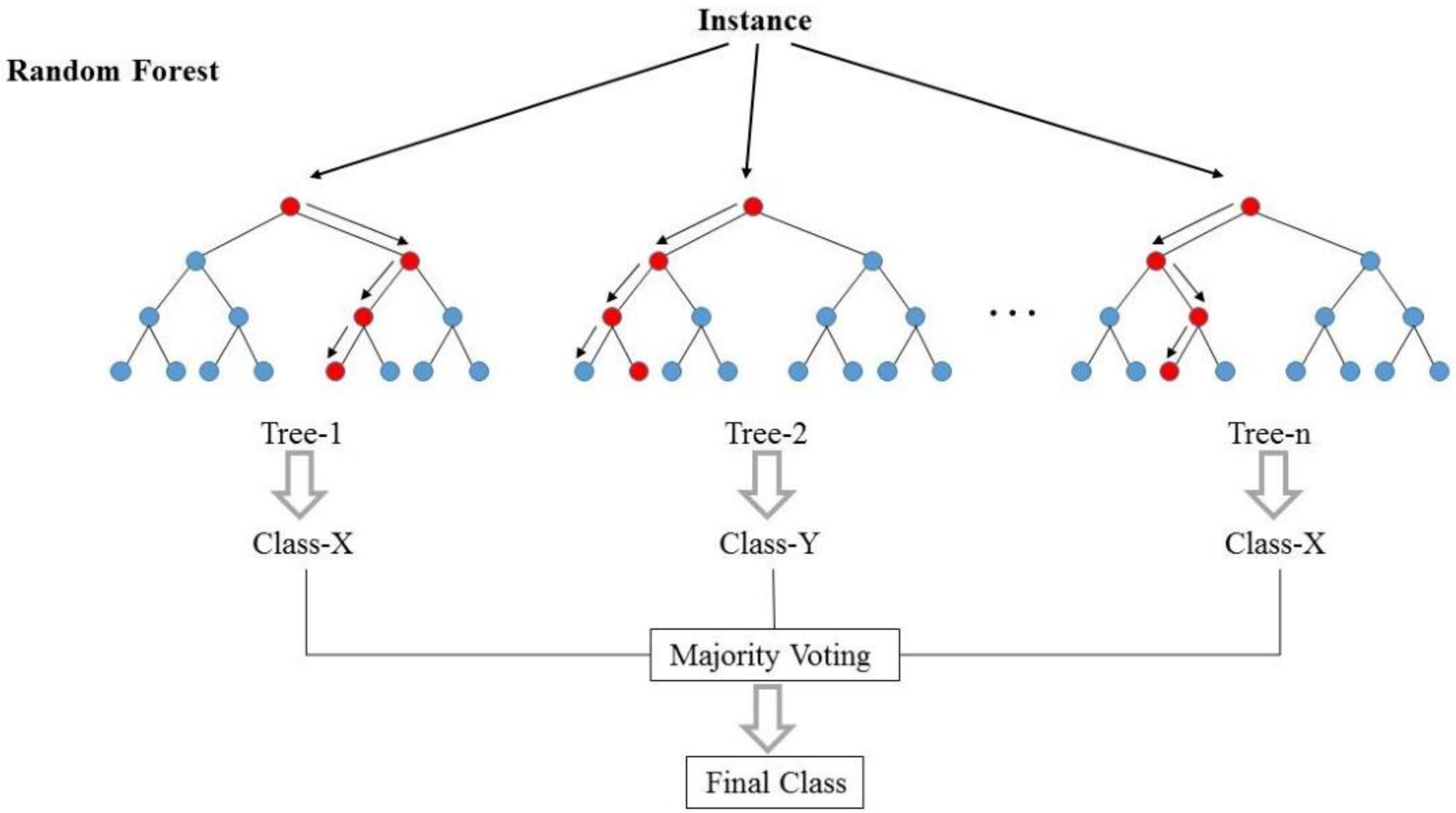
결정트리 (Decision tree)

- 데이터에 있는 규칙을 학습을 통해 자동으로 찾아내 트리(Tree) 기반의 분류 규칙을 만드는 것
- "스무고개 게임": if, else
- 규칙이 많다는 것은 곧 분류를 결정하는 방식이 더욱 복잡해진다는 것이므로 곧 과대적합으로 이어지기 쉽다. 즉, 트리의 깊이(depth)가 깊어질수록 결정 트리의 예측 성능이 저하될 가능성이 높다.
- 장점: 정보의 "균일도"라는 룰을 기반으로 하고 있어서 알고리즘이 쉽고 직관적이다.
- 단점: 과대적합되어 정확도가 떨어질 수 있다. 이를 극복하기 위해 트리의 크기를 사전에 제한하는 튜닝이 필요하다.



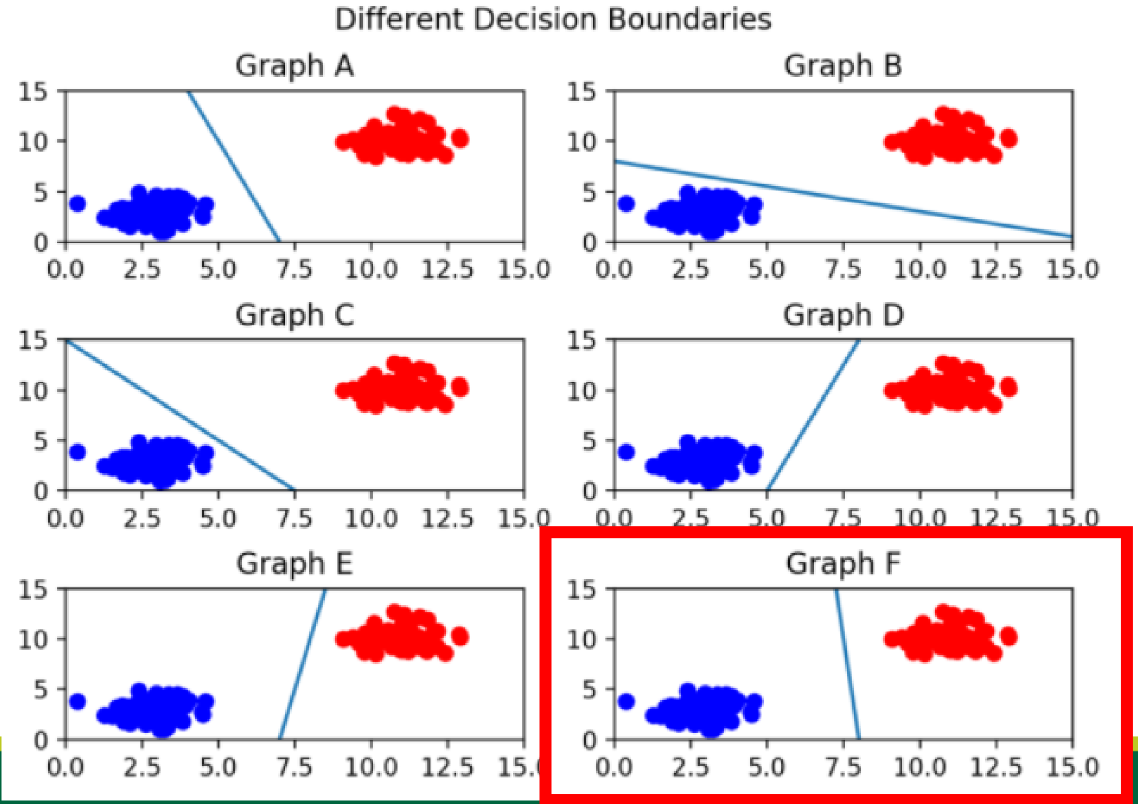
랜덤 포레스트(RF, random forest)

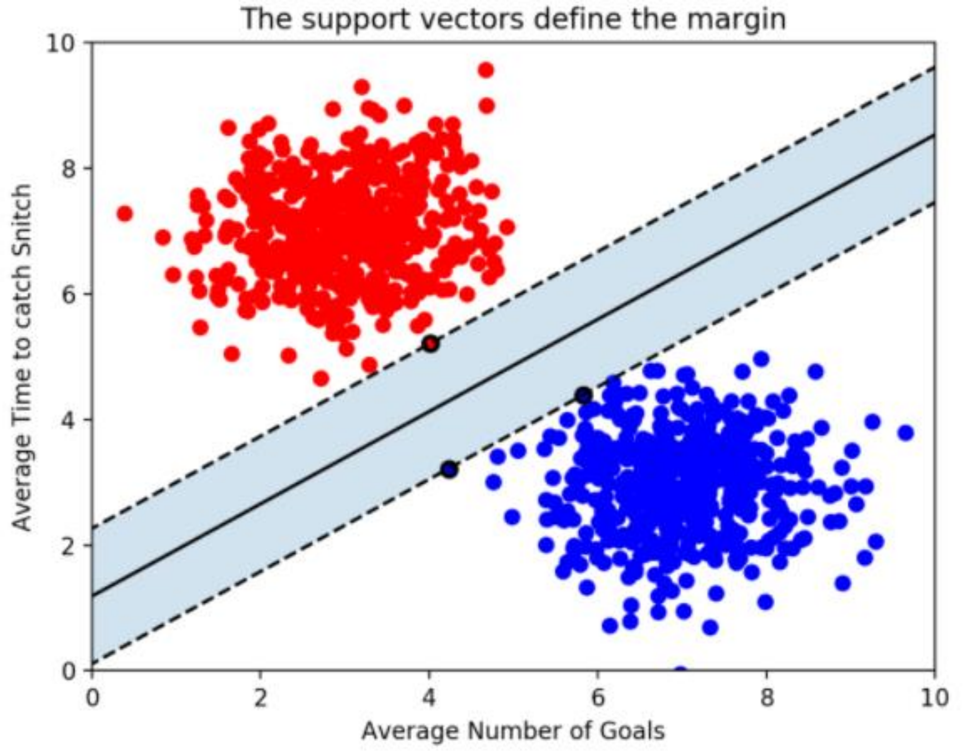
- 여러 개의 결정 트리 분류기가 전체 데이터에서 배깅(Bagging) 방식으로 각자의 데이터를 샘플링해 개별적으로 학습을 수행한 뒤 최종적으로 모든 분류기가 보팅을 통해 예측 결정을 함.
- 수행 성능 빠른 편.



서포트 벡터 머신(SVM, support vector machine)

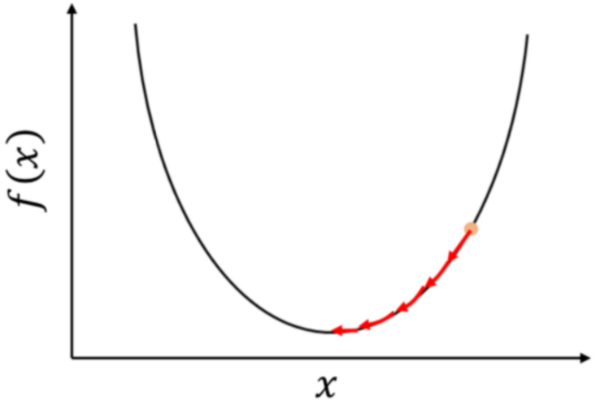
✓ 서포트 벡터 머신(SVM)은 결정 경계(Decision Boundary), 즉 분류를 위한 기준 선을 정의하는 모델. 최적의 결정 경계는 마진(margin)을 최대



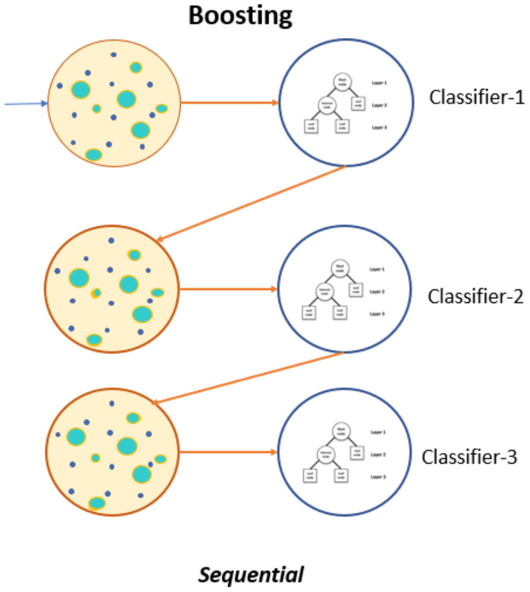


Gradient Boosting Machine (GBM)

- ✓ 부스팅 알고리즘은 여러 개의 약한 학습기(weak learner)를 순차적으로 학습/예측하면서 잘못 예측된 데이터에 가중치 부여를 통해 오류를 개선해 나가면서 학습하는 방식
- ✓ GBM은 가중치 업데이트를 경사 하강법(Gradient Descent) 방법을 이용함.
- ✓ 오류값 = 실제값 - 예측값



Gradient



Boosting

https://machinelearningmastery.com/develop-first-xgboost-model-python-scikit-learn/

XGBoost

- ✓ 트리기반의 앙상블 중 가장 각광받고 있음.
- ✓ GBM 에 기반하고 있음.
- ✓ GBM의 단점인 느린 수행 시간을 병렬 수행 등 다양한 기능으로 해결함.
- ✓ GBM에는 과대적합 규제(regulation) 기능 없으나 XGboost 에는 있음.
- ✓ 뛰어난 예측 성능.

```
# First XGBoost model for Pima Indians dataset
from numpy import loadtxt
from xgboost import XGBClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
# load data
dataset = loadtxt('pima-indians-diabetes.csv', delimiter=",")
# split data into X and y
X = dataset[:,0:8]
Y = dataset[:,8]
# split data into train and test sets
seed = 7
test_size = 0.33
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size=test_size, random_state=seed)
# fit model no training data
model = XGBClassifier()
model.fit(X_train, y_train)
# make predictions for test data
y_pred = model.predict(X_test)
predictions = [round(value) for value in y_pred]
# evaluate predictions
accuracy = accuracy_score(y_test, predictions)
print("Accuracy: %.2f%%" % (accuracy * 100.0))
```

Xgboost vs Neural Network

scikit learn

Neural Network

XGBoost

Binary classification

Xgboost 16:3 Neural Network

Multiclass classification

Xgboost 4:3 Neural Network

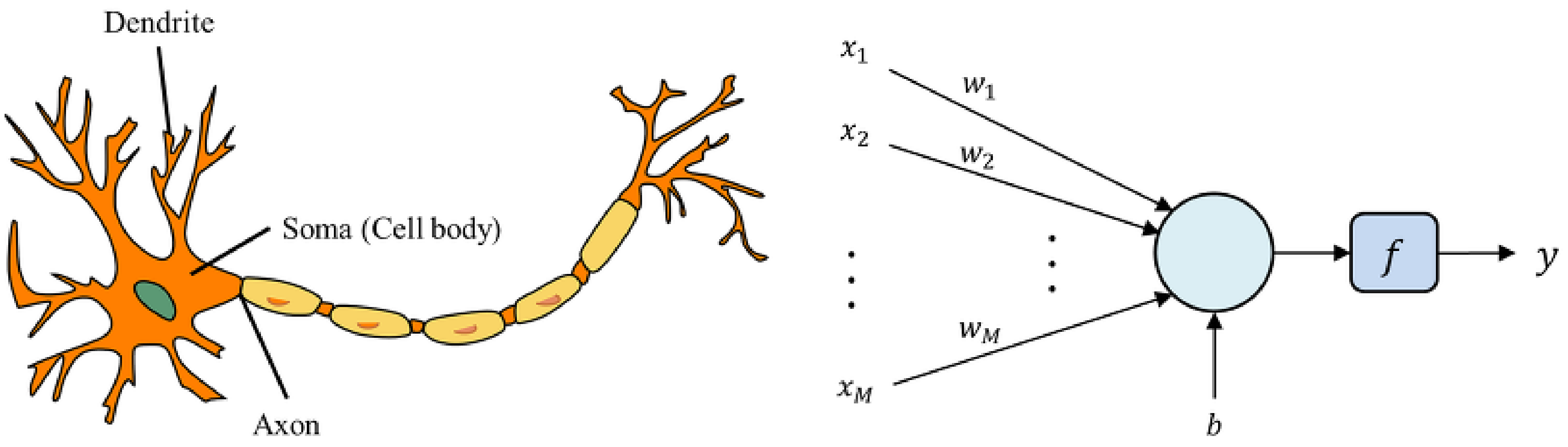
Regression

Xgboost 14:2 Neural Network

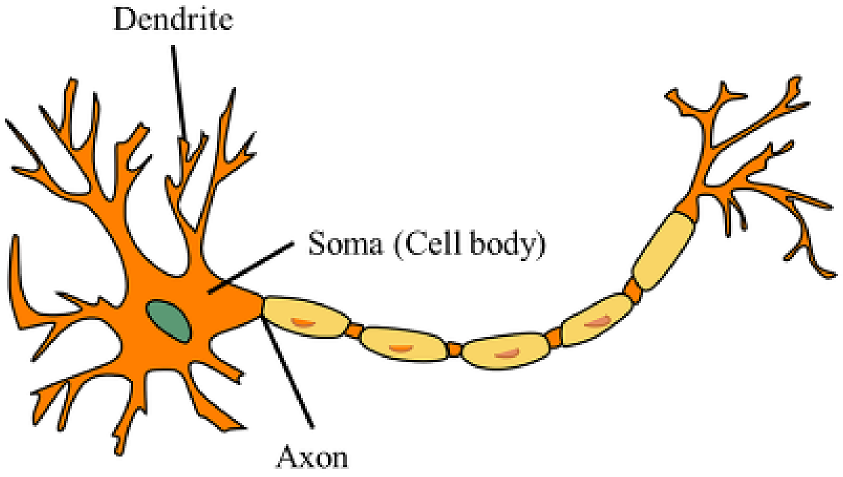
XGBoost

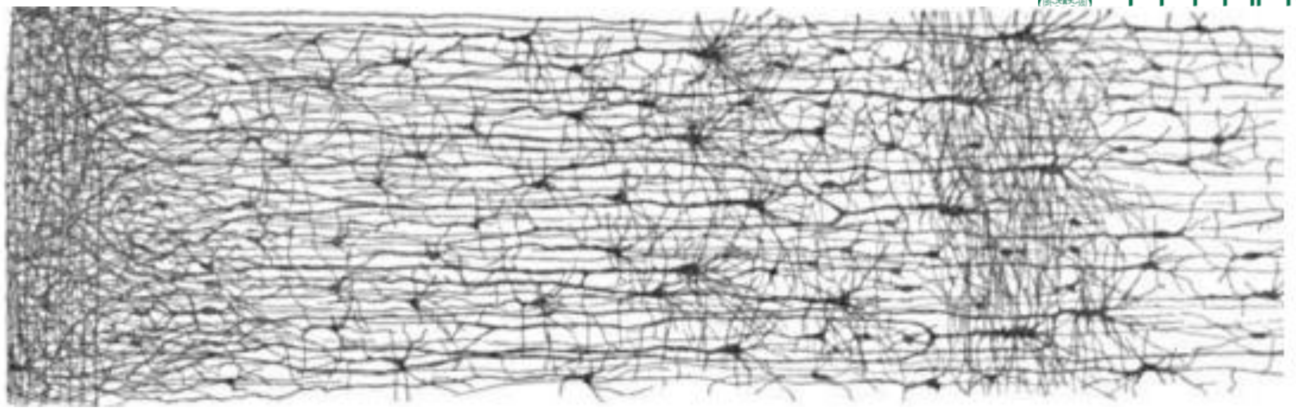
딥러닝 (Deep neural network)

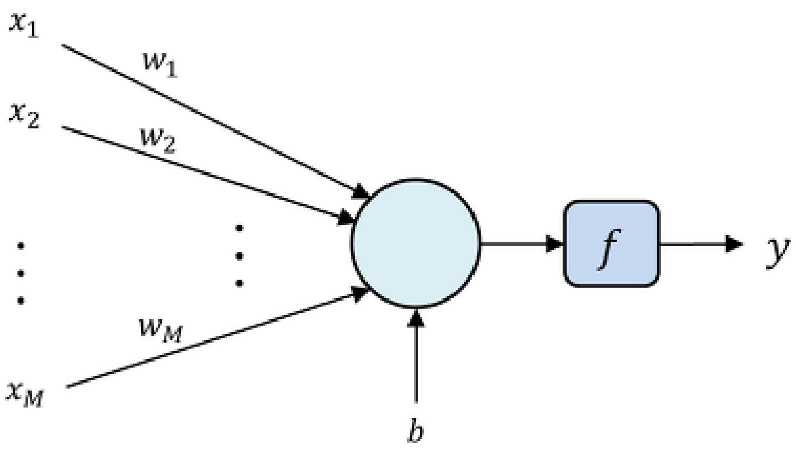
✓ 많은 층으로 구성된 심층 신경망의 학습 구조

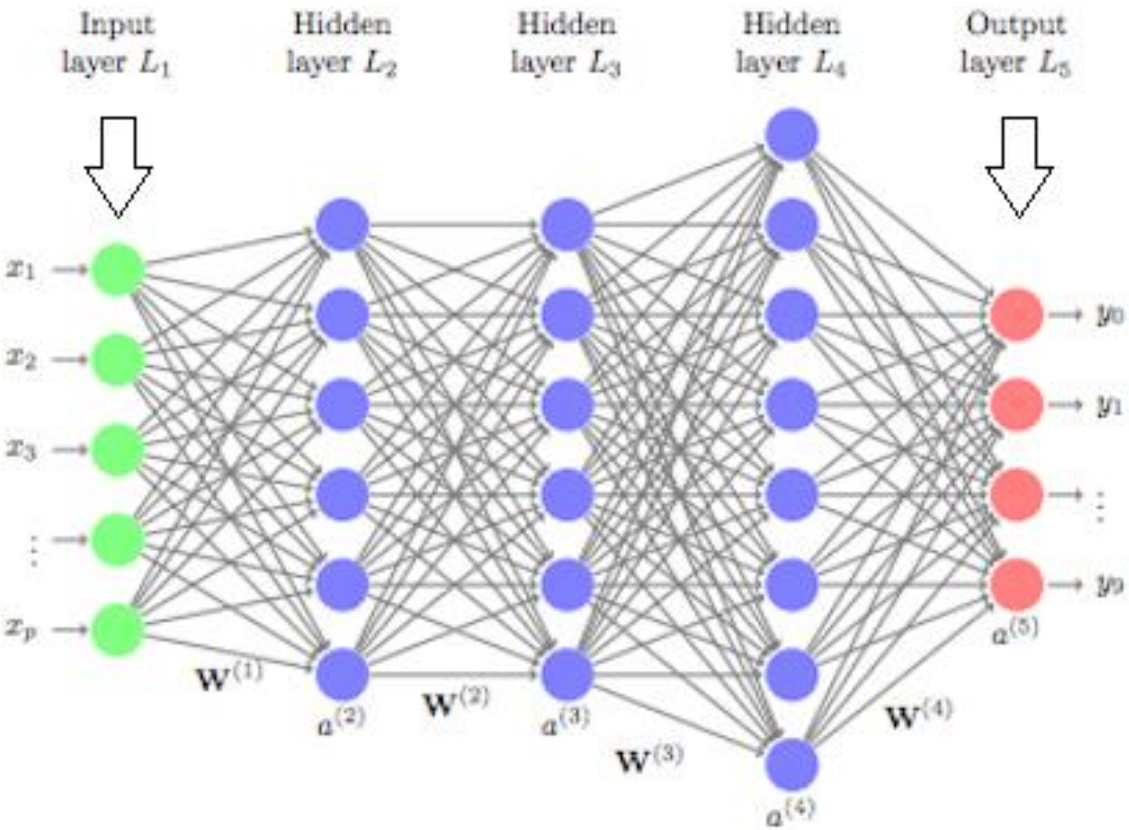


https://untitledtblog.tistory.com/141

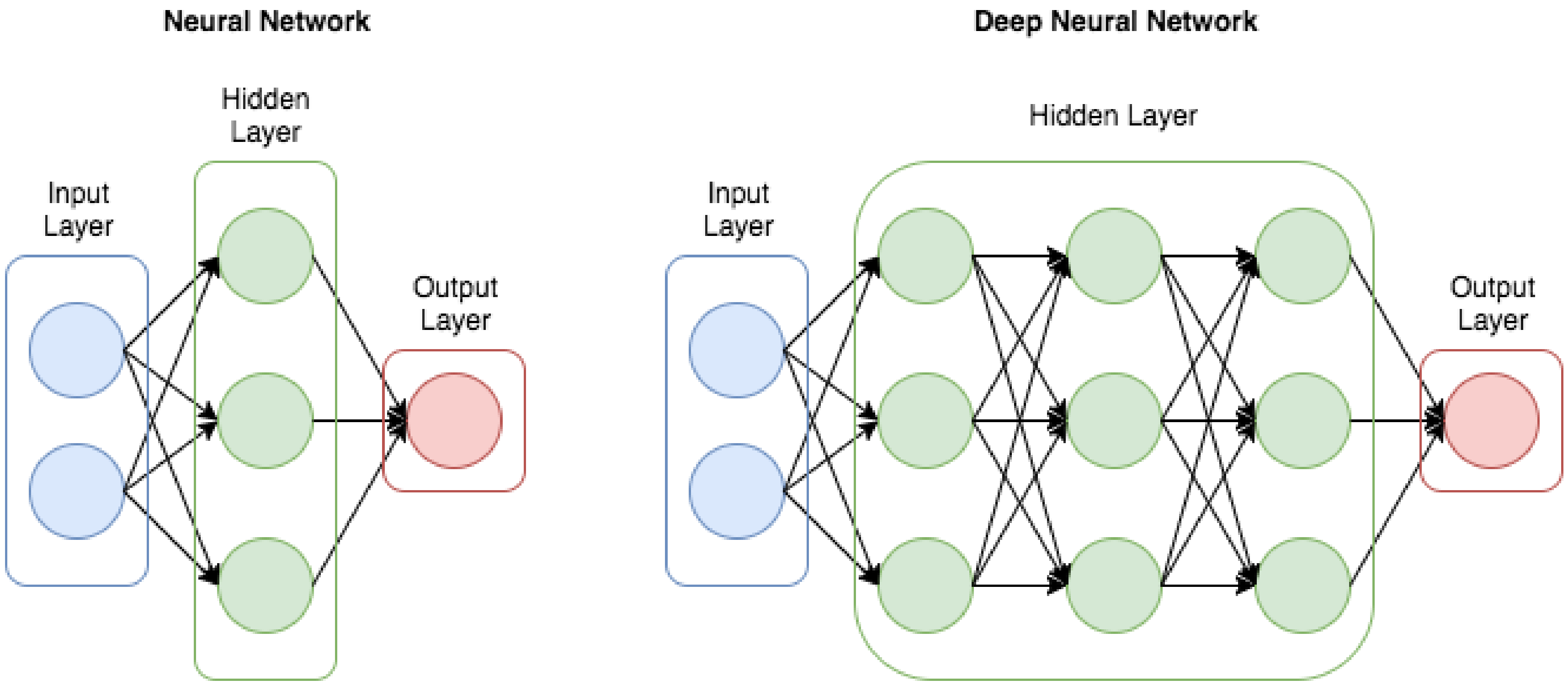

Dendrite
Soma (Cell body)
Axon




x_1
w_1
x_2
w_2
w_M
x_M
f
y
b


Input layer L_1
Hidden layer L_2
Hidden layer L_3
Hidden layer L_4
Output layer L_5
x_1
x_2
x_3
x_p
y_0
y_1
y_9
$\mathbf{W}^{(1)}$
$\mathbf{W}^{(2)}$
$\mathbf{W}^{(3)}$
$\mathbf{W}^{(4)}$
$a^{(2)}$
$a^{(3)}$
$a^{(4)}$
$a^{(5)}$

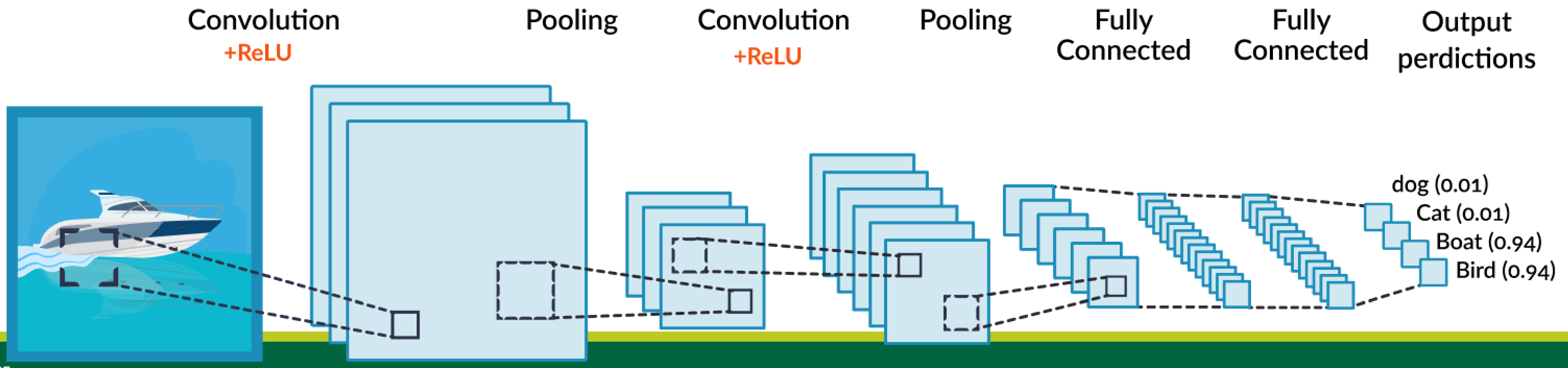
ANN vs. DNN (은닉층 수가 많다.)



https://ebbnflow.tistory.com/119

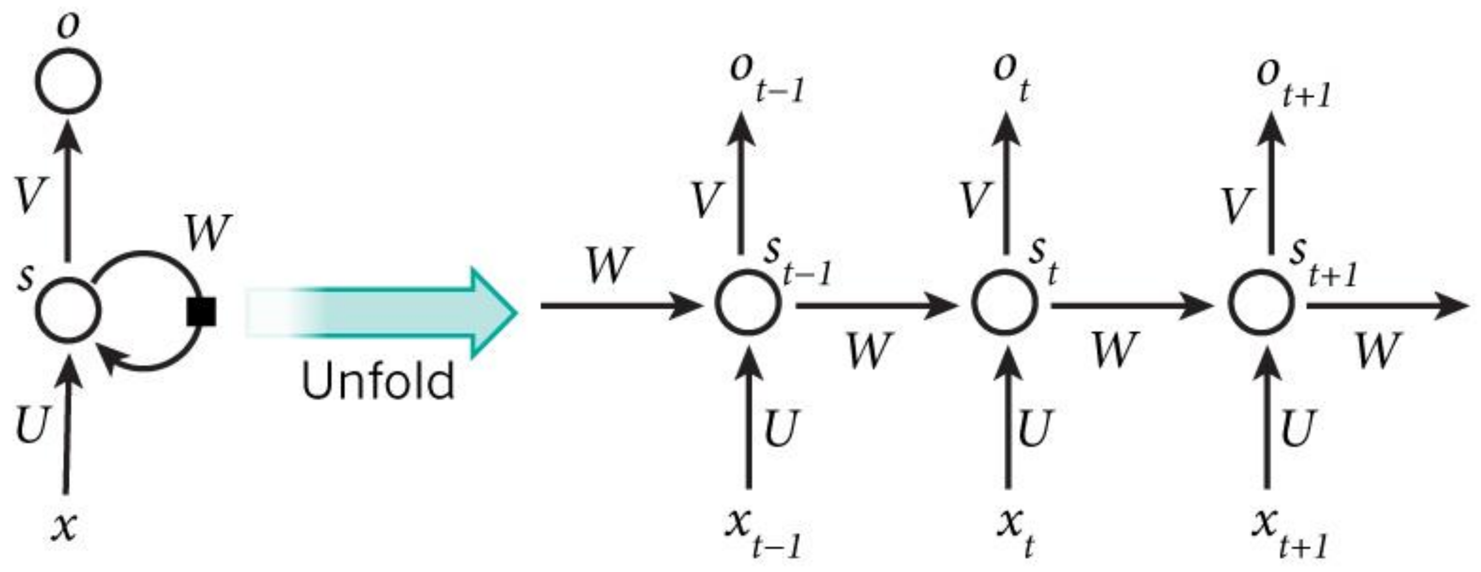
CNN: 합성곱신경망, convolutional neural network

- ✓ 기존의 방식은 데이터에서 지식을 추출해 학습이 이루어졌지만, CNN은 데이터의 특징(feature)을 추출하여 특징들의 패턴을 파악함.
- ✓ Convolution과정과 Pooling과정을 통해 진행
  - Convolution: 데이터의 특징을 추출하는 과정, 데이터에 각 성분의 인접 성분들을 조사해 특징을 파악하고 파악한 특징을 한장으로 도출시키는 과정
  - Pooling: Convolution 과정을 거친 레이어의 사이즈를 줄여주는 과정
- ✓ 예: 개와 고양이의 사진에서 특성을 추출하여 학습하여 (귀모양 등), 개와 고양이를 구별함.



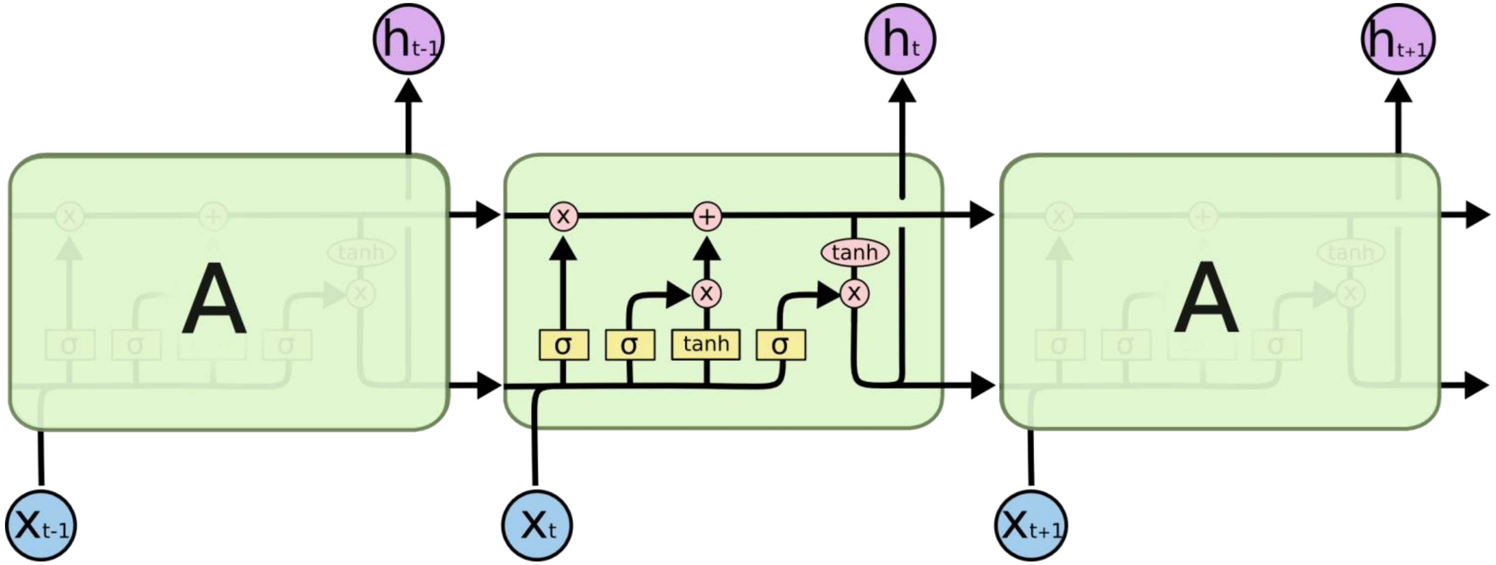
RNN: 순환신경망, recurrent neural network

- 반복적이고 순차적인 데이터(Sequential data)학습에 특화된 인공신경망
- 순환구조를 이용하여 과거의 학습을 Weight를 통해 현재 학습에 반영
- 예: 예전의 날씨 기록을 토대로 내일의 날씨를 예측.
- 단점: 장기적인 인과관계를 학습하기는 어렵다.



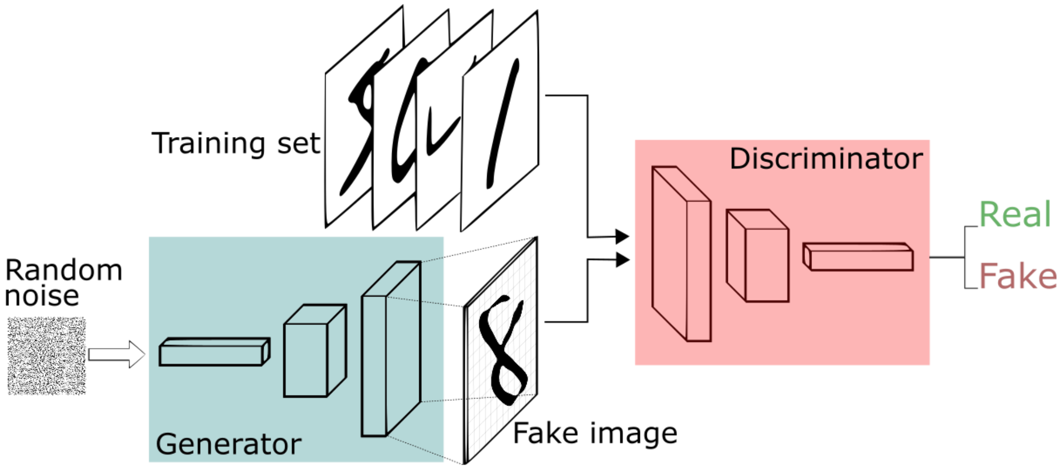
LSTM: Long Short Term Memory Networks

- RNN의 발전된 형태로, RNN의 단점인 장기기억 부분을 극복하여 **장기기억**과 단기기억을 모두 유지할 수 있음.
- RNN의 일종을 게이트(gate)와 셀(cell) 구조를 도입하여, 이 구조를 통해 과거 정보를 어느 정도 기억할지 판단하면서 필요한 정보만을 다음 시점에서 이어받음.



GAN: Generative adversarial networks (생성적 적대 신경망)

- ✓ 생성자(Generator)와 식별자(Discriminator)라는 신경망 2개가 서로 경쟁하면서 학습하는 **생성** 모델.
- ✓ 생성자: 위조지폐범
- ✓ 식별자: 경찰
- ✓ 실제 이미지를 학습하여 실제 이미지와 확률분포가 최대한 비슷하도록 허구 이미지를 만듬. 이 학습 과정에서 판별망의 판단 결과를 활용(이 판별망은 이미지의 확률 분포를 판별함). 생성망이 만든 허구 이미지를 판별망이 실제 이미지로 착각하도록 만드는 방향으로 생성망 학습이 이루어짐.
- ✓ 두 개 이상의 신경망이 서로를 향하고 서로 대항하듯이 훈련함으로써 생성자 모델을 학습해냄.



빅데이터가 새로운 기술을 개발하게 만들기도 한다.

Fei-Fei Li

Geoffrey Hinton

ILSVRC (ImageNet Large Scale Visual Recognition Challenge)

IMAGENET

14,197,122 images, 21841 synsets indexed

Home Download Challenges About

Not logged in. Login | Signup

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

Competition

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) evaluates algorithms for object detection and image classification at large scale. One high level motivation is to allow researchers to compare progress in detection across a wider variety of objects -- taking advantage of the quite expensive labeling effort. Another motivation is to measure the progress of computer vision for large scale image indexing for retrieval and annotation.

For details about each challenge please refer to the corresponding page.

- ILSVRC 2017
- ILSVRC 2016
- ILSVRC 2015
- ILSVRC 2014
- ILSVRC 2013
- ILSVRC 2012
- ILSVRC 2011
- ILSVRC 2010

Workshop

Every year of the challenge there is a corresponding workshop at one of the premier computer vision conferences. The purpose of the workshop is to present the methods and results of the challenge. Challenge participants with the most successful and innovative entries are invited to present. Please visit the corresponding challenge page for workshop schedule and information.

Download

The most popular challenge is the ILSVRC 2012-2017 image classification and localization task. It is available on Kaggle. For all other data please log in or request access.

Evaluation Server

The evaluation server can be used to evaluate image classification results on the test set of ILSVRC 2012-2017. Please see here for our submission policy. Importantly, you should not make more than 2 submissions per week.

Updates

- October 10, 2019: The ILSVRC 2012 classification and localization test set has been updated. The Kaggle challenge and our download page both now contain the updated data.
- June 2, 2015: Follow-up update regarding status of the server
- May 19, 2015: Annoucement regarding the submission server

Citation

When reporting results of the challenges or using the datasets, please cite:

Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (* = equal contribution) **ImageNet Large Scale Visual Recognition Challenge.** *IJCV*, 2015. paper | bibtex | paper content on arxiv | attribute annotations

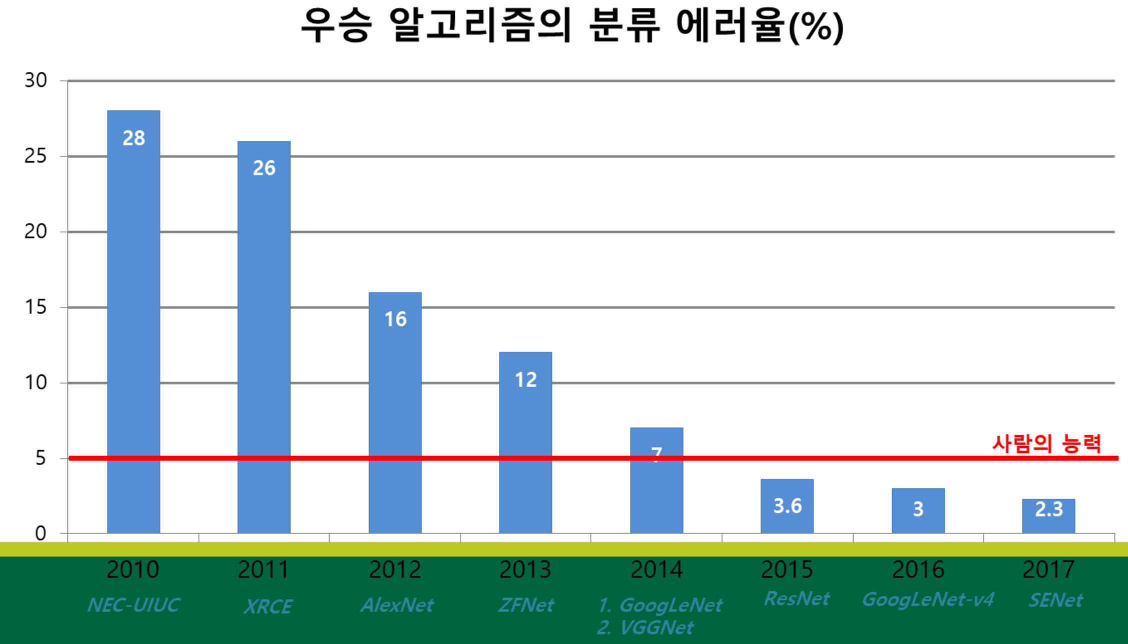
Additional references

These are some additional publications directly related to collecting the challenge dataset and evaluating the results. These papers are all discussed in the main paper above. Please refer to the individual challenge webpages for information about the most successful entries, and to the ImageNet publications page for a complete list of publications.

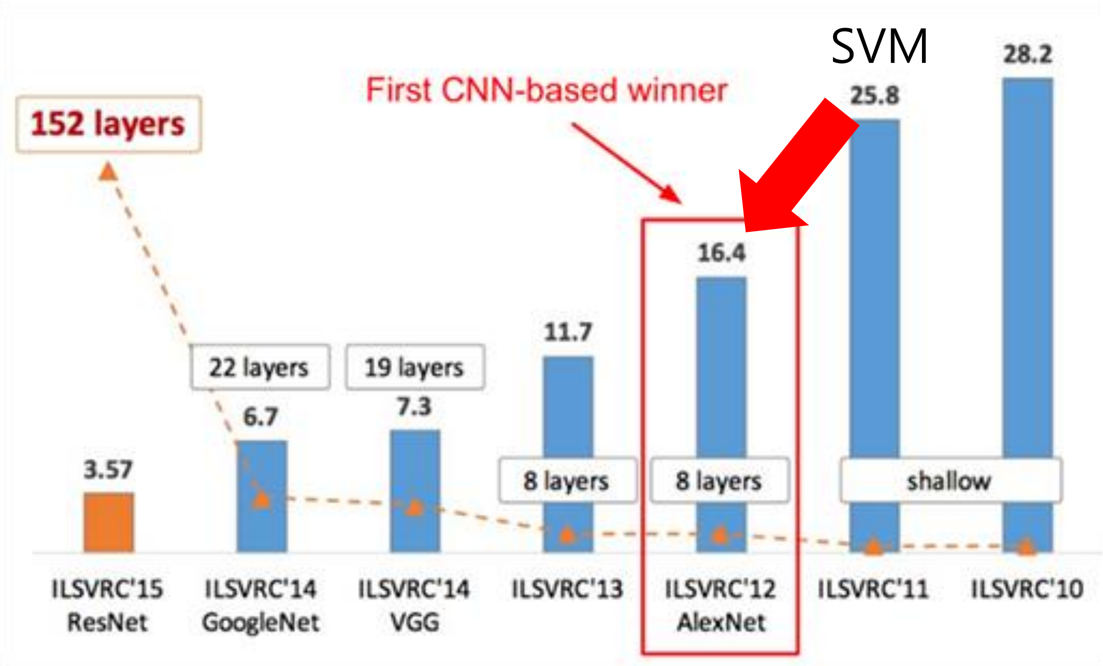
www.image-net.org

페이페이 리 (스탠포드 교수)

"중요한 것은 알고리즘이 아니라 데이터 학습량이다."
이미지넷 대회를 열기 위해 데이터를 많이 모음

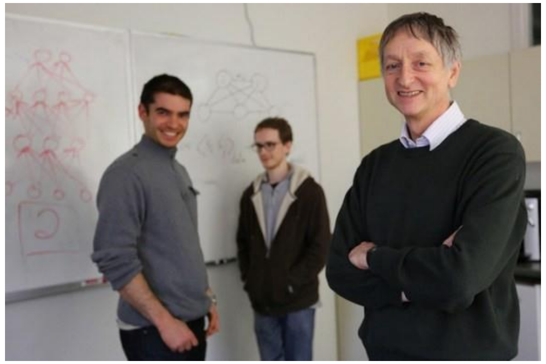


ILSVRC (ImageNet Large Scale Visual Recognition Challenge)



ILSVRC: 이미지 인식(image recognition) 경진대회

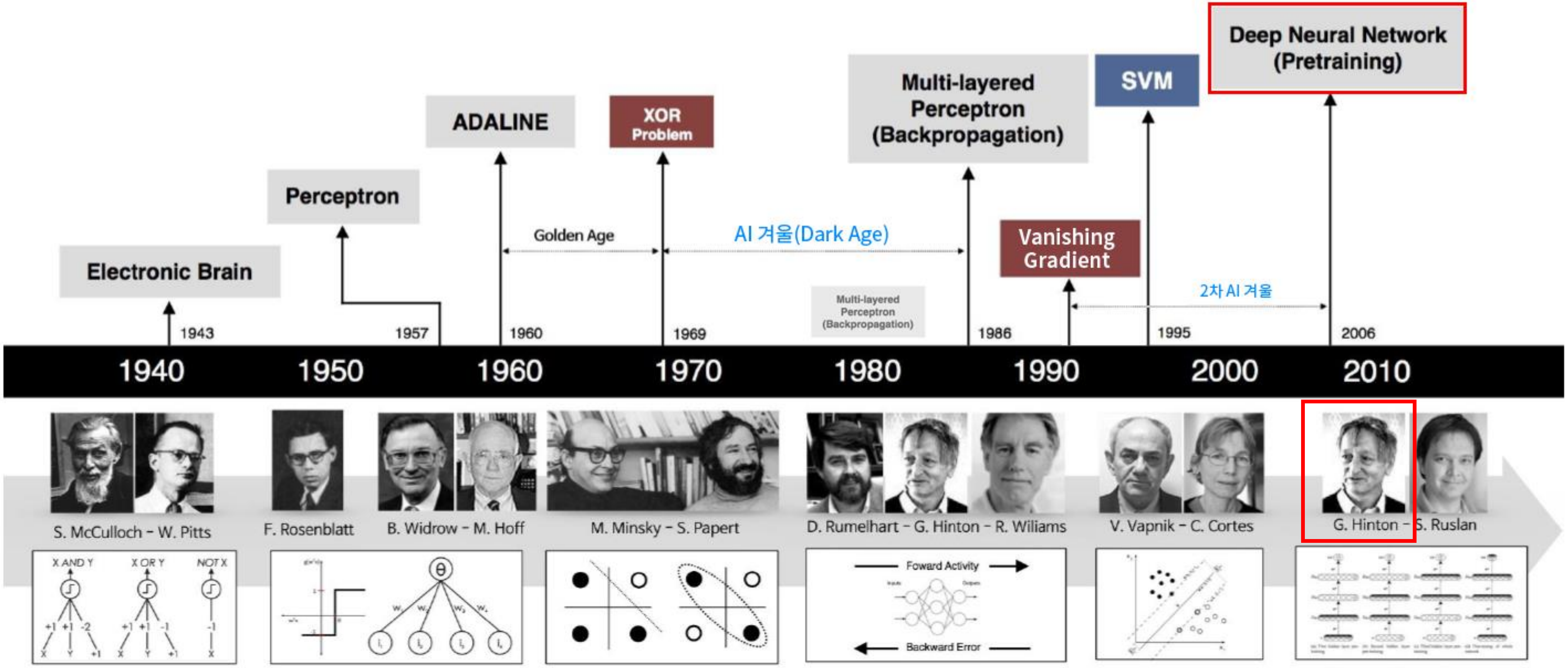
2012년 CNN 기반 딥러닝 알고리즘 AlexNet이 우승 이후, 깊은 구조(deep architecture)를 가진 알고리즘들이 우승을 차지함.



제프리 힌튼, "수퍼비전 팀"
2012년 알렉스넷

알렉스넷
성능 향상 위해 GPU 기능 추가 – 행렬 계산 속도 빨라짐 (제일 처음 쓴 건 독일, 슈미트후버 2010).
활성화함수로 처음으로 ReLU를 사용한 네트워크 (ReLU 개발자 – 요슈아 벤지오).
Data의 양을 늘리기 위해 Data augmentation을 적용함. Data augmentation이란, 원래 이미지를 변환 (좌우 대칭, 늘리기, 회전 등)하여 비슷한 이미지 만드는 과정임.
Dropout 0.5 값을 줘서 overfitting 문제를 해결 (모든 노드를 계산하지 않고 무작위로 일부 노드 계산값 버리기) (니티시 스리바스타바)
Learning rate : 0.01, 이 때 0.001로 점차 줄여나감 (보통 Learning rate는 overfitting을 막기 위해서 처음에는 크게 주고, 점차 줄여나감.)

IT 위키, https://itwiki.kr/w/%EC%9D%B8%EA%B3%B5%EC%A7%80%EB%8A%A5



의학에서의

machine learning 연구 최신지견

Papers	비고	Date	Journal	IF
Prediction of type 2 diabetes using genome-wide polygenic risk score (gPRS) and metabolic profiles: A machine learning analysis of population-based 10-year prospective cohort study	• the Korean Genome and Epidemiology Study (KoGES) Ansan-Ansung cohort (n=1425) • gPRS 계산 • Metabolite 선택 (Boruta algorithm) • **random forest (RF)**-based machine learning models • Type 2 Diabetes prediction	DEC 1, 2022	*EBioMedicine*	11.205
Preoperative data-based deep learning model for predicting postoperative survival in pancreatic cancer patients	• N=229 (training), N=53 (test) • **Clinical data**-based machine learning • **CT data**-based deep learning models • **➔ Ensemble model** • Prediction of OS, RFS • AUC • Comparing with AJCC stage	September 2022	*Int. J. Surg.*	13.400
Artificial intelligence for predicting survival following deceased donor liver transplantation: Retrospective multi-center study	• (n=785) deceased donor liver transplant recipients • Machine learning ✓ random forest ✓ artificial neural networks ✓ decision tree ✓ naïve Bayes ✓ support vector machine • traditional statistical models ✓ Cox regression ✓ MELD score ✓ donor MELD score ✓ balance of risk score • Survival prediction • AUC-ROC	September 2022	*Int. J. Surg.*	13.400

Papers	비고	Date	Journal	IF
Machine Learning-Derived Integer-Based Score and Prediction of **Tertiary Hyperparathyroidism** among **Kidney Transplant Recipients**: An **Integer-Based Score** to Predict Tertiary Hyperparathyroidism	• kidney allograft recipients (n=669) • N=542 (External validation, Korean Cohort Study for Outcome in Patients with Kidney Transplantation) • AUROC	July 2022	*Clin. J. Am. Soc. Nephrol.*	10.624
Network-based machine learning approach to predict immunotherapy response in cancer patients	• N=700 • ICI-treated patient samples • **Network-based** machine learning	*Nat. Commun.*	June 2022	17.694
Prediction of **Hidden** **Coronary Artery Disease** Using Machine Learning in Patients With Acute Ischemic Stroke	• (n=1710) training set • (n=348) test set • **XGBoost** model • AI model 이 예측한 군에서 더 events가 많이 발생한다.	*Neurology*	April 25, 2022	11.360
Development and validation of a prognostic and predictive 32-gene signature for gastric cancer	• N=567 • machine learning algorithm NTriPath to identify a gastric-cancer specific **32-gene signature** • **Support Vector Machine** model (risk score)	*Nat. Commun.*	February 2022	17.694
Machine learning algorithms for predicting **direct-acting antiviral treatment failure** in chronic hepatitis C: An HCV-TARGET analysis	• HCV-TARGET registry data • Training (*n* = 4894) and validation (*n* = 1631) ✓ multivariable logistic regression ✓ elastic net ✓ random forest ✓ gradient boosting machine (GBM) ✓ feedforward neural network machine learning • to predict DAA treatment failure	*Hepatology*	January 2022	17.298

Papers	비고	Date	Journal	IF
Improved prediction of immune checkpoint blockade efficacy across multiple cancer types	• a comprehensively curated cohort (MSK-IMPACT) with 1,479 patients • by **integrating genomic, molecular, demographic and clinical data** • we developed an **ensemble learning** random forest14 classifier with 16 input features (hereafter called RF16) • significantly outperformed predictions based on **tumor mutational burden**	*Nat. Biotechnol.*	November 2021	68.16
Machine learning model for predicting **excessive muscle loss** during neoadjuvant chemoradiotherapy in oesophageal cancer	• N=232 • seven different machine-learning algorithm • None of the clinicopathologic variables differed significantly **between the two groups.** • The **ensemble model** of **logistic regression** and **support vector classifier** showed the highest area under the curve value	*J. Cachexia Sarcopenia Muscle*	October 2021	12.51
An artificial intelligence model to predict hepatocellular carcinoma risk in Korean and Caucasian patients with chronic hepatitis B	• N=6051 (four hospitals in Korea) • External validation (PAGE-B cohorts 등) • **concordance index** (c-index), 0.79 • GBM-based model provides the best predictive power	*J. Hepatol.*	October 2021	30.08

Papers	비고	Date	Journal	IF
Cortical Thickness from MRI to Predict Conversion from Mild Cognitive Impairment to Dementia in Parkinson Disease: A Machine Learning–based Model	• N=42 (progressed) • N=75 (not progressed) • 10000 randomly generated training sets • 10000 randomly resampled test sets • AUROC • External validation set • **RF, SVM** • Models trained with cortical thickness variables (AUC range, 0.75–0.83) showed fair to good performances similar to those trained with **clinical variables** (AUC range, 0.70–0.81). Model performances improved when models were trained with both variables (AUC range, 0.80–0.88).	*Radiology*	AUGUST 2021	29.15
Markers of Myocardial Damage Predict Mortality in Patients With Aortic Stenosis	• n = 440, derivation • n = 359, validation cohort • **random survival forest model** was built using 29 variables (13 CMR) with post-AVR death	*J. Am. Coll. Cardiol.*	2021 Aug	24.09

Development and Validation of Machine Learning–based Model for the Prediction of Malignancy in Multiple Pulmonary Nodules: Analysis from Multicentric Cohorts	• N=520 (1739 nodules)
	• **XGBoost**
	• 10-fold cross-validation
	• Compared with
	    • solitary pulmonary nodule (SPN) models
	    • Clinicians
	    • a computer-aided diagnosis (CADx) system
	    • (Brock, PKU, Mayo, VA models)
	• Validation
	    • an independent transnational cohort
	    • prospective multicentric cohort.
	• AUC
	• **Calibration (Brier score)**
		Clin. Cancer Res.	February 24, 2021	13.801

Papers	비고	Date	Journal	IF
Development and Multiple Validation of the Protein Multi-Marker Panel for Diagnosis of Pancreatic Cancer	• N=959 plasma samples • A **logistic regression analysis with stepwise selection** was performed to build a multi-marker panel using the 44 biomarker candidates. • 14 biomarker proteins • AUC • To construct an optimal, diagnostic, multi-marker panel, we applied data preprocessing procedure to biomarker candidates.	*Clin. Cancer Res.*	January 27, 2021	13.801
Thalamocortical dysrhythmia detected by machine learning	• Thalamocortical dysrhythmia (TCD) is a model proposed to explain divergent neurological disorders • **Support vector machine** learning • analyzing **resting-state electroencephalography oscillatory patterns** in patients with Parkinson's disease, neuropathic pain, tinnitus, and depression	*Nat. Commun.*	March 2018	17.694
Improved Accuracy in Optical Diagnosis of Colorectal Polyps Using Convolutional Neural Networks with Visual Explanations	• Narrow-band imaging (NBI) • **convolutional neural networks (CNNs)** • Train set - 1100 adenomatous polyp, 1050 hyperplastic polyps from 1379 patients • Test set - 300 images of 180 adenomatous polyps and 120 hyperplastic polyps • Comparison – 22 endoscopists • **Accuracy (AI only or +doctors), time saving**	*Gastroenterology*	February 29, 2020	33.883

Papers	비고	Date	Journal	IF
Development and validation of a deep learning model to diagnose COVID-19 using time-series heart rate values before the onset of symptoms	• a **deep learning model** to diagnose COVID-19 before the onset of symptoms using **heart rate (HR) data** obtained from a smartwatch • AUROC	*J. Med. Virol.*	January 2023	20.69
Preoperative data-based deep learning model for predicting **postoperative survival** in pancreatic cancer patients	• Training set = 229 • Test set = 53 • **Ensemble model** • Comparison: AJCC stage	*Int. J. Surg.*	September 2022	13.400
Association of Artificial Intelligence–Aided Chest Radiograph Interpretation With Reader Performance and Efficiency	• **Lunit** • 497 frontal **chest radiograph**s • The data used were collected at 2 quaternary academic hospitals in Boston, Massachusetts: Beth Israel Deaconess Medical Center (The Medical Information Mart for Intensive Care **Chest** X-Ray [MIMIC-CXR]) and Massachusetts General Hospital (MGH). • Comparison: Radiologists	*JAMA Netw. Open*	August 31 2022	13.360
Feasibility of **anomaly score** detected with deep learning in irradiated breast cancer patients with reconstruction	• **generative adversarial network (GAN)** deep learning algorithm • Training = 251 normal breast images • Generated anomaly score (AS)	*npj Digit. Med.*	2022 Aug 23	15.357
Deep Learning-Based Automatic Detection and Grading of Motion-Related Artifacts on Gadoxetic Acid-Enhanced Liver MRI	• **deep learning**-based algorithm (DLA) • for detection and grading of motion-related **artifacts** • Training set = 336 arterial phase images • Sensitivity, specificity	*Invest. Radiol.*	2022 Aug 24	10.065

Papers	비고	Date	Journal	IF
Deep Learning–based Detection of Solid and Cystic Pancreatic Neoplasms at Contrast-enhanced CT	• nnU-Net–based **deep learning** model • Training set = 852 patients • Test set1 = 603 patients • Test set2 = 589 patients	*Radiology*	Aug 23 2022	29.146
Deep Learning Prediction of Survival in Patients with Chronic Obstructive Pulmonary Disease Using Chest Radiographs	• deep learning–based **survival** prediction • Training set = 3475 • Validation = 435 • Internal test = 315 • External test = 394, 416, 337 • Chest radiographs = **model1** • Add clinical variables = **model2** • **Time dependent AUC** at 5-year survival • Goodness: **Hosmer-Lemeshow test** • Comparison: clinical indexes: BODE, ADO, CAT, SGRQ	*Radiology*	June 7 2022	29.146
Deep Learning-Based Prediction Model Using Radiography in **Nontuberculous Mycobacterial Pulmonary Disease**	• **Prognostic prediction** • **deep learning** • Training = 1638 CXR (1034 patients) • Test = 566 CXR (200 patients) • DL-driven radiographic score • AUC • 10-, 5, and 3-year mortality	*Chest*	2022 Jun	11.393

Papers	비고	Date	Journal	IF
Deep Learning for Detecting Pneumothorax on Chest Radiographs after Needle Biopsy: Clinical Implementation	• deep learning–based (computer aided detection) CAD system • CAD 도입전 사진 • CAD 도입후 사진 • Standard by 2 radiologists • Accuracy 비교: **generalized estimating equations** • **그룹 matching**: radiograph reade와 PTNB operator 기준으로 greedy matching 시행	*Radiology*	Jan 25 2022	29.146
Deep Learning for Detection of Pulmonary Metastasis on Chest Radiographs	• 위와 비슷한 모델 • 도입전 = 5681 CXR • 도입후 = 2916 CXR • True positive rate, False positive rate • Accuracy 비교: **generalized estimating equations** • **PSM**: age, sex, primary cancer • Non-inferiority study	*Radiology*	Aug 31 2021	29.146
Development and validation of a deep learning algorithm detecting 10 common abnormalities on chest radiographs	• **Lunit** • a ResNet34-based neural network with lesion-specific channels • 146717 radiographs • 10 common radiological abnormalities • AUC	*Eur. Resp. J.*	May 20, 2021	33.801
End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography	• **Google AI** • deep learning algorithm • Low dose CT • predict the risk of lung cancer • **AUC** • 6,716 National Lung Cancer Screening Trial cases • External validation = 1139 • Comparison - radiologists	*Nat. Med.*	May 2019	87.24

전체적 평가

- ✓ 아직까지 diagnosis, or prognosis prediction에 중점을 둔 지도학습
- ✓ Dataset → 다양한 machine learning (Xgboost, SVM, LF, Deep learning…)
- ✓ Image → Deep learning
- ✓ Ensemble model
- ✓ AI가 기존의 통계기법을 대체한다는 느낌
- ✓ AI 연구 기법은 정형화되어가는 중
- ✓ 기법적 측면보다 임상적 질문이 더욱 중요

Machine learning 연구 실제

한국보건복지인력개발원
-2020년도 의료 인공지능 전문가 양성과정, 제3팀

이형철
코치

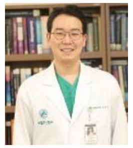
의료진

데이터 사이언티스트

김이준 이대목동병원

서수량 서울대학교 보라매병원

서준교 서울아산병원



이상욱 서울아산병원

이승욱 경북대학교 의과대학

이경헌 삼성융합의과학원

이헌이 서울대학교병원





의료 인공지능 개발 전 생각할 점 1

Question

- 의료 임상에서 가장 필요한 부분이 무엇인가?

Answer

- 인간이 예측하기 어려운 환자의 심각한 예후 예측
- 인간과 비슷하거나 더 뛰어난 분석 능력

Solution

- 수술 후 30일 이내 사망

의료 인공지능 개발 전 생각할 점 2

Question	• 인공지능에서 가장 중요한 것은 무엇인가?
Answer	• 데이터의 양과 질
Solution	• 빅데이터 다기관 연구

의료 인공지능 개발 전 생각할 점 3

Question

- 인공지능 개발까지의 현실적인 문제는 무엇인가?

Answer

- 인공지능 개발 능력
- 임상에 대한 지식

Solution

- 엔지니어 + 의료진
- 직접 공부한다!

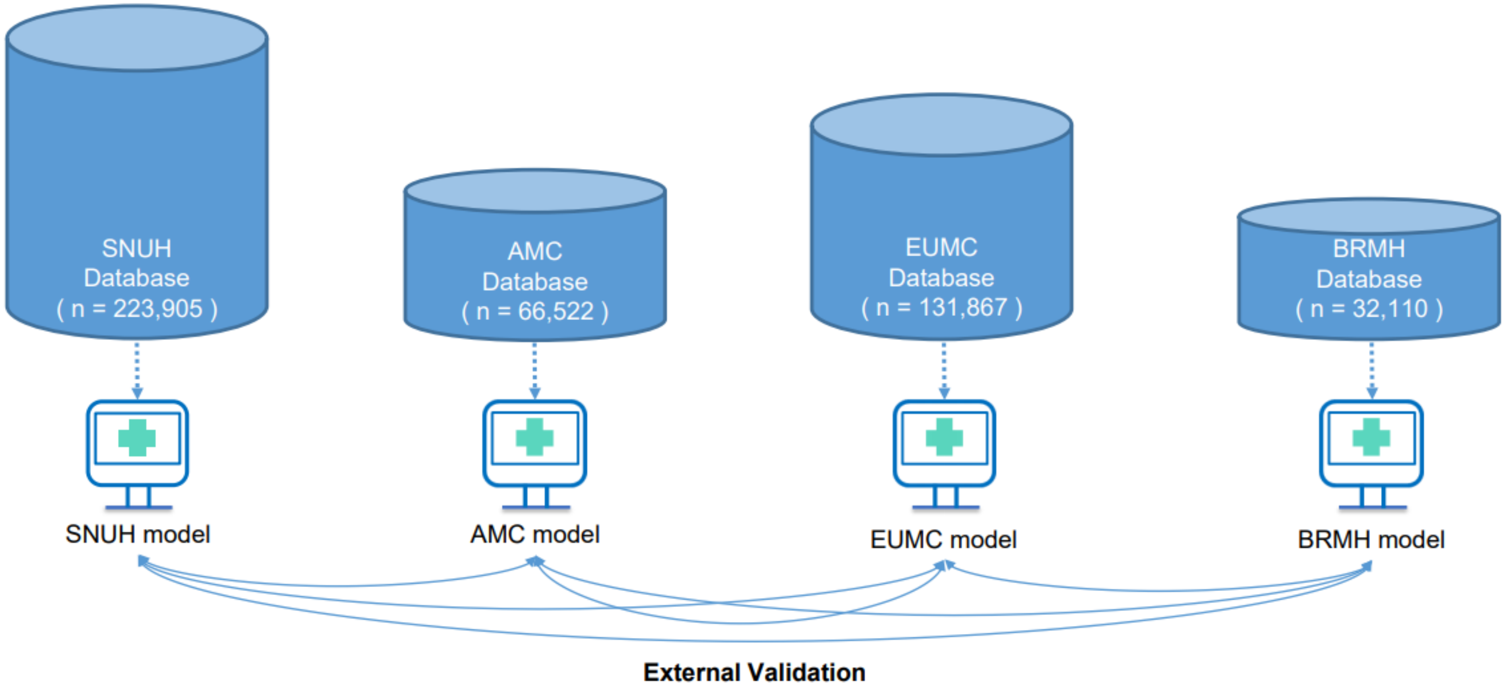
의료 인공지능 개발 전 생각할 점 4

Question
- 기존의 개발된 인공지능의 문제점은 무엇인가?

Answer
- 필요한 파라미터가 너무 많다.
- 한 병원에서 개발한 모델이 다른 병원에서는 사용하기 어렵다.

Solution
- 파라미터 최소화 - 자동 수집 파라미터만 이용
- Transportability 확인 및 개선!

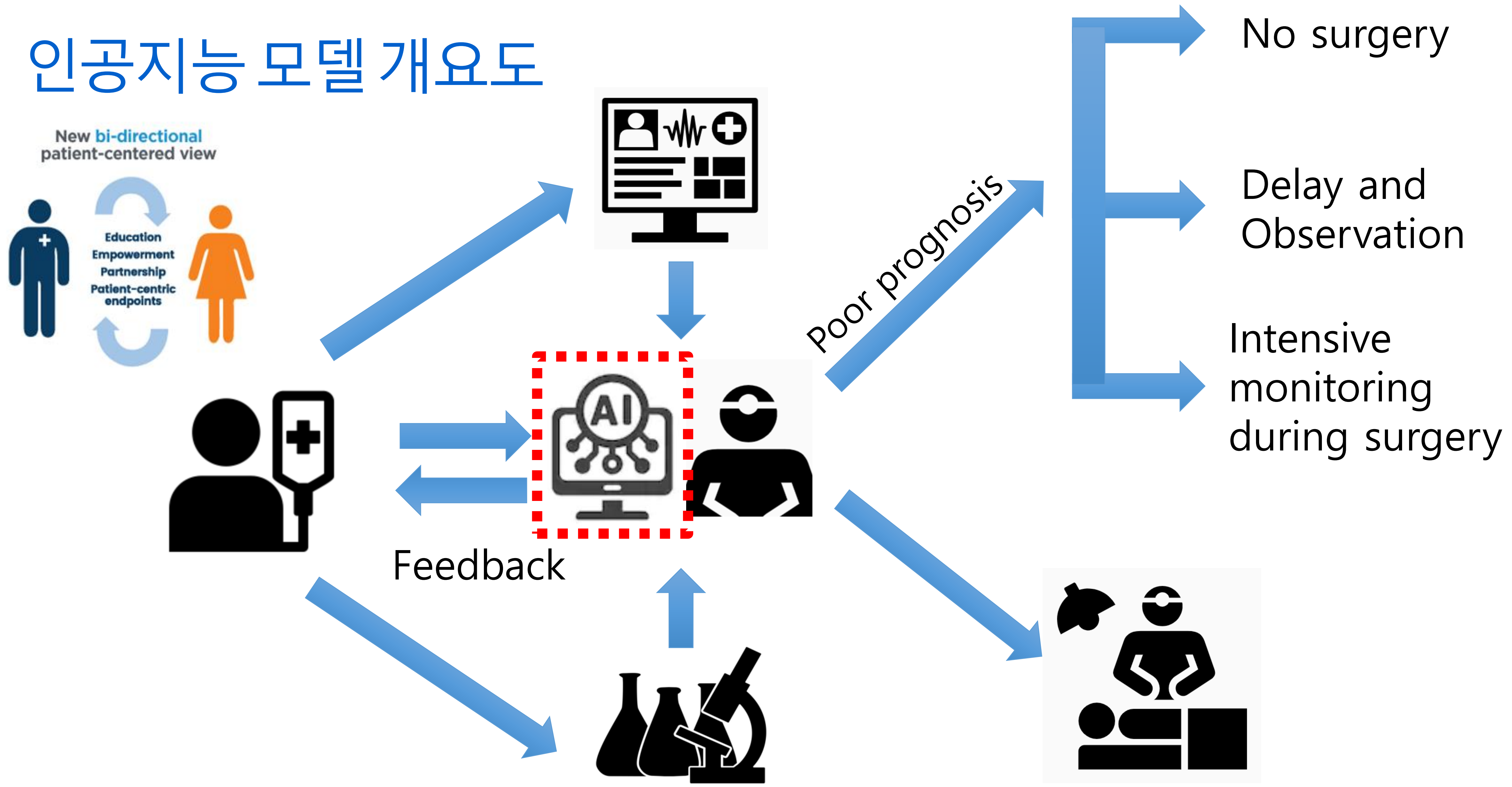
- 한국보건복지인력개발원 – 2020년도 의료 인공지능 전문가 양성과정
- 다기관 인공지능 연구 – 수술 후 30일 이내의 조기사망 예측
- 서울대학교병원, 서울아산병원, 서울특별시 보라매병원, 이화의대부속목동병원 (7명)



➢ Data

- 서울대병원 – VitalDB, EMR
- 서울아산병원 – ABLE (CDW)
- 이대목동병원 – Common data model (CDM)
- 보라매병원 – 병원의무기록실 청구데이터

인공지능 모델 개요도


New bi-directional patient-centered view
Education
Empowerment
Partnership
Patient-centric endpoints
AI
Feedback
Poor prognosis
No surgery
Delay and Observation
Intensive monitoring during surgery

우리 인공지능 모델의 매력적인 포인트

Practical!

- **Light model**
  - Only objective and automatically extractable data
  - Only lab data
- **Transportability**
  - Multicenter study
  - ASA class adjusting
  - Attention method
  - Transfer learning
  - Federated learning
- **Explainable AI**
  - Feature importance analysis

High quality!

- **Reducing Overfitting**
  - Bootstrapping
  - Cross validation
- **Improving accuracy**
  - Ensemble methods - XGBoost
  - Grid search

연구 진행 과정

Simple model

Team A: 이대목동병원, 서울아산병원 마취과, 경북대학교 의과대학, and others

Attention model

Team B: 서울대학교병원 마취과, 서울대학교 의생명연구원, and others

Federated model

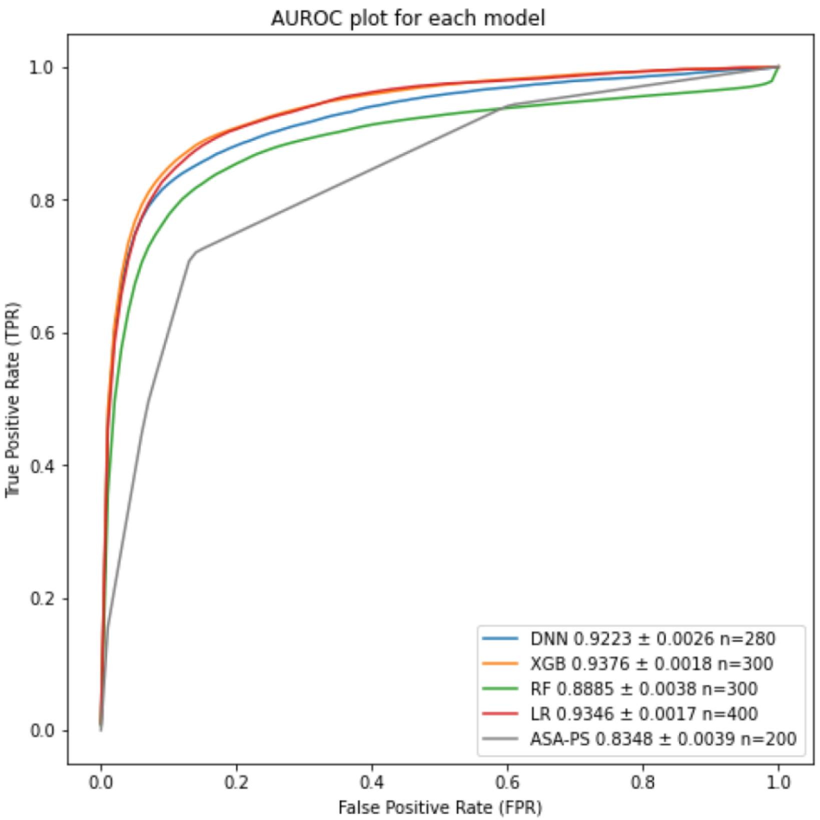
Team C: 서울아산병원 비뇨기과, 보라매병원, 삼성의과학연구원, and others

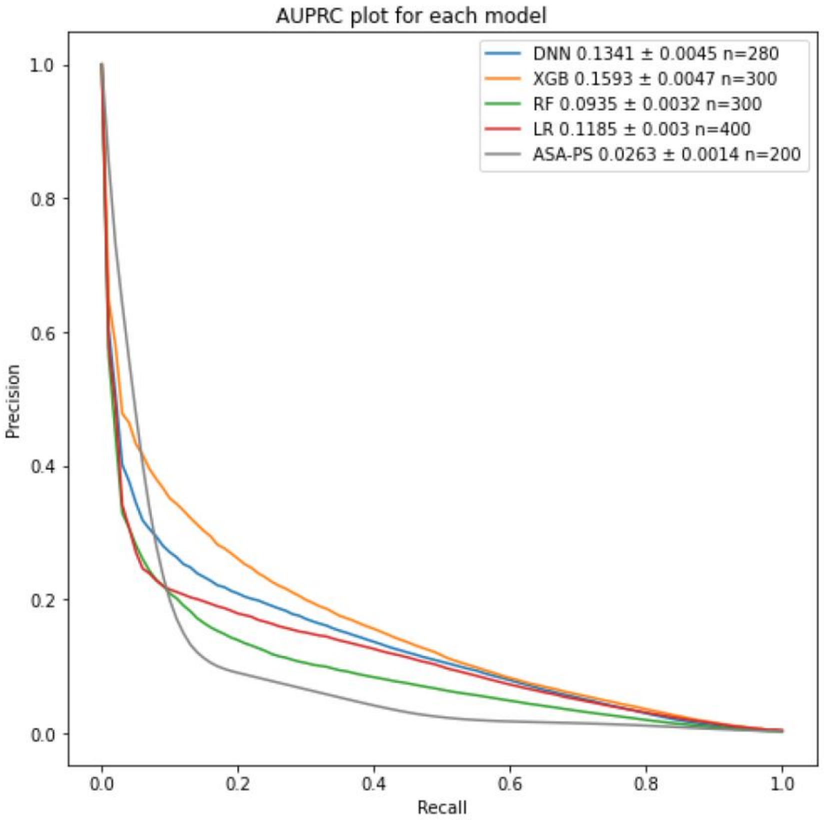
Prospective study

Team D: everybody…

Results – Bootstrap (bagging)

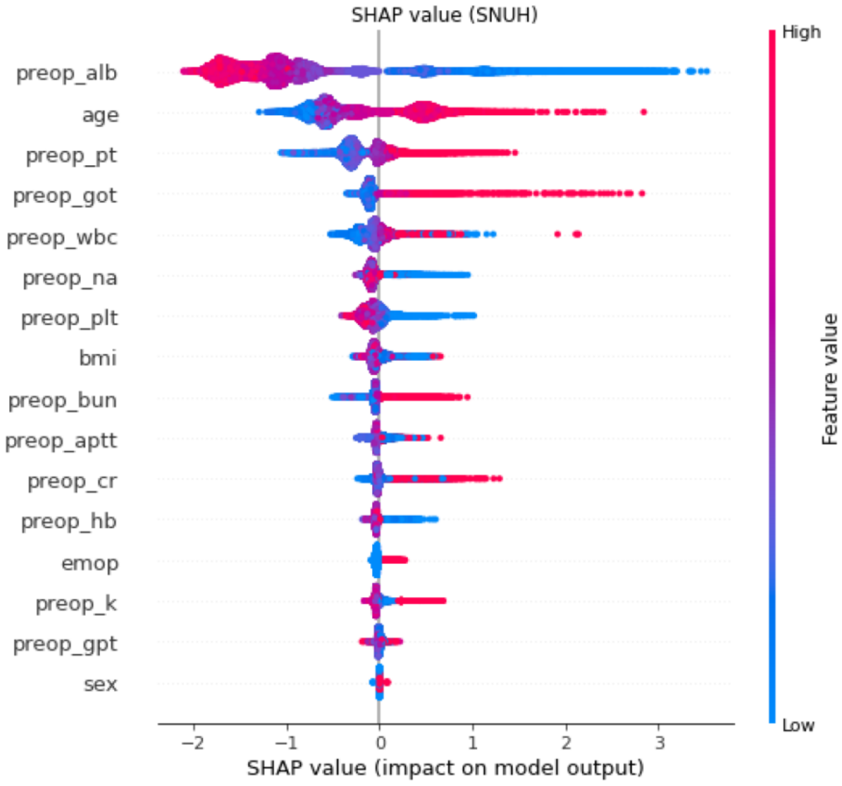
Original data	1	2	3	4	5	6	7	8	9	10	11	12
Set, D1	10	7	4	7	3	4	3	9	11	9	9	10
Set, D2	9	7	4	5	4	4	12	10	11	3	6	2
Set, D3	6	5	1	9	12	5	5	1	9	7	2	9

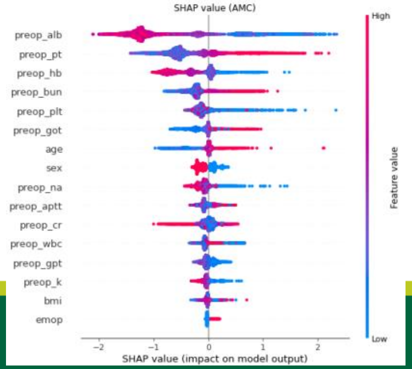


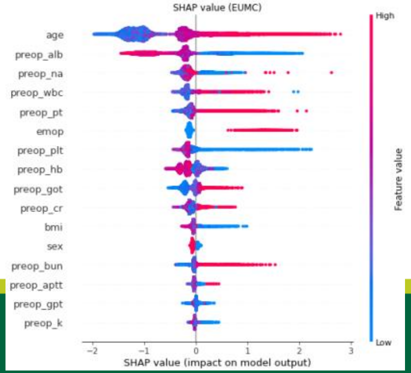


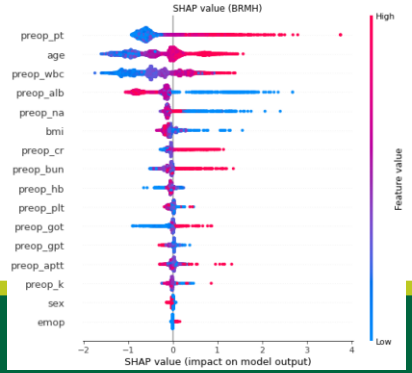
DNN : deep neural network
XGB : XGBoost
RF: Random forest
LR: Linear regression
ASA-PS: ASA performance status

Results – Feature importance







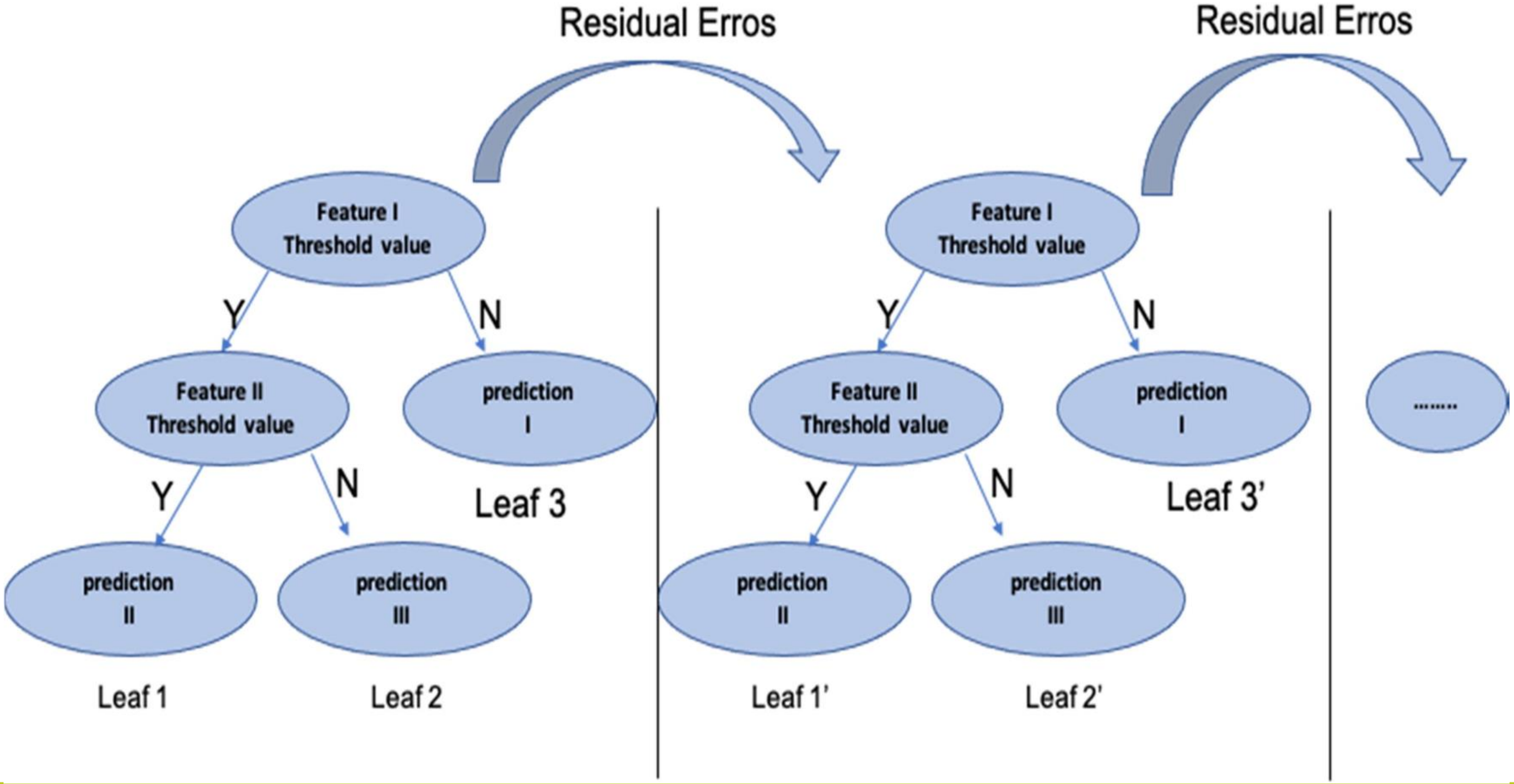


XGBoost

Boosting: 여러 개의 약한 학습기(weak learner)를 순차적으로 학습-예측하면서
잘못 예측한 데이터에 가중치 부여를 통해 오류를 개선해 나가면서 학습하는 방식

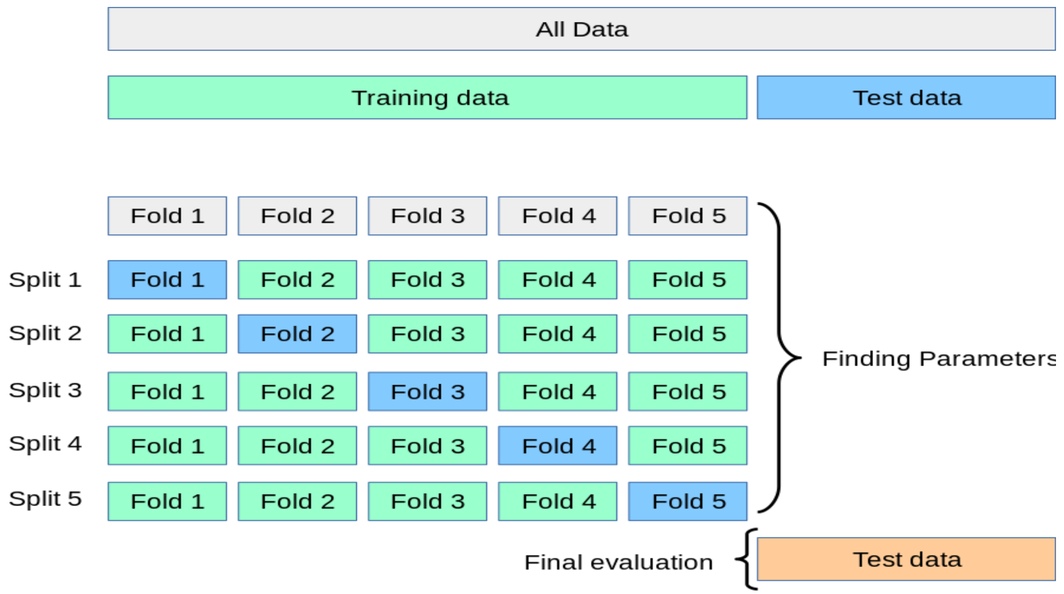
Gradiant Boost Machine (GBM): 가중치 업데이트를 경사 하강법(Gradient Descent)를 이용함

XGBoost (eXtra Gradient Boost) : GBM에 기반하고 있지만, 병렬 수행 및 다양한 기능으로 GBM에 비해 빠른 수행 성능을 보임.

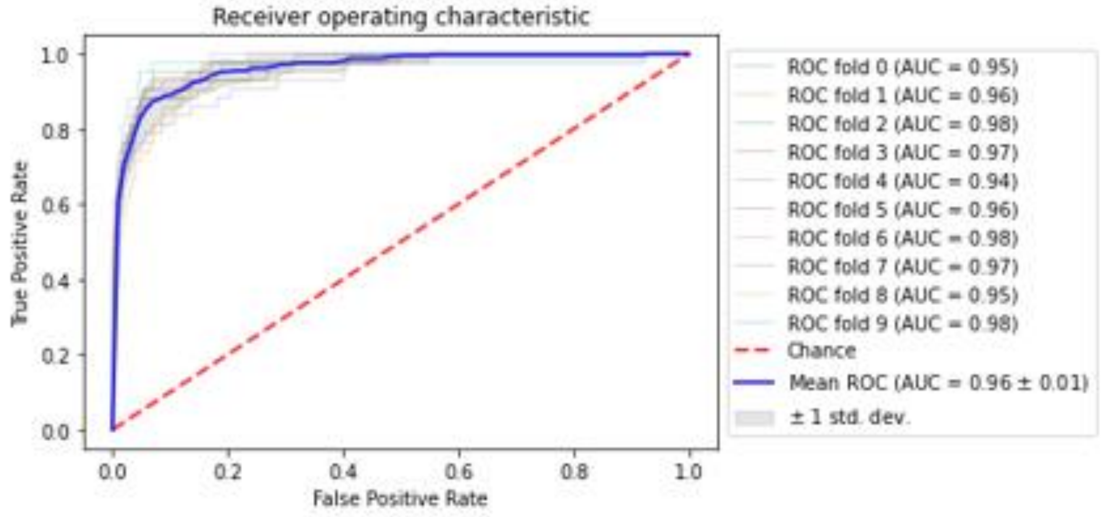


Cross validation

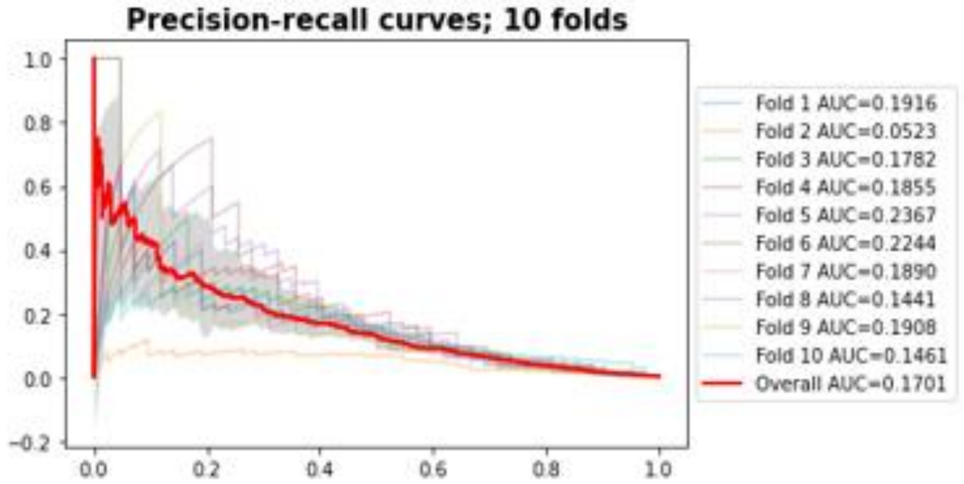
https://scikit-learn.org/stable/modules/cross_validation.html



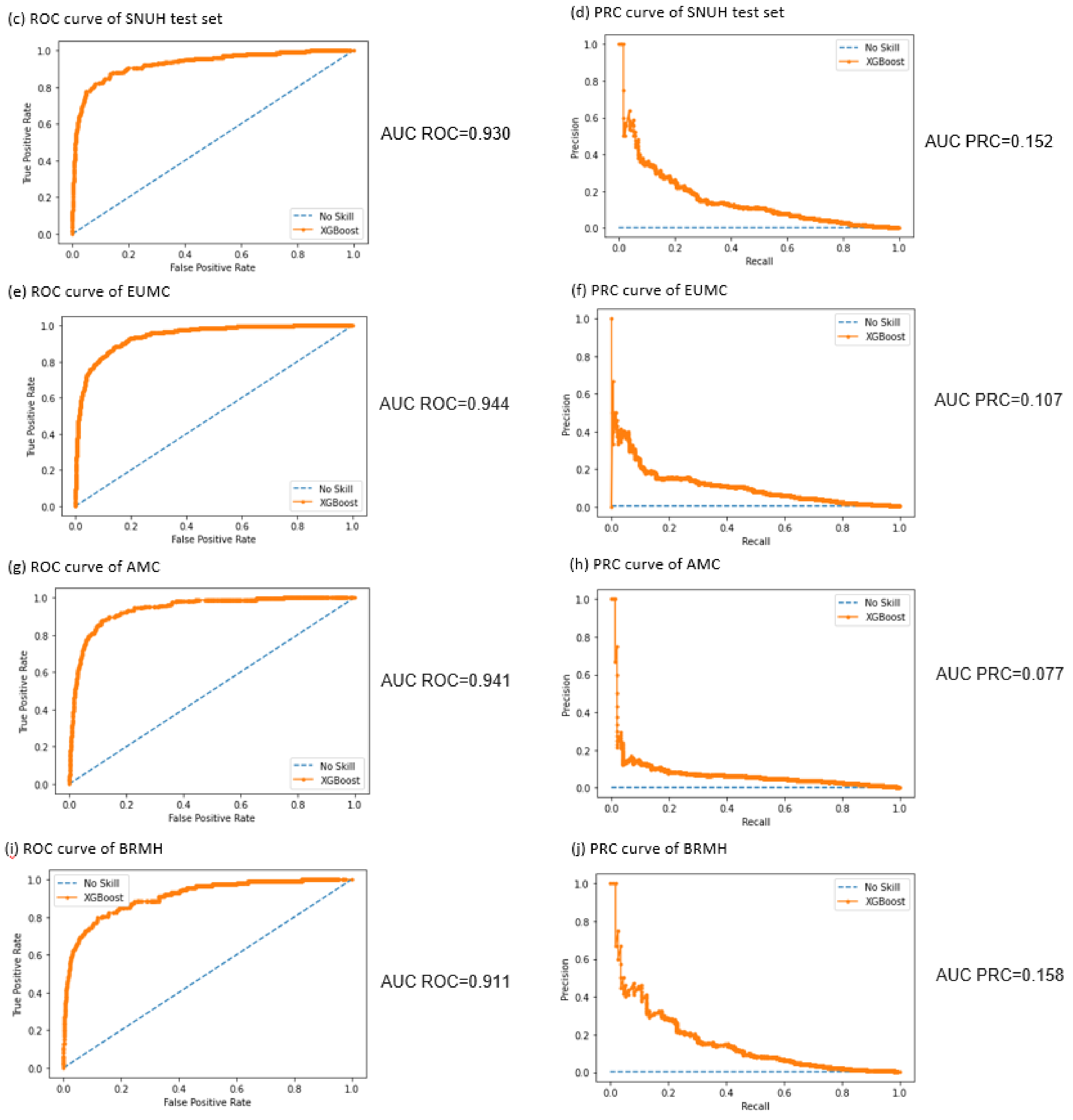
(a) ROC curve of SNUH with 10-folds cross validation



(b) PRC curve of SNUH with 10-folds cross validation



Internal External validation



모델의 평가1

$$sensitivity(true\ positive\ rate) = \frac{TP}{TP + FN}$$

$$specificity(1 - false\ positive\ rate) = \frac{TN}{TN + FP}$$

$$Precision(\text{정밀도}) = \frac{TP}{TP + FP}$$

$$recall(\text{재현율}) = sensitivity = \frac{TP}{TP + FN}$$

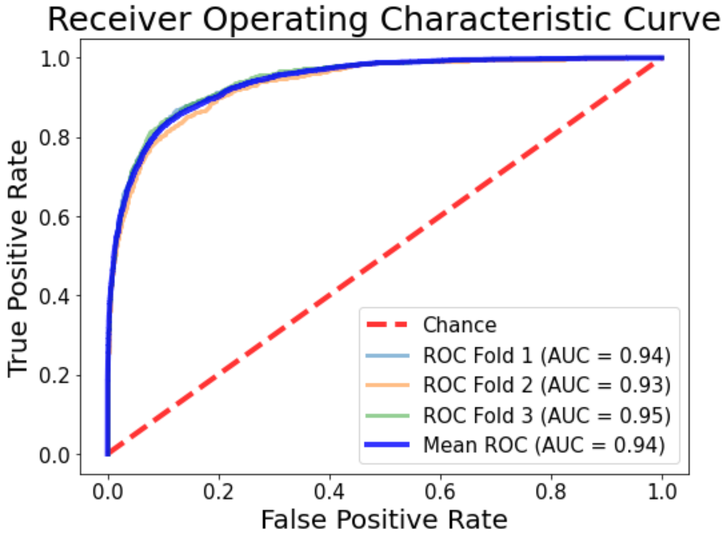
$$F1 = \frac{2\ \times precision\ \times recall}{precision + recall}$$

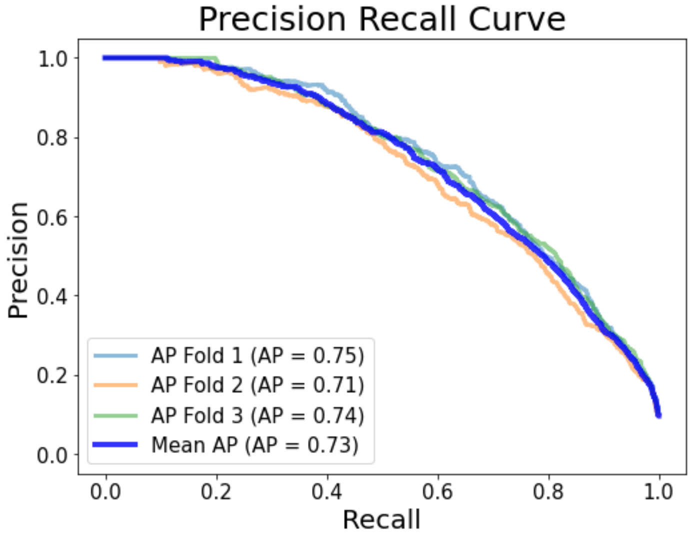
$$accuracy(\text{정확도}) = \frac{TP + TN}{TP + FN + TN + FP}$$

*정밀도/재현율의 트레이드오프

Confusion matrix

		Prediction 1	Prediction 0	
Actual	1	True Positive (TP)	False Negative (FN)	$TP + FN =$ 일정
	0	False Positive (FP)	True Negative (TN)	$FP + TN =$ 일정

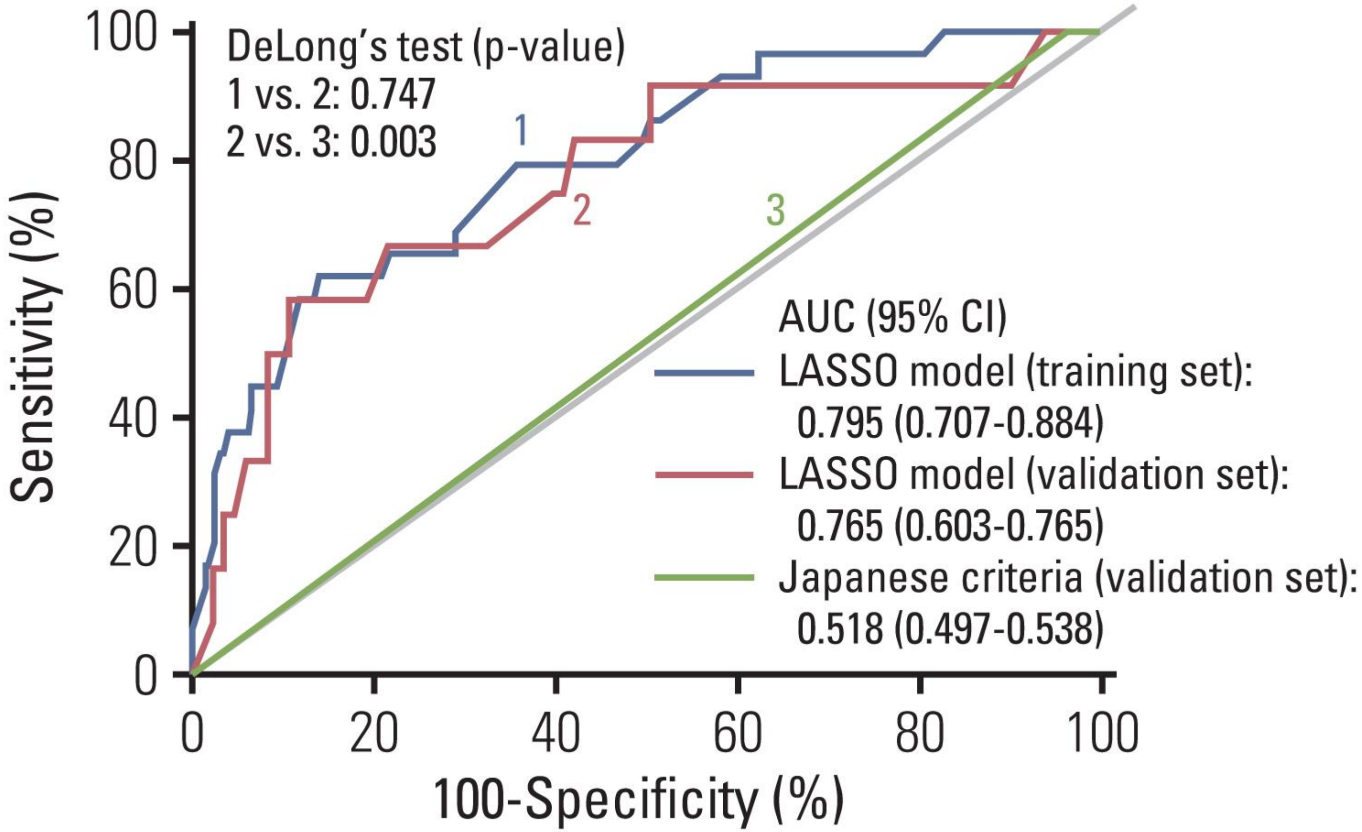




https://www.kaggle.com/c/siim-isic-melanoma-classification/discussion/172453

머신러닝 모델들 간의 비교

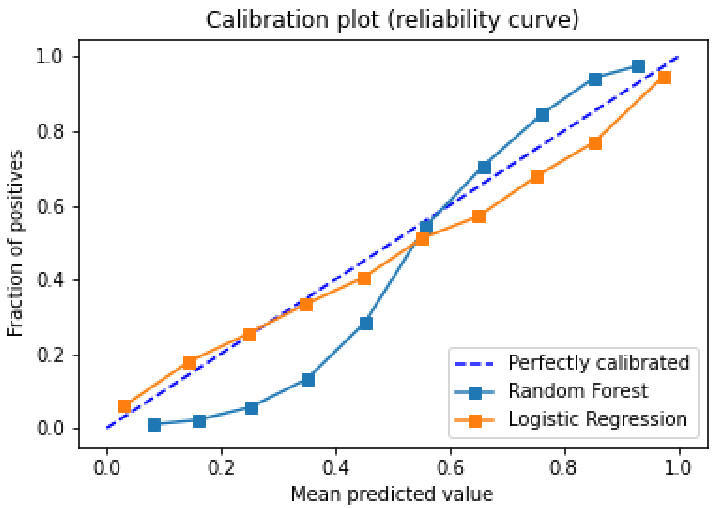
- ✓ Delong test
  - A widely used test to compare the difference between two AUCs relies on the method developed in a seminal paper by DeLong et al. (henceforth 'the DeLong test').
  - It provides a confidence interval and standard error of the difference between two (or more) correlated AUCs.

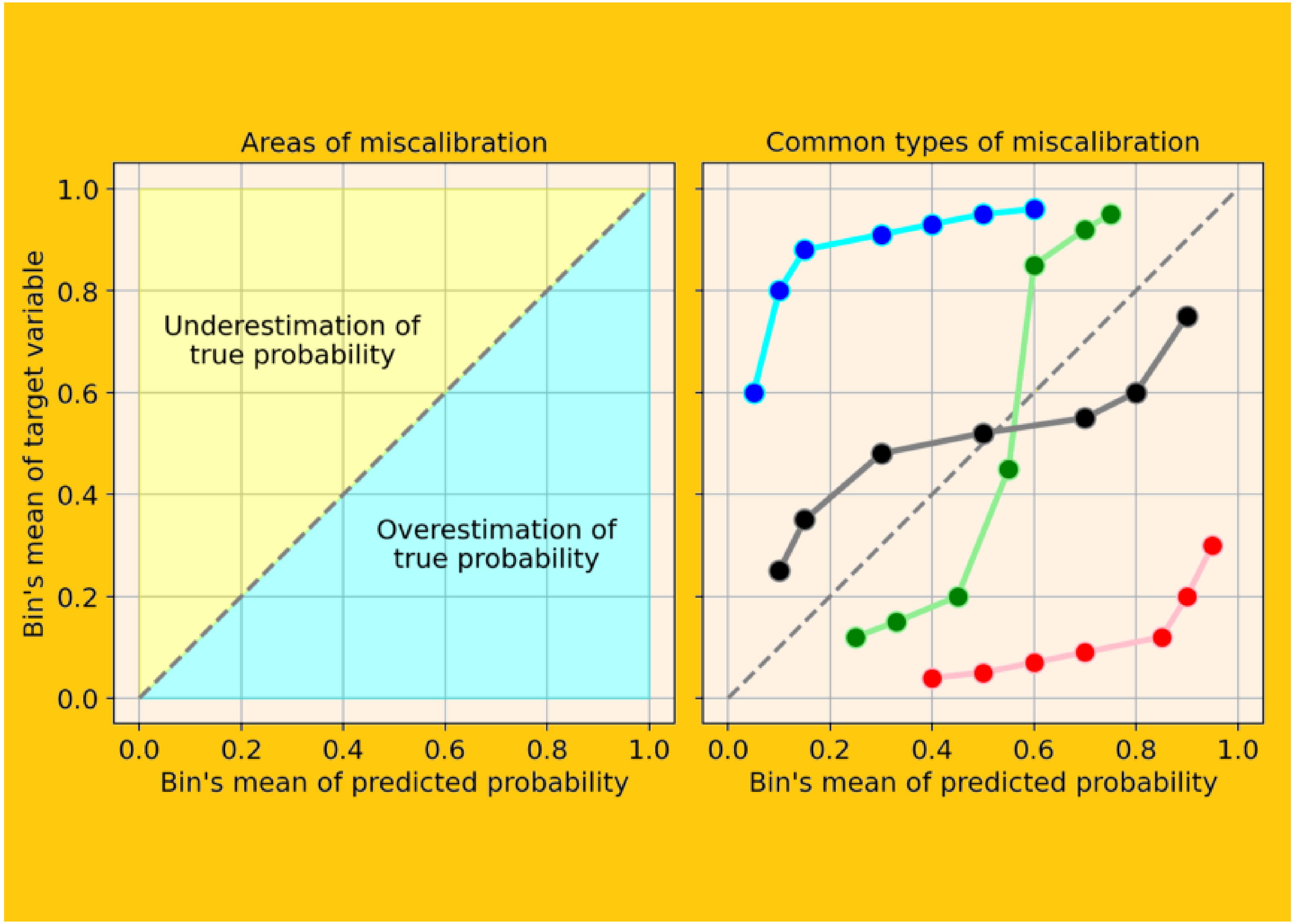


Cancer Res Treat. 2021; 53(3): 773-783

모델의 평가2

- ✓ Calibration Plot





https://towardsdatascience.com/pythons-predict-proba-doesn-t-actually-predict-probabilities-and-how-to-fix-it-f582c21d63fc

npj | digital medicine

www.nature.com/npjdigitalmed

ARTICLE OPEN

Check for updates

Multi-center validation of machine learning model for preoperative prediction of postoperative mortality

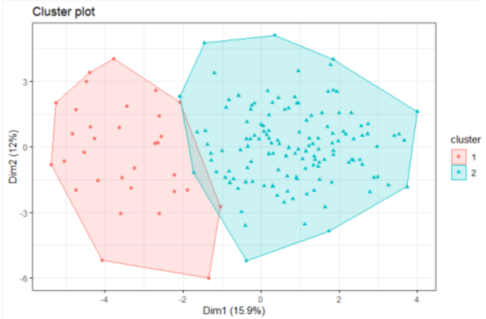
Seung Wook Lee[1], Hyung-Chul Lee[2], Jungyo Suh[3], Kyung Hyun Lee[4], Heonyi Lee[5], Suryang Seo[6], Tae Kyong Kim[7], Sang-Wook Lee[8] ✉ and Yi-Jun Kim[9] ✉

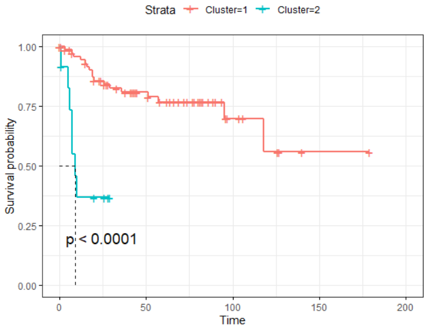
Accurate prediction of postoperative mortality is important for not only successful postoperative patient care but also for information-based shared decision-making with patients and efficient allocation of medical resources. This study aimed to create a machine-learning prediction model for 30-day mortality after a non-cardiac surgery that adapts to the manageable amount of clinical information as input features and is validated against multi-centered rather than single-centered data. Data were collected from 454,404 patients over 18 years of age who underwent non-cardiac surgeries from four independent institutions. We performed a retrospective analysis of the retrieved data. Only 12–18 clinical variables were used for model training. Logistic regression, random forest classifier, extreme gradient boosting (XGBoost), and deep neural network methods were applied to compare the prediction performances. To reduce overfitting and create a robust model, bootstrapping and grid search with tenfold cross-validation were performed. The XGBoost method in Seoul National University Hospital (SNUH) data delivers the best performance in terms of the area under receiver operating characteristic curve (AUROC) (0.9376) and the area under the precision-recall curve (0.1593). The predictive performance was the best when the SNUH model was validated with Ewha Womans University Medical Center data (AUROC, 0.941). Preoperative albumin, prothrombin time, and age were the most important features in the model for each hospital. It is possible to create a robust artificial intelligence prediction model applicable to multiple institutions through a light predictive model using only minimal preoperative information that can be automatically extracted from each hospital.

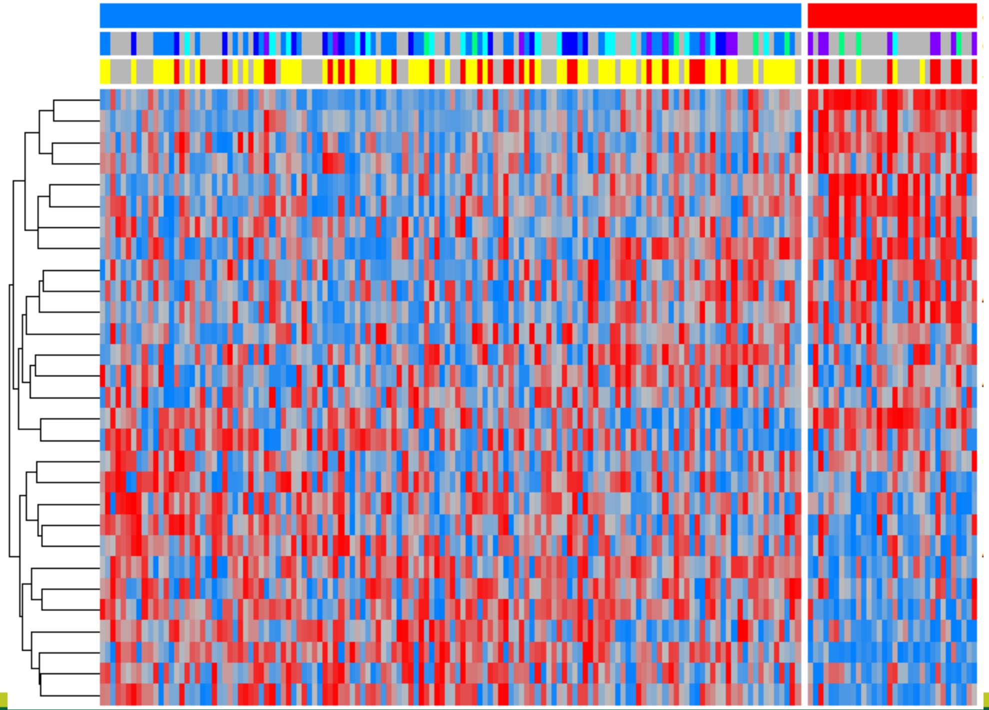
npj Digital Medicine (2022)5:91 ; https://doi.org/10.1038/s41746-022-00625-6

그 밖의 machine learning 사용 예

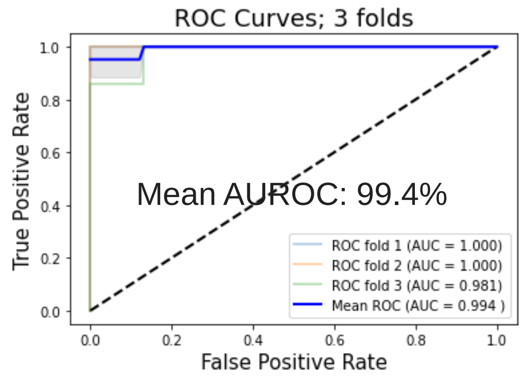
Prognostic analysis

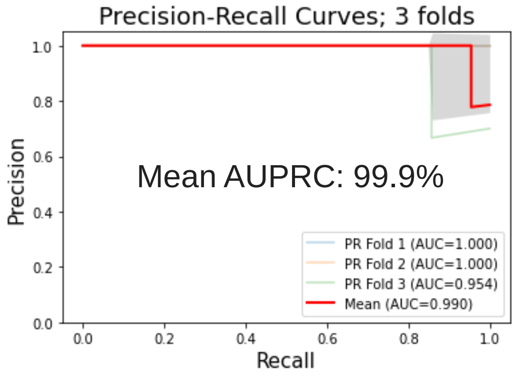




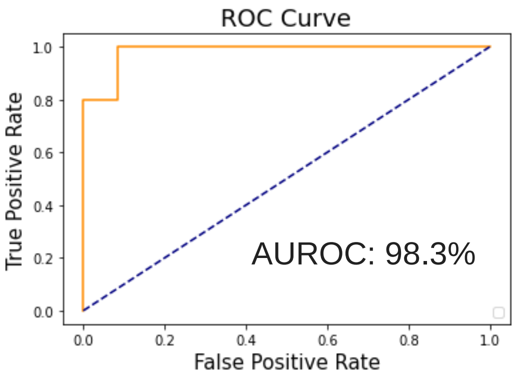


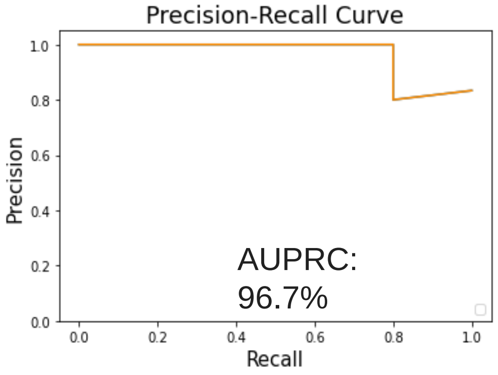
Training set (3-folds cross validation)





Test set





Take home messages

- ✓ 임상적 질문을 생각한다 (제일 중요).
- ✓ Training set, Test set, Internal validation, External validation data를 모으기
- ✓ Retrospective or Prospective?
- ✓ 데이터 preprocessing 까지 되면 80%는 된 것
- ✓ Machine learning - 직접 하거나 (생각보다 easy), collaborator 를 찾는다.
- ✓ Events 0/1 군 간의 matching?
- ✓ 어떤 AI 모델을 쓸 것인지?
- ✓ 어떤 비교 분석 tool 을 쓸 것인지?

이화여자대학교
EWHA WOMANS UNIVERSITY

Thank you!

yijunkim@ewha.ac.kr

EWHA,
THE FUTURE
WE CREATE